



DE LA SENSOMETRIE A L'ETUDE DES PREFERENCES ET DES COMPORTEMENTS ALIMENTAIRES

Pascal Schlich

► To cite this version:

Pascal Schlich. DE LA SENSOMETRIE A L'ETUDE DES PREFERENCES ET DES COMPORTEMENTS ALIMENTAIRES. domain_other. Université de Bourgogne, 2007. tel-00165528

HAL Id: tel-00165528

<https://theses.hal.science/tel-00165528>

Submitted on 26 Jul 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE DE BOURGOGNE

ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

HABILITATION A DIRIGER DES RECHERCHES

DE LA SENSOMETRIE A L'ÉTUDE DES
PREFERENCES ET DES COMPORTEMENTS ALIMENTAIRES

PASCAL SCHLICH

Soutenue le 2 juillet 2007 devant le jury composé de :

Pr. Catherine Dacremont

Examineur

Pr. Marc Danzart

Rapporteur

Pr. Yves Escoufier

Examineur

Dr. Sylvie Issanchou

Examineur

Pr. Ep Köster

Rapporteur

Pr. Georges Le Calvé

Examineur

Pr. El Mostapha Qannari

Rapporteur

RESUME

Ce mémoire relate un parcours de 14 années menant de la sensométrie à la coordination de recherches sur les préférences et les comportements alimentaires de sujets humains. De 1993 à 2000, ce parcours a été réalisé à l'INRA de Dijon dans le Laboratoire de Recherche sur les Arômes, d'abord en tant qu'ingénieur statisticien, puis à partir de 1998 en tant que chargé de recherche. A partir de 2001, ce parcours s'est poursuivi au Centre Européen des Sciences du Goût à Dijon en tant que responsable du Laboratoire d'Interface Recherche-Industrie-Sensométrie.

Après un rappel des acquis de la thèse de Doctorat soutenue en 1993, qui a contribué à l'émergence du domaine de la sensométrie, la liste des travaux de sensométrie menés de 1993 à 2000 est établie en suivant un ordre thématique plutôt que chronologique. Ainsi, la prise en compte des répétitions en tests de différences, des développements originaux en cartographie des préférences (thèse de Philippe Callier), l'utilisation du coefficient RV et des méthodes de type STATIS pour la validation et le traitement du profil sensoriel, une présentation synthétique des résultats d'analyse de la variance des descripteurs du profil sensoriel sous la forme de la table FLASH, une proposition de l'analyse en variables canoniques pour le traitement des données de profil sensoriel, un système intégré de contrôle des performances uni et multidimensionnelles des panels de l'analyse sensorielle (CAP) sont présentés.

Ces développements ont conduit à vouloir accumuler un grand nombre de jeux de données de profils sensoriels dans une base. On explique comment le projet Sensobase, initié en 2003, a réussi à rassembler plus de 500 jeux de données. Trois exploitations de cette base (thèse de Nicolas Pineau) sont présentées. La première propose un système de suivi dans le temps des performances des panels d'analyse sensorielle basé sur le modèle mixte et les cartes de contrôle. La seconde est une méta-analyse de la Sensobase visant à identifier les variables responsables de la performance en évaluation sensorielle. La troisième établit la réalité de l'hétérogénéité des panels en termes de répétabilité et de dispersion des notes, remettant ainsi en cause l'utilisation du modèle classique d'analyse de la variance. L'adjonction d'une covariable à ce modèle permet d'accommoder cette hétérogénéité et offre un gain de puissance.

Le mémoire décrit ensuite des travaux menés au LIRIS ayant tous en commun le développement de mesures sensorielles originales. L'étude des sensations trigéminales causées par la consommation de condiments et épices conduit à un système d'acquisition de données où le sujet doit relever aussi bien l'intensité que la localisation de ces sensations, tout en se concentrant à chaque instant sur l'identification de la sensation dominante à ce moment précis. Cette réflexion aboutit au système de la Dominance Temporelle des Sensations (DTS) dont les liens au profil classique et à la méthode de temps-intensité sont étudiés avec un grand nombre de modèles différents. En particulier, la méthode DTS est adaptée au vin (thèse de Rita Pessina) et son potentiel pour la modélisation des préférences selon le niveau d'expertise en matière de vin est étudié.

Puis, deux projets récents auxquels le LIRIS est associé, et qui lui permettent de poursuivre un développement méthodologique en analyse sensorielle, sont présentés. Le premier porte sur l'effet de la réduction d'alcool dans le vin sur la perception sensorielle et sur les préférences des consommateurs ; le second porte sur les goûts et les préférences en matière d'eau à boire. Chacun de ces projets vient de donner lieu au démarrage d'une thèse CIFRE au LIRIS.

Enfin, deux projets fédérateurs pour le pôle sensoriel dijonnais et visant à mieux comprendre l'origine et l'évolution des préférences et comportements alimentaires des enfants sont présentés. Il s'agit des projets OPALINE et EduSens tout deux soutenus par l'Agence Nationale de la Recherche. OPALINE est l'« Observatoire de Préférences Alimentaires du Nourrisson et de l'Enfant » qui, par le suivi de cohortes mère-enfant depuis le 7^{ème} mois de grossesse de la femme jusqu'aux deux ans de l'enfant, se propose d'étudier les liens entre alimentation pré et post-natale, réactivité olfactive et gustative de l'enfant et préférences et aversions alimentaires de l'enfant. Le projet EduSens se propose quant à lui d'établir dans quelle mesure une éducation sensorielle, c'est-à-dire une sensibilisation au goût plutôt qu'aux propriétés nutritionnelles, peut impacter les préférences et comportements alimentaires d'enfants de 8 à 10 ans. Le LIRIS a initié le projet OPALINE en 2002 et le coordonne maintenant avec l'Unité Mixte de Recherche FLAVIC. Le projet EduSens est, lui, coordonné par le LIRIS. Chacun de ces projets associe une dizaine de partenaires académiques et OPALINE est, de plus, soutenu par un club de mécènes privés.

Remerciements

Je dédie ce mémoire à mes collègues, amis et parents qui m'ont tanné inlassablement pour que j'accouche enfin de cette HDR. Je leur suis reconnaissant pour cette insistance.

Je remercie chaleureusement les membres du jury pour le temps qu'ils ont bien voulu accorder à la lecture et à l'évaluation de ce mémoire.

Le parcours que je relate dans ce mémoire n'aurait pu avoir lieu sans le Laboratoire de Recherche sur les Arômes (UMR FLAVIC) de l'INRA de Dijon. Je tiens à saluer ici l'ensemble de son personnel, sans oublier ceux aujourd'hui ailleurs ou retraités.

Enfin, sans la cohésion de mon équipe, les projets que je coordonne ne pourraient pas fonctionner. Merci au LIRIS !

Sommaire

Introduction	1
Contexte et parcours du candidat	1
Organisation du mémoire	5
1. De la sensométrie à la Sensobase	7
1.1. De la thèse de Doctorat (1993) à la recherche en sensométrie (1994-1998).....	7
1.2. Les répétitions en tests de différences	8
1.3. La cartographie des préférences	10
1.4. L'analyse du profil sensoriel	15
1.5. Le contrôle des performances des panels	20
1.6. Le projet Sensobase	22
2. Concevoir des nouvelles données sensorielles.....	31
2.1. Décrire les sensations trigéminales, noter leur intensité et relever leur succession	31
2.2. Mesurer le « confort digestif » et le « masquage du goût »	32
2.3. Les limites de la méthode de Temps-Intensité	33
2.4. La méthode de la Dominance Temporelle des Sensations (DTS).....	34
2.5. La méthode « Instinctive Temps-Intensité-Profil » (ITIP).....	38
2.6. Le DTS et le vin	39
3. Développer les méthodes au sein de grands projets stratégiques	41
3.1. Justification	41
3.2. Le projet VDQA.....	42
3.3. Le projet SensEaux.....	43
4. Structurer la recherche sur les préférences et les comportements alimentaires des enfants	49
4.1. La Fête de la Science 2002 et 2004.....	49
4.2. Le projet Opaline.....	50
4.3. Le projet EduSens	52
Réflexions et perspectives	55
Références citées dans le texte	60
Bibliographie personnelle	62

Annexes.....	73
Curriculum Vitae.....	74
Organigramme du LIRIS.....	76
Organigramme du CESG	77
Liste des publications annexées au mémoire	79
Article 1	80
Article 2	88
Article 3	105
Article 4.....	154
Article 5.....	173
Article 6.....	180
Article 7	187
Article 8.....	202
Article 9	216
Article 10.....	223
Article 11	239
Article 12.....	251
Article 13.....	263
Article 14.....	271
Article 15.....	302

Introduction

Contexte et parcours du candidat

Ce mémoire relate les activités de recherche et d'encadrement de la recherche que j'ai menées depuis ma thèse de Doctorat soutenue en 1993.

De 1993 à 2000, ces activités se sont déroulées au LRSA (Laboratoire de Recherche sur les Arômes) de l'INRA de Dijon, devenu depuis l'Unité Mixte de Recherche FLAVIC (INRA-ENESAD). Puis à partir de février 2001, ces activités se sont déroulées au Centre Européen des Sciences du Goût (CESG) à Dijon qui est une Unité Mixte de Recherche entre le CNRS, l'Université de Bourgogne et l'INRA.

Recruté en 1985, comme ingénieur d'étude par le LRSA, où je terminais alors mon stage de DEA de statistique de l'Université Paris XI (Orsay), ma carrière de statisticien appliqué commençait avec l'analyse des données de chromatographie en phase gazeuse et les premières tentatives pour les relier aux données de l'analyse sensorielle, qui prenait alors en France son essor sous l'impulsion, entre autres, du LRSA. La nouveauté de cette problématique me conduisit à soutenir une thèse de Doctorat à l'Université Paris XI en 1993. Son titre, « Contributions à la sensométrie », signa la première apparition du mot « sensométrie » qui est aujourd'hui largement accepté, en France comme ailleurs, pour désigner l'ensemble des problèmes statistiques liés à la planification, l'acquisition et l'analyse des données d'évaluation sensorielle.

Cette thèse ouvrit un certain nombre de voies de recherche qui furent, et sont toujours, exploitées par la communauté des sensométriciens. Cette communauté s'est organisée en France autour du « Groupe Agro-Industrie » de la Société Française de Statistique (je suis l'actuel Président de ce groupe) et internationalement autour de la Sensometrics Society (dont je suis membre fondateur).

Immédiatement après ma thèse, j'ai encadré, avec le Professeur Yves Escoufier, la thèse CIFRE (Groupe Danone) de Philippe Callier sur l'extension de la cartographie des préférences aux plans incomplets. Cette thèse a été soutenue en 1996, alors que j'effectuais une année sabbatique dans le Département de Viticulture et Œnologie de l'Université de Californie (Davis) pour travailler avec le Professeur Ann Noble.

En 1998, j'ai été reçu à un concours de Chargé de Recherche de 1^{ère} Classe de l'INRA pour poursuivre mes travaux de sensométrie au LRSA et les orienter progressivement vers la recherche sur les préférences des consommateurs. L'idée à l'époque était de développer le thème des différences interculturelles en termes de préférence et comportement alimentaires. Cette orientation était dictée par des projets européens dans lesquels le LRSA était impliqué.

Mais début 2001, je devais abandonner cette orientation (dont je ne parlerai donc plus dans ce mémoire), car le CNRS et l'INRA me confiaient la direction du LIRIS (Laboratoire d'Interface Recherche-Industrie-Service) qui était une Unité Mixte de Service du CESC. La vocation de cette Unité était à l'époque d'assurer le transfert des résultats du CESC vers l'industrie. Dans les faits, cette Unité fonctionnait exclusivement grâce à des contrats industriels de recherche appliquée en analyse sensorielle. Début 2003, à l'occasion d'une nouvelle contractualisation du CESC en UMR avec l'INRA, il a été décidé que le LIRIS devenait une équipe de recherche à part entière du CESC et que le « S » de LIRIS désignait désormais « Sensométrie » et non plus « Service ».

L'équipe du LIRIS se composait à mon arrivée de 2 ingénieurs contractuels dans le cadre de collaborations industrielles, auxquels venaient s'ajouter des vacataires de courte durée au fil des besoins des expérimentations. L'essentiel de ces contrats venaient de la société Amora-Maille qui avait été à l'origine, avec le Conseil Régional de Bourgogne et le CNRS, de la création du CESC en 1998. Mais suite au rachat de cette société par le Groupe Unilever, les thèmes de recherche développés au CESC avec Amora-Maille n'étaient plus prioritaires, ce qui aurait pu compromettre l'avenir du LIRIS. Fort heureusement, en parallèle de cette évolution négative, j'obtenais deux postes statutaires du CNRS : un ingénieur d'étude (Sylvie Cordelle) en 2001 et un ingénieur de recherche (Christine Lange) en 2002. De plus, l'INRA affecta en 2002 un agent technique (Catherine Pédrón) au LIRIS. Fin 2002, la nouvelle équipe de 4 statutaires était donc composée et elle permit la réorientation du LIRIS du service vers la recherche.

Pour bâtir début 2003 le projet de recherche du LIRIS, j'étais confronté à une triple contrainte. La première contrainte était personnelle, c'était celle de ne pas renier mon passé, de ne pas perdre mon âme de statisticien, de ne pas couper les ponts avec la communauté des sensométriciens que j'avais contribué à créer. La seconde contrainte était mon devoir du moment, il ne fallait pas perdre les acquis du LIRIS obtenus au fil des contrats industriels négociés avec Amora-Maille, le Groupe Danone et les Fromageries Bel par mes deux prédécesseurs (Ep. Köster et Patrick Etiévant). La troisième contrainte enfin était celle de l'avenir, il fallait choisir des thématiques fédératrices au plan dijonnais qui voyait germer les premières discussions de pôle d'excellence autour du sensoriel, il fallait anticiper cette évolution en proposant des projets ambitieux et en assumant les risques qui y étaient attachés.

Ma réponse à la première contrainte a été la thèse de Nicolas Pineau soutenue à l'Université de Bourgogne en 2006 que j'ai encadrée avec le Professeur Mostapha Qannari. Cette thèse a permis la création de la « Sensobase » qui, je l'espère, maintiendra un lien entre mon groupe et la communauté des sensométriciens. Ce projet a pu être réalisé grâce à une collaboration étroite avec l'Association des Centres Techniques des Industries Agroalimentaires (ACTIA) et le soutien de la Région Bourgogne. Le cœur de ce projet consiste à accumuler des données d'analyse sensorielle afin de mieux comprendre le fonctionnement des panels de dégustation

et afin de permettre des comparaisons robustes des méthodes de la sensométrie. Je tiens à spécifier que l'idée de cette base de données sensorielles est l'un des deux projets que j'avais en tête dès la fin des années 90, l'autre projet étant un « observatoire des préférences » dont je parlerai plus loin, et qui ont motivé ma volonté de prendre la direction du LIRIS.

Ma réponse à la seconde contrainte a été la thèse de Rita Pessina soutenue en 2006 à l'Université de Foggia (Italie) (en cotutelle avec l'Université de Bourgogne) que j'ai encadrée avec Benoist Schaal. Cette thèse a permis d'adapter au vin la technique de la « Dominance Temporelle des Sensations » (DTS) mise au point au LIRIS. Depuis fin 2006, et toujours sur le vin, Sophie Meillon a démarré au LIRIS une thèse CIFRE (Pernod-Ricard) dans le cadre du projet « VDQA » (Vins De Qualité à teneurs réduites en Alcool) soutenue par le Programme National de Recherche en Alimentation (PNRA) de l'Agence Nationale de la Recherche (ANR). J'encadre cette nouvelle thèse avec le Professeur Marc Danzart. Ce travail sera l'occasion de poursuivre le développement de la méthode DTS et en particulier de l'utiliser pour mieux comprendre la perception et les préférences des consommateurs. De plus, ce projet a permis l'embauche d'un ingénieur d'étude contractuel affecté à mon équipe et mutualisé entre le projet VDQA et un autre projet sur le goût de l'eau du robinet présenté plus loin.

Le premier volet de ma réponse à la troisième contrainte trouve son origine dans le projet d'un « observatoire des préférences ». L'idée est simple : suivre l'évolution des préférences de cohortes d'individus afin d'en mieux comprendre les mécanismes. Sylvie Issanchou (responsable de l'équipe sensorielle de l'UMR FLAVIC), qui encadrait à l'époque la thèse de Sophie Nicklaus sur les préférences des enfants, et Benoist Schaal (directeur du CESG), qui avait déjà publié de nombreux travaux sur les préférences olfactives du nouveau-né, me donnèrent le point de départ de cet observatoire : les bébés ! Notre décision collégiale était prise dès 2002, il fallait démarrer cet observatoire des préférences en étudiant celles des bébés. Ce choix était à la fois logique et opportuniste. Je fus chargé de réunir les partenaires et de trouver les financements de ce projet. Ainsi naquit « OPALINE » (Observatoire des Préférences ALimentaires du Nourrisson et de l'Enfant). C'est la Région Bourgogne et deux mécènes industriels (Symrise et Nestlé) qui permirent le lancement du projet, puis l'INRA et l'INSERM en 2004 dans le cadre du Programme de Recherche en Nutrition Humaine (PRNH). Dès la fin 2004, ces moyens ont permis le recrutement d'un ingénieur d'étude contractuel dédié à la logistique du projet et basé dans mon équipe. Plus récemment deux nouveaux mécènes (Cedus et Blédina) nous ont rejoints. Mais surtout, OPALINE a pris une nouvelle envergure dans le cadre d'un projet retenu par le PNRA 2006 et coordonné par Sophie Nicklaus (UMR FLAVIC). Ce projet a permis d'adjoindre début 2007 à mon équipe un technicien de recherche contractuel. Au total, OPALINE aura rassemblé un budget de 850 K€ (hors salaire publics), qui pérennise le projet jusqu'à fin 2010. La coordination du projet est assurée depuis début 2007 conjointement par Sophie Nicklaus et moi-même.

Le second volet de ma réponse à la troisième contrainte a été d'aborder le thème des préférences et comportements alimentaires des enfants à un âge plus accessible à l'éducation. Toutefois, l'idée n'était pas celle d'une éducation nutritionnelle – dont l'efficacité est discutable, eu égard au taux de prévalence galopant de l'obésité infantile – mais plutôt celle d'une éducation au goût, dite éducation sensorielle. Il se trouve que, sous la forme des Classes du Goût de Jacques Puisais (Puisais and Pierre, 1987), cette idée de l'éducation sensorielle était déjà disponible sous forme pratique, mais plutôt confidentielle. Le projet « EduSens » est né de l'idée qu'une éducation sensorielle pourrait modifier les préférences et les comportements alimentaires des enfants. Il se propose de tester expérimentalement cette hypothèse en dispensant l'éducation sensorielle en classe, en restauration scolaire ou à la maison par le biais des parents. Il a été soutenu successivement par le Programme de Recherche en Alimentation (PRA) de l'INRA et par le PNRA en 2005. Ces moyens ont permis le recrutement depuis début 2006 d'un ingénieur d'étude contractuel dédié à ce projet et basé dans mon équipe. Je coordonne ce projet qui rassemble 8 équipes de recherche à Dijon et Paris. Son budget total (hors salaires publics) est de 379 K€ pour la période 2006-2008.

Enfin, il y a le projet « SensEaux » qui implique mon équipe et qui allie la volonté de développement de l'analyse sensorielle (seconde contrainte) à la volonté de structuration de la thématique sensorielle dijonnaise (troisième contrainte). Ce projet a été initié sous l'impulsion du pôle de compétitivité dijonnais Vitagora « Goût, nutrition, santé » et a pour slogan « SensEaux, une eau du robinet agréable à boire ». Le projet est porté par Lyonnaise des Eaux qui a lancé en 2006 deux thèses CIFRE à Dijon sur ce sujet. La première a lieu dans l'UMR FLAVIC et consiste à étudier des stratégies de masquage du goût de chlore. La seconde, que j'encadre avec Elisabeth Guichard, est menée par Eric Teillet au LIRIS et consiste à décrire le goût des eaux à boire, qu'elles soient embouteillées ou qu'elles proviennent des réseaux de distribution, et à comprendre les préférences des consommateurs en fonction de cette diversité de goût et de leurs habitudes de consommation. L'ingénieur contractuel recruté sur le projet VDQA travaille à mi-temps sur le projet SensEaux ce qui nous permettra de prolonger son contrat jusqu'en août 2009.

Mon parcours professionnel m'a donc, dans un premier temps, conduit de la consultance en statistique à la recherche en sensométrie. J'ai eu la chance de pouvoir contribuer à l'émergence de cette discipline, qui constitue un champ nouveau et fécond d'application de la statistique. Puis dans un second temps, évoluant dans un milieu où l'objet de la recherche n'est pas la statistique mais les préférences et comportements alimentaires des sujets humains et ayant fait le choix du passage du statut d'ingénieur vers celui de chercheur, mon parcours a évolué de la sensométrie vers l'encadrement de la recherche dans ce domaine des sciences du consommateur. En conséquence, mes activités se sont orientées depuis 6 années vers la coordination de projets et la direction d'une équipe qui compte maintenant 6 chercheurs ou ingénieurs, 2 techniciens, 3 doctorants et 2 à 4 étudiants stagiaires selon la période de l'année.

Aujourd'hui, l'obtention du diplôme d'Habilitation à Diriger des Recherches est devenue nécessaire à la reconnaissance de mon équipe et de ses travaux.

Organisation du mémoire

Nous donnons en annexe de ce mémoire un certain nombre de documents qui étayent l'exposé du parcours qui vient d'être retracé ci-dessus :

- Liste des publications et communications du candidat
- Curriculum Vitae
- Organigramme du LIRIS établi par nos soins
- Organigramme officiel du CESG

Il est de coutume d'accompagner un mémoire d'HDR de quelques exemples de publications du candidat. Après quelques hésitations, nous avons pensé judicieux de sélectionner un et un seul article par année depuis 1994 (année postdoctorale) jusqu'à 2007. De plus, nous avons essayé d'échantillonner ces 14 articles de telle sorte que la diversité des thématiques de recherche et des activités d'application soit restituée. On retrace ainsi assez fidèlement le parcours (et non le projet) du candidat, aussi bien dans sa dimension temporelle que thématique. On a ajouté une dernière publication importante en cours de révision.

En ce qui concerne le mémoire en lui-même, il nous a paru intéressant de le structurer grâce au fil conducteur du parcours exposé ci-dessus.

Ainsi, nous traiterons dans un premier chapitre de la sensométrie, en prenant comme point de départ un résumé de notre thèse de Doctorat soutenue en 1993 et en listant ensuite la poursuite de ces travaux sensométriques qui nous ont menés au projet Sensobase. On terminera naturellement ce chapitre par une présentation du projet Sensobase en expliquant la genèse de ce projet, sa conception et son fonctionnement actuel. Nous en donnerons quelques résultats obtenus récemment.

Le second chapitre rassemblera différents travaux qui concourent tous à proposer de nouveaux types de données d'évaluation sensorielle. On donnera d'abord un résumé des premiers travaux menés au LIRIS en 2001-2002 avec la société Amora-Maille sur la mesure des sensations trigéminales et du confort digestif. On rappellera ensuite notre intérêt pour la méthode de Temps-Intensité qui remonte à des travaux menés lors de notre séjour sabbatique à l'Université de Californie (Davis) en 1996-1997. Cet intérêt a pu rebondir dans le cadre du second volet de recherche méthodologique dans lequel était engagé le LIRIS lors de notre prise de fonction : la mesure de la temporalité des sensations dans le cadre d'un contrat avec le Groupe Danone et les Fromageries Bel. Alors que le premier volet avec Amora-Maille a dû être abandonné pour les raisons économiques déjà évoquées, le second volet sur la dominance temporelle des sensations est fécond et a donné lieu à de nombreuses applications et

développements que nous décrirons. Enfin, nous présenterons des travaux menés en 2003-2004 sur la mesure de la complexité perçue lors de la dégustation d'un produit. Bien que ces travaux n'aient pas encore été publiés, ils sont actuellement repris dans le cadre du projet EduSens et le seront prochainement dans le cadre du projet VDQA.

Ce souci d'inscrire notre recherche méthodologique dans le cadre de projets de recherche pluri-partenaires et pluridisciplinaires est maintenant devenu une idée directrice des orientations de notre équipe. En faisant ce choix, on s'inscrit dans une dynamique où l'objet premier de la recherche n'est pas la méthodologie sensorielle, mais soit la connaissance de la perception du produit par le consommateur, soit la connaissance du comportement du consommateur. Nous avons engagé notre équipe dans deux projets, VDQA et SensEaux, orientés respectivement vers la perception du vin et de l'eau à boire par le consommateur. Nous décrirons donc dans un troisième chapitre ces deux projets en insistant sur les développements méthodologiques que nous avons, ou allons, mettre en œuvre dans le cadre de ces deux projets.

Dans le domaine de la connaissance des préférences et du comportement alimentaire du consommateur, nous décrirons dans un quatrième chapitre les deux grands projets fédérateurs OPALINE et EduSens dont nous pouvons légitimement revendiquer la paternité. Nous entamerons ce chapitre par un rappel d'expérimentations consommateurs originales réalisées au LIRIS dans le cadre des Fêtes de la Science en octobre 2002 et octobre 2004 qui ont démontré comment on pouvait tirer parti d'un souci de vulgarisation de la recherche pour mener des « terrains consommateurs » intéressants.

Enfin, on conclura ce mémoire par quelques réflexions sur notre parcours et quelques perspectives de notre projet.

Chapitre 1. De la sensométrie à la Sensobase

1.1. De la thèse de Doctorat (1993) à la recherche en sensométrie (1994-1998)

Les apports de la thèse par ordre chronologique

- a. Promotion de l'analyse des correspondances pour le traitement des données issues de la recherche sur les arômes [4, 7, 9, 11]
- b. Utilisation du coefficient RV pour l'analyse des données de chromatographie en phase gazeuse [6, 56, 64, 66]
- c. Utilisation de la méthode d'ACPVI pour la modélisation du profil sensoriel par des mesures physico-chimiques [6, 66]
- d. Production de tables pour le contrôle de la puissance des tests de différences [16]
- e. Promotion des plans de voisinage pour la planification des dégustations et analyse des effets d'ordre et de report [17]
- f. Un système graphique pour le contrôle des performances des sujets en analyse sensorielle [20]
- g. Introduction en France de la technique du profil libre-choix couplée à l'analyse procustéenne généralisée [10, 12, 65, 67] et comparaison avec la méthode STATIS [178]
- h. Introduction en France des méthodes de cartographie des préférences [14]

Thèmes poursuivis après la thèse

Après la thèse, nous avons poursuivi des travaux sur les points **d, f, g et h** en développant (ou en faisant développer) respectivement :

- La prise en compte des répétitions dans les tests de différences
- Le système CAP pour le Contrôle des Performances des Assesleurs
- L'utilisation des tests de permutations dans le contexte de STATIS pour évaluer l'homogénéité du panel en termes de différences multidimensionnelles entre produits, corrélations entre descripteurs et scores des produits par descripteur
- La table FLASH et l'utilisation de l'analyse en variables canoniques pour l'analyse des données de profils sensoriels classiques
- L'extension des méthodes de cartographie des préférences aux plans incomplets (méthode CLIP) et la nouvelle méthode PrefMaX

Nous allons donc passer en revue ces différentes réalisations et montrer comment celles portant sur le profil sensoriel, que ce soit le contrôle des performances des sujets ou l'analyse des différences entre produits, nous ont conduits à concevoir le projet Sensobase. La dernière partie de ce chapitre présentera donc ce projet, ses acquis et ses perspectives.

Autres activités d'application de la sensométrie

En parallèle de ces travaux de développement, nous avons continué des travaux d'application de la sensométrie à l'analyse des données couplant évaluation sensorielle et données de chromatographie en phase gazeuse. Nous allons juste mentionner les principaux travaux ci-dessous.

Avec Luigi Moio (Université de Naples) nous avons étudié l'arôme des vins blancs de Bourgogne issus du cépage Chardonnay. Dans un premier temps, nous avons établi grâce à l'analyse des correspondances une liste de descripteurs clefs de cet arôme [15]. Puis, nous avons acquis des aromagrammes de ces vins (description olfactive des composés aromatiques présents dans le vin) que nous avons analysés par ACP [18]. Enfin, nous avons corrélé ces aromagrammes au profil aromatique des vins établis par des panels d'analyse sensorielle [21].

Nous avons ensuite apporté notre concours de statisticien aux travaux de nos collègues de l'Institut Jules Guyot de Dijon sur la caractérisation de la région de production des vins de Bourgogne du cépage Chardonnay par l'analyse des acides aminés libres [29, 36].

Avec la profession du fromage de Comté, nous avons réalisé une typologie aromatique des fromages de Comté [27] à partir d'une batterie de 91 descripteurs de l'arôme du Comté que nous avons pu réduire à 28 descripteurs clefs [24].

Enfin, si nous n'avons plus développé autour des points **b** et **c**, les méthodes proposées durant la thèse ont continué à nous être utiles dans des travaux d'application, tels que [31] par exemple, que nous ne développerons pas ici. Avec Mostapha Qannari, nous avons publié récemment [62] une synthèse méthodologique et bibliographique sur ce sujet.

1.2. Les répétitions en tests de différence

La méthode de correction des degrés de liberté par le coefficient de sur-dispersion

La publication des tables de puissance pour les tests de différence [16] ont conduit des analystes sensoriels à poser la question de la prise en compte des répétitions. Peut-on considérer que 10 sujets ayant réalisé 3 fois le même test triangulaire sont équivalents à 30 sujets ayant chacun réalisé une seule fois le test triangulaire ? Evidemment, non. Néanmoins, les raisons ne sont pas si triviales et ont fait couler beaucoup d'encre, le débat n'étant d'ailleurs toujours pas clos. L'utilisation d'un modèle binomial suppose que chaque test

triangulaire est réalisé avec la même probabilité inconnue de réussite. Etant donné la complexité des processus psychologiques mis en jeu dans un test de différence et la variabilité des sensibilités individuelles, cette hypothèse a peu de chance d'être vraie.

Ainsi, nous avons proposé avec Per Brockhoff [28] une méthode basée sur une correction du modèle binomial au prorata de l'hétérogénéité des sujets d'un panel. Face à un panel complètement homogène, la méthode considère que les répétitions sont autant de degrés de liberté supplémentaires. Par contre, face à un panel très hétérogène la méthode consiste à ne conserver que la proportion du nombre de bonnes réponses par sujet, soit un seul degré de liberté par sujet. La méthode nécessite tout de même une estimation a priori du niveau d'hétérogénéité de la population visée. Grâce à des données intensément répétées, mises à disposition par le Professeur Catherine Dacremont, nous avons pu suggérer qu'une estimation robuste du niveau d'hétérogénéité requiert une première expérimentation basée sur au moins une dizaine de répétitions [43]. La faisabilité d'une telle opération n'est pas assurée dans la pratique.

Nouveaux développements

Depuis, cette méthode a été critiquée par d'autres auteurs (Ennis and Bi, 1998; Kunert and Meyners, 1999; Brockhoff, 2003) qui ont fait d'autres propositions, notamment basées sur le modèle beta-binomial qui correspond mieux à la nature des données. Les réponses apportées par ces différentes approches ne sont pas consensuelles et ne clôturent pas le débat (Meyners, 2007), duquel nous sommes personnellement sortis par manque de temps pour s'y impliquer davantage. Par contre, nous avons souhaité annexer l'article 10 à ce mémoire [54], qui nous avait été commandé sur ce sujet en 2003 par le Professeur Escoufier, alors chargé de coordonner un rapport sur le rôle de la statistique dans l'enseignement des mathématiques. Nous avons laissé l'introduction et la postface de ce rapport afin de bien situer le cadre dans lequel ce texte nous avait été commandé. Le rapport complet, composé de plusieurs autres textes invités, est disponible en ligne¹. Il nous a semblé intéressant de constater que 10 années après la publication des tables de puissance, ce débat, en apparence très technique, revêt aussi une valeur pédagogique pour l'enseignement de la statistique.

Aujourd'hui, la discussion sur les tests de différence de l'analyse sensorielle, presque toujours employés pour montrer une similitude, rebondit avec l'introduction dans notre domaine (Bi, 2005; Bi, 2007) d'une approche, venant de la pharmacologie, qui consiste à tester directement la significativité d'une égalité plutôt que celle d'une différence en contrôlant la puissance du test. Nous sommes sceptique quant aux chances pour cette nouvelle approche de se populariser en analyse sensorielle, eu égard au très fort poids des normes (AFNOR ou ISO)

¹ <http://smf.emath.fr/Enseignement/CommissionKahane/RapportStatistiqueProba/Statistique-annexe.pdf>

publiées sur ce sujet. De plus, les articles cités ci-dessus ne montrent pas clairement, selon nous, que cette nouvelle approche conduit à des résultats sensiblement différents.

1.3. La cartographie des préférences

Rappelons brièvement que la cartographie des préférences est un ensemble de techniques statistiques dédiées à la modélisation des préférences d'un panel de consommateurs par les profils sensoriels des produits établis par un groupe d'analyse sensorielle descriptive. On parle dans ce cas de cartographie externe des préférences, par opposition à la cartographie interne qui ne s'intéresse qu'à l'analyse des notes de préférence pour elles-mêmes sans référence aux profils sensoriels.

La stratégie AUTOFIT

Les 4 modèles classiques (vectoriel, circulaire, elliptique et quadratique) de cartographie externe des préférences [14, 68] conduisent à des difficultés d'interprétation dues à plusieurs causes. D'une part, le meilleur modèle peut être différent d'un consommateur à l'autre. D'autre part, les préférences de certains consommateurs peuvent ne s'ajuster à aucun des 4 modèles. Face à ces difficultés, nous avons proposé une stratégie de choix de modèles appelée AUTOFIT basée sur des tests emboîtés du modèle linéaire. Cette stratégie accompagnée d'un texte didactique sur la cartographie des préférences est proposée en annexe (article 2, [58]) de ce mémoire. Nous tenons à signaler que cette stratégie a été incluse dans l'un des deux principaux logiciels d'acquisition et de traitement des données d'analyse sensorielle (FIZZ) et également dans le module MX du logiciel d'analyse statistique XLSTAT. Un tutoriel disponible en ligne² de ce logiciel reprend les données sur les chips anglaises de l'article d'origine [14].

La stratégie AUTOFIT permet aussi de tester la significativité statistique du meilleur des 4 modèles choisi individuellement pour chaque consommateur. Or, comme les études de cartographie des préférences sont souvent menées avec un nombre restreint de produits (entre 4 et 8), ce test est souvent non significatif. De plus, lorsque l'espace sensoriel des produits ne se résume pas facilement à deux dimensions, tout se complique aussi bien en raison de l'inflation des degrés de liberté du modèle, que par la difficulté de la présentation des résultats. L'article 8 [44], annexé à ce mémoire, présente une étude où au contraire tout se passait bien en raison d'un grand nombre de produits (24 bières) et en raison d'un espace sensoriel qui se résumait bien en deux dimensions. Un autre mérite de cette étude est d'avoir inclus une seconde notation des produits accompagnés chacun de sa bouteille et de son prix, et ce, de manière indépendante de la première. Nous avons ainsi pu mettre en évidence que les

² <http://www.xlstat.com/fr/support/tutorials/prefmap.htm>

préférences des consommateurs les plus jeunes étaient les plus sensibles à l'information externe. De plus, cet effet était différent entre les hommes et les femmes, les jeunes hommes ayant davantage tendance à systématiquement surcoter les bières d'importation qui étaient aussi les plus chères.

La thèse de Philippe Callier

Parallèlement au développement de la stratégie AUTOFIT (1993-1995), nous encadrions aussi notre première thèse, celle de Philippe Callier en contrat CIFRE avec le Groupe Danone. Il s'agissait de généraliser la cartographie des préférences aux plans incomplets dans lesquels chaque consommateur ne goûte qu'une partie des produits de l'étude.

La solution retenue a consisté à proposer l'utilisation de plans en carrés latins mutuellement orthogonaux (MOLS) pour la planification des ordres de dégustation assurant l'équilibre des effets de rangs de dégustation et de report de l'effet du produit précédent sur le produit suivant. Il faut signaler que cette stratégie avait été trouvée en même temps par nos collègues anglais (Wakeling and Macfie, 1995), qui ont été plus rapides que nous à la publication...

Pour l'analyse des données, Philippe Callier a développé une méthode originale, nommée CLIP pour « CLustering of Incomplete Preferences ». Cette méthode est décrite dans l'article 4 [25] annexé à ce mémoire qui comporte toute la bibliographie relative à la méthode. Rappelons simplement qu'elle utilise l'indice de Gower (Gower, 1971) qui permet d'affecter une similarité entre deux consommateurs n'ayant dégusté que quelques produits en commun. La matrice de similarité entre consommateurs, ainsi construite, est alors décomposée par une technique de « multidimensional scaling » (MDS), dans laquelle chaque couple de consommateurs est pondéré par une quantité proportionnelle au nombre de produits dégustés en commun. Ainsi, la cellule de cette matrice correspondant à deux consommateurs ayant dégusté des produits tous différents sera affectée d'un poids nul et ne participera donc pas au calcul des axes du MDS. Une fois ces axes obtenus et leur nombre choisi par l'observation des valeurs de STRESS, CLIP effectue une classification des consommateurs en groupes de préférences homogènes.

Philippe Callier a étudié dans sa thèse les propriétés de robustesse de cette technique selon le degré d'incomplétude de l'étude (rapport du nombre de produits dégustés par sujet au nombre de produits de l'étude). Pour ce faire, il a utilisé aussi bien des données simulées que des données réelles. De plus, la planification en MOLS d'un plan complet fait que la simple suppression d'une partie des ordres de dégustation, rendant le plan incomplet, conserve les bonnes propriétés d'équilibre du plan initial. Ainsi, on a pu étudier la robustesse de deux aspects essentiels : les différences de préférences entre classes de consommateurs et la composition elle-même de ces classes. Pour ce dernier point, nous avons utilisé l'indice de Rand ajusté (Hubert, 1985) pour lequel nous avons développé un test de permutation

permettant d'évaluer si deux partitions d'un même ensemble de consommateurs sont significativement différentes.

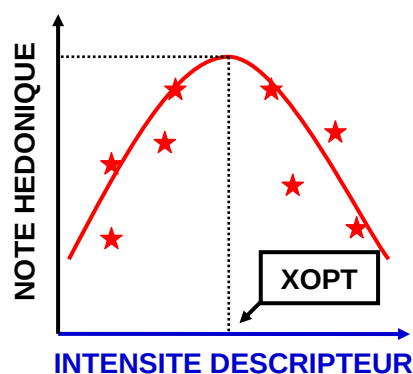
Les présentations que Philippe a pu faire dans des congrès [71, 72, 75, 129] ont retenu l'attention des professionnels car elles débouchaient sur des recommandations pratiques. Ainsi, sur le niveau d'incomplétude, il a suggéré que chaque consommateur ait au moins dégusté la moitié (ou l'entier immédiatement supérieur à la moitié) des produits de l'étude. Ainsi, deux consommateurs auraient toujours au moins un produit en commun. Beaucoup, nous ont dit après coup « *cela va de soi, c'est une affaire de bon sens* ». Ici les statistiques ont au moins permis de confirmer que le « bon sens » avait raison. Après sa thèse, Philippe a été embauché comme statisticien au Centre de Recherche de Nestlé à Lausanne. Il a travaillé une dizaine d'années dans ce poste avant d'opérer en 2005 une reconversion dans le « marketing stratégique » au siège du Groupe Nestlé.

La méthode PrefMaX

L'approche de la méthode CLIP consiste à segmenter les consommateurs sur la base de leurs préférences. Lorsque le profil sensoriel des produits est disponible et que les modèles de cartographie externe des préférences évoqués au début de cette section sont appliqués, la question de la segmentation des consommateurs se pose a posteriori des modélisations individuelles. Il est alors difficile de réaliser cette segmentation. Seules des solutions empiriques peu précises étaient décrites dans la littérature. Nous avons donc mis au point une méthode, appelée PrefMaX, qui conserve l'idée de la modélisation individuelle et de la segmentation a posteriori. La figure 1 décrit le principe de cette méthode.

Dans cette méthode, contrairement aux méthodes classiques de cartographie externe des préférences, on ne résume plus le profil sensoriel par un plan factoriel, obtenu par exemple par analyse en composantes principales des moyennes des intensités produit x descripteur. Cette réduction présente l'avantage de pouvoir plus facilement traiter les jeux de données avec peu de produits et/ou un grand nombre de descripteurs. Elle permet également de traiter les plans incomplets. Elle peut paraître réductrice dans la mesure où elle ignore les liaisons entre descripteurs. En fait, celle-ci seront prises en compte à l'étape ultérieure de segmentation des consommateurs sur la base de leur point idéal dans l'espace des descripteurs. L'obtention de cet espace où les objets sont les consommateurs et où les dimensions sont les descripteurs est l'intérêt essentiel de cette méthode. Il permet d'interpréter la position de chaque produit de l'étude, ou encore de chaque formulation potentielle, comme celle d'un consommateur dont l'idéal correspondrait exactement aux intensités des descripteurs relevés dans ce produit. Pratiquement, cela permet de tester l'acceptabilité d'un nouveau produit grâce à la seule connaissance de son profil sensoriel.

PrefMaX



Pour chaque couple consommateur/descripteur, régression quadratique des notes hédoniques des produits sur les moyennes d'intensité du descripteur. Stockage des niveaux optimaux XOPT qui deviennent les entrées des analyses suivantes pondérées par les R^2 de chaque régression.

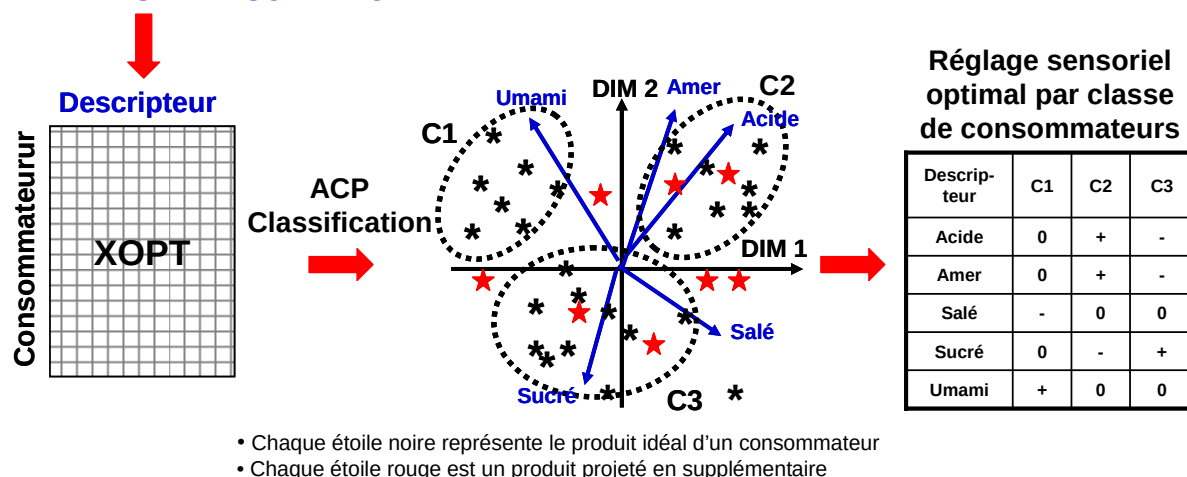


Figure 1 : Principe de la méthode PrefMaX

Nous n'avons jamais publié une présentation exhaustive de cette méthode car elle mériterait encore un certain nombre d'améliorations techniques – par exemple, pour le traitement des idéaux négatifs résultant d'une régression quadratique conduisant à une courbe convexe. Néanmoins, nous avons employé cette méthode dans un grand nombre de situations réelles où elle a donné d'excellents résultats, facilement interprétables et présentables. Quelques unes de ces applications ont été publiées [39, 42, 53, 77]. Les références [42, 53] sont annexées à ce mémoire (articles 7 et 9 respectivement) et nous allons les décrire brièvement car nous estimons qu'elles ont une vraie valeur applicative.

La première application de l'analyse sensorielle dans le domaine automobile

L'article 7 en collaboration avec Renault a une valeur historique, car il relate la première application de l'analyse sensorielle dans le monde de l'automobile, qui eut lieu entre 1994 et 1996. A l'époque il s'agissait surtout d'évaluer dans quelle mesure les concepts de l'analyse sensorielle des aliments pouvaient être transposés à l'essai automobile par des pilotes professionnels (nos dégustateurs). Deux années plus tard, ces données d'essai automobile jouèrent le rôle de profils sensoriels explicatifs des préférences déclarées par des consommateurs en situation d'essai de plusieurs véhicules. Même si cette étude souffre d'un certain nombre de limites détaillées dans l'article, elle a vraiment montré que le domaine de l'automobile, comme tout autre domaine mettant en jeu la perception de produit à usage

personnel, peut bénéficier de l'analyse sensorielle. Pour revenir à PrefMaX, l'article compare cette approche à celle de la cartographie des préférences classiques et montre bien ses avantages pratiques. Aujourd'hui, aussi bien chez Renault que chez PSA, il existe des départements d'analyse sensorielle employant plusieurs dizaines de personnes. Nous n'avons jamais vraiment réussi à savoir s'il en était de même chez les constructeurs automobiles étrangers. Leur absence des congrès de l'analyse sensorielle peut laisser supposer le contraire.

La première application de la cartographie des préférences aux produits frais non industriels

L'article 9 est le fruit d'une collaboration avec le CTIFL (Centre technique interprofessionnel des fruits et légumes) qui visait à appliquer, pour la première fois, les techniques de cartographie des préférences aux produits frais non industriels, tels que les fruits et légumes. Les contraintes inhérentes à ce domaine ont fait que sur les deux modèles testés, tomate et abricot, seule la première expérimentation, menée en 1999 sur la tomate, a donné de bons résultats. Nous avons mis en évidence une segmentation des préférences des consommateurs en 4 groupes interprétables aussi bien du point de vue sensoriel, que du point de vue de la typologie des consommateurs composant ces 4 groupes. Mus par un souci de validation, nous relançons une expérimentation similaire en 2001, qui donna sensiblement les mêmes résultats détaillés dans l'article 9. Depuis, le CTIL a réalisé d'autres cartographies des préférences sur d'autres modèles.

Quelques autres développements importants

La cartographie des préférences est vraiment un chapitre essentiel de l'analyse sensorielle et c'est peut-être le plus stratégique dans la mesure où il permet le dialogue entre les services de marketing de l'entreprise et le laboratoire d'analyse sensorielle. Nous avons arrêté notre recherche dans ce domaine vers la fin des années 90, c'est-à-dire à peu près lorsque l'objet de nos recherches évoluait des méthodes d'analyse vers l'acquisition de nouveaux types de données. Ce n'est pas le lieu ici pour faire la bibliographie de toutes les nouveautés en matière de cartographie des préférences, mais nous pourrions tout de même en citer quelques unes avant de clore cette section. Tout d'abord, l'extension des méthodes de cartographie des préférences aux plans de comparaison par paire (Courcoux and Séménou, 1997; Qannari, Courcoux et al., 2000) est essentiel car on sait que ce type de données a la faveur de beaucoup de départements de marketing. Ensuite, l'utilisation de modèles en classes latentes (Poulsen, Brockhoff et al., 1997; Vigneau, Courcoux et al., 1999; Courcoux and Chavanne, 2001) est une avancée décisive pour la détermination du nombre de segments de préférences. Enfin, l'identification dans le modèle d'une classe de consommateurs dont les scores de préférences ne seraient que du bruit (Cleaver and Wedel, 2001) est une idée intéressante que les faibles taux d'ajustement obtenus par la stratégie AUTOFIT valide.

1.4. L'analyse du profil sensoriel

Coefficient RV, STATIS et tests de permutation

En analyse sensorielle, c'est la méthode du profil libre-choix (Williams and Langron, 1983; Williams and Langron, 1984) qui aura permis la promotion de l'analyse procustéenne généralisée (Gower, 1975). Autrement dit, c'est bien un nouveau protocole d'acquisition de données qui stimule l'application d'une méthode statistique nouvelle. Après avoir présenté dans notre thèse ces techniques anglo-saxonnes, tout en soulignant les potentialités des méthodes françaises basées sur le coefficient RV (Escoufier, 1973), l'heure était venue d'écrire une synthèse sur cette question à l'adresse de la communauté de l'analyse sensorielle.

Une invitation à écrire un chapitre pour l'ouvrage « *Multivariate Analysis of Data in Sensory Science* » m'en donna l'occasion. Dans ce texte (article 3 [59] annexé à ce mémoire), nous montrons tout l'intérêt de la méthode STATIS (Lavit, 1988) pour le traitement des données de profil sensoriel libre-choix, ainsi que de profil classique. Dans ce chapitre, nous mettons également au service de l'analyse sensorielle des outils nouveaux de combinatoire (Kazi-Aoual, Hitier et al., 1995) permettant de tester l'homogénéité d'un panel de dégustateurs vis à vis de la configuration multidimensionnelle des produits. Grâce à la variante STATIS dual, nous avons montré comment l'on pouvait aussi tester l'homogénéité de ce panel vis à vis de la structure de corrélation entre les descripteurs. Enfin, en présence de plusieurs panels les mêmes concepts (tests de permutation analytiques) permettent de tester dans quelle mesure échanger les sujets entre les panels ne modifie pas les différences entre produits. Ce dernier aspect, pas encore vraiment intégré par les utilisateurs, pourrait être très intéressant dans le cadre d'un grand groupe possédant des panels de dégustation aux quatre coins du monde.

Pour illustrer ce chapitre, nous avons utilisé les données de l'expérimentation sur le café menée par le *European Sensory Network* (ESN) auquel nous appartenions à l'époque. C'est Jean McEwan du *Campden Food & Drink Research Association* (Chipping Campden, UK) et moi-même qui avons établi le plan d'expérience de cette opération et en avons analysé les données. Les résultats ont été consignés dans un ouvrage collectif publié par l'ESN [55]. Cette expérimentation consistait à faire réaliser par 11 panels d'analyse sensorielle descriptive situés dans 8 pays européens le profil sensoriel des mêmes 16 cafés préparés exactement selon le même protocole. Seules les procédures de sélection et entraînement des sujets étaient propres à chaque panel, ainsi que le choix des descripteurs bien entendu. Le résultat majeur de cette étude fut d'établir que la carte sensorielle des produits était sensiblement la même pour tous les panels, validant ainsi l'analyse sensorielle descriptive comme un outil de mesure objectif. Les outils statistiques basés sur le coefficient RV furent essentiels pour établir ce diagnostic.

Un nombre conséquent d'études d'analyse sensorielle utilisant ces outils ont depuis été publiées. Avec notre participation, nous pouvons citer [30, 35, 40] (le premier étant l'article 5

annexé). A chaque fois, le coefficient RV est employé pour décider si deux configurations multidimensionnelles sont similaires ou pas. Mais la diversité des situations d'application est vaste : vin de base versus Champagne [30] pour établir l'apport aromatique de la méthode de vinification des vins de Champagne, odeur versus arôme [35] pour discriminer des vins rouges de Bourgogne et panel d'étudiants versus panel de sujets rompus à l'analyse sensorielle [40].

Aujourd'hui, sur ce sujet nous collaborons avec le groupe de sensométrie de Nantes qui est leader dans le domaine de l'analyse multi-tableaux appliquée à l'évaluation sensorielle. [82] et [84] sont deux exemples de propositions qui, si elles n'ont pas encore reçu d'application démonstrative en analyse sensorielle, présentent un fort potentiel. Dans la nouvelle méthode STATIS de [82], on peut forcer le compromis à un nombre fixé de dimensions (un peu dans l'esprit du *multidimensional scaling*), ou encore le contraindre à être lié de manière optimale à des données externes. Cette dernière possibilité mériterait d'être étudiée dans le cadre de la cartographie des préférences.

Analyse de la variance et table FLASH

Avant d'être un objet multidimensionnel, le profil sensoriel est une collection de descripteurs dont chacun est analysé séparément par analyse de la variance (ANOVA). Le statut fixe ou aléatoire du sujet dans les modèles d'ANOVA a fait couler beaucoup d'encre et il a fallu se battre [33, 34] pour faire accepter l'idée que c'est presque toujours un effet aléatoire.

Une fois le modèle choisi, très souvent « produit + sujet + produit*sujet », on procède à la comparaison multiple des moyennes des produits. Hormis le problème du choix de la méthode de contrôle des risques d'erreurs dues à l'inférence multiple, si le nombre de produits est conséquent, disons supérieur à 4, et si le nombre de descripteurs est également conséquent, disons supérieur à 10, la présentation de ces comparaisons de moyennes et leur interprétation est délicate. Dans la pratique, on a presque toujours plus de 4 produits et plus de 10 descripteurs. Confronté dès le début de ma carrière d'ingénieur statisticien à ce problème pratique, aussi bête qu'embêtant, et réalisant que, finalement, dans la plupart des applications l'utilisateur a juste besoin de savoir quels sont les produits les plus et les moins intenses pour chaque dimension sensorielle, nous avons mis au point une analyse allégée conduisant à une présentation synthétique des résultats sous la forme d'une table, appelée FLASH [32, 41, 74] car au premier coup d'œil à cette table l'histoire principale contenue dans les données devrait « sauter aux yeux ».

ATTRIBUTE	F	PROB	GMEAN	16	3	10	1	6	2	8	5	11	4	13	9	7	12	15	14
FIRM	17.20	0.0000	37	19-	21-	21-	22-	22-	24-							49+	63+	69+	77+
COMPACT	8.58	0.0000	49		29-		40-	27-	39-	36-			59+				60+	73+	73+
GRAINY	11.60	0.0000	27	6-	14-	12-	15-	18-		52+							39+	51+	46+
PASTRY	4.43	0.0000	19				10-			10-						4-	28+	44+	
SWEET	9.49	0.0000	42	54+	30-	56+	27-	32-		29-	33-			54+	56+	22-	51+	69+	
BUTTER	1.74	0.0470	17								25+							25+	9-
CHESNUT	6.02	0.0000	41						29-	21-		30-		53+	58+			63+	58+
CEREAL	2.33	0.0040	26				17-												44+
ARTICHOK	2.63	0.0010	40			52+			23-										28-
CELERY	1.86	0.0290	17			29+									24+				
HERBACE	3.23	0.0000	31						20-			49+				42+		20-	13-
EARTHY	5.78	0.0000	30				40+	42+		41+		43+		21-		47+	19-	14-	
ASTRING	2.58	0.0010	29													44+	18-		
BITTER	2.88	0.0000	27						18-	36+				17-		46+			
RAWPOTAT	2.81	0.0010	23				32+			38+								10-	
METALLIC	1.78	0.0400	15				23+	26+											
PASTY	3.68	0.0000	35				22-	47+	56+								23-	24-	
MEALY	11.40	0.0000	42	29-	60+			54+	78+	55+	31-	54+		32-			26-	27-	31-
STICKY	2.95	0.0000	28	37+	16-							17-			47+				
MASHABLE	13.90	0.0000	40	69+		58+	66+		29-	25-	51+					26-	27-	16-	11-
HOMOGEN	7.41	0.0000	59	78+		72+	70+		42-	31-	69+		70+		69+			45-	
GREASY	5.27	0.0000	15	26+	7-			8-	2-	7-	26+	7-	25+		25+	8-			
MOIST	11.70	0.0000	47	68+		63+	69+	38-	17-	29-	58+	35-				33-	56+		34-

Figure 2 : Table FLASH pour 16 variétés de pommes de terre cuites à la vapeur (16 dernières colonnes du tableau) évaluées chacune 2 fois par un panel de 14 sujets pour 30 descripteurs sur une échelle d'intensité allant de 0 à 100. F et PROB sont les valeurs de la statistique de Fisher et de son risque α associé pour tester l'existence de différences entre les produits. GMEAN est la moyenne générale du descripteur.

L'idée de la table FLASH consiste à remplacer les comparaisons multiples de moyennes par la comparaison de chaque moyenne à la moyenne des autres produits et de la faire ressortir dans la table si et seulement si elle est significativement plus grande (en jaune suivie d'un signe «+») ou plus petite (en vert suivie d'un signe «-») que celles des autres produits. Ainsi, la table fait ressortir les « gagnants » et les « perdants » seulement. Une case vide signifie que l'on peut considérer que la moyenne était égale à celle de la colonne GMEAN (moyenne générale).

De plus, l'ordre des lignes (descripteurs) et des colonnes (produits) de la table est choisi de manière à optimiser la lisibilité du tableau. Les lignes sont rangées selon une classification de variables (procédure VARCLUS option CENTROID de SAS®) qui assure que les descripteurs d'une même classe (même couleur), donc bien corrélés entre eux, se suivront dans la table. La première classe en haut de la table est celle qui apporte le plus de discrimination entre les produits. Puis les classes sont rangées selon la meilleure corrélation multiple avec la classe précédente. Au sein d'une même classe, le premier descripteur sera celui qui représente le mieux la classe au sens de la corrélation multiple et ainsi de suite. Les colonnes (produits) seront ordonnées selon les moyennes croissantes du premier descripteur de la table.

Grâce à cette série d'astuces, la table FLASH produit vraiment cette lisibilité recherchée et ce même avec beaucoup de produits et/ou beaucoup de descripteurs. Nous avons établi cette représentation très tôt dans notre carrière d'ingénieur statisticien et depuis, nous n'avons jamais cessé de l'utiliser aussi bien pour du profil sensoriel que pour d'autres types de données. Nous touchons ici un aspect de l'analyse des données, qui peut paraître secondaire, mais qui est en fait essentiel : savoir présenter les résultats des analyses statistiques de manière synthétique et pratique. Il faut toujours garder à l'esprit que le client du consultant statisticien doit pouvoir prendre ses décisions d'action au vu du rapport d'analyse sans être obligé lui-même de devoir retravailler sa présentation. Dans mon activité d'enseignant, la table FLASH m'est très utile pour faire passer cette notion.

L'analyse en variables canoniques

Pour l'analyse du profil sensoriel conventionnel (c'est-à-dire lorsque tous les sujets notent les mêmes descripteurs), après avoir analysé chaque descripteur en ANOVA, la pratique courante des analystes sensoriels est de réaliser une ACP du tableau des moyennes des produits. Cette analyse est parfois réalisée après avoir pris soin de retirer les descripteurs non significatifs en ANOVA. Cette précaution est en effet souhaitable surtout lorsque l'on utilise l'ACP normée. Néanmoins, il reste que la variabilité des sujets et/ou des répétitions n'est pas exploitée par cette analyse en ACP. En analyse unidimensionnelle, cela consisterait à montrer les moyennes des produits sans utiliser leur écart-type afin de les comparer par un test statistique... Alors que tout le monde s'accorderait à rejeter une telle pratique, peu de gens semblent réaliser que leur ACP fait la même chose au niveau multidimensionnel.

Nous avons donc promu l'utilisation de la MANOVA et de l'analyse en variables canoniques (AVC) de l'effet produit dans le modèle de MANOVA ad hoc [41, 80]. Celle-ci consiste à rechercher les vecteurs propres du produit matriciel $B.W^{-1}$, où B est la matrice de covariance inter-produits et W la matrice de covariance intra-produits. En fonction du plan d'expérience et du statut fixe ou aléatoire que l'on accorde à l'effet sujet, la matrice W ne sera pas la même. En clair, il faut en MANOVA, comme en ANOVA, savoir quelles sont les unités expérimentales dont la variabilité va être la base de comparaison des produits. Dans la plupart des cas, W sera la matrice de covariance du terme d'interaction produit*sujet.

Ce qui est couramment appelé en France l'analyse factorielle discriminante (AFD) correspond en réalité à l'AVC du facteur groupe dans un modèle de MANOVA à un seul facteur groupe. Ainsi, ceux qui utiliseraient un logiciel d'AFD pour analyser leurs données de profils sensoriels devrait absolument effectuer au préalable les deux transformations suivantes :

1. Moyenner les répétitions issues d'un même couple produit*sujet
2. Centrer les descripteurs par sujet (afin de retirer l'effet sujet)

Pour un jeu de données équilibré (tous les sujets ont dégusté chaque produit un même nombre de fois), ce procédé conduirait aux mêmes axes que ceux de notre analyse. Par contre, les tests statistiques associés à la MANOVA ne seraient pas corrects.

Ces tests permettent d'une part de décider s'il existe dans l'espace sensoriel tout entier des différences entre produits – c'est toujours le cas, autrement on n'aurait pas réalisé un profil sensoriel – et, d'autre part, de se prononcer sur le nombre de dimensions nécessaires pour restituer la discrimination des produits. On verra dans la section suivante comment ces deux statistiques peuvent être mises à profit pour le contrôle des performances des sujets.

Sur l'exemple des données de la table FLASH précédente, cette analyse conduit à la figure 3. Cette figure raconte de manière géométrique la même histoire que la table FLASH nous avait contée de manière plus numérique. Les ellipses de confiance permettent d'estimer visuellement la significativité des différences entre produits, exprimées dans cette carte sensorielle par leur distance.

En résumé, une table FLASH et un plan d'AVC nous ont en général toujours suffi pour analyser un jeu de données de profils sensoriels. Néanmoins, il faut parfois réaliser ce couple d'analyse par type de descripteurs (odeur, texture, saveur et arôme ...). C'est un autre débat.

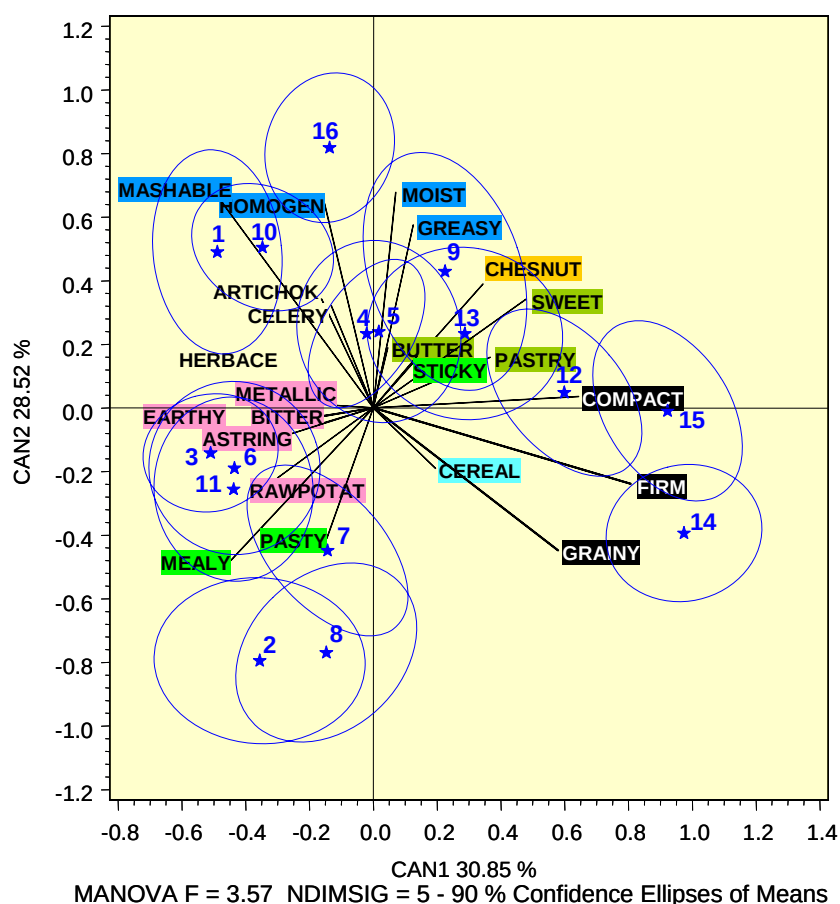


Figure 3 : premier plan de l'analyse en variables canoniques de l'effet produit pour les données de la figure 2. Les couleurs des descripteurs renvoient à la classification de la table FLASH. Les ellipses de confiance des moyennes des produits sont basées sur la loi bi-normale en utilisant les projections individuelles (une seule par sujet) pour estimer la matrice de covariance. Les descripteurs sont représentés à l'aide de leur coefficient de corrélation total avec les axes.

1.5. Le contrôle des performances des panels

En tant qu'outil de mesure, le sujet et le panel d'analyse sensorielle se doivent d'être répétables (fiabilité), discriminants (efficacité) et justes (validité). La sensométrie a proposé de nombreux outils pour contrôler ces performances de la mesure sensorielle. Le système GRAPES [20] proposé dans le cadre de notre thèse consistait à présenter pour chaque descripteur 6 statistiques de performances sous forme de 3 graphiques. A l'usage, ce système n'est pas apparu assez lisible et pratique, surtout en présence d'un grand nombre de sujets et descripteurs.

La table CAP

Ainsi, nous avons mis au point le système CAP (« Control of Assessor Performances ») qui a évolué d'une première version en noir et blanc en 1997 [76] vers une version rénovée aussi bien au point de vue de la méthode que des couleurs de l'internet en 2006 [86].

ATTRIBUTE	FPROD	FINTER	MEAN	RMSE	7	5	11	3	1	8	12	4	10	14	6	9	2	13
FIRM	17.17	1.51	37.5	18.8	+	+	+	+	+	+	+	+	+	+	=!	+	+	=
MASHABLE	13.91	0.88	40.2	25.5	+	=	+	+	=	=!	=	=	=	+	=	+	+	=
MOIST	11.65	1.20	46.9	21.1	+	+	+	+	=	+	=	+	+	=!	+	+	=	=
GRAINY	11.57	1.10	26.9	20.6	+	+	+	+	=!	=	+	=	=	+	=!	+	=	=
MEALY	11.44	0.83	42.3	25.2	+	=	=	=	=!	=	=	=	=	=	=	=	+	=
SWEET	9.49	1.10	42.4	21.9	=	+	=	+	=	=!	=	+	=!	+	+	=	=	=
COMPACT	8.57	1.77	48.6	18.6	+	+	=	+	=	+	=!	+	+	+	=!	=	+	+
HOMOGEN	7.40	1.17	59.0	22.4	+	+	+	=	=	+	=	+	=	=!	=	+	=	+
CHESNUT	6.02	1.10	41.1	24.3	+	+	=	=	-	+	=!	=	=!	=	=	+	=!	=
EARTHY	5.78	1.07	29.8	21.7	+	+	=	=	=	=!	=	=	=	=	-	=!	=	=
GREASY	5.26	1.47	15.3	15.7	+	+	=	=!	=	=!	=	=	+	=	=	=!	=	=
PASTRY	4.42	1.51	19.2	20.1	=	+	=	=!	+	-	+	=	+	=	=	=!	=	=
PASTY	3.67	1.32	34.7	22.8	+	=!	+	+	+	=	=!	+	-	=!	=!	=!	=	=
HERBACE	3.23	1.55	30.5	21.2	=!	+	+	+	+	=	=!	=	+	-	+	=	=	=
STICKY	2.95	1.30	27.6	22.5	-	+	=	=	=	+	=	+	-	=!	=	=	=	=
BITTER	2.87	1.28	26.9	20.2	=!	=	=	=	+	=	+	=	=	=	-	=!	=	=
RAWPOTAT	2.81	1.09	22.7	21.9	=!	=	+	=	+	=!	=	=	+	=	=	=!	=	=
ARTICHOK	2.62	0.95	39.7	24.7	=	=!	=	-	=	=!	=	=	=!	=	=	-	=	=
ASTRING	2.58	1.00	29.3	21.9	-	=!	+	=	-	=!	=!	-	=	=	=	=!	-	+
CEREAL	2.32	1.05	26.3	22.5	=!	=!	=	=	+	=!	+	=	=	=!	=	=	=	=
CELERY	1.86	1.38	16.9	15.9	+	=!	-	=	=	+	=!	=	=	=	=	=!	=	=
METALLIC	1.77	1.13	14.9	17.7	=	=!	=	-	=	=	=	=	-	=!	=	=!	=	=
BUTTER	1.73	1.52	16.8	17.3	=	=	-	=	=!	+	+	=	=!	=	=	=	=	=
NUTS	1.63	1.65	16.3	16.5	=!	=!	=	+	=	-	+	=	=	=	-	=	=	=
STRAW	1.62	1.05	29.0	21.5	-	=!	=	=	=	=!	=	=	=	=!	=	=!	=	-
CARDBOAR	1.45	1.30	8.07	13.8	=!	+	=	=	-	=!	+	=	=	=	+	=	=!	=
TOASTEDB	1.43	1.06	8.71	15.8	=	=	=	=	=	=!	=	=	=!	=!	=	=	=	=
MUSHROOM	1.38	1.29	12.0	14.9	=!	=!	=	-	=	=!	-	=	=!	=	-	=	=	=
ACID	1.34	1.39	14.7	15.8	=!	=!	=	=!	=	+	=	=	=	=	=!	=!	=	=
UMAMI	0.99	1.08	14.8	17.4	=!	=!	=	+	=	=!	=	=	=	=	=	=!	=	=
rankf	.	.	.	.	5.57	6.13	6.50	6.67	6.73	7.03	7.10	7.37	7.60	7.90	8.23	8.87	9.20	10.10

Figure 4 : Table CAP des données des figures 2 et 3. FPROD et FINTER sont les statistiques F de test des effets produit et produit*sujet. MEAN est la moyenne générale. RMSE est l'écart-type résiduel du modèle d'ANOVA employé au niveau du panel.

Les 14 dernières colonnes de cette table sont les sujets ordonnés du plus discriminant au moins discriminant sur l'ensemble des descripteurs. Les descripteurs (lignes du tableau) sont

également ordonnés du plus discriminant au moins discriminant. Pour chaque couple de sujet et de descripteur, un diagnostic est posé :

- « + » (vert) : discrimine les produits en accord avec le groupe
- « - » (rouge) : discrimine les produits, mais en désaccord avec le groupe
- « = » (orange) : ne discrimine pas les produits

De plus, à chacun de ces trois diagnostics de base peut éventuellement s'ajouter un « ! » (avec une couleur de la case plus foncée) si la répétabilité du sujet était significativement moins bonne que celle du groupe.

Les outils statistiques employés pour poser ces diagnostics sont simples :

- ANOVA à un facteur produit par sujet pour la discrimination ($p=0.10$)
- Test unilatéral à droite de nullité du coefficient de corrélation de Kendall entre le vecteur des moyennes produit d'un sujet discriminant avec le vecteur correspondant calculé pour la moyenne des sujets discriminants ($p=0.20$)
- ANOVA à deux facteurs produit et sujet sur les écart-types des répétitions, calculés pour chaque paire associant un produit à un sujet. Puis, test de contraste unilatéral gauche d'égalité de la moyenne de chaque sujet à celle des autres. Ceci pour le diagnostic de manque de répétabilité ($p=0.001$)

Les valeurs données entre parenthèses sont les niveaux du risque α auxquels ces tests sont réalisés par défaut.

Dans notre exemple de pommes de terre, cette table résume 30 ANOVA au niveau du panel, 420 ANOVA individuelles, plus de 100 coefficients de corrélation de Kendall et 420 tests de contraste sur les écart-types des répétitions. Evidemment, le risque global que cette table comporte des diagnostics erronés est proche de 1. Mais l'important est qu'elle n'en comporte pas trop. Il faut comprendre cet outil comme une aide à l'animateur de panel pour évaluer les performances de son jury à un instant donné afin de pouvoir réorienter l'entraînement de tout ou partie du panel.

Comme pour la table FLASH, l'idée forte qui a gouverné la conception de cette table est la recherche d'une synthèse lisible à partir de laquelle on puisse rapidement prendre des décisions.

Les performances multidimensionnelles

Le profil sensoriel est intrinsèquement un objet multidimensionnel. La liste de descripteurs n'est pas une fin en soi, mais juste un moyen pour approcher l'espace perceptif qui est multidimensionnel. C'est pourquoi, lorsque l'on restitue l'analyse des différences entre produits, une ou plusieurs cartes sensorielles sont nécessaires en complément des ANOVA par descripteur. Pour le contrôle des performances des sujets, on a pu penser qu'il en était autrement, car l'entraînement du panel se fait descripteur par descripteur. Néanmoins, même

si la littérature ne met pas cet aspect en avant, il reste intéressant de définir des notions de performances multidimensionnelles, aussi bien du sujet que du panel. Cet aspect vaut surtout pour la discrimination et pour l'accord avec le groupe, pour lesquels le passage au multidimensionnel induit de nouveaux concepts. Nous avons essayé de mettre cela en avant à différentes occasions [80, 90, 165, 166] et surtout en encadrant le stage de Delphine Brajon [187] qui a conduit à un outil SAS[®] dédié à cette problématique.

En termes de discrimination, on distinguera l'intensité des différences entre produits de la dimensionnalité de l'espace générant ces différences. Le contexte de la MANOVA et de l'analyse en variables canoniques nous offre respectivement les statistiques de tests permettant de quantifier ces deux aspects. En réalisant ces analyses par sujet, on pourra donc comparer les sujets d'un même panel aussi bien sur l'importance de la discrimination des produits que sur la complexité (dimensionnalité) de cette discrimination. On peut bien entendu utiliser les mêmes outils pour comparer des panels ayant profilé les mêmes produits.

En termes d'accord avec le groupe, on a recommandé l'utilisation du coefficient RV pour évaluer l'accord de chaque sujet avec le panel quant à la configuration multidimensionnelle - «espace sensoriel» - des produits et quant à la structure de corrélation - «associations perceptives ou cognitives» - des descripteurs.

1.6. Le projet Sensobase

Origine du projet

Les différentes écoles de sensométrie, et nous avec, ont chacune passé de nombreuses années à développer et promouvoir leurs propres méthodes sans toujours comparer les méthodes entre elles. Des comparaisons théoriques ont bien entendu été données dans les écrits, mais les comparaisons pratiques se limitent souvent à un ou deux jeux de données bien connus et pas du tout représentatifs de la variété des études d'analyse sensorielle. En arrivant au LIRIS en 2001, nous avons déjà fait ce constat et avons pour projet de rassembler un grand nombre d'études dans une base de données, afin de permettre la validation des mérites respectifs de chaque méthode basée sur un grand nombre d'études.

La seconde motivation était aussi de pouvoir mener des analyses sur un très grand nombre de cas réels, afin de pouvoir calibrer les performances des panels en fonction des types de produits et de descripteurs et afin de répondre à des questions méthodologiques telles que :

- L'entraînement permet-il vraiment d'améliorer les performances des sujets ?
- Les performances dépendent-elles des méthodes d'évaluation sensorielle ?
- Le sexe et l'âge ont-ils une influence sur les performances des sujets ?

Enfin, à terme, cette base pourrait devenir un centre de documentation des pratiques de l'évaluation sensorielle. A notre connaissance, une telle base n'existe nulle part ailleurs.

Même les grands groupes industriels ont du mal à constituer une telle base entre leurs différentes implantations, trop occupés qu'ils sont par des programmes à court terme. Le LIRIS nous paraissait être l'endroit idéal pour développer une telle base, car si celle-ci devenait conséquente, elle permettrait d'identifier clairement notre laboratoire comme un centre de ressources de l'évaluation sensorielle, ce qui serait fortement bénéfique à l'image du CESC et du pôle sensoriel dijonnais.

Ainsi, le LIRIS s'est engagé dès 2002 dans la construction de cette base de données d'analyse sensorielle en rassemblant d'abord tous les jeux de données existants au CESC et à l'INRA de Dijon.

A cette même époque, l'Association des Centres Techniques des Industries Agroalimentaires (ACTIA) me contactait pour les accompagner dans une réflexion sur l'évaluation et le suivi dans le temps des performances des panels situés dans les différents centres techniques (environ une dizaine) pratiquant l'analyse sensorielle en France. Je vis là une opportunité unique de faire décoller notre base de données en lançant, avec le soutien de la Région Bourgogne, une thèse visant à définir les méthodes statistiques de suivi des performances dans le temps. Le versant pratique de cette thèse consisterait à organiser l'acquisition, le stockage et la gestion des données des centres ACTIA dans une base de données relationnelle. Et évidemment, cette base ne serait pas réservée uniquement aux données de l'ACTIA. Elle resterait la propriété du LIRIS qui pourrait l'abonder avec des données de tout autre partenaire. Le projet Sensobase put donc véritablement démarrer grâce au lancement de la thèse de Nicolas Pineau [181].

Dans un premier temps, les centres ACTIA avaient un intérêt immédiat à abonder la base, car une convention entre l'ACTIA et le CNRS stipulait que pour chaque jeu de données apporté à la base, le LIRIS retournerait gratuitement au centre technique un « état des performances » des sujets.

Etat des performances pour chaque jeu de données déposé dans Sensobase

Après une première année de thèse, il a été décidé que cet état des performances serait composé des éléments suivants agencés sous la forme d'un petit site internet composé des pages HTML suivantes :

- La table CAP
- La carte CAP. Une nouveauté développée par Nicolas Pineau [86, 181] pour représenter les performances individuelles sous la forme d'une carte d'ACP interactive
- Une table de statistiques de performances de discrimination multidimensionnelle
- Une table de statistiques d'accord multidimensionnel avec le groupe
- La table FLASH pour l'analyse des différences entre produits
- Une carte d'analyse en variables canoniques des produits

La référence [86] est annexée à ce mémoire (article 13). On y trouvera notamment un exemple de la carte CAP.

Mes anciens programmes SAS[®] ont été repris par Delphine Brajon et Nicolas Pineau, principalement afin d'améliorer la présentation des résultats à l'aide des outils informatiques récents (résultats en couleurs, utilisation de pages html). Quelques améliorations ont également été apportées au niveau des calculs statistiques afin de prendre en compte des situations plus diverses. Par exemple, les études comportant des sujets n'ayant pas effectué de répétitions peuvent être prises en compte dans la table CAP.

De manière capitale, cette restitution des analyses est accompagnée d'une notice explicative (aide à l'interprétation) destinée aux utilisateurs non statisticiens qui a été rédigée par Nicolas Pineau au cours de sa thèse. L'ensemble de ce système est réuni dans une macro SAS[®] appelée PERF qui permet à l'utilisateur de régler les paramètres de chaque analyse (choix des analyses, risques des différents tests, options graphiques...).

En fonction des remarques des utilisateurs et des développements méthodologiques nouveaux, cet outil sera amené à évoluer dans les années à venir afin de simplifier davantage la lecture de certains résultats ou afin d'apporter des informations complémentaires grâce à d'autres analyses.

Le fonctionnement du projet Sensobase

Le projet Sensobase s'articule autour d'un site Internet³ qui permet aux fournisseurs de données de s'inscrire à ce programme d'échange, d'envoyer leurs études de profil sensoriel, puis de demander des analyses statistiques. Afin de faciliter le stockage dans une structure unique, les données relatives à chaque étude doivent être saisies dans un fichier Excel[®] spécifiquement formaté afin de recevoir ce type de données. Ce fichier demande à l'utilisateur de fournir divers renseignements sur l'étude sensorielle (âge et sexe des panélistes, type de descripteurs...) puis de saisir les données à proprement parler selon une structure précise. Le fichier est ensuite envoyé via le site Internet et stocké dans une base de données. Suite à l'insertion des données dans la base, l'utilisateur peut effectuer une demande d'analyse de ce jeu de données dans laquelle certains paramètres des analyses sont réglables. L'utilisateur reçoit ultérieurement les résultats par courrier électronique au format html. Un schéma du fonctionnement de ce système est présenté dans la figure ci-dessous.

³ <https://liris.cesg.cnrs.fr>

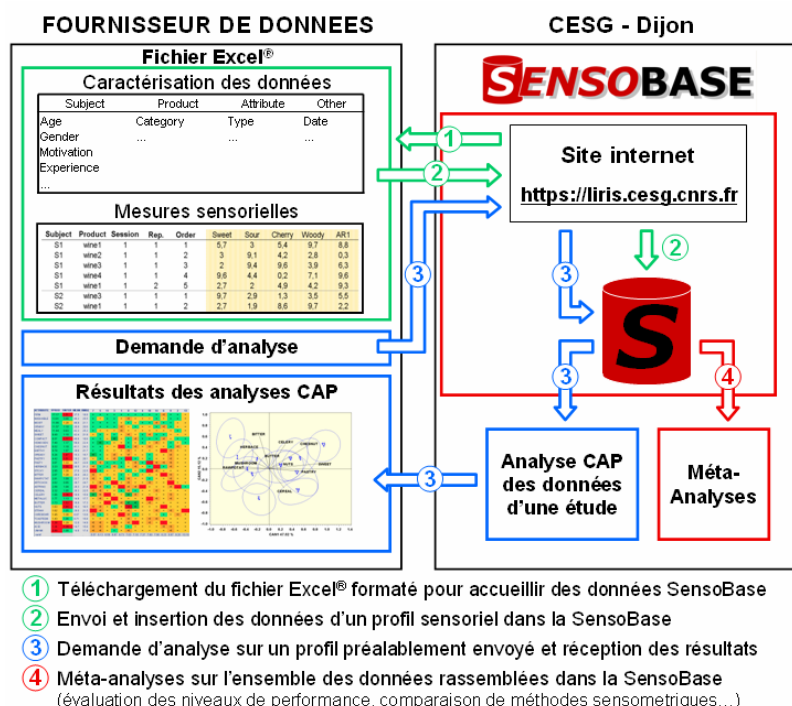


Figure 5 : Schéma synoptique du fonctionnement du projet Sensobase

L'architecture de la base de données a été développée initialement par Sylvie Cordelle, puis implémentée en MySQL en utilisant le logiciel phpMyAdmin par Delphine Brajon au cours de son contrat au LIRIS. Les programmes de vérification des données ont été développés par Delphine Brajon, puis par Julien Delaborde, stagiaire informatique encadré par Nicolas Pineau au cours de sa thèse.

Le site web de la Sensobase³ est l'interface dédié aux utilisateurs. Il a été majoritairement développé par Delphine Brajon, puis mis à jour par Nicolas Pineau au cours de sa thèse. Ce site est constitué d'une partie accessible à tous les visiteurs et d'une partie réservée aux utilisateurs inscrits au programme Sensobase. La partie en libre accès présente le projet, le déroulement de l'envoi des données, la restitution des analyses, un extrait de la notice explicative des résultats statistiques et le formulaire d'inscription. Le site est disponible en français et en anglais. La partie réservée aux utilisateurs est accessible à l'aide d'un identifiant et d'un mot de passe fournis lors de l'inscription. L'utilisateur peut alors télécharger le fichier Excel® permettant d'envoyer des données de profil sensoriel et de demander une analyse sur une étude déjà envoyée (ou sur une sélection des données de cette étude). Cette procédure d'insertion automatique des données était le sujet principal du stage de Licence2 de Julien Delaborde encadré par Nicolas Pineau.

Une fois la demande d'analyse effectuée, une procédure génère le code SAS qui devra être exécuté (appel de la macro PERF) pour réaliser les analyses statistiques correspondantes et l'administrateur de la Sensobase reçoit un courrier lui indiquant qu'une demande vient d'être effectuée. Ce dernier doit alors manuellement exécuter le code SAS et vérifier que les

résultats obtenus sont corrects. Les analyses, produites au format HTML, doivent ensuite être compressées (format ZIP) et envoyées manuellement par courrier électronique au fournisseur. Cette dernière étape de traitement des données et d'envoi des analyses n'est pas encore automatisée car il existe de nombreux cas pour lesquels certaines analyses ne peuvent pas être exécutées (absence de répétitions, plan de présentation trop déséquilibré...). Il faut alors expliquer à l'utilisateur la raison de ce problème. La grande diversité des cas à prévoir est donc un frein à la réalisation de cette étape. L'objectif du laboratoire est cependant d'automatiser complètement le processus d'échange de données contre des analyses statistiques dans les mois à venir.

Aujourd'hui, le délai moyen d'attente entre l'envoi d'une étude sensorielle et la réception des analyses est inférieur à une journée, pourvu qu'il y ait au laboratoire une personne disponible et compétente pour traiter la demande. Depuis la fin de la thèse de Nicolas, il n'y a plus que Sylvie Cordelle et moi-même qui pouvons traiter ces demandes.

L'état actuel de la Sensobase

Jusqu'à présent, le système proposé a permis de rassembler plus de 500 études sensorielles provenant d'une quarantaine de laboratoires d'analyse sensorielle situés dans 8 pays différents. Cela représente approximativement 3000 produits, 8000 descripteurs, 1500 panélistes, soit plus de trois millions de scores d'intensité sensorielle.

Parmi ces jeux de données, 135 proviennent des centres ACTIA. Il est donc évident que le projet a su dépasser son cadre d'origine. Nous continuons à recevoir plusieurs jeux de données par semaine, malgré une absence de publicité volontaire autour de ce projet, pour cause d'un manque de personnel en ce moment pour le poursuivre et le développer.

Le suivi des performances dans le temps

En parallèle de la création de la Sensobase et du choix des analyses pour chaque jeu de données, le sujet méthodologique de la thèse de Nicolas Pineau était de définir des méthodes de suivi de l'évolution de la performance d'un panel et de ses panélistes au fil des séances d'évaluation réalisées sur un même univers produit avec les mêmes descripteurs.

Nicolas a basé sa démarche sur l'utilisation du modèle mixte en analyse de la variance dans lequel le sujet est vu comme un effet aléatoire dont les mesures de performances sont répétées au fil du temps. Cette méthode tient compte simultanément des effets panel et panéliste dans un même modèle. Cette approche est complétée par l'utilisation de cartes de contrôle, utilisant les paramètres définis dans le modèle mixte, afin d'identifier les temps de mesure auxquels le panel ou un panéliste s'éloigne significativement de la modélisation retenue.

Cette méthodologie a été mise en œuvre pour une demi-douzaine d'études longitudinales. La thèse [181] détaille les résultats de deux de ces études. La première, fournie par le centre

technique AERIAL, consistait en 8 profils sensoriels de plats préparés congelés réalisés entre 1997 et 2001, à intervalle de six mois, avec la participation de 9 à 11 panélistes parmi un ensemble de 13 dégustateurs. La seconde, fournie par le centre de recherche Unilever de Vlaardingen, consistait en 43 profils de sauces soja répartis sur une année, avec la participation de 7 à 12 panélistes parmi un ensemble de 13 dégustateurs.

La présentation de cette méthodologie et son application à l'étude AERIAL vient juste d'être publiée [52] et est annexé à ce mémoire (article 14).

Méta-analyse de la Sensobase

Pour l'étude des niveaux de performance et l'étude de plusieurs facteurs pouvant influencer ces niveaux de performances, des méta-analyses ont été effectuées sur tout ou une partie des jeux de données de la Sensobase. L'objectif est de définir des niveaux moyens de performance dans le domaine de l'analyse sensorielle pour plusieurs indices et d'étudier l'effet de plusieurs variables sur ces niveaux. Nous avons étudié l'effet de cinq variables « intra-études » (âge, sexe, niveau scolaire, niveau d'expérience, motivation, type de descripteur), dont les niveaux varient à l'intérieur de chaque expérimentation d'un sujet à l'autre ou d'un descripteur à l'autre, et l'effet de cinq variables « inter-études » (fournisseur, type de produits, type d'échelle, pays d'origine, laboratoire public ou privé), dont le niveau est constant à l'intérieur de chaque étude.

Concernant les variables intra-étude, les résultats montrent que le niveau d'accord et de discrimination augmente avec le niveau d'expérience en analyse sensorielle. Le niveau de discrimination s'élève également avec le niveau scolaire, mais ce dernier ne modifie pas les niveaux d'accord et de répétabilité. L'âge agit significativement sur l'indice de discrimination uniquement, mettant en avant des performances plus élevées pour les personnes entre 30 et 45 ans. L'effet du type de descripteur est significatif pour tous les indices de performance. La répétabilité observée est meilleure pour les descripteurs de saveurs, d'arômes et d'odeurs, mais ce résultat pourrait être lié à l'utilisation d'une portion plus restreinte de l'échelle de notation pour ces descripteurs. Les niveaux d'accord et de discrimination sont beaucoup plus élevés pour les descripteurs visuels. Le niveau d'accord est également meilleur pour les descripteurs de texture que pour les termes d'arôme, de saveur et d'odeur, mais la discrimination de la texture est relativement faible. L'effet sexe n'est jamais significatif parmi les panélistes sélectionnés pour participer à des profils sensoriels, mais les hommes auraient tendance à utiliser une portion de l'échelle plus importante que les femmes, lors de la notation de descripteurs visuels uniquement.

L'analyse des variables inter-études montre que la source majeure de variabilité entre les performances des études réunies dans la SensoBase provient de l'effet « fournisseur », hautement significatif ($p < 0,001$) quel que soit l'indice de performance considéré. Cette variabilité de la performance est en outre plus importante au regard de l'indice de répétabilité.

Pour les quatre autres variables étudiées (type d'échelle, laboratoire privé ou public, pays et type de produit), seul le niveau de répétabilité est significativement modifié selon les catégories de chaque facteur. L'effet est plus marqué concernant le facteur pays et montre un niveau de répétabilité plus faible des fournisseurs français. Puisque cet écart de performance n'est pas significatif concernant le niveau de discrimination, cela semble indiquer que l'utilisation de l'échelle est différente entre les fournisseurs français et étrangers. Une approche différente de l'entraînement des panélistes pourrait alors être en cause. Les différences observées sur le niveau de répétabilité pour les variables « type d'échelle » et « laboratoire privé ou public » sont moins flagrantes. Elles pourraient en outre être liées à la corrélation entre ces facteurs et les variables « fournisseurs » et « pays ». Les niveaux d'accord et de discrimination ne sont pas modifiés en fonction des variables inter-études.

Un premier article de présentation de ces résultats est soumis pour publication. En fonction de l'enclin des fournisseurs à envoyer des données toujours plus complètes, nous espérons à l'avenir pouvoir étudier l'effet d'autres variables, comme la fréquence de consommation du produit de l'étude, la motivation des panélistes (curiosité, occupation du temps libre ou rémunération), le fait qu'ils soient fumeurs ou non... Afin d'améliorer la portée de ces résultats, en particulier concernant les variables inter-études, nous devons à présent diversifier l'origine des données et créer des partenariats avec un plus grand nombre de fournisseurs différents. Il serait en effet intéressant de pouvoir considérer le fournisseur en tant qu'unité expérimentale de nos analyses.

Utilisation de la Sensobase pour l'étude de l' « Assessor Model » de Brockhoff

L'analyse de la variance des données de profil sensoriel suppose que les sujets sont homogènes quant à la dispersion de leurs notes sur l'échelle de notation, autrement dit qu'ils utilisent tous une portion de l'échelle plus ou moins de même taille. On désigne, à plus ou moins juste titre, cette dispersion par le mot « scaling ». Pour de nombreuses raisons, cette hypothèse a peu de chance d'être vraie dans la réalité. Brockhoff a proposé l' « Assessor Model » (Brockhoff and Skovgaard, 1994) qui prend en compte explicitement cet aspect en rajoutant un terme individuel multiplicatif du paramètre produit dans le modèle d'ANOVA à deux facteurs produit et sujet sans interaction. Ce modèle permet également d'accommoder des répétabilité différentes entre sujets. Il permet en outre de tester dans quelle mesure le panel est hétérogène vis à vis de ces deux aspects : scaling et répétabilité. Même si Brockhoff a montré plus tard la réalité de ces effets (Brockhoff, 1998) en utilisant son modèle sur les cafés de l'ESN dont on a déjà parlé, il reste que l' « Assessor Model » n'est pas utilisé par la communauté de l'analyse sensorielle. La raison principale de ce désintérêt tient certainement à l'absence dans ce cadre d'un véritable test statistique de différences entre les produits.

Ainsi, nous avons encadré en 2006, le stage d'ingénieur ENSAI de Bérengère Fournier [195] sur ce sujet. Il s'agissait d'abord d'évaluer l'importance de l'hétérogénéité des panels d'analyse sensorielle en matière de répétabilité et de scaling. En utilisant 264 jeux de données

issus de la Sensobase, nous avons pu établir [176] que cette hétérogénéité existait de manière significative dans respectivement 73 et 76 % des cas. Ce résultat pourrait remettre en cause des décennies d'utilisation de l'analyse de la variance avec les données sensorielles !

Néanmoins, nous avons voulu ensuite étudier dans quelle mesure une transformation préalable des données visant à retirer l'effet de scaling conduisait à une conclusion différente quant à l'existence de différences entre produits. Toujours sur la base de ces 264 études, nous n'avons trouvé que 0,3% de descripteurs qui devenaient significatifs après transformation et que 1,5% d'autres descripteurs qui perdaient leur significativité après transformation. On peut donc avancer que la réponse en termes de « oui/non » à la question de l'existence de différences entre produits n'est globalement pas modifiée par la prise en compte de l'hétérogénéité des sujets en termes de scaling. Par contre, nous avons aussi calculé des coefficients de corrélation entre les vecteurs des effets produit avec et sans transformation et il est apparu que dans 28 % des cas cette corrélation n'était pas significativement non nulle. Ainsi, on peut avancer que l'hétérogénéité des sujets en matière de scaling modifie l'ordonnement des produits une fois sur quatre.

Plutôt que de transformer les données a priori, Brockhoff (communication personnelle) propose de « rassembler » son modèle et le modèle classique dans un seul modèle où le terme multiplicatif issu de l' « Assessor Model » apparaît finalement comme une covariable (l'effet produit de l' « Assessor Model » étant la covariable). Néanmoins, cela nécessite d'ajuster d'abord, par maximum de vraisemblance, l' « Assessor Model ». Dans la pratique, cette solution complique beaucoup les choses. Sur une idée développée conjointement avec Per Brockhoff lors du dernier *Sensometrics Meeting* de 2006 en Norvège, nous avons testé dans quelle mesure on pourrait remplacer la covariable issue de l' « Assessor Model » par la simple moyenne des produits. L'intérêt étant que ce nouveau modèle (appelé CAM, comme « Covariance Assessor Model ») peut s'ajuster avec une procédure classique de modèle linéaire, comme le GLM de SAS®. Toujours sur la base des 264 études tirées de la Sensobase, nous avons pu établir que si le modèle classique ne présentait que 59,36 % de descripteurs significatifs pour l'effet produit, le nouveau modèle de Brockhoff en comportait 68,46% et le nouveau modèle simplifié (CAM) 63,26 %. Ces chiffres sont éloquentes. Ils montrent d'abord un gain net de significativité (+ 9,10 %) grâce à la sophistication du modèle. Rappelons que par transformation préalable des données, nous n'avons pas réussi à mettre en évidence de tels effets. Ils montrent ensuite que la concession du modèle CAM fait perdre 5,20 % par rapport au modèle le plus sophistiqué et en gagne tout de même 3,90 % par rapport au modèle classique.

Face à l'ensemble de ces résultats, que recommanderons-nous aux praticiens quant à l'utilisation de l'analyse de la variance pour traiter leurs données ? C'est la question essentielle que l'on est en train de traiter dans la rédaction d'un article sur ce sujet... qui a pris

un peu de retard par rapport à ces travaux. Néanmoins, cet article sera réalisé et peut-être enrichi des nouveaux jeux de données arrivés depuis dans la Sensobase.

Chapitre 2. Concevoir des nouvelles données sensorielles

2.1. Décrire les sensations trigéminales, noter leur intensité et relever leur succession

Dans le cadre d'un contrat liant un consortium de moutardiers européens au LIRIS, un premier travail de caractérisation des sensations trigéminales (piquant, chaleur, brûlure, frais,...) provoquées par des modèles irritants tels que la moutarde et les épices a été initié par Ep Köster et réalisé par Françoise Chauvin sous l'encadrement de Patrick Etiévant, Sylvie Issanchou et moi-même, juste avant que je ne prenne la direction du LIRIS. Il a permis de distinguer 36 sensations trigéminales différentes, chacune localisée en une ou plusieurs des 30 zones anatomiques différenciées par les sujets. Un programme informatique a été mis au point pour capturer de manière dynamique au fil de la dégustation la nature des sensations perçues, leur localisation et leur intensité (figure 5). Cet outil, couplé à l'analyse multidimensionnelle a permis de dresser une carte sensorielle de 37 moutardes fournies par un consortium de moutardiers européens partenaires du programme. Ce même outil appliqué à d'autres modèles de stimulus irritants (épices) a permis de montrer que la perception trigéminal est différente d'un agent irritant à l'autre aussi bien en terme de qualité perçue qu'en terme de localisation des sensations [148].

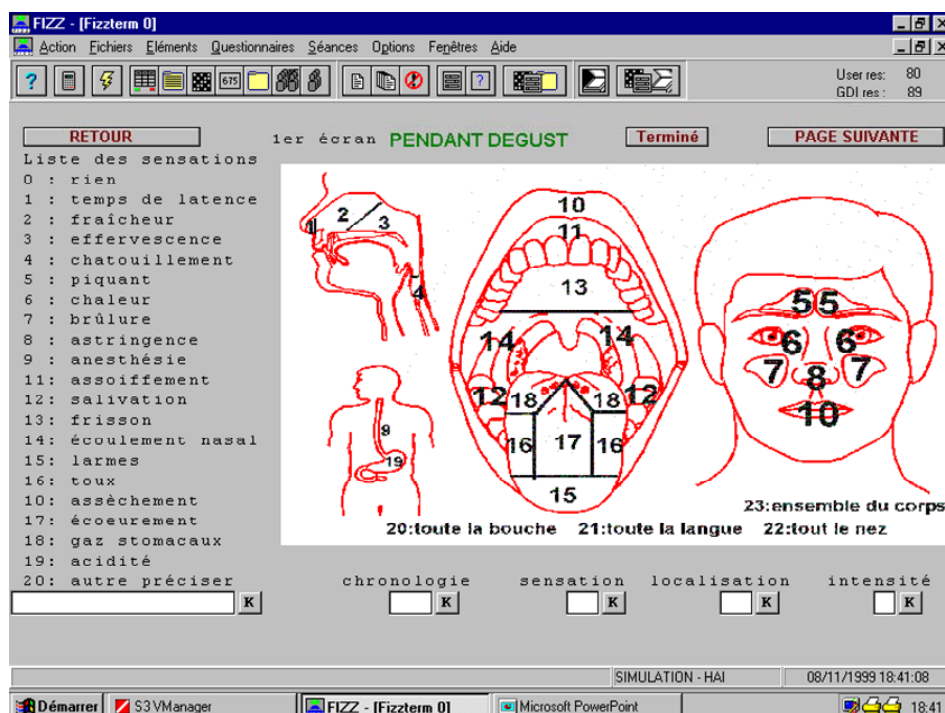


Figure 6 : Ecran du module Fizz dédié à l'acquisition des données sensorielles trigéminales

2.2. Mesurer le « confort digestif » et le « masquage du goût »

Il s'agit d'une ancienne collaboration avec la société Amora-Maille qui remonte à la création du CSG. Notre partenaire désirait comprendre pourquoi certaines personnes apprécient beaucoup les condiments forts, tels que la moutarde, alors que d'autres les rejettent radicalement. Dans la mesure où les non-consommateurs de moutarde invoquaient une nuisance digestive et un masquage du goût des aliments dus à la consommation de moutarde, notre partenaire industriel désirait confirmer ou infirmer ces deux allégations.

Une première étude menée au LIRIS 2001-2002 avait permis de réfuter la nuisance digestive de la moutarde et de suggérer même que ses consommateurs réguliers auraient une perception plus courte de la durée de digestion lorsque le repas est accompagné de moutarde. Ce résultat n'ayant été obtenu que de manière déclarative, nous avons convenu avec notre partenaire qu'il devait être validé par des mesures physiologiques qui auraient été réalisées par l'équipe « Métabolisme » du CSG. Mais le rachat de la société Amora-Maille par le groupe Unilever-BestFood-France (UBF) nous a placés en face de nouveaux interlocuteurs qui ne semblaient pas intéressés par cette problématique.

Toutefois, une dernière expérimentation que nous avons déjà contractualisée avec Amora-Maille a permis d'étudier la réalité de la seconde allégation relative au masquage du goût des aliments par la moutarde. Nous avons fait confectionner des saucisses qui différaient entre elles selon un plan d'expérience contrôlant le niveau de gras et les arômes fumé et épicé. Une centaine de sujets ont dégusté ces produits au cours de 4 repas qui pouvaient être ou ne pas être accompagnés de moutarde. Les sujets ont noté leurs préférences pour ces produits et leur perception de la sensation de gras et des arômes. Bien que juste au seuil de significativité, d'intéressants effets de la moutarde ont été détectés :

- ✓ une augmentation des scores de préférence donnés aux saucisses,
- ✓ une perception accrue du caractère épicé et poivré des saucisses,
- ✓ un masquage de l'arôme fumé,
- ✓ une perception accrue du gras et de la jutosité pour certaines saucisses estimées trop peu grasses et trop peu juteuses,
- ✓ une perception diminuée du gras et de la jutosité pour certaines saucisses estimées trop grasses et trop juteuses,
- ✓ enfin et surtout, ces effets sont amplifiés chez les consommateurs réguliers de moutarde.

Ces derniers résultats avaient vocation à être publiés, mais la fin de la collaboration avec Amora-Maille depuis son rachat par UBF a provoqué le départ des deux ingénieurs contractuels du LIRIS qui menaient ces travaux et ainsi la rédaction de ces publications a été délaissée et a depuis perdu son caractère prioritaire.

Une autre étude importante a été menée en 2002-2003 pour UBF sur l'amélioration de la qualité gustative et aromatique des vinaigrettes industrielles. Cette étude a comparé la perception sensorielle des vinaigrettes « maison » » versus « industrielle » en prenant en compte la connaissance ou non de cette origine et en étudiant différents groupes de consommateurs différant selon leurs habitudes de consommation de chacun des deux types de produits. Les résultats ont donné lieu à de nouvelles gammes de produits, mais seront difficilement publiables sans nuire aux clauses de confidentialité de cette étude. Il avait été discuté d'étendre cette étude à différents pays européens, ce qui aurait pu donner lieu à des comparaisons entre cultures dont les résultats auraient, eux, eu vocation à être publiés. Néanmoins, ce projet n'a jamais vu le jour sans que l'on ait vraiment su pourquoi.

En conclusion, notre collaboration avec le groupe Amora-Maille s'est terminée d'une manière inattendue et plutôt désagréable et ne nous a pas laissé l'occasion de valoriser nos travaux. La leçon que le LIRIS a tiré de cette expérience est qu'il vaut mieux définir ses propres thèmes de recherche et trouver ensuite des partenaires industriels, plutôt que d'adapter de manière systématique sa thématique à la demande industrielle. L'évolution du LIRIS d'une structure de service vers une équipe de recherche est en accord avec ce constat et nous pensons que les choix de thématiques et de programmes que nous avons faits ensuite répondent à cette analyse critique.

2.3. Les limites de la méthode de Temps-Intensité

La méthode de temps-intensité (TI) consiste à noter l'intensité perçue d'une sensation au fil du temps que dure la dégustation d'une bouchée (ou d'une gorgée) du produit. Pratiquement, le sujet déplace de manière continue un curseur au prorata de l'intensité perçue et le résultat est une courbe dont les abscisses sont les temps et les ordonnées les intensités.

L'analyse statistique des données de TI est un chapitre à part entière de la sensométrie et on pourra trouver une bibliographie complète sur ce sujet dans la thèse de Stéphanie Le Dauphin (Dauphin, 2007) récemment soutenue à l'Université de Nantes et que nous avons eu la chance d'examiner.

Lors de notre séjour sabbatique à l'Université de Californie (Davis), nous avons travaillé avec Ann Noble sur des données de temps-intensité qu'elle avait accumulées afin de mieux comprendre la perception de l'amertume et de l'astringence du vin, critères souvent cités comme responsables de possibles rejets du consommateur. La perception de l'amertume est clairement temporelle, elle se développe en général après un certain temps de latence et devient problématique lorsqu'elle tend à persister en fin de bouche. La persistance de l'astringence d'un vin est tout autant problématique. Beaucoup d'études visent à déterminer comment on peut « faire passer au second plan de la perception » ces sensations. Augmenter la douceur (sucre) du vin et/ou son degré alcoolique aura-t-il une influence sur la perception

de l'amertume et de l'astringence ? En guise d'exemple de ce type d'étude, nous avons annexé à ce mémoire (article 6) un travail [38] mené lors de notre séjour à Davis.

C'est ainsi que la méthode TI se trouve bien souvent utilisée successivement avec plusieurs descripteurs pour caractériser la perception d'un même ensemble de produits. Nous pensons qu'elle est mal adaptée à ce type de situation où plusieurs descripteurs doivent être étudiés simultanément, pour la simple raison qu'elle ne permet pas d'évaluer ces descripteurs simultanément. D'ailleurs, à l'origine la méthode TI a été développée pour étudier un simple descripteur (perception de la sensation sucrée en fonction de différents édulcorants). Fort de cette constatation, nous n'étions que plus intéressé par la méthode du paragraphe suivant qui était en cours de développement lors de notre arrivée au LIRIS.

2.4. La méthode de la Dominance Temporelle des Sensations (DTS)

Origine de la méthode

Comme le profil descriptif d'un aliment comporte souvent 20 à 40 descripteurs pertinents, et comme l'intensité de chacun de ces descripteurs évolue avec le temps, il est fort improbable que notre attention se focalise sur l'ensemble de ces changements lors de la mastication et après la déglutition. Il est plus probable que seul un petit nombre de sensations puisse capter notre attention à un instant donné, et que la nature de ces sensations change avec le temps. Il y aurait donc une dominance temporelle des sensations perçues au cours de l'ingestion d'un aliment.

L'objectif de ce travail était de mettre en évidence ce phénomène, de développer une méthodologie permettant de le mesurer, de comprendre la relation existant entre cette dominance et le suivi de l'intensité au cours du temps (TI) des principales caractéristiques sensorielles, et d'estimer son importance sur les préférences des consommateurs.

L'origine de cette idée revient à Ep Köster, premier directeur du LIRIS. Le travail a été réalisé dans le cadre d'un contrat tripartite (CNRS/Groupe Danone/Fromageries Bel) négocié en 2000 par Patrick Etiévant pour une durée de 3 ans. En arrivant au LIRIS en 2001, j'ai donc pris en route la responsabilité du déroulement de ce contrat qui comportait 3 phases, chacune censée durer une année. La première devait définir une méthode d'acquisition des données temporelles. La seconde devait comparer les résultats de cette nouvelle méthode avec ceux obtenus par le profil classique et le TI. Enfin, la troisième devait relier ces données temporelles aux préférences des consommateurs

Phase 1. Choix d'une méthode d'acquisition des données

Deux systèmes d'acquisition des données ont été comparés. Le premier prototype a été développé par la société Biosystèmes, à partir du logiciel Fizz. Ce système permet aux sujets

de noter les sensations dominantes perçues au fur et à mesure de la mastication, avec un rafraîchissement d'écran selon une fréquence programmable. L'écran présente 10 échelles pour 10 descripteurs différents. Il est donc possible de relever à la fois les sensations dominantes au cours du temps, et d'en estimer l'intensité. Pratiquement, après la mise en bouche du produit, il est demandé au sujet de choisir le descripteur qu'il perçoit comme dominant et d'en indiquer son intensité en cliquant sur son échelle. Puis, le sujet doit réitérer ce processus jusqu'à l'extinction de sa perception. Il peut donc citer plusieurs fois le même descripteur à des temps différents. Le même descripteur peut aussi être cité plusieurs fois consécutivement s'il reste dominant mais que son intensité change. Certains descripteurs peuvent n'être jamais cités par un sujet. Pour chaque citation, le programme conserve 3 données : le temps auquel cette citation est réalisée, le nom du descripteur cité et son intensité.

Le second prototype permet de collecter les mêmes informations, mais avec une évaluation d'intensité basée non pas sur l'estimation de l'intensité par une grandeur, mais sur l'égalisation avec une autre perception sensorielle : l'écartement des doigts. Le prototype, relié à un PC et commandé par le logiciel Fizz comporte 10 pistes. Le sujet place son pouce dans une saignée du boîtier et se déplace vers la piste correspondant au descripteur à noter. Il place ensuite son index sur la piste et l'écarte du pouce jusqu'à ressentir une tension dans la main qu'il juge équivalente à la perception de la sensation dominante à noter. Un test préalable sur 8 sujets a permis de vérifier l'aptitude des sujets à utiliser la main droite ou la main gauche pour la notation d'intensité. Le résultat de ce test a orienté le développement du prototype vers deux boîtiers de cinq voies chacun. Ces boîtiers sont reliés à un boîtier d'affichage qui permet au sujet de vérifier que la position de ses doigts correspond bien au descripteur qu'il veut évaluer.

Afin de comparer les deux méthodes, deux études ont été mises en place chacune avec un panel de 20 sujets différents. Le modèle de la première était constitué de 8 bières blanches fournies par le Groupe Danone, alors que les Fromageries Bel ont fourni 6 fromages fondus pour la seconde. Pour chaque étude, une liste de 10 descripteurs a été retenue de manière consensuelle par chaque panel, suite à plusieurs séances de familiarisation avec les produits. Quatre séances de mesures ont été réalisées avec chacune des deux méthodes de mesure sur les produits de l'étude.

Pour chaque sensation, 4 variables ont été définies à partir des données obtenues : le temps d'apparition de la sensation, la durée pendant laquelle elle persiste, sa durée de dominance et son intensité maximale pondérée par sa durée. Des MANOVA ont été effectuées pour comprendre la part des différentes variables dans la différenciation des produits et des analyses de correspondances pour étudier les séquences de sensations qui permettent de caractériser les produits. Une méthode d'analyse de la répétabilité et du pouvoir discriminant des sujets a été mise au point. Pour finir, plusieurs types de représentations graphiques des séquences des sensations au sein des produits de l'étude ont été réalisés.

Le traitement des données a montré qu'il existait très peu de différences entre les résultats obtenus avec les deux systèmes. Du coup, il a été décidé de poursuivre l'étude en utilisant le système d'acquisition basé sur la saisie à partir d'un écran de PC car ce système était beaucoup plus pratique à déployer pour un panel composé de 8 à 16 sujets.

Phase 2. Comparaison avec l'approche du profil classique et du TI

L'objectif de cette seconde phase était de mieux comprendre la nature du DTS en le comparant aux profils conventionnels de l'analyse sensorielle et aux mesures de TI. L'hypothèse à tester était la suivante : notre esprit est attiré, lors de la mastication, par les sensations émergentes ou décroissantes à un instant donné. En effet, le cerveau est peu sensible aux stimuli constants, mais amplifie les signaux nouveaux pouvant alerter l'individu sur une modification de son environnement. Ainsi, il devrait exister une relation plus ou moins simple entre le profil DTS et certains paramètres des profils TI.

Comme pour la phase 1, deux études ont été réalisées. La première portait sur 8 laits fermentés fournis par le Groupe Danone et la seconde sur 8 fromages fournis par les Fromageries Bel. Cette fois-ci, les études ont été réalisées chez les partenaires industriels et au CESC. La méthode d'acquisition des données DTS a donc été transférée aux partenaires.

Les données de DTS ont été traitées de la même façon qu'en phase 1. La répétabilité et le pouvoir discriminant des sujets ont été évalués. La répétabilité obtenue est plus élevée que celle obtenue en première phase, mais reste encore très décevante pour les laits fermentés, ce qui montre la difficulté des sujets à réaliser des mesures de DTS sur des produits liquides. L'origine de cette difficulté réside dans la durée de séjour en bouche du produit, très courte par rapport à celle de produits solides. De nouvelles représentations graphiques de la séquentialité des sensations ont été proposées. A la demande des partenaires industriels, l'une de ces représentations (sous forme de camemberts) a été retenue comme la plus adaptée. Des macros ont été développées sous Excel pour permettre la réalisation de ces graphiques en camemberts de façon automatique à partir du fichier des données brutes. Ces macros permettent également d'extraire les 4 variables définies en phase 1, mais aussi de déterminer la répétabilité des sujets. Ces macros ont été distribuées aux partenaires industriels pour leur faciliter le traitement de leurs données DTS.

Enfin, des analyses statistiques ont été effectuées pour étudier la relation entre les mesures de DTS et celles du profil classique et du TI. En ce qui concerne la comparaison des mesures de DTS et des mesures de profil, on constate que les intensités mesurées avec les deux méthodes sont très proches. Toutefois, chaque méthode doit être utilisée en fonction des objectifs de l'étude : le profil permet de décrire de manière précise chaque caractéristique d'un produit alors que les mesures DTS ne portent que sur les sensations dominantes. L'étude des relations entre les mesures de DTS et de TI a permis de mettre en évidence l'importance de la prise en compte de l'aspect temporel dans la dominance des sensations. Les durées des sensations

mesurées en TI et les durées de persistance mesurées en DTS sont corrélées, de même que les intensités notées avec les deux méthodes.

L'ensemble des résultats montre que les données de DTS permettent d'obtenir autant d'information sur les produits que les données TI et que les données de profil classique. La dominance d'une sensation résulte à la fois de son intensité et de ses notions temporelles telles que son temps d'apparition et sa durée de persistance.

Phase 3. Liaison aux préférences des consommateurs

L'ensemble des expérimentations de cette phase a été mené par les partenaires industriels. Ils ont effectué du profil classique, du DTS et des tests consommateurs avec 12 fromages bicouche (Bel) et 9 biscuits de type génoises enrobées de chocolat et fourrées à l'orange (Danone). Pour les fromages, les données de profil et de DTS ont été obtenues avec 17 sujets et 10 descripteurs et les données de préférences avec 119 consommateurs. Pour les biscuits, les données de profil classique et de DTS ont été obtenues respectivement avec 10 et 12 sujets pour respectivement 37 et 10 descripteurs et les données de préférence avec 123 consommateurs.

Le codage des données DTS a été légèrement modifié pour ne plus conserver que 3 paramètres : « Durée » (durée de dominance totale du descripteur considéré), « Temps » (premier temps d'apparition du descripteur) et « Score » (intensité moyenne sur l'ensemble des citations pondérées par leur durée). Les méthodes d'analyse de cartographie des préférences ont alors été adaptées à ce codage. Elles ont permis de mettre en évidence l'importance de la temporalité des produits dans les préférences des consommateurs dans la mesure où certaines différences entre les produits au niveau des préférences ne pouvaient s'expliquer que par des caractéristiques sensorielles temporelles.

La représentation graphique des données retenue en phase 2 (camemberts), n'étant plus considérée comme satisfaisante par les partenaires du projet, elle a dû être étudiée à nouveau. Cela a donné lieu à l'élaboration de courbes de cumul du nombre de citations en fonction de temps, plus lisibles que les représentations précédentes. Ce mode de représentation des données DTS sous forme de courbe de fréquence de citation est toujours celui utilisé à l'heure actuelle.

Cette troisième phase a été réalisée au LIRIS par Nicolas Pineau dans le cadre de son stage de fin d'étude ENITIAA [184]. De part leur confidentialité, les résultats de ces études en terme de produits n'ont pas vocation à être publiés. Par contre, un article méthodologique [51] sur le DTS, cosigné par l'ensemble des partenaires de ce projet, a été soumis et est en cours de révision. Nous l'avons ajouté à la liste des articles annexés à ce mémoire (article 15).

Perspectives ouverte par ce projet

La méthode DTS ayant été présentée dans plusieurs congrès [79, 81, 155], elle a suscité beaucoup d'intérêt de la part des laboratoires d'analyse sensorielle. Depuis, la méthode a été testée par plusieurs d'entre eux avec ou sans notre collaboration. En guise d'exemple, Nestlé et PSA vont chacun présenter au prochain *Pangborn Sensory Science Symposium* une application du DTS que l'on a réalisée en collaboration.

Le projet avait pour objectif de se focaliser sur le fait que, lors de l'évaluation sensorielle d'un produit, le dégustateur ne se concentre probablement que sur quelques sensations, voire sur une seule sensation, à un instant donné. C'est la détermination de cette sensation, dominante à un instant donné, qui a toujours été le cœur du problème. L'approche méthodologique utilisée, qui consiste à demander à chaque instant quel est le descripteur qui domine, semble donc très bien adaptée au problème soulevé par ce projet, puisqu'à chaque instant, l'information recueillie est censée représenter ce qui caractérise le plus le produit.

En revanche, bien que cette méthode donne de bons résultats, quelques points pourraient être discutés. Par exemple, on peut penser que les sujets ont tendance à apprendre une sorte de schéma de réponse lors des séances, qui conduit certains à noter systématiquement un grand nombre de descripteurs alors que d'autres, à l'inverse, en notent peu. Ce genre de considérations laisse donc la porte ouverte à d'autres études ayant comme objectif de récupérer le même type d'informations, mais en contraignant les sujets à répondre de manière plus instinctive.

Enfin, les données de préférences collectées dans le cadre de ce projet n'étaient jamais temporelles. Nous pensons qu'il serait intéressant de réfléchir à des systèmes d'acquisition de données hédoniques dynamiques ; c'est-à-dire des protocoles où les sujets noteraient leurs préférences continûment au fil de la dégustation d'un produit

2.5. La méthode « Instinctive Temps-Intensité-Profil » (ITIP)

Afin d'empêcher tout apprentissage d'une séquentialité qui serait ensuite restituée de manière forcée par le sujet pour tous les produits de l'étude, nous avons, dans le cadre du DEA de Nicolas Pineau [186] et du stage de maîtrise de Laure Pelletier [185], développé, testé et comparé au DTS une nouvelle méthode temporelle appelée « Instinctif Temps-Intensité-Profil » (ITIP).

Après la mise en bouche du produit, ITIP consiste à demander au sujet de noter l'intensité d'un descripteur à des intervalles de temps très rapprochés, de l'ordre de 5 secondes. L'astuce consiste à faire varier la nature de ce descripteur au cours du temps, tout en permettant, par le jeu des répétitions, à chaque descripteur d'être évalué à différents temps répartis de manière uniforme le long de la durée de la perception. De manière pragmatique, le sujet met en

bouche, puis reçoit sur l'écran de son ordinateur une échelle d'intensité à évaluer toutes les 5 secondes, sans jamais savoir quel descripteur sera le suivant. Il est possible d'introduire un descripteur d'appréciation hédonique parmi l'ensemble des descripteurs évalués.

En utilisant comme modèle des chewing-gums comportant plusieurs arômes encapsulés (chewing-gums produits spécialement pour notre étude par la société Symrise), nous avons comparé les résultats d'un groupe de sujets réalisant le DTS, avec un groupe réalisant le ITIP et avec un troisième groupe réalisant le ITIP pour lequel un descripteur d'appréciation hédonique globale était ajouté aux descripteurs sensoriels. Les résultats de ce dernier groupe n'étaient pas moins bons que ceux du groupe ITIP. Par contre, la répétabilité de cette nouvelle méthode ITIP ne semblait pas acquise et, du moins sur la base de cette première étude, elle ne pouvait prétendre à remplacer la méthode DTS. Néanmoins, notre codage des données ITIP aurait peut-être pu être amélioré. Nous n'avons jamais eu l'occasion de reprendre ce travail...

2.6. Le DTS appliqué au vin

Bien que l'on ait vu que le DTS semble plus délicat à réaliser sur les aliments liquides que sur les solides, nous avons tout de même choisi de poursuivre nos travaux sur le DTS avec le modèle vin. En effet, la temporalité des sensations est reconnue comme capitale par les experts du vin sans pour autant qu'ils possèdent une méthode objective pour la mesurer. La perception gustative d'un vin peut se décomposer en au moins 5 composantes qualitatives différentes : acidité, douceur, amertume, astringence et chaleur. La perception de chacune de ces composantes est temporelle dans le sens où son intensité évolue au fil du temps que dure la perception après mise en bouche. Nous faisons l'hypothèse que la notion de « qualité du vin » dépend davantage de la combinaison de ces temporalités plutôt que de l'intensité globale ou maximale que peut atteindre chacune de ces composantes. Toutefois, nous pensons que cette notion de qualité dépend du degré d'expertise du dégustateur.

Une thèse (Rita Pessina) en cotutelle entre l'Université de Foggia (Italie) et l'Université de Bourgogne a donc démarré sur ce sujet en 2003 au LIRIS et a été soutenue en 2006 [180]. Une première expérimentation, utilisant plusieurs versions d'un même vin rouge Italien différant entre elles par l'augmentation de l'intensité d'une seule saveur à la fois, nous a permis d'établir que la méthode DTS fonctionnait bien pour décrire la temporalité de la perception gustative des vins. Une seconde expérimentation, utilisant cette fois-ci des vins blanc Italiens « manipulés » seulement sur l'arôme, a permis d'établir que la méthode fonctionnait également pour la description temporelle aromatique du vin et que cette évaluation pouvait être menée simultanément à l'évaluation gustative.

Une troisième étude avec des vins commerciaux a eu lieu en 2005 dans le cadre du stage d'ingénieur de Denis Goffin [193]. Elle visait à établir l'existence d'un lien entre la temporalité des sensations, les préférences des consommateurs et la qualité du vin, du moins

celle présumée par le marché et exprimée par le prix du produit. Trois gammes de 3 vins ont été choisies chez 3 producteurs : l'un en Bourgogne, l'autre en Campanie (Italie) et le troisième en Australie du sud. Un panel entraîné au DTS, un panel entraîné au profil sensoriel classique, un panel de professionnels Bourguignons et deux panels de consommateurs l'un à Dijon et l'autre à Paris ont évalué ces 9 vins de manière descriptive et hédonique. Cette étude a été financée à 100% grâce à un contrat de recherche avec le Groupe Pernod Ricard.

Un premier résultat de cette expérimentation fut un accueil favorable de la méthode DTS par le panel de professionnels qui a réussi à l'utiliser sans entraînement. Les courbes DTS de ce panel étaient très similaires à celles obtenues avec notre panel entraîné au DTS. Ceci ouvre des opportunités à la méthode DTS hors des laboratoires classiques d'analyse sensorielle.

Un second résultat fut que les préférences des trois types de sujets (sujets entraînés à l'analyse sensorielle, professionnels du vin et consommateurs) étaient différentes entre elles, mais ne reproduisaient pas la gamme de qualité annoncée par les prix des produits. Enfin, pour certaines gammes de vins et certains panels des liaisons entre DTS et préférence sont apparues, mais pas de manière suffisamment systématique pour que l'on puisse vraiment prétendre à une réelle causalité. Des investigations plus fines de ces données doivent encore être menées avant de conclure sur cette étude.

Chapitre 3. Développer les méthodes au sein de grands projets stratégiques

3.1. Justification

Suite à notre passage de l'INRA au CESP, notre activité s'est orientée d'une part du développement en sensométrie vers la Sensobase (chapitre 1 de ce mémoire) et, d'autre part, de l'analyse de données sensorielles classiques vers le développement de méthodes d'acquisition et de traitement de nouveaux types de données sensorielles (chapitre 2 de ce mémoire).

Pour mener de telles recherches méthodologiques, il faut des modèles. On peut les choisir soit seul de manière totalement autonome, soit en fonction des opportunités des demandes industrielles. Nous avons toujours plutôt privilégié la seconde option. Néanmoins, elle présente aussi des risques. Tout d'abord, il n'est pas facile d'obtenir a priori des contrats de longue durée avec l'industrie ; souvent, on enchaîne des périodes de 6 mois, ce qui n'est pas facile à gérer. En effet, il devient alors difficile dans ces conditions de garder un cap méthodologique clair. La diversité des modèles et des questions d'intérêt peut nuire à une continuité de la recherche méthodologique. Enfin, les problèmes de confidentialité sont toujours délicats dans les petits contrats qui n'ont pas fait l'objet d'une réflexion approfondie de cette question avant de démarrer la recherche.

Nous pensons qu'il faut privilégier les collaborations d'envergure associant plusieurs partenaires académiques et un ou plusieurs partenaires industriels. On s'inscrit alors dans des projets qui ont un statut reconnu par la communauté scientifique. Il pourra paraître plus difficile d'y insérer sa recherche méthodologique, mais on aura l'assurance de contribuer ainsi au progrès dans un autre domaine de la connaissance que celui de la méthodologie.

C'est dans cet état d'esprit que nous avons récemment apporté le concours de notre groupe aux deux projets décrits ci-dessous. L'un est soutenu par l'ANR, l'autre a été initié par le pôle de compétitivité Vitagora⁴ « Goût, Nutrition, Santé ». Ce sont là des bonnes garanties pour inscrire sa recherche méthodologique dans un cadre qui garantit déjà son application à un domaine d'intérêt général.

⁴ <http://www.vitagora.com/>

3.2. Le projet VDQA

La teneur en alcool des vins tend à augmenter au fil des progrès de la viticulture et de l'œnologie. Le réchauffement climatique participe également à cette tendance. Cette dérive est préjudiciable à différents points de vue. Tout d'abord, une consommation régulière de vin à table au cours du repas, tout en maintenant une consommation modérée d'alcool, est d'autant plus difficile que le vin consommé est fort en alcool. D'autre part, dans un contexte de crise de la commercialisation du vin Français, il apparaît opportun d'identifier de nouveaux marchés et de produire des vins adaptés à cette demande. Un vin à teneur réduite en alcool – entre 8 et 12° au lieu de 12 à 16° - pourrait séduire différentes catégories de nouveaux consommateurs de vins : jeunes adultes, femmes, consommateurs issus de pays non producteurs de vin... Le marché international du vin propose déjà un certain nombre de références à teneur réduite en alcool. Toutefois, on a très peu de données sur l'acceptabilité de ces produits. Certains consommateurs traditionnels pourraient avoir tendance à rejeter a priori l'idée d'un vin « allégé » en alcool, en invoquant une infériorité gustative et aromatique par rapport aux vins traditionnels, sans que l'on sache vraiment quelles sont les différences de perception sensorielle du vin en fonction de sa teneur en alcool.

Dans ce contexte, le projet VDQA, dont on pourra trouver une description en ligne⁵, ambitionne de développer une recherche intégrée pour « construire » des vins de qualité à teneur réduite en alcool. Il associe des équipes de recherche aux compétences complémentaires, l'institut technique de la filière, des fédérations de producteurs, des industriels et l'interprofession des distilleries. Il a été sélectionné en 2005 par l'ANR dans le cadre du Programme National de Recherche en Alimentation (PNRA). Il comporte trois volets :

1. étudier les conditions de développement du marché des vins à teneur réduite en alcool dans les sociétés modernes
2. étudier le rôle de la teneur en alcool sur la perception chimio-sensorielle et sur l'appréciation de la qualité par le consommateur
3. valider en vraie grandeur, avec les professionnels, un procédé de désalcoolisation partielle performant et économiquement acceptable.

Le LIRIS est chargé de la coordination du second volet qui associe l'UMR FLAVIC, Pernod Ricard, l'Unité Viticole Expérimentale de l'INRA de Narbonne et le LIRIS.

Dans ce cadre, nous avons lancé depuis le 1^{er} novembre 2006 avec le Groupe Pernod Ricard une thèse CIFRE réalisée par Sophie Meillon au LIRIS et encadrée par le Professeur Marc Danzart et moi-même. Elle cherchera à élucider en quoi les équilibres gustatifs et aromatiques du vin sont modifiés par la réduction de sa teneur en alcool.. Elle s'attachera ensuite à évaluer

⁵ http://www.cesg.cnrs.fr/DocumentsPDF/VDQA_Com_ANR.pdf

l'impact de ces modifications sur l'acceptabilité du produit par différents types de consommateurs. Cette connaissance devrait alors permettre d'envisager des stratégies de compensation sensorielle pour l'élaboration de vins de qualité à teneurs réduites en alcool. Enfin, il s'agira de valider l'acceptabilité des nouveaux produits à teneur réduite en alcool par des consommateurs en situation de consommation réelle (domicile, restaurant...), c'est-à-dire aussi face à une information sur cette réduction d'alcool.

D'un point de vue méthodologique, ce projet était propice à un certain nombre de développements qui nous ont conduits à accepter d'y participer :

- Apport du DTS pour évaluer les modifications de l'équilibre gustatif et aromatique des vins par réduction de leur teneur en alcool.
- Comparaison de tests hédonique en laboratoire (au CESG) et à domicile (méthodologie propre à Pernod Ricard).
- Développement d'une mesure dynamique de la perception des consommateurs. Comment adapter le DTS afin qu'il soit réalisable par des consommateurs sans entraînement ?
- Développement de mesures de la complexité perçue par le consommateur en reprenant d'anciens travaux qui sont exposés dans le dernier chapitre de ce mémoire à propos du projet EduSens.
- Adaptation de méthodes issues de la micro-économie expérimentale pour étudier l'arbitrage entre le goût et l'image du produit pour l'acceptabilité et les choix du consommateur en matière de vins à teneur réduite en alcool.

En résumé, il s'agit d'abord d'une thèse orientée vers le produit (le vin) avec une problématique d'actualité associée à ce produit (la réduction du taux d'alcool) qui s'intègre dans un vaste projet de recherche validé par l'ANR. Mais il s'agit aussi d'une thèse qui conduira à réaliser du développement méthodologique en analyse sensorielle. Ces nouvelles méthodes appelleront forcément de l'innovation dans le traitement des données, c'est-à-dire de la recherche en sensométrie.

3.3. Le projet SensEaux

En 2004, la consommation mondiale d'eaux embouteillées s'élevait à 154 milliards de litres, soit 57% de plus qu'en 1999. De plus, la demande en eaux embouteillées augmente même dans les régions où l'eau du robinet est de bonne qualité sanitaire, induisant de grandes quantités de débris inutiles et une grande consommation d'énergie (production des bouteilles, transport).

Ainsi, à l'heure actuelle, seuls 50% des français déclarent consommer de l'eau du robinet à table. Un foyer de 4 personnes ne consommant jamais d'eau du robinet dépensera plus de 500 € par an dans l'achat d'eaux embouteillées. S'il consommait l'eau du robinet, il pourrait allouer ce budget à un autre usage, peut-être plus écologique et probablement plus dynamique

d'un point de vue économique. Lyonnaise des Eaux (LdE), en accord avec sa politique volontariste en matière de développement durable, a décidé de réhabiliter l'image de l'eau du robinet dans l'esprit des consommateurs en tant qu'eau de table. Pour cela, ce groupe industriel a initié en 2005, en s'appuyant sur la dynamique naissante du pôle de compétitivité Bourguignon Vitagora « Goût, Nutrition, Santé », le programme de recherche SensEaux⁶ dont le slogan est « L'eau du robinet, une eau agréable à boire »

Le principal motif de non consommation évoqué est celui du goût de chlore. Réduire et/ou masquer ce goût de chlore est donc une étape nécessaire, mais pas suffisante, pour faire accepter l'eau du robinet comme eau de table. Il apparaît donc nécessaire de mieux connaître les mécanismes de perception du chlore et leur interaction avec la minéralisation de l'eau. Ce sujet sera celui d'une première thèse CIFRE entre LdE et l'UMR FLAVIC de l'INRA de Dijon. Ce travail associe également la société Robertet, chargée de produire des agents potentiellement masquant du goût de chlore.

La seconde thèse CIFRE avec LdE est réalisée au LIRIS par Eric Teillet (depuis le 1^{er} octobre 2006) sous l'encadrement d'Elisabeth Guichard et de moi-même. Cette thèse se concentrera sur la mesure du goût de l'eau et sur la modélisation des différences de goût occasionnées par des différences de minéralisation. Il s'agit véritablement d'établir la carte sensorielle des eaux plates (minérales, comme de source) du marché français et d'y projeter les eaux du réseau de distribution en échantillonnant la variabilité spatiale et temporelle des goûts de ces dernières.

De plus, cette thèse étudiera les préférences des consommateurs vis-à-vis de ces différents goûts de l'eau et notamment l'influence de ces préférences sur la décision finale de consommer ou non l'eau du robinet. Préfère-t-on les eaux auxquelles on a été davantage exposé de part sa consommation ou bien les préférences sont-elles davantage dictées par d'autres facteurs à élucider ? Enfin, même si le goût de l'eau du robinet est apprécié par le consommateur lors d'une dégustation à l'aveugle, rien ne dit qu'il continuera de l'apprécier lorsqu'il la tirera lui-même du robinet. Identifier les verrous qui bloquent cette consommation, proposer, en collaboration étroite avec le département de marketing de la LdE, des stratégies de communication visant à débloquent ces verrous et concevoir des expérimentations visant à démontrer l'efficacité de ces stratégies est l'étape terminale de ce projet.

Mise à part mon adhésion personnelle à l'aspect citoyen de ce projet, les raisons qui nous ont conduits à accepter d'y participer sont d'ordre politique, historique et méthodologique.

Politique, parce qu'en 2005 lors du montage du pôle de compétitivité Vitagora, il était judicieux d'identifier des futurs projets de collaboration entre le tissu industriel local et le CESC. Parmi les différentes discussions que nous avons pu avoir à l'époque avec les

⁶ http://www.cesg.cnrs.fr/DocumentsPDF/SensEaux_Com_LIRIS_Fr.pdf

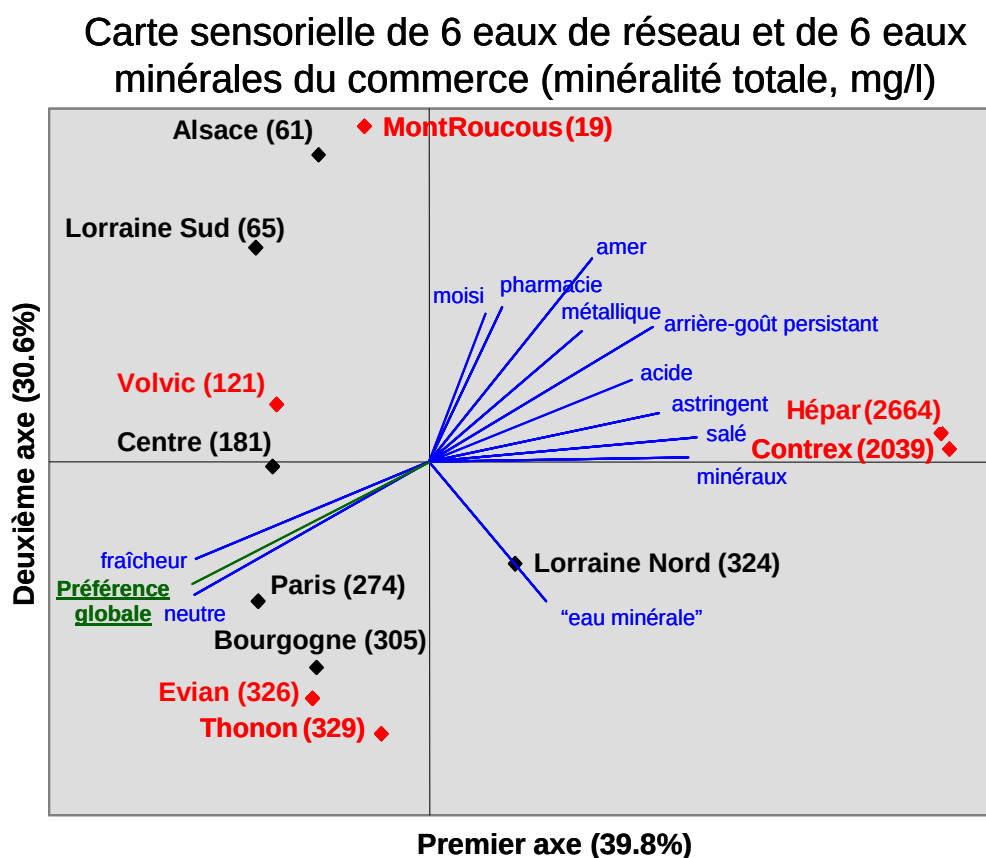
différentes filières produits (vin, fromage, arts culinaires...) du pôle en construction, celles avec la filière eau, représentée exclusivement par LdE, étaient beaucoup plus consistantes d'un point de vue scientifique.

Historique, parce nous avons déjà travaillé sur l'eau au début des années 1990 à l'INRA de Dijon. C'est précisément avec le modèle eau que nous avons introduit en France la méthode du profil libre-choix. Nous n'avions pu publier que des résultats portant sur les eaux gazeuses [12], mais nous en avons d'autres sur les eaux plates dans nos tiroirs. J'étais très intéressé de confronter ces cartes sensorielles, obtenues il y a plus de 15 années avec le profil libre-choix, avec celles que nous allions obtenir au cours de ce projet avec d'autres méthodes. Déjà à l'époque, nous avons mis en évidence un continuum de goûts d'eaux allant des « eaux douces » (peu minéralisées) au goût métallique et piquant (exemple Volvic) aux « eaux lourdes » au goût terreux et salé (exemple Contrexéville) en passant par des eaux au goût plus neutre (exemple Evian). Par contre, nous ne savions rien de la répartition des préférences des consommateurs entre ces différents goûts d'eau.

Méthodologique, parce que nous savions par expérience que l'analyse sensorielle de l'eau est un véritable challenge. On entend dire couramment qu'une bonne eau est une eau sans goût. Décrire le goût de quelque chose qui ne doit pas en avoir est assurément un challenge. Nous savions avant de démarrer le projet qu'il faudrait innover sur le plan méthodologique pour réellement réussir à décrire les goûts de l'eau. Enfin, nous n'étions pas du tout persuadés de la véracité de cette assertion sur la bonne eau qui ne doit pas avoir de goût. Comme pour beaucoup d'autres produits alimentaires, l'existence d'une segmentation des préférences est possible. Rechercher cette segmentation avec un produit de consommation courante, comme l'eau (sans jeu de mots), conduirait nécessairement à des interrogations stimulantes sur la prise en compte des habitudes de consommation.

Afin de clore ce chapitre avec quelques résultats, la figure 7 présente celui d'une expérimentation menée par le LIRIS dans le cadre de la Convention d'Encadrement de Lyonnaise des Eaux à la Cité des Sciences et de L'industrie de la Villette en octobre 2006. Ce fut l'occasion de voir 389 cadres LdE déguster à l'aveugle des eaux issues aussi bien du commerce que du réseau de distribution français de la LdE. Ils ont réalisé d'abord une épreuve de tri libre qui a consisté à classer les 12 eaux en groupes d'eaux à goût similaires, le choix du nombre de groupes étant laissé libre à chaque sujet. Ces données s'analysent en MDS et permettent d'obtenir la carte des eaux de la figure 7 où les eaux sont d'autant plus proches qu'elles ont souvent été jugées similaires en goût. Puis, les sujets devaient décrire chacun des groupes par le choix d'items dans une liste qui leur était proposée. Enfin, ils devaient déclarer pour chaque groupe d'eaux leur niveau de préférence, avec la possibilité de ne pas répondre ou de répondre que le groupe était hétérogène en termes d'appréciation hédoniques des eaux. Les items descriptifs et la préférence globale sont représentés dans la carte par une flèche pointant dans la direction d'intensité ou de préférence croissante. Les

coordonnées de ces flèches sont obtenues par le calcul de coefficients de corrélation entre la fréquence de citation et les coordonnées des eaux dans la carte.



Etablie en octobre 2006 par un panel de 389 consommateurs ayant effectué une épreuve de tri libre, suivie d'une description sensorielle et hédonique

Figure 7 : Résultat de l'expérimentation SensEaux de La Villette en octobre 2005

On a pu déjà conclure grâce à cette carte que :

- La carte sensorielle des eaux obtenue au début des années 1990 à l'INRA de Dijon était quasi-identique à celle-ci, validant l'existence des trois grands types de goûts d'eau évoquée précédemment.
- Cette carte sensorielle est fortement structurée par le taux de résidu sec. Les eaux à faible minéralité et les eaux à forte minéralité sont celles qui ont le plus de goût, plutôt amer pour les premières et plutôt salé pour les secondes.
- Les eaux de réseau (une fois le chlore retiré) décrivent à peu près le même espace sensoriel que celui des eaux minérales du commerce.
- Globalement, plus la minéralité d'une eau est moyenne (environ 300 à 500 mg/L), plus l'eau est appréciée en raison de son aspect neutre et rafraîchissant.

Ainsi, l'adage qui disait « une bonne eau est une eau sans goût » semble vérifié, du moins au niveau de la moyenne du groupe. L'enjeu est maintenant d'étudier dans quelle mesure il existe ou non une segmentation des préférences derrière cette préférence globale pour les eaux de minéralité équilibrée.

La technique du tri libre, dont l'utilisation est relativement récente en analyse sensorielle, n'avait encore jamais été utilisée avec autant de sujets. Capitalisant sur ce grand nombre, Eric Teillet a mené des simulations visant à établir la robustesse de la carte sensorielle en fonction du nombre de sujets. Ces résultats sont en cours de rédaction.

D'autre part, nous avons en 2006 recruté et entraîné des panels au profil sensoriel de l'eau et il s'est effectivement avéré que l'exercice était très difficile. Ainsi, nous sommes en train de développer, toujours dans le cadre de la thèse d'Eric Teillet, une nouvelle méthode visant à positionner les eaux par rapport à des références fixées qui seront 3 eaux du commerce représentant chacune l'un de ces trois groupes de goût bien identifié. L'intérêt de la méthode est qu'elle permettrait d'agréger des données d'évaluation sensorielle obtenues lors d'expérimentations différentes et avec des sujets de types différents (consommateurs et experts). Le traitement de ces données d'un type nouveau nécessitera également un développement spécifique en sensométrie.

Chapitre 4. Structurer la recherche sur les préférences et les comportements alimentaires des enfants

4.1. La Fête de la Science 2002 et 2004

L'acceptabilité sensorielle des aliments est traditionnellement mesurée à l'aide d'échelles hédoniques sans que l'on sache vraiment quelle est la fiabilité de telles mesures, surtout lorsqu'elles sont effectuées avec des jeunes enfants. En ligne avec notre objectif d'améliorer les méthodes de l'analyse sensorielle et parce que le projet de construction d'un observatoire des préférences nous conduira à beaucoup travailler avec des enfants, nous avons décidé de lancer des recherches pour étudier la fiabilité de ces mesures.

Pour étudier cet aspect, nous avons utilisé en 2002 la Fête de la Science pour faire venir au LIRIS un grand nombre de sujets de tous âges pour effectuer les mêmes tests sensoriels, notamment en accueillant des classes scolaires et en leur offrant à l'issue de la dégustation une présentation didactique sur le goût et une analyse en temps réel de leur propres résultats de dégustation. Une première expérimentation sur le jus d'orange, nous a permis d'étudier la répétabilité intra-séance des notes hédoniques avec 917 sujets âgés de 10 à 80 ans (article 11 annexé à ce mémoire, [45]) et de comparer les préférences pour les différents saveurs présentes dans le jus d'orange en fonction de l'âge des sujets. Les résultats obtenus ont ensuite été validés par une seconde étude des mêmes produits avec 315 nouveaux sujets en étendant les âges étudiés aux enfants de 7 à 10 ans et en comparant un échantillon de sujets français et allemands [47]. Parmi les résultats les plus marquants de cette étude figurent d'une part le fait que l'amertume est moins rejetée par les seniors, alors que l'acidité est davantage acceptée par les enfants, du moins dans le jus d'orange, et d'autre part que la consistance des notes de préférence ne dépend pas de l'âge et du sexe du sujet, même si les enfants et les seniors ont ensuite beaucoup plus de difficultés à décrire les raisons de leurs préférences.

Les bons résultats obtenus sur le jus d'orange nous ont conduit à relancer une opération Fête de la Science en 2004 qui a réuni 619 consommateurs âgés de 6 à 69 ans qui ont dégusté 8 yaourts à la fraise différant entre eux selon leur niveau de sucre, d'acidité et selon l'épaisseur de leur texture. Parmi les résultats les plus intéressants figure le fait que l'ajout de sucre augmente moins les préférences des enfants que celles des adultes. La méthode DTS a également été appliquée à ces produits avec un panel de sujets entraînés. La publication de ces résultats a pris du retard, car à l'époque c'était le projet vin qui était devenu prioritaire.

Nous envisageons de relancer une opération Fête de la Science en octobre 2007 sur l'eau dans le cadre du projet SensEaux.

4.2. Le projet Opaline

L'origine du projet a été détaillée dans l'introduction de ce mémoire. Rappelons juste qu'il est le premier volet d'une opération plus large intitulée « Observatoire des préférences » qui a motivé, avec le projet de la SensoBase, ma venue au CESG en 2001.

L'acronyme OPALINE, que l'on doit à Vincent Boggio notre partenaire de la Faculté de Médecine de Dijon dans ce projet, signifie « Observatoire des Préférences Alimentaires du Nourrisson et de l'Enfant ». Il vise à comprendre l'origine de la variété des préférences alimentaires en observant l'évolution du goût de la naissance à l'âge de deux ans.

Sa justification se trouve dans les travaux de la thèse de Sophie Nicklaus, fondée sur une base de données de choix alimentaires d'enfants âgés de 2 à 4 ans accumulées par Vincent Boggio durant près de 20 ans, qui ont permis de mettre en évidence une grande variabilité des préférences alimentaires des enfants dès l'âge de 2 ans et d'établir un lien entre ces préférences précoces et le comportement alimentaire au cours de l'enfance et de l'adolescence. OPALINE cherche donc à déterminer l'origine de ces préférences précoces et à comprendre les mécanismes de leur évolution.

L'origine pourrait remonter à la période d'introduction des aliments solides chez le nourrisson, ou à la période d'allaitement maternel, voire aux dernières semaines de la vie in utero. Dans ces deux dernières situations, c'est l'alimentation de la mère qui est le vecteur de transmission des arômes et des saveurs. Il est donc essentiel de contrôler le profil aromatique et gustatif de cette alimentation maternelle et de le mettre en regard des réactions de préférence, ou de rejet, du nourrisson. L'un des enjeux essentiels est d'établir si une plus large variété alimentaire à l'étape de la diversification favorise l'acceptation d'aliments nouveaux et variés à l'âge de 2 ans. Mais il est aussi important de mesurer la réactivité sensorielle des nourrissons et des enfants à des stimulations, aussi bien olfactives que gustatives, car une plus grande sensibilité pourrait expliquer les aversions précoces. Enfin, le contexte familial joue certainement un rôle important dans l'évolution des préférences et aversions alimentaires des enfants, il est donc nécessaire de le caractériser au fil du temps.

L'ensemble de ces mesures est en cours de réalisation avec un premier groupe de 120 mères recrutées progressivement depuis janvier 2005 au 7^{ème} mois de leur grossesse. Ces femmes ont relevé la composition de leurs repas dans des semainiers dont l'analyse permet de dresser le profil aromatique et gustatif de leur alimentation au fil du temps.

La constitution d'une seconde cohorte de 180 mères a commencé début 2007 avec lesquelles en plus des mesures déjà décrites, nous effectuerons des analyses physico-chimiques et sensorielles des laits maternels des femmes allaitantes et des laits maternisés.

Alors que les études antérieures sur l'alimentation des nourrissons portaient surtout sur la composition des laits et aliments et sur l'établissement de recommandations nutritionnelles, OPALINE introduit une rupture originale en se focalisant sur la perception et l'acceptabilité des aliments par le nourrisson et l'enfant. Dans le contexte actuel de l'augmentation inquiétante de la prévalence de l'obésité infantile, la connaissance apportée par OPALINE pourrait favoriser des actions visant à un meilleur comportement alimentaire des enfants.

OPALINE rassemble une vingtaine de chercheurs de disciplines variées (évaluation sensorielle, biologie, médecine, psychologie, sociologie, éthologie et statistique) issus du Centre Européen des Sciences du Goût (Dijon), de l'Unité Flaveur-Vision-Consommateur (INRA-ENESAD, Dijon), de la Faculté de Médecine de Dijon, de l'Unité de Recherche sur la Consommation (INRA, Ivry), de l'Unité de Sociologie Rurale (INA-PG, Paris), de l'Unité de Psychologie de l'Université de Paris X, de l'Unité de Neurobiologie de l'Olfaction et de la Prise Alimentaire (INRA, Jouy-en-Josas) et du Centre d'Etude de Physiologie Appliquée (CNRS, Strasbourg).

OPALINE est financé par l'INRA et l'INSERM dans le cadre du Programme de Recherche en Nutrition Humaine (PRNH), par l'Agence Nationale de la Recherche dans le cadre du Programme National de Recherche en Alimentation (PNRA), par le Conseil Régional de Bourgogne et par l'Institut Fédératif de Recherche 92 « Qualité des Aliments ». De plus, des mécènes industriels (Blédina, CEDUS, Nestlé et Symrise) apportent un soutien financier régulier à Opaline. Enfin, Opaline est labellisé par le pôle de compétitivité Vitagora « Goût, Nutrition, Santé ».

OPALINE est donc un projet d'envergure que j'ai eu la chance d'initier et que je coordonne maintenant avec Sophie Nicklaus. Cette activité d'initiateur et de coordonateur a peu de choses à voir avec mon métier de statisticien et avec ma spécialisation en analyse sensorielle. Je ne peux donc pas, pour le moment, revendiquer une synergie quelconque de cette activité avec celles présentées jusqu'à maintenant dans ce mémoire. Mais, cette activité a été suffisamment chronophage pour expliquer les différents retards pris dans la publication des résultats des autres activités...

Christine Lange au LIRIS est responsable du volet suivi alimentaire de ce projet, alors que l'UMR FLAVIC prend en charge toutes les mesures de réactivité sensorielle. Le volet sociologique est réalisé grâce aux collaborations avec le CORELA, l'INA-PG et l'Université Paris X. La logistique du recrutement et de la gestion des cohortes est assurée par le LIRIS grâce au recrutement depuis fin 2004 de Caroline Laval sur différents postes en contrat à durée déterminée. Un poste d'ingénieur d'étude statutaire INRA est actuellement ouvert au concours externe pour pérenniser cette fonction au LIRIS.

OPALINE commence à donner lieu à la création de données de type nouveau, aussi bien pour les mesures de réactivité sensorielle que pour celles de suivi alimentaire. Claire Chabanet,

biométricienne de l'UMR FLAVIC, développe des méthodes de traitement pour ces données. Notre souhait serait de pouvoir s'impliquer davantage dans ces travaux à l'avenir. De nouvelles ouvertures pour la sensométrie pourraient en découler.

4.3. Le projet EduSens

Origine du projet

L'origine de ce projet remonte à 2003-2004, lorsqu'Adeline Thibaut [190] et Caroline Reverdy [189] effectuaient respectivement leur stage de DESS et de DEA au LIRIS sur la mesure de la complexité perçue des aliments et l'effet d'une éducation sensorielle auprès de jeunes adultes.

Avec Adeline Thibaut, nous avons développé un questionnaire de mesure de la complexité perçue [170]. Le principe de ce questionnaire consiste à décomposer la complexité selon plusieurs dimensions : nombre de composantes du produit, relations entre ces composantes entre elles et avec le produit, difficulté à décrire le produit et à citer ses composantes, familiarité et sophistication du produit. Ce questionnaire est composé de 12 échelles dont les extrémités sont illustrées par les images de la figure 8. L'idée est de faciliter la tâche du sujet en lui permettant d'aligner sa perception aromatique ou gustative sur ces supports visuels. Ce questionnaire a été appliqué à différents modèles et son analyse selon la méthode STATIS duale a permis de dégager 4 composantes du concept de complexité et d'étudier leurs relations [171]. Comme il a été indiqué lors de la description du projet VDQA, cet outil sera utilisé prochainement avec des consommateurs de vin pour étudier l'influence de la réduction d'alcool sur la perception sensorielle du produit.

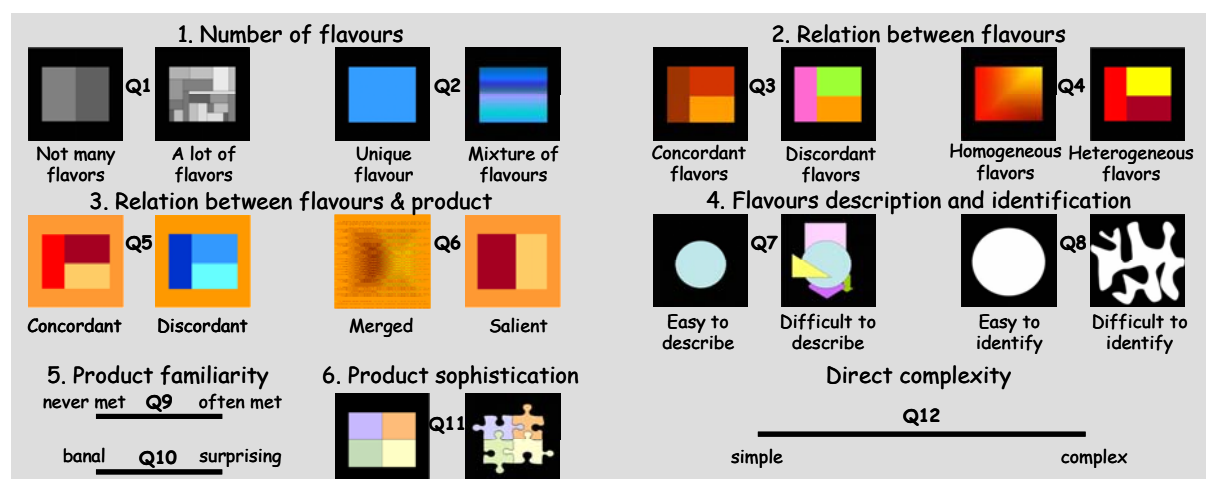


Figure 8 : Images servant de références pour les extrémités des échelles des items du questionnaire mis au point au LIRIS et visant à mesurer la complexité perçue d'un produit lors de sa dégustation

Dans le cadre du DEA de Caroline Reverdy, nous utilisons en parallèle cet outil pour une étude sur l'effet de l'éducation sensorielle sur les préférences et la perception de la complexité

aromatique de jeunes adultes. Nous supposons que l'éducation à « l'écoute du goût » déplacerait les préférences des aliments simples vers des aliments complexes, dû à une augmentation des performances individuelles et à une diminution de la complexité perçue des variantes complexes.

Finalement, cette étude a effectivement permis de montrer qu'il était possible de déplacer les préférences de jeunes consommateurs du simple vers le complexe grâce à un apprentissage sensoriel. Ayant développé leurs capacités descriptives lors des séances d'apprentissage, les produits les plus complexes leur semblent alors moins difficiles à décrire et à identifier, plus familiers et sont donc préférés. Forts de ces résultats préliminaires, nous pouvions nous lancer dans le montage du projet EduSens avec des enfants.

Justification du projet

Face à l'inquiétante augmentation de la prévalence de l'obésité infantile et devant le peu de résultats des actions en matière d'éducation nutritionnelle, il semble primordial d'étudier l'intérêt d'une éducation sensorielle. Une éducation au goût apprenant aux enfants à être davantage à l'écoute de leur perception sensorielle, à mieux verbaliser leur ressenti et à enrichir leur répertoire alimentaire pourrait bien se révéler plus efficace pour modifier leur comportement alimentaire. Cette hypothèse repose sur le fait que les préférences sont modulables sous l'effet de l'expérience et que le plaisir alimentaire peut s'éduquer. Il est possible d'apprendre à aimer des stimuli rejetés à la naissance (par exemple le goût amer). Il est peut-être également possible d'enrichir le répertoire alimentaire des enfants et de réduire leur attirance pour le sucre et le sel en leur apprenant à découvrir d'autres saveurs plus complexes.

Organisation du projet

Structuré en 3 expériences d'une durée de 18 mois chacune, le projet permet de tester successivement l'effet d'une éducation sensorielle dispensée en classe, puis au restaurant scolaire et enfin à la maison par les parents ayant suivi une formation au LIRIS. Chaque expérience inclut environ 180 enfants de 8 à 10 ans, dont la moitié fait office de groupe contrôle ne recevant pas d'éducation au goût. Des tests sensoriels sont réalisés avant et après la période d'éducation au goût, puis encore 9 mois plus tard afin d'évaluer la persistance et/ou l'évolution des effets observés. Les tests sensoriels mesurent la capacité à catégoriser et à décrire les sensations perçues en mangeant et évaluent les modifications des préférences et des choix des enfants. Ces tests sensoriels sont complétés par des mesures attitudinales et comportementales visant à évaluer l'évolution des croyances et des comportements alimentaires des enfants due à l'éducation sensorielle. Enfin, l'évolution de l'ambiance au sein de la classe et de la famille est observée car il est attendu un fort pouvoir socialisant de l'éducation sensorielle. Dans le même ordre d'idée, l'évolution des performances scolaires des classes ayant suivi l'éducation sensorielle est contrôlée.

Le projet, piloté par le LIRIS, rassemble des chercheurs de l'UMR FLAVIC, l'Institut du Goût, fondateur français de l'éducation sensorielle mise en œuvre dans les "Classes du Goût", un psychologue de l'Université d'Utrecht (Pays-Bas), spécialiste de la mesure des préférences chez les enfants, une sociologue de l'INA-PG, un chercheur en science de l'éducation de l'IREDU, un médecin spécialiste de l'alimentation infantile de la Faculté de Médecine de Dijon et deux statisticiens de l'INRA. L'Académie et la Ville de Dijon sont également deux partenaires essentiels qui permettent au projet EduSens de dispenser l'éducation sensorielle dans des classes et dans des restaurants scolaires.

EduSens a été sélectionné en 2005 par le Programme de Recherche en Alimentation (PRA) de l'INRA et par le programme National de Recherche en Alimentation (PNRA) de l'Agence National de la Recherche (ANR). J'ai construit ces deux dossiers en 2005 et je coordonne maintenant le projet EduSens.

En résumé, ce projet se propose donc d'observer expérimentalement dans quelle mesure une éducation sensorielle peut modifier les préférences alimentaires des enfants et d'étudier l'impact de cette éducation et des modifications de préférences qu'elle génère sur les comportements alimentaires des enfants. Si les résultats de ce projet sont probants, sa valorisation devrait conduire à une intégration de l'éducation sensorielle dans les programmes scolaires.

Premiers résultats

Pour le moment, seule la première expérimentation en classe est terminée, la seconde est en cours et la troisième débutera début 2008. Le résultat le plus marquant de la première expérimentation est la mise en évidence d'une diminution de la néophobie alimentaire (peur de manger des aliments inconnus) grâce à l'éducation sensorielle. De plus, il a été établi que la catégorisation d'odeurs inconnues est moins guidée par les aspects hédoniques suite à une éducation sensorielle. Ces deux premiers résultats suggèrent donc que l'éducation sensorielle aura permis d'éveiller la curiosité des enfants vis-à-vis de l'alimentation. Cet éveil est susceptible d'être le moteur d'une modification des préférences et des comportements alimentaires. Néanmoins, un premier niveau d'analyse des résultats des tests de dégustation en laboratoire n'a pas permis d'établir l'existence de telles modifications pour l'ensemble des enfants. Les analyses complètes de la première expérimentation seront disponibles à l'issue de la thèse de Caroline Reverdy en fin d'année 2007.

Réflexions et perspectives

A propos de la sensométrie

Parmi les 15 articles annexés à ce mémoire, seul le premier [88] n'a pas encore été cité. Il s'agit du texte d'une conférence invitée au congrès annuel de la Société Française de Statistique en 1994, où il m'avait été demandé, à l'issue de ma thèse de Doctorat, de présenter une réflexion sur les liens entre l'analyse sensorielle et les statistiques. De manière un peu provocatrice, ce texte s'intitule « L'évaluation sensorielle : une technologie statistique ? ». Nous citons ci-dessous in extenso la conclusion de cette réflexion :

A l'issue de ce panorama, non exhaustif, de l'implication des statistiques dans l'évaluation sensorielle, nous proposons une réponse positive à la question soulevée dans le titre de cette communication, tout en rappelant que si les statistiques sont nécessaires à l'évaluation sensorielle, elles ne lui sont pas suffisantes. Mais, les progrès en évaluation sensorielle ont souvent eu lieu grâce à l'adaptation d'un outil statistique, connu ailleurs, à une idée nouvelle de protocole sensoriel. Ainsi, deux conditions nécessaires à une activité de recherche en méthodologie de l'évaluation sensorielle sont donc d'une part une culture statistique profonde et étendue, et d'autre part une bonne compréhension des besoins des clients de l'évaluation sensorielle. La première condition ne peut être remplie que par une véritable formation de statisticien, alors que la seconde s'apprend forcément sur le tas. Notre message à l'attention des décideurs de l'industrie agro-alimentaire est donc de confier une part importante de la responsabilité de l'évaluation sensorielle à des statisticiens ouverts à l'application. L'enseignement des statistiques dans notre pays doit se tenir prêt à pouvoir répondre efficacement à de telles offres.

Aujourd'hui, 13 années plus tard, nous sommes toujours d'accord avec cette analyse. Avec une telle analyse en tête dès 1994, il n'est pas étonnant que notre parcours ait évolué lentement de la sensométrie pure (analyse des données) vers le développement de nouveaux protocoles d'acquisition de données sensorielles. Même si ce parcours a toujours été dicté par les impératifs et les priorités du groupe de recherche auquel nous appartenions et de l'Institut auquel nous avions à rendre des comptes, on a presque envie de dire que nous avons forcé le destin...

En ce qui concerne l'aspect enseignement et débouchés industriels évoqué dans la conclusion de cet article, nous pensons qu'effectivement il y a eu ces dix dernières années une augmentation nette des postes de statisticiens pour l'analyse sensorielle dans l'entreprise. Ce ne sont pas les deux doctorants que nous avons encadrés en sensométrie qui nous contrediront. Néanmoins, le nombre de ces postes reste faible en comparaison avec d'autres domaines tels que la pharmacie. Toutefois, il faut relativiser ce constat car il s'est développé

en France des formations hybrides entre les sciences de l'aliment et les statistiques qui placent beaucoup de leurs étudiants aux postes de responsabilité en analyse sensorielle. Nous tenons à citer ici les spécialités de 3ème année d'école d'ingénieur « Sensométrie et chimiométrie » de l'ENITIAA de Nantes et « Statistique appliquée » de l'ENSA de Rennes qui proposent un cursus commun aux étudiants. On peut trouver en ligne⁷ des informations intéressantes sur le devenir de ces étudiants. Doté d'un bon bagage en statistique, on doit pouvoir compter sur ces étudiants pour transférer dans l'industrie les acquis de la recherche en sensométrie. L'évolution entrevue en 1994 a donc bien eu lieu, mais il faut noter qu'elle n'est pas venue du milieu académique de la statistique, mais plutôt de celui de l'application et des besoins industriels.

Cela pose le problème du statut de la sensométrie dans la communauté statistique. Sommes-nous reconnus par cette communauté au même titre que peut l'être la psychométrie par exemple ? La réponse est clairement négative. Invités en 2005 au congrès de la *Psychometric Society* [91], les membres du bureau de la *Sensometrics Society* ont pu mesurer l'écart qui sépare nos deux disciplines, aussi bien en termes de nombres d'équipes, qu'en termes de production théorique au niveau statistique. Toutefois, est-il légitime de comparer la jeune sensométrie à l'ancienne psychométrie dont l'origine se confond avec celle du champ de la statistique multidimensionnelle au début du siècle dernier ? Certes, non. De plus, une évolution vers la théorie pour la théorie - comme la psychométrie me semble parfois être devenue l'objet - est-elle souhaitable pour la sensométrie ? Eu égard à l'origine délibérément appliquée à l'industrie de nos objets d'étude, la réponse est à nouveau négative.

Pourtant, sommes-nous aujourd'hui capable d'identifier les champs thématiques de la statistique que nous voudrions développer, voire même créer, pour assurer un développement pérenne de notre discipline ? C'est un débat incontournable qu'il faudra avoir, faute de quoi nous risquons de tourner en rond...

A propos de la Sensobase

C'était mon rêve... mon équipe l'a réalisé. Ce projet fonctionne parce que les fournisseurs de données y trouvent une réponse à un problème concret : obtenir rapidement un traitement statistique sophistiqué de leurs données sans avoir eux-mêmes à naviguer dans un ou plusieurs logiciels statistiques. Evidemment, ce système devra se tenir à l'écoute des nouveaux développements en sensométrie afin d'adapter son contenu aux progrès les plus prometteurs de notre discipline. Il y aurait véritablement dans ce service apporté par la Sensobase, jusqu'à présent offert gracieusement, la possibilité de développer une entreprise innovante...

⁷ <http://www.agrocampus-rennes.fr/math/enseignement/master-pro/memoires.html>

Pour enrichir les méta-analyses, deux voies sont à privilégier.

La première consiste à diversifier le nombre de fournisseurs de données plutôt que le nombre d'études, en raison de la forte variabilité des performances due au facteur fournisseur que nous avons pu observer dans nos premières méta-analyses. Pour cela, il faut faire davantage connaître l'existence de la Sensobase par la publication d'articles, la présentation de conférences et aussi l'organisation de formation à la lecture et à l'exploitation des résultats.

La seconde voie consiste à obtenir davantage d'information sur les sujets, leur entraînement et le mode d'évaluation des produits. Pour cela, il faut nouer des partenariats privilégiés avec des grands groupes qui nous sous-traiteraient en quelque sorte l'ensemble du contrôle des performances de leurs panels et qui s'engageraient dans une démarche de qualité pour la gestion, le stockage et la documentation de leur pratique d'analyse sensorielle. Des discussions ont déjà eu lieu avec certains, mais elles n'ont pas encore abouti. De plus, est-il stratégique pour un institut de recherche public de s'engager dans un tel programme ? Ne serait-il pas plus judicieux de confier une recherche si finalisée à une structure privée ?

Une perspective est récemment apparue avec la volonté de l'ACTIAA de relancer avec nous un projet similaire à Sensobase qui viserait à créer une « Prefbase » qui collectionnerait des études de préférence sur tout type de produit. Ainsi, nous aurions le descriptif avec la Sensobase et les préférences avec la Prefbase. Ce projet pourrait voir le jour dans le cadre d'un Réseau Mixte de Technologie en analyse sensorielle piloté par l'ACTIA et auquel l'INRA apporterait son expertise. L'ACTIA dépose en ce moment des dossiers de demande de subvention pour lancer un tel projet. Si ces financements pouvaient me permettre d'obtenir un ingénieur en informatique-statistique pour une durée significative (2-3 ans), alors le développement autour de la Sensobase pourrait reprendre. Car depuis la fin de la thèse de Nicolas Pineau, je reste un peu seul, assis sur cette mine d'or qu'est la Sensobase, sans le temps nécessaire pour la fouiller par moi-même.

A propos du DTS

Rose-Marie Pangborn, elle-même, écrivait en 1964 (Pangborn, 1964):

“If an appropriate method could be developed, a modification of the so-called time-intensity study (Neilson, 1958) could be used to establish the temporal and sequential changes in apparent taste intensity of both compounds in a mixture.”

Elle n'avait que 40 ans d'avance sur le DTS !

Nous pensons effectivement que le DTS répond à un fort besoin en analyse sensorielle que le long développement du profil sensoriel a quelque peu mis dans l'ombre. La perception d'un produit n'est pas analytique comme l'est le profil sensoriel. On perçoit une image, puis une

autre et encore une autre ; c'est finalement un film qui se joue dans la tête du goûteur et le DTS tente d'en capturer les moments clefs.

Nous avons désormais appliqué cette méthode à de nombreux modèles variés avec pas moins de 8 partenaires industriels différents. Nous sommes au courant de l'existence d'au moins une demi-douzaine d'autres entreprises qui ont testé la méthode DTS. Il est maintenant nécessaire de faire le bilan de toutes ces applications. Pour cela, il convient de rattraper le retard pris dans la publication de nos propres applications et d'attendre les communications des autres qui ne devraient plus tarder.

Du point de vue de la sensométrie, il reste beaucoup de choses à faire avec les données DTS puisque la représentation des données DTS que nous utilisons actuellement n'exploite que la fréquence de citation des descripteurs au fil du temps et pas leur intensité. Les paramètres D, T et S que nous extrayons des courbes afin de les analyser comme un triplet de variables statistiques ne constituent peut-être pas le meilleur codage de cette information.

Plus généralement, la mesure DTS n'est peut-être qu'une première tentative vers l'acquisition de profil dynamique, c'est-à-dire vers la mesure d'un grand nombre de sensations au fil du temps que dure l'ingestion d'une bouchée, ou même la consommation d'une ration. La tentation est grande aussi d'essayer d'adapter ce type de dispositif dynamique pour les consommateurs sans entraînement préalable et, pourquoi pas, dans des conditions de dégustation plus écologiques que le laboratoire d'analyse sensorielle. Il faudra alors s'interroger sur la manière d'obtenir, simultanément à ces mesures, une mesure de la valeur hédonique du produit, elle aussi variable au fil du temps.

A propos des projets VDQA et SensEaux

Comme on l'a déjà expliqué dans ce mémoire, nous voyons ces projets comme un tremplin pour la recherche méthodologique. Il devient de plus en plus difficile de financer une recherche méthodologique hors du contexte de projets de ce type.

Néanmoins, cette situation est stimulante car elle oblige le méthodologiste à produire des résultats. Un peu comme un statisticien qui n'analyserait jamais ou rarement des données réelles, un méthodologiste qui ne participerait pas à la mise en valeur de ses outils manquerait finalement une part essentielle de la démarche globale à laquelle il contribue en amont.

Pour notre part, nous espérons que le projet sur le vin permettra de poursuivre le développement du DTS dans les directions que nous venons d'évoquer ci-dessus et de reprendre le développement des outils de mesure de la complexité perçue, initié à l'origine du projet EduSens.

Enfin, nous souhaitons que le projet sur l'eau conduise au développement d'une nouvelle forme de profil sensoriel léger, réalisable par des consommateurs et permettant d'agréger les données issues de panels et d'études différentes.

A propos des projets OPALINE et EduSens

Depuis 2004, une grande partie de mon temps a été investi dans le montage des projets OPALINE et EduSens. Ce fut un travail intéressant et nouveau pour moi, je ne regrette absolument pas de l'avoir accompli, surtout dans la mesure où OPALINE et EduSens sont sortis lauréats de tous les appels d'offre nationaux où ils ont été présentés.

Ce travail de coordonnateur a donc consisté à transformer une idée en germe en deux projets constitués de partenaires issus de disciplines différentes et à leur trouver des financements conséquents. Les études proprement dites n'ont débuté que l'année dernière et commencent seulement maintenant à produire leurs premiers résultats. Je considère ce travail de montage de projets comme une réalisation à part entière qui sert maintenant à d'autres chercheurs. Je suis curieux de voir la portée des résultats que nous obtiendrons. Je suis impatient aussi de pouvoir dégager un peu de temps pour participer maintenant à ces projets en tant que statisticien, mon métier d'origine qui continue à me faire vibrer.

Références bibliographiques

- Pangborn R. M.; Chrisp, R. B. (1964). Taste Interrelationships. Vi. Sucrose, Sodium Chloride, And Citric Acid In Canned Tomato Juice. *Journal of Food Science*, (29), 490.
- Gower J. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27, 857-874.
- Escoufier Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, 751-760.
- Gower J.C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1), 33-50.
- Williams A.A. & Langron S.P. (1983). A new approach to sensory profile analysis. *Flavour of distilled beverages, origin and development*. J.R. Piggott. Chichester, E. Horwood: 219-224.
- Williams A.A. & Langron S.P. (1984). The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food and Agriculture*, 35, 558-568.
- Hubert L., Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2, 193-218.
- Puisais J. & Pierre C. (1987). *Le goût et l'enfant*. Paris: Flammarion.
- Lavit C. (1988). *Analyse conjointe de tableaux quantitatifs*. Paris: Masson.
- Brockhoff P.M. & Skovgaard Ib M. (1994). Modelling individual differences between assessors in sensory evaluations. *Food Quality and Preferences*, 5, 215-224.
- Kazi-Aoual F., Hitier S., Sabatier R. & Le Breton J.D. (1995). Refined approximations to permutation tests for multivariate inference. *Computational statistics and data analysis*, 20, 643-656.
- Wakeling I.N. & Macfie H.J.H. (1995). Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of k samples from t may be tested. *Food Quality and Preference*, 6(4), 299-308.
- Courcoux P. & Séménou M. (1997). Preference data analysis using a paired comparison model. *Food Quality and Preference*, 8(5/6), 353-358.
- Poulsen C.S., Brockhoff P.M. & Erichsen L. (1997). Heterogeneity in consumer preference data - A combined approach. *Food Quality and Preference*, 8(5/6), 409-417.
- Brockhoff P.M. (1998). Assessor modelling. *Food Quality and Preference*, 9(3), 87-89.
- Ennis D.M. & Bi J. (1998). The beta-binomial model : accounting for inter-trial variation in replicated difference and preference tests. *Journal of Sensory Studies*, 13(4), 389-412.

- Kunert J. & Meyners M. (1999). On the triangle test with replications. *Food Quality and Preference*, **10**(6), 477-482.
- Vigneau E., Courcoux Ph. & Séménou M. (1999). Analysis of ranked preference data using latent class models. *Food Quality and Preference*, **10**(3), 201-207.
- Qannari E.M., Courcoux P. & Séménou M. (2000). New models for the analysis of paired comparison data: segmentation of the panel. *Food Quality and Preference*, **11**(1/2), 71-76.
- Cleaver G. & Wedel M. (2001). Identifying random-scoring respondents in sensory research using finite mixture regression models. *Food Quality and Preference*, **12**(5-7), 373-384.
- Courcoux P. & Chavanne P.C. (2001). Preference mapping using a latent class vector model. *Food Quality and Preference*, **12**(5-7), 369-372.
- Brockhoff P. B. (2003). The statistical power of replications in difference tests. *Food Quality and Preference*, **14**(5-6), 405-417.
- Bi Jian (2005). Similarity testing in sensory and consumer research. *Food Quality and Preference*, **16**(2), 139-149.
- Bi Jian (2007). Similarity testing using paired comparison method. *Food Quality and Preference*, **18**(3), 500-507.
- Meyners M. (2007). Proper and improper use and interpretation of Beta-binomial models in the analysis of replicated difference and preference tests. *Food Quality and Preference*, **18**(5), 741-750.
- Dauphin Stéphanie Le (2007). Analyse statistique d'évaluations sensorielles au cours du temps. Nantes, Université de Nantes: 110.

Bibliographie de Pascal Schlich

Articles dans des revues internationales à comité de lecture

1. Etiévant P., Schlich P., Bertrand A., Symonds P., Bouvier J.C. (1988). Varietal and geographic classification of french red wines in terms of pigments and flavonoid compounds. *Journal of the Science of Food and Agriculture*, **42**, 39-54.
2. Etiévant P., Schlich P., Bouvier J.C., Symonds P., Bertrand A. (1988). Varietal and geographic classification of french red wines in terms of elements, amino acids and aromatic alcohols. *Journal of the Science of Food and Agriculture*, **45**, 25-41.
3. Etiévant P., Schlich P., Cantagrel R., Bertrand A., Bouvier J.C. (1989). Varietal and geographic classification of french red wines in terms of major acids. *Journal of the Science of Food and Agriculture*, **46**, 421-438.
4. Issanchou S., Schlich P. & Guichard E. (1989). Profils d'odeur des constituants de l'arôme de l'abricot: description par analyse des correspondances. *Sciences des Aliments*, **9**, 351-370.
5. Rousset S., Schlich P. (1989). Amylase production in submerged culture using principal component analysis. *Journal of Fermentation and Bioengineering*, **68**, 339-343.
6. Schlich P., & Guichard E. (1989). Selection and classification of volatile compounds of apricot using the RV coefficient. *Journal of Agricultural and Food Chemistry*, **37**, 142-150.
7. Guichard E., Schlich P. & Issanchou S. (1990). Composition of apricot aroma: correlations between sensory and instrumental data. *Journal of Food Science*, **55**(3), 735-738.
8. Guichard E., Schlich P., Issanchou S. (1990). Tipicity of apricot aroma: correlation between sensory and instrumental data. *Journal of Food Science*, **55**, 735-738.
9. Latrasse A., Schlich P., Le Quéré J.L. (1990). Chemotaxonomy of the essential oils of the blackcurrant buds (*Ribes* sp.) from major terpenes (hydrocarbons and oxygenated compounds). *Lebensmittel Wissenschaft und Technologie*, **23**, 377-385.
10. Rousset S., Schlich P. & Touraille C. (1990). Traitement statistique de profil sensoriel des jambons secs par analyse procustéenne généralisée. *Viande & Produits Carnés*, **11**(6), 317-318.
11. McEwan J. A. & Schlich P. (1991/92). Correspondence analysis in sensory evaluation. *Food Quality and Preference*, **3**(1), 23-36.
12. Issanchou S., Schlich P. & Lesschaeve I. (1992). Carte sensorielle d'eaux gazeuses : une application du vocabulaire libre et de l'analyse procustéenne. *Industries Alimentaires et Agricoles*, **109**, 181-186.
13. Martin B., Etiévant P.X., Le Quéré J.L., Schlich P. (1992). More clues about sensory impact of Sotolon in some flor sherry wines. *Journal of Agricultural and Food Chemistry*, **40**, 475-478.
14. Schlich P. & McEwan J.A. (1992). Cartographie des préférences. Un outil statistique pour l'industrie agro-alimentaire. *Sciences des Aliments*, **12**(3), 339-355.
15. Moio L., Schlich P., Issanchou S., Etiévant P.X. & Feuillat M. (1993). Description de la typicité aromatique de vins issus du cépage Chardonnay. *Journal International des Sciences de la Vigne et du Vin*, **27**(3), 179-189.
16. Schlich P. (1993). Risk tables for discrimination tests. *Food Quality and Preference*, **4**(3), 141-151.
17. Schlich P. (1993). Uses of change over designs and repeated measurements in sensory and consumer studies. *Food Quality and Preference*, **4**(4), 223-235.
18. Moio L., Schlich P. & Etiévant P. (1994). Acquisition et analyse d'aromagrammes de vins de Bourgogne issus du cépage Chardonnay. *Sciences des Aliments*, **14**(5), 601-608.
19. Molimard P., Lesschaeve I., Bouvier I., Vassal L., Schlich P., Issanchou S. & Spinnler H.E. (1994). Amertume et fractions azotées de fromages à pâte molle de type Camembert: rôle de l'association de *Penicillium camemberti* avec *Geotrichum candidum*. *Le Lait*, **74**, 361-374.

20. Schlich P. (1994). Grapes: a method and a SAS[®] program for graphical representations of assessor performances. *Journal of Sensory Studies*, **9**(2), 157-169.
21. Schlich P. & Moio L. (1994). Corrélation entre les profils aromatiques et les aromagrammes de vins de Bourgogne issus du cépage Chardonnay. *Sciences des Aliments*, **14**(5), 609-615.
22. Moio L., Chambellant E., Lesschaeve I., Issanchou S., Schlich P. & Etiévant P.X. (1995). Production of representative wine extracts for chemical and olfactory analysis. *Italian Journal of Food Sciences*, **3**, 625-278.
23. Pham T., Guichard E., Schlich P., Charpentier C. (1995). Optimal conditions for the formation of sotolon from alpha-ketobutyric acid in the french "Vin Jaune". *Journal of Agricultural and Food Chemistry*, **43**, 2616-2619.
24. Bérodiér F., Stévenot C. & Schlich P. (1997). Descripteurs de l'arôme de Comté. *Lebensmittel-Wissenschaft und-Technologie*, **30**, 298-304.
25. Callier P. & Schlich P. (1997). La cartographie des préférences incomplètes. Validation par simulation. *Sciences des Aliments*, **17**(2), 155-172.
26. Issanchou S., Schlich P. & Lesschaeve I. (1997). Sensory analysis: methodological aspects relevant to the study of cheese. *Lait*, **77**, 5-12.
27. Stévenot C., Bérodiér F. & Schlich P. (1997). Typologie aromatique des fromages de Comté. *Sciences des Aliments*, **17**, 547-553.
28. Brockhoff P.B. & Schlich P. (1998). Handling replications in discrimination tests. *Food Quality and Preference*, **9**(5), 303-312.
29. Carnevillier V., Feuillat M., Schlich P. (1998). Caractérisation de la région de production des vins du cépage Chardonnay par l'analyse des acides aminés libres. *Revue des Oenologues*, **89**, 11-13.
30. De La Presa-Owens C., Schlich P., Davies H., Noble A.C. (1998). Effect of "Méthode Champenoise" process on aroma of four V. vinifera varieties. *American Journal of Enology and Viticulture*, **49**, 289-294.
31. De La Presa-Owens C., Schlich P., Wada K., Noble A.C. (1998). Using sensory and instrumental data to interpret the effect of storage at elevated temperatures on aroma of Chardonnay wines. *Annals of the New York Academy of Sciences*, **855**, 854-859.
32. Schlich P. (1998). What are the sensory differences among coffees? Multi-panel analysis of variance and FLASH analysis. *Food Quality and Preference*, **9**(3), 103-106.
33. Schlich P. (1998). Comments from Pascal Schlich on the Næs & Langsrud paper. *Food Quality and Preference*, **9**(3), 167-168.
34. Schlich P. (1998). Comments from Pascal Schlich on the Steinsholt's paper. *Food Quality and Preference*, **9**(3), 172-173.
35. Aubry V., Schlich P., Issanchou S. & Etiévant P. (1999). Comparison of wine discrimination with orthonasal and retronasal profilings. Application to Burgundy Pinot Noir wines. *Food Quality and Preference*, **10**(4/5), 253-259.
36. Carnevillier V., Schlich P., Guerreau J., Charpentier C., Feuillat M. (1999). Characterisation of the production region of Chardonnay wines by analysis of free aminoacids. *Vitis*, **38**, 37-42.
37. Courregelongue S., Schlich P. & Noble A.C. (1999). Using repeated ingestion to determine the effect of sweetness, viscosity and oiliness on temporal perception of soymilk astringency. *Food Quality and Preference*, **10**(4/5), 273-279.
38. Peleg H., Gacon K., Schlich P. & Noble A.C. (1999). Bitterness and astringency of flavan-3-ol monomers, dimers and trimers. *Journal of the Science of Food and Agriculture*, **79**, 1123-1128.
39. Kälviäinen N., Schlich P. & Tuorila H. (2000). Consumer texture preferences: Effect of age, gender, and previous experience. *Journal of Texture Studies*, **31**, 593-607.
40. Martin N., Molimard P., Spinnler H.E. & Schlich P. (2000). Comparison of odour sensory profiles performed by two independent trained panels following the same descriptive analysis procedures. *Food Quality and Preference*, **11**(6), 487-495.

41. Porcherot C. & Schlich P. (2000). FLASH table and canonical mapping of potato varieties. *Food Quality and Preference*, **11**(1/2), 163-165.
42. Schlich P. & Chabanon C. (2000). L'analyse sensorielle appliquée au confort de freinage. *Revue Française de Marketing*, **179/180**, 129-142.
43. Schlich P., Dacremont C. & Brockhoff P.B. (2000). Application of replicated difference testing. *Food Quality and Preference*, **11**(1/2), 43-46.
44. Guinard J.X., Uotani B. & Schlich P. (2001). Internal and external mapping of preferences for commercial lager beers: comparison of hedonic rating by consumers blind versus with knowledge of brand and price. *Food Quality and Preference*, **12**(4), 243-255.
45. Cordelle S., Lange C. & Schlich P. (2004). On the consistency of liking scores: Insights from a study including 917 consumers from 10 to 80 years old. *Food Quality and Preference*, **15**(7-8), 831-841.
46. Schlich P. (2004). Just about right analysis of Just-About-Right scales. *Food Quality and Preference*, **15**, 892-893.
47. Cordelle S., Piper D. & Schlich P. (2005). On the consistency of liking scores: A validation study run in France and Germany. *Food Quality and Preference*, **16**(6), 493-503.
48. Rousset S., Deiss V., Juillard E., Schlich P. & Droit-Volet S. (2005). Emotions generated by meat and other food products in women. *British Journal of Nutrition*, **94**(4), 609-19.
49. Courcoux P., Qannari E.M. & Schlich P. (2006). Introduction to the sensometrics workshop of the Sixth Rose Marie Pangborn Sensory Science Symposium. *Food Quality and Preference*, **17**(7-8), 3-5.
50. Findlay C.J., Castura J.C., Schlich P. & Lesschaeve I. (2006). Use of feedback calibration to reduce the training time for wine panels. *Food Quality and Preference*, **17**(3-4), 266-276.
51. Pineau N., Cordelle S., Mathonière C., Etiévant P., Issanchou I., Imbert A., Rogeaux M., Schlich P. (2006). Temporal dominance of sensations: a new technique to record several sensory attributes simultaneously over time. *Food Quality and Preference*. Under review.
52. Pineau N. Chabanet C., Schlich P. (2007). Modeling the evolution of a sensory panel: a mixed-model and control chart approach. *Journal of Sensory Studies*, **22**, 212-241.

Articles dans des revues sans comité de lecture et expertises

53. Scandella D., Navez B., Lespinasse N., Vénien S., Jost M., Schlich P. (2002). La segmentation gustative. Cibler les envies du consommateur. L'exemple de la tomate. *Infos-Ctifl*, **185**, 9-13.
54. Schlich P. (2003). Tests de différences en analyse sensorielle. In *Commission de réflexion sur l'enseignement des mathématiques*. Y. Escoufier. Paris: 18-25

Ouvrages ou chapitres d'ouvrage

55. European Sensory Network (1996). *A European sensory and consumer study. A case study on coffee*. Chipping Campden, Gloucestershire, GL55 6LD UK: European Sensory Network. Available from Campden & Chorleywood Food Research Association.
56. Fors S.M. & Schlich P. (1989). Flavor composition of oil obtained from crude and roasted oats. *Thermal Generation of Aromas*. McGorrrin R.J. Parliment T.H., Ho C.T. Washington, American Chemical Society Symposium Series: 121-131.
57. Etiévant P.X., Moio L., Guichard E., Langlois D., Lesschaeve I., Schlich P., Chambellant E. (1993). Aroma extract dilution analysis (AEDA) and the representativeness of the odour of food extracts. *Trends in Flavour Research*. Heij. Noordwijkerhout D.G.V.D. Maarse H., Elsevier: 179-190.
58. Schlich P. (1995). Preference mapping: relating consumer preferences to sensory or instrumental measurements. *Bioflavour 95. Analysis/Precursor studies/Biotechnology*. P. Etiévant and P. Schreier. Paris, INRA Editions: 135-150.

59. Schlich P. (1996). Designing and analysing an European sensory and consumer study. *Consumer preference & sensory analysis*. S. Porretta. Maarsen, Miller Freeman: 40-49.
60. Schlich P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. *Multivariate Analysis of Data in Sensory Science*. T. Naes and E. Risvik. Amsterdam, Elsevier Science: 259-306.
61. Schlich P. (1998). L'analyse multidimensionnelle de profils sensoriels de vins. *Oenologie. Fondements scientifiques et technologiques*. C. Flanzy. Paris, Lavoisier: 265-278.
62. Qannari E.M. & Schlich P. (2006). Matching sensory and instrumental data. *Flavour in food*. A. Voilley and P. Etiévant. Cambridge, CB1 6AH, UK, Woodhead Publishing Limited: 62-80.

Communications avec article complet dans les actes du congrès

63. Adda J. & Schlich P. (1988). Les arômes des aliments. In AGORAL, Nancy (France). 16-39.
64. Schlich P., Issanchou S., Guichard E., Etiévant P. & Adda J. (1987). RV Coefficient: a new approach to select variables in PCA and to get correlations between sensory and instrumental data. In *Flavour Science and Technology*, Martens, M., Dalen G.A. & Russwurm H.Jr.. Chichester, John Wiley & Sons: 469-474.
65. Schlich P. (1989). A SAS/IML program for generalised procrustes analysis. In "Seugi 89" *Proceedings of the SAS European Users Group International conference.*, Cologne, SAS Institute GmbH. 529-537.
66. Schlich P. (1990). RV coefficient & principal component analysis of instrumental variables using SAS/IML software. In "Seugi 90". *Proceedings of the SAS European Users Group International Conference*, Paris, SAS Institute GmbH. 566-573.
67. Schlich P. & Issanchou S. (1990). Les méthodes de profils et leurs alternatives. In *1ères journées Agro-industrie et méthodes statistiques*, Angers. 254-272.
68. Schlich P. & Mcewan J.A. (1991). Cartographie des préférences. In *2èmes Journées Agro-Industrie et Méthodes Statistiques*, Nantes, 79-93.
69. Schlich P. (1992). A propos du profil sensoriel, de sa planification et de son analyse. In *3èmes journées européennes. Agro-industrie et Méthodes Statistiques.*, Montpellier, 123-126.
70. Schlich P. (1993). RV coefficient and PCAIV. In *Assessing food quality*, As, FLAIR SENS project, c/o MATFORSK, 1430 As, Norway. 50-51.
71. Callier Ph. & Schlich P. (1995). Extension de la cartographie des préférences aux plans incomplets. In *4ièmes Journées Agro-Industries et Méthodes Statistiques*, Dijon, 149-159.
72. Callier P. & Schlich P. (1995). Preference mapping in progress. In *4ièmes Journées Agro-Industries et Méthodes Statistiques*, Dijon, 283-287.
73. Lesschaeve I., Issanchou S. & Schlich P. (1995). Importance relative de la sélection et de l'entraînement sur les performances d'un jury d'analyse sensorielle descriptive. In *4èmes Journées Agro-Industries et Méthodes Statistiques*, Dijon, 97-107.
74. Schlich P. & Degoud S. (1995). Flash analysis of profiling data. In *4^{èmes} Journées Agro-Industries et Méthodes Statistiques*, Dijon. 305-309.
75. Callier P. & Schlich P. (1996). Pratique de la cartographie des préférences. In AGORAL 96. *Production industrielle et qualité sensorielle*, Paris, Lavoisier Technique & Documentation. 33-45.
76. Schlich P. (1997). CAP : une méthode et un outil de contrôle rapide et synthétique des performances des sujets en évaluation sensorielle descriptive. In *5èmes Journées Agro-industries et Méthodes Statistiques*, Versailles, 8:1-10.
77. Sibiril S., Schlich P., Lemesle S., Delange R. & Issanchou S. (1997). De la technique du profil idéal à la cartographie des préférences : la quête du cidre idéal. In *5èmes Journées Agro-Industries et Méthodes Statistiques*, Versailles, 11:1-13.
78. Schlich P. (1998). Analysis of the apple profiling data. In *Statistical modelling of consumer data*, Ås, Norway, AIR-CAT project, c/o MATFORSK. 39-45.

79. Pineau N., Cordelle S., Imbert A., Rogeaux M., Schlich P. (2003). Dominance temporelle des sensations. Codage et analyse d'un nouveau type de données sensorielles. In *XXXVèmes Journées de Statistiques*, Lyon, 777-780.
80. Schlich P. (2004). L'analyse en variables canoniques des données de profils sensoriels. Application pour le contrôle des performances des sujets et pour la comparaison des produits. In *8ème Journées Agro-Industries et Méthodes Statistiques*, Rennes. 33-43.
81. Pineau N., Cordelle S., Schlich P. (2004). Méthodes d'acquisition, de codage et d'analyse de profils sensoriels temporels. In *8ème Journées Agro-Industries et Méthodes Statistiques*, Rennes. 87-93.
82. Qannari E.M., Hannafi M., Schlich P. (2004). Extension de la méthode STATIS. Application en évaluation sensorielle. In *XXXVIèmes Journées de Statistiques*, Montpellier.
83. Pineau N., Chabanet C., Schlich P. (2006). Utilisation du modèle mixte et des cartes de contrôle pour le suivi des performances d'un panel d'analyse sensorielle au cours du temps. In *9èmes Journées Agro-industries et Méthodes Statistiques*, Montpellier, 101-106.
84. Sauzay L., Hanafi M., Qannari E.M., Schlich P. (2006). Analyse de K+1 tableaux à l'aide de la méthode STATIS. Application en évaluation sensorielle. In *9èmes Journées Agro-industrie et Méthodes Statistiques*. Montpellier. 157-160.
85. Schlich P. (2006). Atelier d'analyse de données. Influence de l'origine du consommateur sur ses préférences en matière de saumon fumé. In *9èmes Journées Agro-industrie et Méthodes Statistiques*, Montpellier. 7-10.
86. Schlich P. Pineau N., Brajon D., Cordelle S. (2006). Le projet Sensobase: construire une base de profils sensoriels pour documenter les performances des panels d'analyse sensorielle. In *9èmes Journées Agro-industrie et Méthodes Statistiques*, Montpellier. 101-106.

Conférences invitées dans des congrès internationaux

87. Schlich P. (1993). RV Coefficient and STATIS : useful multivariate statistical methods for sensory evaluation. In *Food 2000 Preservation*, Boston (USA), 19-21 October 1993. 419-430.
88. Schlich P. (1994). L'évaluation sensorielle : une technologie statistique? In *XXVI^e journées de statistique*, Neuchâtel, Groupe de statistique Pierre-à-Mazel 7, CH 2000 Neuchâtel, Suisse. 44-49.
89. Schlich P. (1999). Panelist performances versus product differences: from univariate to multivariate approach. In *Chemical Senses Day*, Santa Rosa, USA.
90. Schlich P. (2004). Panel and panelists performances. In *Annual Meeting of the Institute of Food Technology*, Las Vegas, USA.
91. Schlich P. (2005). Sensory profiling. In *14th International Meeting of the Psychometric Society*, Tilburg, The Netherlands.
92. Schlich P. (2005). Sensometrics overview. Challenges and opportunities. In *6th Pangborn Sensory Sciences Symposium*, Harrogate, UK.
93. Schlich P. (2006). Multivariate analysis of a sensory database to learn about panel performances. In *2nd European User Meeting on Multivariate Analysis*, Como, Italy.

Séminaires académiques invités

94. Schlich P. (1992). GPA ou STATIS, consensus ou compromis ? In *Sensory evaluation : European seminar*, Bruxelles, 17-20 november 1992. 151-156.
95. Schlich P. (1993). Relating data sets. The big deal in food research? In *FLAIR SENS Meeting*, Practical models for relationships between data sets,. 59-60.
96. Schlich P. (1994). Méthodes d'analyse et de validation du profil sensoriel. In *Séminaire Européen sur les corrélations entre les préférences et les données de l'analyse sensorielle*, Lyon.
97. Schlich P. (1994). Profil sensoriel et cartographie des préférences. In *Séminaire de statistique de l'Ecole Polytechnique Fédérale de Lausanne*, Lausanne.

98. Schlich P. (1995). Problèmes statistiques en évaluation sensorielle. In *Séminaire de statistique de l'ENSAI*, Malakoff.
99. Schlich P. (1995). Some STATS inputs to improve your sensory data analysis. In *Seminar at Department of Viticulture and Enology, University of California*, Davis, C
100. Schlich P. (1995). Results of the descriptive and consumer tests on coffee carried out by ESN members. In *A matter of taste. ESN Seminar*, Nordwijkterhous.
101. Schlich P. (1997). RV coefficient : the companion of my all life of statistician analysing sensory data. In *Seminar of the Statistical Department of the Royal Veterinary and Agricultural University of Denmark*, Copenhagen.
102. Schlich P. (2002). De l'analyse sensorielle vers la sensométrie. Acquis et perspectives. Séminaire de l'Institut de Statistique de l'Université Catholique Louvain-La-Neuve, Belgique.
103. Schlich P. (2003). Cartographie des préférences par la méthode de PrefMaX. In *Réunion conjointe du Club PLS et du Groupe Agro-Industrie de la SFdS*, Paris.

Conférences invitées à vocation de transfert

104. Schlich P. Risvik E. (1994). A pan-european study on coffees. In *ASTM Meeting*, Sarasota.
105. Schlich P. (1999). SAS on sensory data. In *Colgate-Palmolive Global Technology Center*, New Jersey, USA.
106. Schlich P. (2000). Nouvelle méthode statistique pour interpréter les résultats des tests discriminatifs ou de la puissance des tests de différence vers la prise en compte de répétitions. In *L'évaluation sensorielle au service des consommateurs*, Paris, AFNOR.
107. Scandella D. Schlich P. (2000). Cartographie des préférences pour la tomate. In *Journée interprofessionnelle de la Tomate: "Tomate, l'enjeu qualité"*, Baladran.
108. Schlich P. (2000). Panel performance. In *Nestlé Sensory Analysis Workshop*, Beauvais.
109. Schlich P. (2000). Le contrôle des performances des sujets. In *Analyse Sensorielle*, Sigma +, Paris.
110. Schlich P. Navez B. (2000). Critères de qualité et préférences des consommateurs. In *Journée interprofessionnelle de la Tomate: "Tomate, l'enjeu qualité"*, Baladran.
111. Schlich P. (2002). La cartographie des préférences: peut-on définir le produit idéal pour les différents consommateurs? In *Construction de la qualité des aliments*, Ecole Chercheurs Inra, La Grande Motte.
112. Schlich P. (2002). Faire répéter une épreuve de différence à des sujets permet-il de pallier un manque de puissance statistique? In *Commission Internationale des Industries Agricoles et Alimentaires, "Les nouveautés dans le domaine de la métrologie sensorielle"*, Dijon.
113. Schlich P. (2003). Quelques éléments sur le traitement statistique des données de profil sensoriel. In *Session de formation INAO*, ENSBANA, Dijon.
114. Lange C. & Schlich P. (2005). Le comportement d'achat du consommateur: comment évaluer l'impact de l'information disponible sur l'étiquette. In *Séminaire CRECEP*, Beaune.
115. Schlich P. (2005). Preference mapping: methods and applications. In *ESN Conference "Sensory Evaluation more than just food"*, Madrid, Spain.
116. Schlich P. (2005). The value of sensory analysis to understand consumer preferences in beverages. In *Drinktec, World Fair N°1 for Beverage and Liquid Food Technology*, München, Germany.
117. Schlich P. (2005). Effet de la variété et du terroir sur le goût et la texture de la pomme de terre. In *Journée technique de la pomme de terre*, ARVALIS-Institut du Végétal/ITPT, Paris.
118. Schlich P. (2005). Temporalité des sensations dans le vin. In *Assemblée Générale de l'Association des Oenologues Universitaires Bourguignons (ADOUB)*, Dijon.
119. Schlich P. (2005). Temporalité des sensations dans le vin. In *Séminaire CRECEP*, Beaune.

Communications avec résumé dans les actes du congrès

120. Latrasse A., Le Quéré J.L., Schlich P., Lantin B. (1988). Chémotaxonomie des huiles essentielles du bourgeon de cassis (*Ribes nigrum* L.). In *7èmes journées internationales Huiles Essentielles*, Digne-les-Bains.
121. Latrasse A., Piffaut C., Schlich P., Spinnler H.E. (1990). Production of gamma-decalactone and cis, 6-gamma-dodecanolactone by *Fusarium poae*. Influence of some nutritional factors and relation to the fungal fatty acids. In *Bioformation of Flavours*, Londres.
122. Molimard P., Lesschaeve I., Schlich P., Vassal L., Issanchou I., Soubbker H.E. (1993). Effet de l'association de souches de *Penicillium camemberti* avec *Geotrichum candidum* sur la qualité sensorielle de camemberts. In *Colloque Aliment 2002*, Paris.
123. Rousset S., Schlich P., Martin J.F., Cadic A.L., Touraille C. (1993). Influence of product consumer group and presentation order on the hedonic assessment of steaks. In *Advances in Sensory Food Science, Rose Marie Pangborn memorial Symposium*, Helsinki, Food Quality and Preference. 96.
124. Schlich P. (1993). Change-over designs and repeated measurements. In *Advances in Sensory Food Science, Rose Marie Pangborn memorial Symposium*, Helsinki, Food Quality and Preference. 83.
125. Schlich P. (1993). Power tables for discrimination tests. In *Advances in Sensory Food Science, Rose Marie Pangborn memorial*, Helsinki, Food Quality and Preference. 102.
126. Spinnler H.E., Molimard P., Lesschaeve I., Vassal L., Bouvier I., Schlich P. & Issanchou S. (1993). Etude physiologique et caractérisation de la production d'arômes par des champignons filamenteux participant à l'affinage des fromages de type Camembert. In *Colloque Aliment 2002*, Paris.
127. Schlich P. (1994). Design and some analyses of the coffee experiment organized by the European Sensory Network (ESN). In *2nd Sensometrics Meeting*, Edinburgh.
128. Spinnler H.E., Molimard P., Lesschaeve I., Bouvier I., Vassal L., Schlich P. & Issanchou S. (1994). Etude de la qualité de camemberts réalisés avec des associations de *Penicillium camemberti* avec *Geotrichum candidum*: effet désamérisant de *Geotrichum candidum*. In *Gestion des populations microbiennes dans les industries agro-alimentaires*, DIJON, 9-10/03. 56.
129. Callier P. & Schlich P. (1995). Preference mapping in progress. In *2nd Pangborn Sensory Science Symposium*, Davis.
130. Schlich P. & Callier Ph. (1995). Progress in preference mapping. In *2nd Pangborn Sensory Science Symposium*, Davis.
131. Schlich P. & Degoud S. (1995). Flash analysis of profiling data. In *2nd Pangborn Sensory Science Symposium*, Davis.
132. Schlich P., Callier P. & Degoud S. (1996). The dark side of preference mapping. In *Third Sensometrics Meeting*, Nantes, 19-21 June 1996.
133. De La Presa Owens C., Schlich P., Wada K., Noble A.C. (1997). Using sensory and instrumental data to interpret the effect of storage at elevated temperatures on aroma of Chardonnay wines. In *ISOT XII Proceedings*, New York Academy of Science.
134. De La Presa-Owens C., Schlich P., Davies H., Noble A.C. (1997). Changes in aroma produced by a second fermentation on 28 base wines to their corresponding sparkling wines. In *48th Annual Meeting of the American Society of Enology and Viticulture*, San Diego, CA.
135. Guinard J.X., Uotani B., Schlich P. (1997). Consumer testing of commercial lager beers: effect of knowledge of brand and price, and relation with descriptive analysis and expert quality ratings. In *Institute of Food Technologists Annual Meeting*, Orlando, FL.
136. Guinard J.X., Uotani B., Schlich P. (1997). An investigation of the preferences of American beer drinkers. In *Twelfth Annual Brewers and Vintners Technical Seminar*, Shell Beach, CA.
137. Noble A. & Schlich P. (1997). Comparison of configuration from sorting with those from sensory profiles. In *Chemical Senses Day XIII*, Dublin, CA.

138. Schlich P. (1997). CAP: A statistical system for the control of assessor performances. In *ISOT XII*, San Diego, New York Academy of Science.
139. Schlich P. & Callier Ph. (1997). Moving from complete to incomplete designs in preference mapping studies. In *Chemical Senses Day XIII*, Dublin, CA
140. Brockhoff P.B. & Schlich P. (1998). Handling replications in discrimination tests. In *3rd Pangborn Sensory Science Symposium*, Alesund.
141. Porcherot C., Issanchou S. & Schlich P. (1998). Influence of cooking modes on potato sensory properties. In *Fourth Sensometrics Meeting*, Copenhagen.
142. Schlich P. (1998). CAP: a user-friendly system for the control of assessor performances. In *3rd Pangborn Sensory Science Symposium*, Alesund.
143. Schlich P. & Chabanon C. (1998). La cartographie des préférences ou l'optimisation du plaisir subjectif à l'aide d'un réglage sensoriel objectif. In *4 èmes Journées Modulad, applications industrielles de l'analyse des données*, Paris.
144. Schlich P., Dacremont C., Brockhoff P.B. (1998). Application of replicated difference testing. In *3rd Pangborn Sensory Science Symposium*, Alesund.
145. Dupin P., Schlich, U. Fisher (1999). Differentiation of wines produced by organic or conventional viticulture according to their sensory profiles and aroma composition. In *24th OIV World Vine and Wine Congress*, Mainz.
146. Schlich P. (2000). Control of assessor performances. In *5th Sensometrics Meeting*, Columbia, MO.
147. Schlich P. (2000). From some classical ANOVA models to some less classical preference mapping techniques. In *5th Sensometrics Meeting*, Columbia, MO.
148. Chauvin F., Piper D., Schlich P., Köster E., Etiévant P. (2001). Comparison of the trigeminal effects of several mustards and spices. In *4th Pangborn Sensory Science Symposium*, Dijon.
149. Gourillon S., Issanchou S., Schlich P., Hansen G., Kramer Vig K., Nielsen R., Hoyer S., Seppelt B., Simchen U., Sheehan E. & Delahunty C. (2001). Impact of age and culture on food liking in three European countries. In *4th Pangborn Sensory Science Symposium*, Dijon.
150. Randall L., Heymann H., Schlich P. (2001). Integration of analytical and consumer sensory data in cross-cultural studies for the prediction of consumer acceptance. In *4th Pangborn Sensory Science Symposium*, Dijon.
151. Lange C., Schlich P. & Issanchou S. (2002). Impact of the information provided to consumers on their preference and purchase behaviour for Burgundy red wines. In *Xth Food Choice Conference*, Wageningen.
152. Cordelle S., Lange C. & Schlich P. (2003). On the consistency of liking scores: Insights from a study including 917 consumers from 10 to 80 years old. In *5th Pangborn Sensory Science*, Boston, MA.
153. Lange C., Monnery-Patris S., Marlier L., Boggio V., Nicklaus S., Chabanet C., Issanchou S., Gojard S., Schaal B. & Schlich P. (2003). Développement précoce des préférences et aversions alimentaires : un projet d'étude longitudinale. In *Colloques scientifiques, XIIèmes Rencontres Scientifiques de Nutrition, Goûts et conduites alimentaires de l'enfant*, Paris.
154. Lange C., Carré E., Schlich P. (2003). The impact of hedonic versus ethical values of a food product on consumer willingness to pay. The effect of earlier sensory exposure and/or knowledge of fair trade principle. In *5th Pangborn Sensory Science Symposium*, Boston, MA.
155. Pineau N., Cordelle S., Schlich P. (2003). Temporal dominance of sensations: a new technique to record several sensory attributes simultaneously over time. In *5th Pangborn Sensory Science Symposium*, Boston, MA.
156. Schlich P. (2003). On the multivariate analysis of sensory profiling data. In *5th Pangborn Sensory Science Symposium*, Boston, MA.
157. Schlich P. (2003). Just about simple analysis of just about right scales. In *5th Pangborn Sensory Science Symposium*, Boston, MA.

158. Schlich P., Lespinasse N., Navez B., Scandella D. (2003). Preference mapping of tomato varieties from the French market. A PrefMaX consumer segmentation validated by a two-year study. In *5th Pangborn Sensory Science Symposium*, Boston, MA.
159. Simchen U., Gourillon-Cordelle S., Issanchou S., Schlich P., Chabanet C., Hoyer S., Zunft F., Jeppsen L., Kramer Vig K., Sheehan E. & Delahunty C. (2003). Contribution of culture, gender and age to preferences to different rice samples. In *5th Pangborn Sensory Science Symposium*, Boston, MA.
160. Brajon D., Pineau N., Schlich P. (2004). Le contrôle des performances en analyse sensorielle. In *SAS Forum*, Paris.
161. Lange C., Rousseau F., Nicklaus S., Schlich P. & Issanchou S. (2004). The effect of expected versus perceived quality on consumer preference and choice according to the level of typicality of the product. In *A Sense of Identity. European Conference on Sensory Science of Food and Beverages*, Florence.
162. Pessina R., Patron C., Piombino P., Moio L., Schlich P. (2004). Measuring temporality of sensations in wine. In *A Sense of Identity, European Conference on Sensory Science of Food and Beverages*, Florence.
163. Pineau N. Pessina R., Cordelle S., Imbert A., Rogeaux M., Schlich P. (2004). Temporal dominance of sensations: comparison with conventional profiling and time-intensity. In *7th Sensometrics meeting*, Davis, CA.
164. Reverdy C., Thibaut A., Lange C., Schlich P. (2004). Effect of sensory learning on individual sensory performances, perceived complexity of food and consumer preference. In *A Sense of Identity, European Conference on Sensory Science of Food and Beverage*, Florence.
165. Schlich P., Brajon D., Pineau N. (2004). Le contrôle des performances en analyse sensorielle. In *SAS Forum*, Paris.
166. Schlich P., Pineau N., Brajon D., Qannari E.M. (2004). Multivariate control of panelist performances. In *7th Sensometrics meeting*, Davis, CA.
167. Cordelle S. & Schlich P. (2005). Effect of sweetness, acidity and texture on preference for fruit yoghurts in children, adolescents and adults. In *6th Pangborn Sensory Science Symposium*, Harrogate, UK.
168. Juillard E., Rousset S., Deiss V., Droit-Volet S., Schlich P. (2005). Influence de l'expérience sensorielle et du contexte social sur l'acceptation de la viande. In *Symposium sur les Territoire et Enjeux du Développement Régional*, Lyon.
169. Lange C. (2005). Effect of sensory and cognitive factors on the early development of food preferences. Workshop "Understanding the development of food preferences early in life: focus on follow-up studies". In *6th Pangborn Sensory Science Symposium*, Harrogate, UK.
170. Lange C., Thibaut A., Schlich P. (2005). A questionnaire of perceived aromatic complexity. In *6th Pangborn Sensory Science Symposium*, Harrogate, UK.
171. Lange C., Thibaut A., Schlich P. (2005). Effect of the perception of aromatic complexity. In *6th Pangborn Sensory Science Symposium*, Harrogate, UK.
172. Pessina R., Boivin L., Moio L., Schlich P. (2005). Temporal dominance of sensations in wine. In *In Vino Analytica Scientia*, Montpellier.
173. Pineau N., Brajon D., Cordelle S., Schlich P. (2005). Mining a database of sensory descriptive analyses to understand panellist performances. In *6th Pangborn Sensory Science Symposium*, Harrogate, UK.
174. Reverdy C., Lange C., Schlich P. (2005). Effect of sensory learning on consumer preference. In *6th Pangborn Sensory Science Symposium*, Harrogate, UK.
175. Schlich P., Pineau N., Brajon D. (2005). Control of assessor performances: display of the CAP-V3 outputs. In *6th Pangborn Sensory Science Symposium*, Harrogate, UK.
176. Fournier B., Schlich P., Pineau N., Brockhoff P.B. (2006). Confronting the Brockhoff's assessor model to the usual ANOVA model on 264 datasets. In *8th Sensometrics meeting*, As, Norway.

177. Pineau N., Schlich P., Brajon D., Cordelle S. (2006). Meta-analysis of panel performances from a database of hundreds of datasets. In *8th Sensometric Meeting*, Norway. 102.

Thèse du candidat

178. Schlich P. (1993). Contributions à la sensométrie. Thèse de Doctorat. Université de Paris-Sud (Centre d'Orsay). Soutenue le 12 juillet 1993. 284 pages

Thèses co-encadrées par le candidat

179. Callier P. (1996). La cartographie des préférences. Son application en milieu industriel et son extension aux plans incomplets. Thèse de Doctorat. Université Montpellier II. Soutenue le 5 novembre 1996. 181 pages
180. Pessina R. (2006). Dominance temporelle des sensations gustatives et aromatiques du vin. Università degli Studi di Foggia et Université de Bourgogne. Soutenue le 12 juillet 2006. 205 pages
181. Pineau N. (2006). La performance en analyse sensorielle. Une approche base de données. Thèse de Doctorat. Université de Bourgogne. Soutenue le 13 décembre 2006. 190 pages

Autres travaux d'étudiants encadrés ou co-encadrés par le candidat (depuis 2001)

182. Carré E. (2002). Evaluation des importances relatives des déterminants sensoriels et informationnels dans les préférences des consommateurs pour des cafés issus du commerce équitable. Mémoire de stage de seconde année de l'INA-PG.
183. Hollender T. (2002). Etude des préférences et du comportement d'achat du consommateur pour le café. Diplôme Universitaire de Technologie de Génie Biologique. Université de Metz.
184. Pineau N. (2002). Traitement des données de Dominance Temporelle des Sensations. Mémoire de fin d'étude ingénieur ENITIAA.
185. Pelletier L. (2003). Acquisition et analyse d'une méthode de profil sensoriel temporel, le Temps Intensité Profil. Maîtrise des Sciences et Techniques « Le Goût et son environnement » de l'Université de Tours.
186. Pineau N. (2003). Acquisition et analyse de profils sensoriels temporels. DEA des Sciences de l'Alimentation de l'ENSBANA. Université de Bourgogne.
187. Brajon D. (2004). Evaluation des performances multidimensionnelles des sujets et des panels d'analyse sensorielle. DESS d'Ingénierie Statistique du CNAM. Université de Versailles Saint Quentin en Yvelines.
188. Leclerc N. (2004). Mise en place d'une analyse sensorielle permettant d'étudier l'effet d'un apprentissage sensoriel sur les préférences et les performances des consommateurs. GRETA Dijon.
189. Reverdy C. (2004). Effet d'un apprentissage sensoriel sur les performances et les préférences sensorielles. DEA des Sciences de l'Alimentation de l'ENSBANA. Université de Bourgogne.
190. Thibaut A. (2004). Élaboration d'un questionnaire de complexité. Mesure des préférences en fonction de la complexité perçue. DESS de l'Institut Supérieur International de la Parfumerie, de la Cosmétique et de l'Aromatique alimentaire. Université de Versailles Saint Quentin en Yvelines.
191. Chesnel F. (2005). Mesure de catégorisation des odeurs, de néophobie alimentaire et de capacité de description sensorielle auprès d'enfants de 8 à 10 ans. Maîtrise des Sciences et Techniques « Le Goût et son environnement » de l'Université de Tours.
192. Ginon E. (2005). Mesure des préférences et des choix alimentaires d'enfants âgés de 8 à 10 ans. Maîtrise des Sciences et Techniques « Le Goût et son environnement » de l'Université de Tours.
193. Goffin D. (2005). Signature temporelle de la qualité des vins et influence de cette temporalité sur les préférences. Mémoire d'ingénieur de l'Institut Meurice. Bruxelles, Haute Ecole Lucia de Brouckère.

194. Descours V. (2006). Les classes du Goût et la famille. Mémoire d'ingénieur ENESAD. Université de Bourgogne.
195. Fournier B. (2006). Analyse statistique de données sensorielles. Application du modèle de Brockhoff à une base de données sensorielles et comparaison de ses performances à celle de l'approche classique. Mémoire d'ingénieur de l'ENSAI.
196. Garret R. (2006). Organisation et animation de séances de mesures sensorielles. GRETA Dijon.

Rapports de contrats de recherche gérés par le candidat (depuis 2001)

197. Schlich P. (2001). PrefMap cross culturel. Contrat avec le Groupe Danone.
198. Mathonière C. (2001). Rôle de la dominance temporelle des sensations organoleptiques sur les préférences des aliments. Première phase. Rapport contrat CNRS-Danone-Bel. Dijon, CESG: 174.
199. Mathonière C. (2001). Application du module « Temps Intensité multiple à attributs simultanés » du logiciel Fizz à l'acquisition de mesures de Dominance Temporelle des Sensations. Rapport contrat CNRS-Danone-Bel. Dijon, CESG: 20.
200. Mathonière C. (2001). Description et mode d'utilisation des macros développées sous le logiciel Excel pour analyser les mesures de Dominance Temporelle des Sensations. Rapport contrat CNRS-Danone-Bel. Dijon, CESG: 101.
201. Mathonière C. (2001). Rôle de la dominance temporelle des sensations organoleptiques sur les préférences des aliments. Deuxième phase. Rapport contrat CNRS-Danone-Bel. Dijon, CESG: 152.
202. Chauvin F., Favre C., Schlich P. (2003). Gold standard salad dressing. Contrat avec Unilever BestFood France.
203. Chauvin F., Favre C., Schlich P. (2003). Gold standard white-salad dressing. Contrat avec Unilever BestFood France.
204. Chauvin F., Schlich P. (2003). Moutarde et confort digestif. Contrat avec Amora-Maille.
205. Chauvin F., Schlich P. (2003). Effet de la moutarde sur la perception du gras et des arômes. Contrat avec Amora-Maille.
206. Favre C., Schlich P. (2003). Gold standard vinaigrette validation. Contrat avec Unilever BestFood France.
207. Pineau N., Cordelle S., Schlich P. (2003). Corrélation des données de Dominance Temporelle des Sensations avec les données de préférences. Contrat avec Groupe Danone et Fromageries Bel
208. Cordelle S. (2004). Mesure de la Dominance Temporelle des Sensations. Contrat de recherche n°01/5162/05/00011 avec le Conseil Régional de Bourgogne.
209. Schlich P., Cordelle, S. (2004). Confort digestif et aspects hédoniques de la consommation de moutarde et de condiments vinaigrés. Contrat de recherche n°02516AAO2S640 avec le Conseil Régional de Bourgogne.
210. Schlich P., Cordelle S. (2004). Confort digestif et aspects hédoniques de la consommation de moutarde et de condiments vinaigrés. Contrat avec Amora-Maille.
211. Pineau N., Schlich P., Cordelle S. (2007). La performance en analyse sensorielle. Une approche Base de Données. Allocation de thèse, Hors contrat de plan, Année 2003 avec le Conseil Régional de Bourgogne.

Annexes

Curriculum Vitae	74
Organigramme du LIRIS.....	76
Organigramme du CESG.....	77
Liste des publications annexées au mémoire	79
Article 1	80
Article 2	88
Article 3	105
Article 4	154
Article 5	173
Article 6	180
Article 7	187
Article 8	202
Article 9	216
Article 10	223
Article 11	239
Article 12	251
Article 13	263
Article 14	271
Article 15	302

Curriculum Vitae

Pascal SCHLICH

Adresse personnelle

20 impasse Cardinal Petit de Julleville
21000 Dijon
Tel : 03 80 36 25 70

Date de naissance : 26/11/1960

Lieu de naissance: Juvisy sur Orge

Nationalité : française

Situation familiale : marié, 4 enfants

Service militaire : effectué en 85-86

Adresse professionnelle

Centre Européen des Sciences du Goût (CESG)
Laboratoire d'Interface Recherche-Industrie-Sensométrie (LIRIS)
15 rue Hugues Picardet. 21 000 Dijon
Tél : 03 80 68 16 38 – Fax : 03 80 68 16 32 – E-mail : schlich@cesg.cnrs.fr

Diplômes

1979	Baccalauréat C	Académie de Créteil
1983	Maîtrise de Mathématiques et Appl. Fond.	Université Louis Pasteur, Strasbourg
1984	DEA de Statistiques et Modèles	Université Paris XI, Orsay
1993	Doctorat de Statistiques	Université Paris XI, Orsay
1996	Certificate of Proficiency in English	University of Cambridge, UK

Titre professionnel : Chargé de Recherche de 1^{ère} classe à l'INRA

Postes :

1984	Ingénieur d'Etudes. INRA, Laboratoire de Recherches sur les Arômes
1998	Chargé de Recherches de 1 ^{ère} classe. INRA, Laboratoire de Recherches sur les Arômes
2001	Directeur du LIRIS (UMS INRA-CNRS) au CESG
2003	Responsable de l'équipe LIRIS dans l'UMR des Sciences du Goût (CNRS-U.Bourgogne-INRA)

Année sabbatique : University of California (Davis). Pr Ann C. Noble. De juin 1995 à août 1996

Responsabilité et coordination de projets de recherche à partir de 2001

2001-2002	Contrats de recherche avec Amora-Maille, Groupe Danone, Fromageries Bel et Conseil Régional de Bourgogne. Budget annuel moyen de 165 K€
2003-2007	Projets ANR EduSens, OPALINE et VDQA. Contrats de recherches avec Pernod Ricard, Barilla et Lyonnaise des Eaux. Budget annuel moyen de 152 K€

Recrutement et encadrement de l'équipe LIRIS

2001-2002 Recrutement de 3 titulaires : un ingénieur d'étude (CNRS), un ingénieur de recherche (CNRS) et un agent technique (INRA)

2005-2006 Recrutement de 3 ingénieurs d'étude et d'un technicien contractuels pour les projets Opaline, EduSens, VDQA et SensEaux (chacun pour une durée d'environ 3 années)

Etudiants stagiaires : entre 3 et 5 étudiants selon l'époque de l'année, en général de niveau M1 ou M2

Main d'œuvre occasionnelle : entre 2 et 8 vacataires à temps partiel en période d'analyse sensorielle

Encadrement ou co-encadrement de thèse – Devenir des étudiants

1993-1996 Philippe Callier. *La cartographie des préférences. Son application en milieu industriel et son extension aux plans incomplets*. CIFRE, Danone. Université de Montpellier II.
Recruté comme ingénieur statisticien au Centre de Recherche de Nestlé (Lausanne) en 1997, puis affecté en stratégie marketing au siège de Nestlé (Vevey) depuis 2005

2004-2006 Rita Pessina. *Dominance temporelle des sensations gustatives et aromatiques du vin*. Soutenue le 12/07/2006 à l'Université de Foggia (Italie) en cotutelle avec l'Université de Bourgogne. Assistante de recherche à l'Université de Naples.

- 2003-2006 Nicolas Pineau. *La performance en analyse sensorielle. Une approche base de données*. Cofinancement ACTIA-Région Bourgogne. Soutenue le 13/12/2006 à l'Université de Bourgogne. Ingénieur statisticien au Centre de Recherche de Nestlé (Lausanne) depuis 1/10/2006.
- 2004-2007 Caroline Reverdy. *Effet d'une éducation sensorielle délivrée en classe sur les préférences et les comportements alimentaires des enfants*. Bourse du Ministère. Université de Bourgogne. Soutenance prévue fin 2007 (participation à l'encadrement assuré par Christine Lange)
- 2006-2009 Eric Teillet. *Goût et préférences en matière d'eau à boire*. CIFRE, Lyonnaise des Eaux
- 2006-2009 Sophie Meillon. *Impact de la réduction d'alcool dans le vin sur la perception sensorielle et sur l'acceptabilité par le consommateur*. CIFRE, Pernod Ricard

Activités d'enseignement

- Cours réguliers de statistique et analyse des données sensorielles à l'ENSBANA de Dijon (Master Pro « Gestion des Propriétés Sensorielles ») et à l'ENSAI de Rennes (élèves ingénieurs) - 30 heures par an
- Environ 2 à 4 séminaires par an sur l'analyse sensorielle dans différentes écoles ou universités (INA-PG, ENSIA, ENTIAA, CNAM, U. Montpellier II ...) – 15 heures par an
- Formations et séminaires en analyse sensorielle pour l'industrie - 10 heures par an

Membre du Comité éditorial des deux revues principales d'analyse sensorielle

- *Journal of Food Quality and Preference*
- *Journal of Sensory Studies*

Lecteur régulier pour les deux revues

- *American Journal of Enology and Viticulture*
- *Journal of Agricultural and Food Chemistry*
- *Journal of Food Quality and Preference*
- *Journal of the Science of Food and Agriculture*
- *Journal of Sensory Studies*
- *Lebensmittel Wissenschaft und Technologie*

Responsabilités dans des sociétés savantes

- Membre fondateur de la *Sensometric Society*
- Président du bureau du *Groupe Agro-industrie* de la *Société Française de Statistique*

Congrès. Organisation et comité scientifique

- Organisateur des 4^{ème} *Journées Agro-industries et Méthodes Statistiques*. Dijon, 1995
- Coorganisateur du 4th *Pangborn Sensory Science Symposium*. Dijon, 2001
- Président du Comité scientifique du 4th *Sensometrics Meeting*. Copenhagen, 1998
- Président du Comité scientifique des 6^{ème} *Journées Agro-industries et Méthodes Statistiques*. Pau, 2000
- Président du Comité scientifique des 8^{ème} *Journées Agro-industries et Méthodes Statistiques*. Montpellier, 2006
- Membre du comité scientifique du *Pangborn Sensory Science Symposium* (2001, 2003, 2005)
- Membre du comité scientifique du *Sensometrics Meeting* (depuis sa création)

Publications et communications

- 52 publications dans des revues à comité de lecture
- 8 ouvrages ou chapitres d'ouvrages
- 7 conférences invitées dans des congrès internationaux
- 82 communications dont 24 avec article complet dans les actes du congrès
- 10 séminaires académiques invités
- 18 conférences ou articles invités à vocation de transfert

Organigramme du LIRIS

Laboratoire d'Interface Recherche - Industrie - Sensométrie

Equipe de recherche de l'UMR (CNRS - Univ. Bourgogne - INRA) des Sciences du Goût

Code CNRS 5170 - Code INRA 1214 - Localisation: Centre Européen des Sciences du Goût (Dijon)



Nom	Grade	Institut	Projet ou fonction	Dates
-----	-------	----------	--------------------	-------

Personnel titulaire

Pascal Schlich	CR1	INRA	Direction	mars-01
Christine Lange	IR2	CNRS	Responsable OPALINE et EduSens	juil.-02
Sylvie Cordelle	IE2	CNRS	Responsable SensoBase et DTS	jan-02
Catherine Pédron	AJT	INRA	Aide technique	mars-02

Personnel contractuel

Caroline Laval	IE	INRA	Coordination logistique OPALINE	nov04-dec09
Christine Urbano	IE	INRA	Suivi des projets VDQA et SensEaux	fev06-aout09
Aude Gaignaire	IE	INRA	Education sensorielle pour EduSens	mars06-dec08
Sandrine Jacob	TR	INRA	Suivi alimentaire pour OPALINE	fev07-dec07

Doctorants	Sujet de thèse. Université. Financement	Dates
------------	---	-------

Caroline Reverdy	Effet des "Classes du Goût" sur les préférences et comportements alimentaires des enfants. Université de Bourgogne. Bourse du Ministère	nov04-oct07
Eric Teillet	Goût et préférences en matière d'eau à boire. Université de Bourgogne. CIFRE, Lyonnaise des Eaux.	oct06-sep09
Sophie Meillon	Impact de la réduction d'alcool dans le vin sur la perception sensorielle et sur l'acceptabilité par le consommateur. AgroParisTech. CIFRE, Pernod Ricard	nov06-oct09

Projets	Description. Sources de financement.	Dates
---------	--------------------------------------	-------

SensoBase	Constitution et exploitation d'une base de données comportant plus de 500 études de profils sensoriels. ACTIA, Région Bourgogne	depuis 2003
DTS	Développement et application à différentes études de la méthode de la Dominance Temporelle des Sensations. Financements par contrats industriels	depuis 2001
OPALINE	Observatoire des Préférences Alimentaires du Nourisson et de l'Enfant. Coordination logistique, communication et suivi alimentaire. Région Bourgogne, PRNH 2004, PNRA 2006, 4 mécènes privés	jan04-dec10
EduSens	Effet d'une éducation sensorielle sur les préférences et les comportements alimentaires des enfants. Coordination générale. PRA 2005, PNRA 2005	nov04-dec08
VDQA	Vin De Qualité à teneur réduite en Alcool. Coordination du volet sensoriel du projet. PNRA 2005, Pernod Ricard	jan06-dec08
SensEaux	"L'eau du robinet: une eau agréable à boire". Lyonnaise des Eaux	fev06-sept09

Type	Collaborations
------	----------------

Locales	INRA-FLAVIC, Faculté de Médecine, IREDU
Françaises	ACTIA, ENITIAA (Sensométrie), ParisAgroTech (Sensométrie, Sociologie rurale), Montpellier SupAgro
Internationales	European Sensory Network (ESN), Sensometrics Society, Technical University of Denmark, University of California (Davis, Viticulture & Enology)
Industrielles	Amora-Maille, Barilla, Blédina, Cédus, Danone, Fromageries Bel, Lyonnaise des Eaux, Nestlé, Pernod Ricard, PSA, Puratos, Senoble, Symrise, Unilever

Périodes	Budget géré (tous projets confondus, hors taxes, hors salaires publics et hors équipement)
----------	--

1999-2002	204 K€ par an. Contrats industriels (80%), Région Bourgogne (17%), INRA (2%), CNRS (1%)
2003-2006	152 K€ par an. PRA & PNRA (53%), Contrats industriels (35%), Région Bourgogne (4%), CNRS (4%), INRA (3%), Université de Bourgogne (1%)



CENTRE EUROPEEN DES SCIENCES DU GOÛT - UMR 5170

Directeur : Benoist SCHAL



Axe I Physiologie des sens chimiques

Axe II Des représentations aux comportements

Service Administratif et Logistique

Thème : Neurophysiologie de l'olfaction

G. SICARD CR1/CNRS
I. JAKOB IR2/CNRS
Doctorante : S. VELOSO

Thème : Développement du système olfactif

J. GASQUEL CR1/INRA
B. QUENNEDEY TCE/CNRS (50%)
Post-Doctorant : T. AMANO (18/04/07)
Doctorante : J. MILLERY

Thème : Biologie moléculaire des sens chimiques

J.P. MONTMAYEUR CR1/CNRS
S. GRALL TCE/CNRS (50%) *
Post-Doctorant : Z. LIU (01/02/07)

Thème : Neurosciences de la mémoire olfactive

F. DATICHE MC/UB
F. LIENARD IE2/CNRS
Doctorant : D. DARDOU

Services communs

Biologie Moléculaire
C. FENECH IR2/CNRS
Traitement du Signal
V. GIGOT IR2/CNRS

Ethologie et psychobiologie sensorielle

B. SCHAAAL DR2/CNRS
G. COUREAUD CR2/CNRS
K. DURAND MC/UB
A. HOLLEY Pr/UL

A. SEIGNEURIC MC/UB
B. PATRIS IR2/CNRS (50%)
N. MERMET IE CDD/CNRS
B. QUENNEDEY TCE/CNRS (50%)
S. MARTIN TCN/CNRS

Doctorants : M. DELAUNAY
S. DOUCET, C. FERDENZI
D. MONTIGNY, A. MAIER (50%)
M2R : G. BAUDON

Psychologie cognitive des sens chimiques

D. VALENTIN MC/UB
J. BALLESTER MC/UB
V. GUFONI MC/UL,
C. DACREMONT Pr/UB
D. PEYRON MC/UB

F. JACQUIER-CHAUVIN CR CDD/ADEPT
B. PATRIS IR2/CNRS (50%)
Post-Doctorant : E. CAMPO (01/01/07)
Doctorants : B. DO, J. JAFFRE,
J. LANGLOIS, M. LELIEVRE
M2R : P. TU

Laboratoire d'Interface Recherche-Industrie et de Sensorimétrie

P. SCHLICH CR1/INRA
C. LANGE IR2/CNRS
S. CORDELLE IE2/CNRS (80%)*
A. GAIGNAIRE IE CDD/INRA
C. LAVAL IE CDD/INRA (80%)*
C. URBANO IE CDD/INRA
C. PEDRON AJT/INRA (80%)*
S. JACOB TR CDD/INRA
Doctorants : C. REVERDY, E. TELLETT,
S. MELLON

Axe III Déterminants physiologiques des comportements alimentaires

Métabolisme et comportement alimentaire

D. RIGAUD PUPH/UB
L. BRONDEL MC/UB
T. JIANG MC/UB
V. VANWYMELEBEKE IR/CHU

Nadine MUTIN IE1/CNRS
Administrateur

Gestion administrative, financière et comptable

V. BUE TCN/CNRS
C. CHABERT AJT/INRA

Bibliothèque, Documentation, Accueil,
S. GROS AJT/CNRS (*contingent handicap*)

Maintenance et aménagement intérieur du Bâtiment, Logistique
C. SERRANO AJT/CNRS (ACMO)

Mécanique
M. TAVAN AJT/CNRS

Assurance Qualité
V. GIGOT IR2/CNRS

Informatique et Réseaux, Assistance micro, Gestion du parc informatique
G. ROYER AI CDD/CNRS (*jusqu'au 31/07/2007*)
F. LEGER TCS/CNRS (ACMO)-(site Web-assistance Micro)

Animalerie
O. BERNARDINI AJT/CNRS

* Personnel à temps partiel

Liste des publications annexées au mémoire

Le choix des publications annexées a été réalisé sous la contrainte de sélectionner une et une seule publication par année depuis 1994 (année post-thèse) jusqu'à 2007 et avec le souci d'échantillonner la diversité des thématiques de recherche et des activités d'application. On retrace ainsi assez fidèlement le parcours (et non le projet) du candidat. On a ajouté une dernière publication en cours de révision.

1. Schlich P. (1994). L'évaluation sensorielle : une technologie statistique? In *XXVI^e journées de statistique*, Neuchâtel, Groupe de statistique Pierre-à-Mazel 7, CH 2000 Neuchâtel, Suisse. 44-49.**80**
2. Schlich P. (1995). Preference mapping: relating consumer preferences to sensory or instrumental measurements. *Bioflavour 95. Analysis/Precursor studies/Biotechnology*. P. Etiévant and P. Schreier. Paris, INRA Editions: 135-150.**88**
3. Schlich P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. *Multivariate Analysis of Data in Sensory Science*. T. Naes and E. Risvik. Amsterdam, Elsevier Science: 259-306.**105**
4. Callier P. & Schlich P. (1997). La cartographie des préférences incomplètes. Validation par simulation. *Sciences des Aliments*, **17**(2), 155-172.**154**
5. De La Presa-Owens C., Schlich P., Davies H., Noble A.C. (1998). Effect of "Méthode Champenoise" process on aroma of four V. vinifera varieties. *American Journal of Enology and Viticulture*, **49**, 289-294.....**173**
6. Peleg H., Gacon K., Schlich P. & Noble A.C. (1999). Bitterness and astringency of flavan-3-ol monomers, dimers and trimers. *Journal of the Science of Food and Agriculture*, **79**, 1123-1128.**180**
7. Schlich P. & Chabanon C. (2000). L'analyse sensorielle appliquée au confort de freinage. *Revue Française de Marketing*, **179/180**, 129-142.**187**
8. Guinard J.X., Uotani B. & Schlich P. (2001). Internal and external mapping of preferences for commercial lager beers: comparison of hedonic rating by consumers blind versus with knowledge of brand and price. *Food Quality and Preference*, **12**(4), 243-255.....**202**
9. Scandella D., Navez B., Lespinasse N., Vénien S., Jost M., Schlich P. (2002). La segmentation gustative. Cibler les envies du consommateur. L'exemple de la tomate. *Infos-Ctifl*, **185**, 9-13.....**216**
10. Schlich P. (2003). Tests de différences en analyse sensorielle. In *Commission de réflexion sur l'enseignement des mathématiques*. Y. Escoufier. Paris: 18-25.....**223**
11. Cordelle S., Lange C. & Schlich P. (2004). On the consistency of liking scores: Insights from a study including 917 consumers from 10 to 80 years old. *Food Quality and Preference*, **15**(7-8), 831-841.**239**
12. Rousset S., Deiss V., Juillard E., Schlich P. & Droit-Volet S. (2005). Emotions generated by meat and other food products in women. *British Journal of Nutrition*, **94**(4), 609-19.....**251**
13. Schlich P. Pineau N., Brajon D., Cordelle S. (2006). Le projet Sensobase: construire une base de profils sensoriels pour documenter les performances des panels d'analyse sensorielle. In *9^{èmes} Journées Agro-industrie et Méthodes Statistiques*, Montpellier. 101-106.**263**
14. Pineau N. Chabanet C., Schlich P. (2007). Modeling the evolution of a sensory panel: a mixed-model and control chart approach. *Journal of Sensory Studies*, **22**, 212-241.....**271**
15. Pineau N., Cordelle S., Mathonière C., Etiévant P., Issanchou I., Imbert A., Rogeaux M., Schlich P. (2006). Temporal dominance of sensations: a new technique to record several sensory attributes simultaneously over time. *Food Quality and Preference*. Under review.**302**

Article 1

1. Schlich P. (1994). L'évaluation sensorielle : une technologie statistique? In *XXVI^e journées de statistique*, Neuchâtel, Groupe de statistique Pierre-à-Mazel 7, CH 2000 Neuchâtel, Suisse. 44-49.

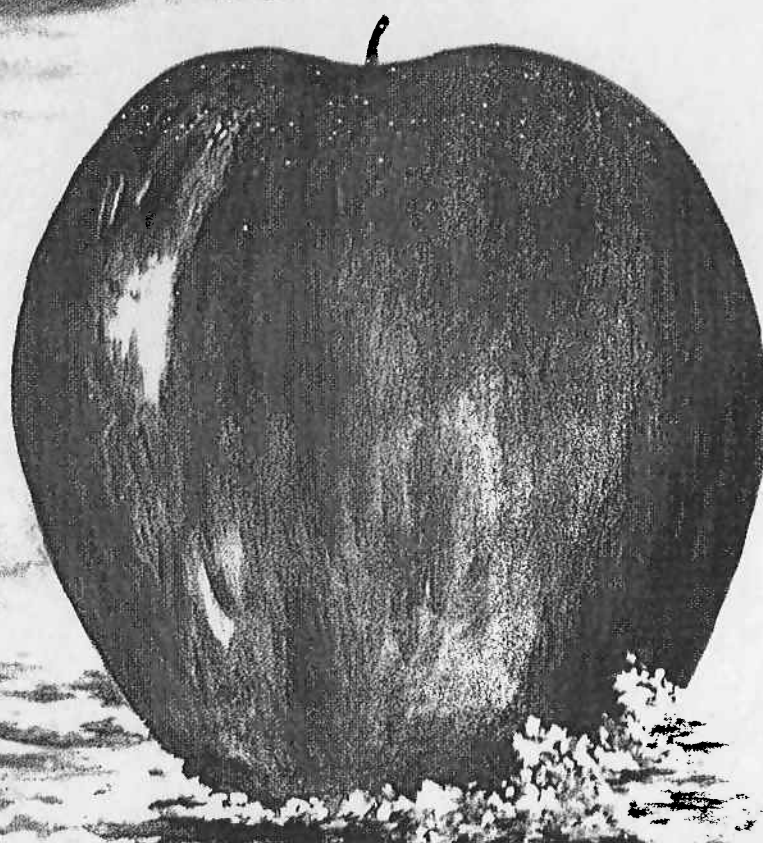
Université de Neuchâtel 24 - 27 mai 1994

XXVI^e Journées de Statistique

Association pour la Statistique et ses Utilisations
Sous l'égide de la Société de Statistique de France

Avec la participation de:

Société Française de Biométrie, Société Francophone de Classification,
Société de Statistique de Paris, Groupe des Membres Français de l'I.I.S.,
Groupe M.A.S. de la S.M.F.



Groupe de Statistique

Pierre-à-Mazel 7, CH - 2000 Neuchâtel, Suisse

Téléphone: +41 (38) 21 14 67 - Fax: +41 (38) 21 17 30

L'EVALUATION SENSORIELLE : UNE TECHNOLOGIE STATISTIQUE ?

Pascal SCHLICH

INRA, Laboratoire de Recherches sur les Arômes
17 rue Sully, BV 1540, 21034 Dijon Cedex, France

Abstract

The paper is a survey of several achievements, flaws and needs in the application of statistics in sensory evaluation. On one hand, it aims to convince the statistical community that sensory evaluation is an interesting field allowing the application of a wide range of statistical technics, and, on the other hand, it argues that statistic is the core of the sensory methodology.

1. Introduction

L'industrie agro-alimentaire, désireuse de mieux contrôler la qualité de ses produits, utilise de manière croissante l'évaluation sensorielle, ou analyse sensorielle [1, 22, 32]. Celle-ci consiste à objectiver autant que possible la dégustation, en sélectionnant et en entraînant un groupe de sujets à détecter et à décrire des différences sensorielles fines entre des produits alimentaires. A terme, l'espoir de prédire les caractéristiques sensorielles à l'aide de mesures instrumentales, réputées plus fiables et moins coûteuses, est bien présent. Enfin et surtout, il s'agit d'évaluer le poids des critères sensoriels dans les préférences des consommateurs, car, in fine de l'acceptabilité des aliments dépend la croissance de l'entreprise. L'origine de l'évaluation sensorielle est à chercher dans la psychophysique [10], mais son objet réside délibérément dans la mesure des caractéristiques sensorielles des aliments, plutôt que dans l'étude des lois régissant la perception de ces caractéristiques. En ce sens, il s'agit bien d'une technologie plutôt que d'une science, même si la démarche est celle de la science expérimentale. Le contenu de cette technologie est un ensemble de protocoles de dégustation, ou plans d'expérience, assortis d'une méthode d'analyse statistique des résultats.

L'objectif de cette communication est de justifier par l'exemple ces lignes introductives auprès de la communauté statistique, afin de la convaincre que l'évaluation sensorielle est un vaste et passionnant terrain d'application des statistiques dans lequel beaucoup de progrès restent à faire.

2. La mesure sensorielle

2.1. Amélioration de produit ou réduction de coût : contrôle de risques

Lorsqu'une industrie agro-alimentaire compare un produit de base à une variante de ce produit, elle désire toujours remplacer le produit de base par la variante. Toutefois, deux motifs radicalement différents existent pour justifier la substitution. Le premier est l'amélioration sensorielle du produit de base alors que le second est la réduction de son coût de production. Dans le premier cas, il s'agit d'établir une différence sensorielle significative

entre la variante et le produit de base; dans le second cas, au contraire, il convient d'établir une identité sensorielle significative entre les deux produits, afin d'éviter de décevoir les consommateurs habituels de ce produit. Les mêmes tests de différences (tests binomiaux), outils de base de l'évaluation sensorielle, sont mis en oeuvre dans ces deux situations. Toutefois, l'amélioration de produit nécessite de contrôler le risque de première espèce de ces tests alors que la réduction de coût requiert le contrôle du risque de seconde espèce, c'est-à-dire de la puissance du test. Or, le concept de puissance d'un test statistique et les moyens d'évaluation de la puissance [5] ne sont pas, en général, bien maîtrisés par les responsables de l'évaluation sensorielle. Des tables de risques pour les tests de différences ont donc été proposées [24]. Au delà des tests de différences de l'évaluation sensorielle, ce problème du contrôle des deux risques nous apparaît comme une question-clef lors de l'application de l'inférence statistique. Malheureusement, les manuels et logiciels de statistiques n'accordent pas à ce problème l'importance qu'il mérite.

2.2. Planification du profil sensoriel : les effets d'ordre et de report

Le rôle du profil sensoriel est de décrire et quantifier les différences entre des produits. Successivement pour chaque produit, le sujet note un certain nombre de descripteurs sensoriels (variables statistiques) présentés dans une fiche de dégustation. Avec ce type de plan en *cross-over*, où les mêmes unités expérimentales (sujets) reçoivent successivement les traitements (produits), il est souhaitable d'équilibrer les ordres de présentation des produits, afin que chacun soit dégusté un même nombre de fois dans chaque position et qu'un produit précède un même nombre de fois chaque autre produit. Certains carrés latins [36] sont de la plus grande utilité pour réaliser, autant que possible, cet équilibre. Bien que les statisticiens enrôlés dans le domaine de l'évaluation sensorielle recommandent périodiquement [11, 18, 25] l'utilisation de ces outils de planification, les manuels et logiciels d'analyse sensorielle ne suivent malheureusement pas.

2.3. Analyse unidimensionnelle du profil sensoriel : variations sur le thème de l'ANOVA

L'analyse des données relatives à un seul descripteur peut être menée dans deux buts différents : le contrôle des performances du panel de dégustation ou la comparaison des produits. Dans le premier cas, le client de l'analyse statistique est le responsable du panel alors que dans le second cas il s'agit plutôt d'un *chef produit*.

Le premier désire évaluer pour chaque sujet son utilisation de l'échelle (localisation moyenne et étendue de la dispersion des notes), sa répétabilité intra et inter séance (aptitude à donner la même note à des répétitions du même produit présentées à l'aveugle au sein d'une même séance ou au cours de séances différentes), son pouvoir discriminant (aptitude à séparer le plus possible des produits différents) et son accord avec le groupe sur les différences entre les produits. Chacune de ces performances est quantifiable via un modèle d'analyse de la variance adapté au plan d'expérience et assorti de décompositions sur les sujets de somme de carrés résiduelle ou d'interaction (accord avec le groupe). Dans le cas où chaque sujet a dégusté une seule fois chaque produit au cours de chaque séance, nous avons donné une procédure automatique pour l'exploration graphique de ces performances [28]. Dans les cas déséquilibrés, parfois les plus fréquents, un véritable travail de transfert de nos techniques statistiques reste à faire et serait très favorablement accueilli par les responsables de l'évaluation sensorielle.

Le second désire obtenir un classement des produits robuste par rapport à la variabilité due à l'échantillonnage des sujets. Il s'agit évidemment de modèles mixtes, qui peuvent devenir assez compliqués en présence de plusieurs facteurs fixes et aléatoires qui interagissent entre eux. De plus, lorsque l'objet de l'étude est le suivi dans le temps de l'évolution d'un produit ou des préférences de groupes de consommateurs, s'ajoute à ces difficultés celle de l'analyse des mesures répétées. Nous avons tenté d'expliquer à la communauté de l'évaluation sensorielle quelques unes de ces difficultés [25]. De tels textes non mathématiques, mais plus systématiques et accompagnés de logiciels dédiés, seraient nécessaires pour promouvoir

l'emploi des techniques évoluées du modèle linéaire dans le domaine sensoriel, où elles sont tant nécessaires.

2.4. Analyse multidimensionnelle du profil sensoriel : choix de variables et de métriques

Le descripteur sensoriel est-il, à l'instar d'un poids ou d'un rendement en agronomie, une véritable variable statistique ? Est-on sûr que la mesure représente bien un concept invariant par changement d'unité expérimentale ou de bloc (sujet) ? La réponse à cette question dépend du contexte, mais dans la négative, l'analyse unidimensionnelle par descripteur est douteuse et devrait être remplacée par une analyse multidimensionnelle de l'ensemble des descripteurs qui ne capitaliserait pas sur cette hypothèse d'*alignement des descripteurs*. Cette idée est à la base de la définition du profil libre-choix [35], dans lequel chaque sujet est laissé libre de choisir et noter sa propre liste de descripteurs, lequel a, couplé à l'analyse procustéenne généralisée (APG) [2, 12], constitué une avancée significative dans l'analyse du profil sensoriel. L'APG s'affranchit des différences individuelles de signification des descripteurs en appliquant aux configurations individuelles de produits des rotations orthogonales (respect des distances inter-produits par sujet) minimisant la somme des carrés des distances entre produits homologues. Avec des profils sensoriels conventionnels, l'APG reste intéressante pour pallier d'éventuels non-alignements des descripteurs. Si de plus des répétitions sont disponibles, une sélection propre à chaque sujet de ses seuls descripteurs discriminants conduit à un *profil libre-choix simulé* [29]. L'exploitation de ce type de données nécessite aussi l'APG, ou une autre technique multi-tableaux, telle que la méthode STATIS [16, 17], qui exploite les seules distances inter-produits par sujet. STATIS, par l'analyse de la matrice des coefficients RV [8] entre les sujets, permet de calculer pour chacun d'eux un poids proportionnel à son accord quant aux distances inter-produits avec le groupe tout entier. L'avantage de STATIS par rapport à l'APG est ne pas être itérative et de fournir, par son système de pondération des sujets, un traitement automatique des individus aberrants. L'utilisation de STATIS en évaluation sensorielle tend à se développer en France [14, 26-27], alors que nos collègues britanniques et néerlandais restent actifs dans le domaine de l'APG [6-7, 34].

L'APG, comme STATIS, fournit in fine une carte des produits qui résume au mieux les distances inter-produits individuelles. Ce résumé est appelé un consensus en APG et, plus modestement, un compromis chez STATIS. La question cruciale du test de la validité de ce résumé a été abordée en APG [34] et avec STATIS [26] par des tests de permutations. Il s'agit de permuter les labels des produits de manière aléatoire et indépendante dans chaque tableau individuel, de calculer le nouveau compromis, de réitérer l'opération un grand nombre de fois et de comparer la qualité moyenne des compromis ainsi obtenus à celle du compromis obtenu avec les données non permutees. Un test non paramétrique de quantile est alors appliqué afin de décider si le compromis est significativement meilleur que celui que l'on obtient *par chance après permutation*. De plus, un résultat récent [15] permet désormais d'exprimer les premiers moments de la distribution d'un coefficient RV sous l'ensemble des permutations des labels des produits au sein de l'un des deux tableaux comparés par ce coefficient. Moyennant une hypothèse de normalité de cette distribution et quelques adaptations, ce résultat permet de définir un test de validité du compromis de STATIS qui ne requiert le calcul effectif d'aucune permutation et qui s'avère être de la plus grande utilité pour la validation du profil sensoriel [27].

Qu'il s'agisse d'APG ou de STATIS, la clef de l'analyse du profil sensoriel réside donc dans le calcul de distances inter-produits. Ces distances dépendent du choix des descripteurs et de la métrique utilisés pour le calcul des distances. Le profil libre-choix, le profil libre-choix simulé et d'autres stratégies de choix de descripteurs peuvent conduire à des fiches de dégustation fortement redondantes. L'utilisation de la métrique de Mahalanobis permet alors d'équilibrer les poids des différentes dimensions sensorielles sous-jacentes dans le calcul des distances inter-produits. De plus, les tests de validation de compromis par permutation peuvent alors fournir un critère de comparaison de différentes stratégies de choix de variables et de métrique [26]; la meilleure stratégie étant celle qui maximiserait la validité du compromis.

3. Prédire le sensoriel par des mesures instrumentales

Il existe toujours chez les technologues de l'alimentaire le vieux rêve de la machine qui remplacerait la perception humaine. Le problème est qu'aucune théorie nous indique comment construire cette machine. Ainsi, le chercheur se tourne vers l'expérimentation et l'emploi de modèles statistiques de corrélation entre les jeux de données sensorielles et instrumentales. Une liste de conditions [19, 26], toutes nécessaires à la réussite de l'étude, peut être établie :

- Existence d'un lien de causalité entre l'instrumental et le sensoriel
- Identité des produits dans les deux études
- Échantillonnage des produits représentatif du domaine de prédiction
- Produits tests et re-échantillonnage pour validation du modèle
- Corrélation à un ou plusieurs compromis sensoriels valides
- Discussion du niveau de mesure des variables
- Prise en compte de liaisons non linéaires

Les méthodes statistiques les plus employées sont la régression linéaire multiple, la régression sur composantes principales, l'analyse des corrélations canoniques, l'ACPVI (ou *redundancy analysis*), la PLS régression (*Partial Least Square*) [20–21] et enfin, très récemment la *continuum regression* [3, 33]. Un avantage que nous avons trouvé à l'ACPVI est la possibilité de sélectionner un sous-ensemble de variables instrumentales optimal [9, 30] pour la corrélation avec le sensoriel. De plus, ces méthodes peuvent être couplées à un pré-traitement de type codage optimal qui permet de prendre en compte des liaisons plus ou moins non linéaires. Comparer le pouvoir prédictif de chacune de ces stratégies à partir de données réelles afin de définir des conditions et domaines d'applicabilité est un travail attendu par la communauté sensorielle.

4. Modélisation sensorielle des préférences des consommateurs

L'acceptabilité d'un aliment par le consommateur dépend dans une certaine mesure des caractéristiques sensorielles du produit. Mais il est plus ou moins admis que l'on ne peut pas demander à un consommateur de réaliser du profil sensoriel. Ainsi, la description sensorielle sera réalisée par un jury entraîné et seule une note de préférence pour chacun des produits sera demandé à chaque consommateur. Les méthodes proposées dans le paragraphe précédant peuvent évidemment être utilisées dans ce contexte où les préférences joueraient le rôle du sensoriel et le sensoriel celui des mesures instrumentales. Mais dans ce contexte, les méthodes de cartographie des préférences [13, 23, 31] permettent alors de segmenter la population des consommateurs en fonction de leurs préférences et de proposer pour chaque segment des modélisations sensorielles. Ces méthodes proviennent de la psychométrie [4], mais reçoivent depuis quelques années un intérêt croissant dans le domaine sensoriel. Les développements actuels autour de ces méthodes portent sur la validation des modèles et la généralisation aux plans incomplets.

Ces méthodes sont hautement stratégiques pour l'industrie agro-alimentaire, car elles visent l'essentiel : mesurer et expliquer les préférences des consommateurs en vue d'une meilleure formulation des produits. Cette problématique, en tant que lieu de rencontre des secteurs R&D et marketing dans l'entreprise, est certainement la question la plus valorisante que doit aborder la recherche en statistique appliquée à l'évaluation sensorielle.

5. Conclusion

À l'issue de ce panorama, non exhaustif, de l'implication des statistiques dans l'évaluation sensorielle, nous proposons une réponse positive à la question soulevée dans le titre de cette communication, tout en rappelant que si les statistiques sont nécessaires à l'évaluation sensorielle, elles ne lui sont pas suffisantes. Mais, les progrès en évaluation sensorielle ont souvent eu lieu grâce à l'adaptation d'un outil statistique, connu ailleurs, à une idée nouvelle de protocole sensoriel. Ainsi, deux conditions nécessaires à une activité de recherche en méthodologie de l'évaluation sensorielle sont donc d'une part une culture statistique profonde

et étendue, et d'autre part une bonne compréhension des besoins des clients de l'évaluation sensorielle. La première condition ne peut être remplie que par une véritable formation de statisticien, alors que la seconde s'apprend forcément *sur le tas*. Notre message à l'attention des décideurs de l'industrie agro-alimentaire est donc de confier une part importante de la responsabilité de l'évaluation sensorielle à des statisticiens ouverts à l'application. L'enseignement des statistiques dans notre pays doit se tenir prêt à pouvoir répondre efficacement à de telles offres.

Mots-Clefs : Agro-industrie, évaluation sensorielle, mesures instrumentales, études consommateurs, statistique appliquée, analyse de la variance, analyse procustéenne, coefficient RV, ACPVI, STATIS, cartographie des préférences.

6. Références

1. Amerine, M.A., Pangborn, R.M. & Roessler, E.B. (1965). *Principles of Sensory Evaluation of Food*. Academic Press, New York.
2. Arnold, G. M. & Williams, A. A. (1986). The use of generalized procrustes techniques in sensory analysis. In *Statistical procedures in food research*. Pigott, J.R. Ed. Elsevier Applied Science. 233-253
3. Brockhoff, P., Skovgaard, I., Poll, L. & Hansen, K. (1993). A comparison of methods for linear prediction of apple flavour from gas chromatographic measurements. *Food Quality and Preference*, 4, 215-222.
4. Carroll, J.D. (1980). Individual differences and multidimensional scaling. In *Multidimensional Scaling : Theory and Applications in the behavioral Sciences (Voll.)*. Shepard, R.N., Romney, A.K. & Nerlove, S.B. (Eds). Seminar Press, New-York, 105-155.
5. Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Second Edition. Lawrence Erlbaum Associates, Hillsdale.
6. Dijksterhuis, G.B. & Gower, J.C. (1991/2). The interpretation of generalized Procrustes analysis and allied methods. *Food Quality and Preference*, 3, 67-87.
7. Dijksterhuis, G.B. & Punter, P. (1990). Interpreting generalized Procrustes analysis 'Analysis of variance tables'. *Food Quality and preference*, 2, 255-265.
8. Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, 751-760.
9. Escoufier, Y. & Robert, P. (1979). Choosing variables and metrics by optimizing the RV coefficient. In *Optimizing Methods in Statistics*. Rustagi, J.S. (Ed). Academic Press, New York, 205-219.
10. Falmagne, J.C. (1985). *Elements of Psychophysical Theory*. Oxford University Press, Oxford.
11. Gacula, M.C. & Singh, J. (1984). *Statistical Methods in Food and Consumer Research*. Academic Press, Inc., Orlando, 128-140.
12. Gower, J. C. (1975). Generalized Procrustes analysis. *Psychometrika*, 40, 33-51.
13. Greenhoff, K. & MacFie, H.J.H. (1993). Preference Mapping. In *Measurements of Food Preferences*. MacFie, H.J.H. & Thomson, D.M.H. (Eds). Elsevier Applied Sciences, London, In press.
14. Kazi-Aoual, F. (1992). Apport de la méthode STATIS pour étudier l'homogénéité d'un panel. *Proceedings 3èmes journées agro-industrie et méthodes statistiques*, Montpellier, 30 Novembre et 1er Décembre 1992.
15. Kazi-Aoual, F. (1993). Approximations to permutation tests for data analysis. *Rapport de Recherche n° 93-06*. Université Montpellier II, Unité de Biométrie.
16. Lavit, Ch. (1988). *Analyse conjointe de tableaux quantitatifs*. Masson, Paris.

17. Lavit, Ch., Escoufier, Y., Sabatier, R. & Traissac, P. (1993). The ACT (STATIS Method). *Computational Statistics and Data Analysis*, In press.
18. MacFie, H.J.H., Bratchell, N., Greenhoff, K. & Vallis, L.V. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies*, 4, 129-148.
19. MacFie, H.J.H. & Hedderley, D. (1993). Current practice in relating sensory perception to instrumental measurements. *Food Quality and Preference*, 4, 41-49.
20. Martens, M. & Martens, H. (1986). Partial least squares regression. In *Statistical procedures in food research*. Pigott, J.R. (Ed). Elsevier Applied Science. 293-359.
21. Martens, H. & Naes, T. (1989). *Multivariate Calibration*. John Wiley & Sons.
22. Meilgaard, M., Civille, G.V. & Carr, B.T. (1991). *Sensory Evaluation Techniques*. CRC Press Inc., Boca Raton.
23. Nute, G.R., MacFie, H.J.H. & Greenhoff, K. (1988). Practical applications of preference mapping. In *Food Acceptability*, Thomson, D.M.H. (Ed). Elsevier Science Publishers Ltd, London, 377-386.
24. Schlich, P. (1993). Risk tables for discrimination tests. *Food Quality and Preference*, 4, 141-151.
25. Schlich, P. (1993). Uses of change-over designs and repeated measurements in sensory and consumer studies. *Food Quality and Preference*, 4, 223-235.
26. Schlich, P. (1993). Contributions à la Sensométrie. *Thèse de Doctorat en Sciences de l'Université Paris XI Orsay*.
27. Schlich, P. (1993). RV Coefficient & STATIS : Useful multivariate statistical methods for sensory evaluation. In *Food 2000 Preservation: integrating Processing, Packaging and Consumer Research*, Boston, 19-21 October 1993. To be published by Science & Technology Corp.
28. Schlich, P. (1994). GRAPES : a method and a program for Graphical Representation of Assessor PErformanceS. *Journal of Sensory Studies*, 9, will appear in #4.
29. Schlich, P. & Issanchou, I. (1990). Les méthodes de profils et leurs alternatives. In *Agro-industries et méthodes statistiques*, Angers, 14-15 Juin 1990. 254-272.
30. Schlich P., Issanchou S., Guichard E., Etievant P. & Adda J. (1987). RV coefficient: A new approach to select variables in PCA and to get correlations between sensory and instrumental data. In *Flavour Science and Technology*, Martens, M., Dalen, G.A. & Russwurm H.Jr. (Eds), Wiley & Sons Ltd, Chichester, 469-474.
31. Schlich, P. & McEwan, J.A. (1992). Cartographie des préférences. Un outil statistique pour l'industrie agro-alimentaire. *Sci. Alim.*, 12, 339-355.
32. Stone, H. & Sidel, J.L. (1985). *Sensory Evaluation Practices*. Academic Press Inc., Orlando.
33. Stone, M. & Brooks, R.J. (1990). Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *J. R. Statist. Soc. B*, 52, 237-269.
34. Wakeling, I.N., Raats, M.M. & MacFie, H.J.H. (1992). A new significance test for consensus in generalized procrustes analysis. *Journal. of Sensory Studies*, 7, 91-96.
35. Williams, A.A. & Langron, S.P. (1984). The use of free-choice profiling for the evaluation of commercial ports. *J. Sci. Food Agric.*, 35, 558-568.
36. Williams, E.J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Aust. J. Sci. Res. Ser. A*, 2, 149-168.

Article 2

2. Schlich P. (1995). Preference mapping: relating consumer preferences to sensory or instrumental measurements. *Bioflavour 95. Analysis/Precursor studies/Biotechnology*. P. Etiévant and P. Schreier. Paris, INRA Editions: 135-150.

Preference Mapping : Relating Consumer Preferences to Sensory or Instrumental Measurements

P. SCHLICH

*INRA, Laboratoire de Recherches sur les Arômes,
17 rue Sully, BV 1540, 21034 Dijon cedex, France*

ABSTRACT

How can successful new products be developed ? Preference mapping is a set of statistical procedures, based on principal component analysis, cluster analysis and multiple polynomial regression, which can actually contribute to answering this question. It has been used mainly by the food industry in the UK. Any kind of industry where human sensory perception of products is involved could use these techniques. The present paper proposes these techniques, together with some recent developments, to the flavour industry. Preference mapping compares a set of competitive products, whose overall acceptability has been scored by a panel of consumers, and whose sensory properties have been profiled by a trained panel of assessors. Internal preference mapping derives a preference map of the products, effectively describing the main underlying preference dimensions, whereas preference clustering identifies groups of consumers having similar preferences. External preference mapping locates ideal products on the sensory map of the products. These ideal products, tailored for different groups of consumers, should become the target of the formulation.

INTRODUCTION

Sensory analysis is widely used in the R&D departments of the food industry, because it makes it possible to declare whether food products are perceived as different or not (difference testing) and to describe their differences by asking a trained panel of subjects to score a list of sensory attributes (sensory profiling). But in the end, the food industry must formulate products actually expected by the consumers to become successful new products. Although the sensory properties of food are likely to be responsible for food acceptability, it is almost impossible to ask a consumer to do sensory profiling. Therefore, the products to be compared are tasted in blind by a panel of consumers, who are only asked to give an overall preference score to each product. Internal preference mapping analyses these data for themselves, whereas external preference mapping correlates them to the sensory profiles of the same products given by a different panel of trained assessors. Therefore, internal analysis identifies patterns of preference, whereas external analysis proposes sensory explanations of these patterns, which should help either to improve current products or to develop successful new products.

Preference mapping techniques were first originated in psychometry in the USA [1-3]. They were introduced later in food science in the UK [4-6]. Three years ago, these techniques

were presented by the author to the French food industry [7]. The present paper aims to present these techniques to the flavour industry, together with some recent developments achieved by the author since that first paper.

DATA

General data for preference mapping

A preference mapping study includes n products to be compared. They can be either competitive products from the market or different possible developments of a new product. The number n is most often between 6 and 12, but a minimum number of 10 products is strongly advisable. Two different kinds of sensory data are collected from these products : the consumer preference scores and the sensory profiles.

The consumer preference scores are obtained from a panel of p "naïve" consumers being sampled in order to be representative of a targeted population. The number p is usually about a few hundred. Each consumer is asked to give an overall liking score to each product using any kind of ordinal or interval scale. These data can also be the ranks of the products, either directly given by the consumer or derived from his own scores. The preference scores are arranged into a matrix Y having n rows (products) and p columns (consumers).

Sensory profiling is the process by which a group of trained assessors scores a number of attributes for each product of the study. The analysis of profiling data by multivariate techniques, such as Principal Component Analysis (PCA) [8] of the mean score products, Generalized Procrustes Analysis (GPA) [9-11] or Structuration des Tableaux A Trois Indices de la Statistique (STATIS) [12-13] enables to define a sensory consensus about the product differences among the assessors. This consensus is a k -dimensional map of the products. The product coordinates on this map are arranged into a matrix X having n rows (products) and k columns (dimensions). For the purpose of external preference mapping, it is strongly recommended to standardize each column of X to the same variance.

Crisps data

These data are available in the annexes of [7]. They are used in this paper in order to illustrate new developments in preference mapping.

These data include 10 samples of crisps (A, B, ..., J) from the UK market. The texture and the flavour of these samples were profiled by a trained panel of 8 assessors. For each sample, each assessor scored the 11 following attributes listed in alphabetical order : artificial, crisp, dissolve, earthy, fatty, hard, overcooked, salty, stale, sweet, teethclogging. These sensory data were processed by GPA, from which a 2-dimensional consensus map was derived and given in the annex 2 of [7].

The overall acceptability of the same samples were scored on an interval scale (0-30) by 99 consumers. These raw data were given in the annex 1 of [7].

6 competitiveness products

Six competitive products, available on the French market, were scored by 1049 consumers on an ordinal scale from 1 to 10. No sensory profiles of these products will be used in this paper and the type of the products must remain confidential.

METHODS

Internal Preference Mapping

Internal preference mapping is basically a PCA of the Y matrix containing the consumer scores. Therefore, the statistical observations of this PCA are the products, whereas the statistical variables are the consumers. This PCA should be done on the basis of

the covariance matrix between consumers in order to allow the consumer preferences to be either slight or strong. If, on the contrary, it is assumed that these differences of individual variance are artefacts due to individual ways of using scale, then one can perform this PCA on the basis of the correlation matrix between consumers. This latter PCA amounts to first dividing the scores given by each consumer by the standard deviation of them and to secondly performing a classical PCA on the results. Nevertheless, it is the opinion of the author that in most cases this standardization is not sensible.

Interpretation of this PCA is usually done thanks to a biplot [14] representation of both consumers and products on the two first axes. This biplot is defined as the superimposition of the two first vectors of loadings (eigenvectors of the covariance matrix with sums of squares equal to 1) and the two first principal components (centred data matrix times loadings which sum of squares equal to eigenvalue) on the same plot. A useful property of the biplot is that the orthogonal projection of the products onto a consumer arrow gives coordinates which approximate the real scores given by this consumer to the products. The quality of this approximation depends on the amount of information given by the two axes.

The internal preference map allows to visually identify groups of consumers pointing in the same direction, therefore preferring the same products. However, when the number of consumers is large, this visual inspection is not straightforward and there is a need for an automatic tool for identifying homogeneous clusters of consumers.

Although it is external preference mapping which is dedicated to relating sensory properties to preference scores, some scientists, mostly in the UK, prefer using internal preference mapping for that purpose. They are correlating the product coordinates in the internal map to the mean score products of each sensory attribute scored by the trained panel, making it possible to superimpose these sensory attributes onto the internal map. Although this technique enhances the information given by the internal map, the author first argues that it cannot replace external preference mapping and secondly suspects that it is done to overcome technical and graphical difficulties attached to external preference mapping models. It is the aim of this paper to clarify these difficulties.

Internal Preference Clustering

Although clustering the consumers is basically a problem of cluster analysis applied to the variables instead of the observations, any classical technique of cluster analysis of observations can be applied to the transposed matrix of consumer scores, in which the consumers are the rows and the products are the columns, provided that the lines of this matrix have been previously centered. However some dedicated algorithms exist for directly clustering the variables.

The author uses the technique of Oblique Centroid Factor Analysis (OCFA) [15], available through the VARCLUS procedure in the SAS[®] software [16]. This technique defines groups of consumers maximising the sum of the variance of the mean score products within clusters. It is a descending splitting algorithm which starts by making two clusters of consumers, then it splits the less homogeneous cluster into two new clusters allowing exchanges of consumers among clusters. This process is reiterated until a convergence criterion is met.

Although comparing different cluster analysis techniques is out of the scope of this paper, the author found that OCFA behaved slightly better (averaged F ratio of product effect within clusters) than the hierarchical cluster analysis using the Ward's criterion (procedure CLUSTER of the SAS[®] software) on the line-centered transposed data matrix (unpublished results).

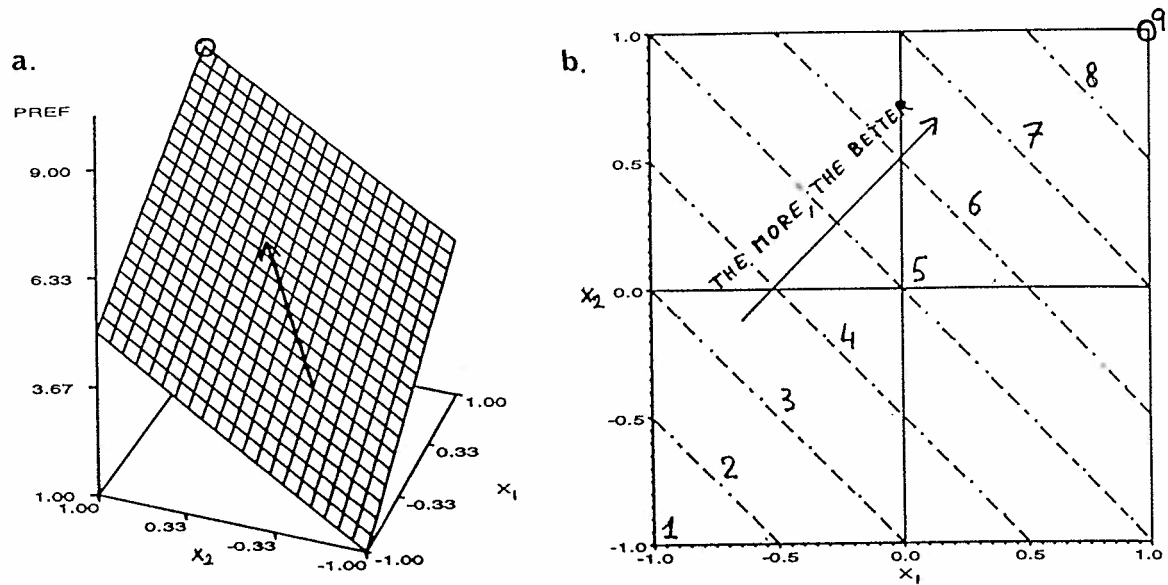


Figure 1 : Vectorial model $PREF = 5 + 2X_1 + 2X_2$; a. response surface, b. isocontours

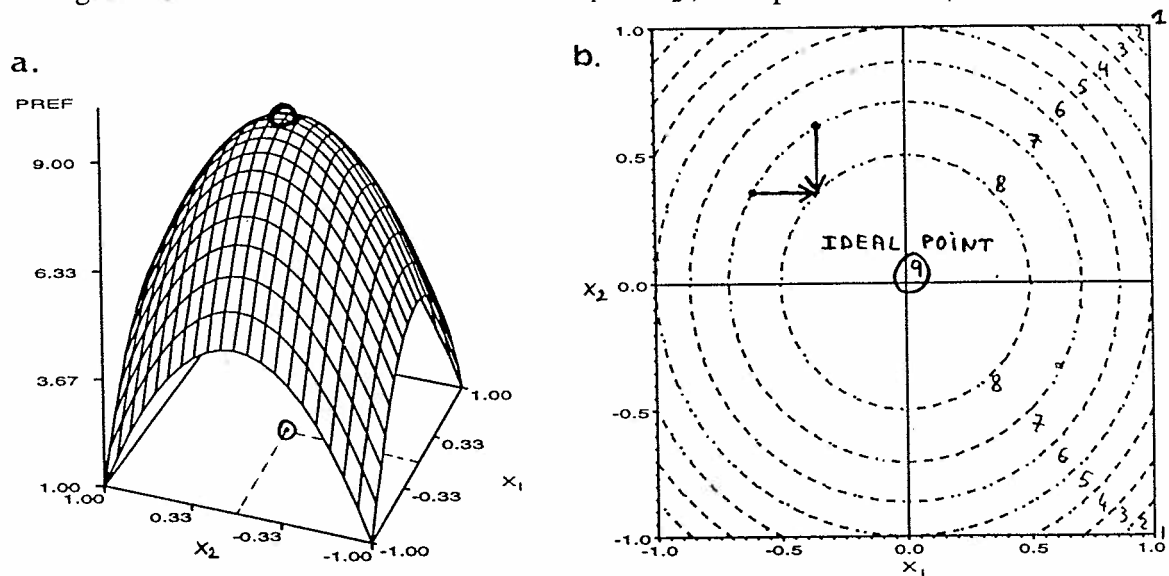


Figure 2 : Circular model $PREF = 9 - 4X_1^2 - 4X_2^2$; a. response surface, b. isocontours

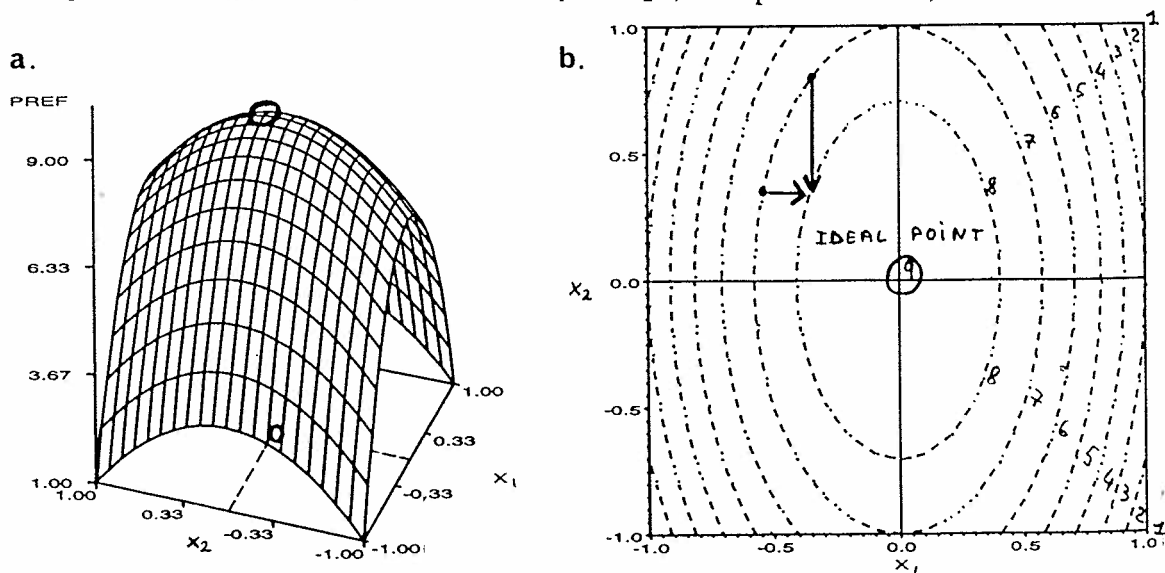


Figure 3 : Elliptical model $PREF = 9 - 6X_1^2 - 2X_2^2$; a. response surface, b. isocontours

External Preference Mapping

In order to simplify the presentation of external preference mapping techniques, it is assumed that the sensory map of the products is bidimensional. Therefore, the $\mathbf{X}$ matrix contains 2 columns $\mathbf{X}_1$ and $\mathbf{X}_2$ giving the coordinates of the products on the sensory map. When the sensory map is more than bidimensional, the following mathematical definitions of external preference mapping models are still possible, providing that the number of products is large enough, however the graphical interpretation is not straightforward and, to the knowledge of the author, there is no published example of that situation.

Basically external preference mapping is a polynomial regression [17] of each individual consumer $\mathbf{Y}_i$, $i=1, \dots, p$, onto the independent variables $\mathbf{X}_1$ and $\mathbf{X}_2$. There are 4 different possible models called vectorial, circular, elliptical or quadratic. These models are nested: the vectorial model is a sub-model of the circular model, which is a sub-model of the elliptical model, which is finally a sub-model of the quadratic model. Each of these models is defined and described below thanks to response surfaces and isocontour plots of examples of preference scores given on a 9-point-scale.

Vectorial model

The vectorial model is the multiple linear regression of $\mathbf{Y}_i$ on $\mathbf{X}_1$ and $\mathbf{X}_2$:

$$\mathbf{Y}_i = \mathbf{a} + \mathbf{b}_1\mathbf{X}_1 + \mathbf{b}_2\mathbf{X}_2 \quad [1]$$

The response surface of an example (fig. 1.a) of such a model is a plane surface with a steepest direction (arrow) which leads to an optimal preference score equal to 9 for the sensory point corresponding to both $\mathbf{X}_1$ and $\mathbf{X}_2$ equal to 1 (circle). The author finds it easier to look at the isocontour plot (fig. 1.b), which in this case consists of parallel straight lines. The vectorial model is often called "the more, the better", indicating that for any consumer fitted by this model, the ideal product should be located as far as possible towards the arrow, therefore that the sensory properties correlated with this direction should be maximized. It is likely that some consumers prefer medium levels of some sensory attributes and therefore present ideal points located inside the sensory map rather than outside or on the border of this map. The following models allow to better fit such consumers.

Circular model

The circular model is the vectorial model plus a quadratic term:

$$\mathbf{Y}_i = \mathbf{a} + \mathbf{b}_1\mathbf{X}_1 + \mathbf{b}_2\mathbf{X}_2 + \mathbf{c}(\mathbf{X}_1^2 + \mathbf{X}_2^2) \quad [2]$$

The response surface of an example (fig. 2.a) of such a model is a bi-dimensional parabola, which leads to an optimal preference score of 9 corresponding to the center of the sensory space and indicating that this consumer is expecting a medium product in terms of the sensory attributes scored. The isocontours (fig. 2.b) of this surface are circles around this ideal point meaning that the weights of the two sensory dimensions are equally important for preference. However, it is likely that this latter assumption does not hold for some consumers, leaving a room for the next models. If the sign of the estimation of the quadratic coefficient $\mathbf{c}$ is positive, then the ideal point is a negative ideal point, meaning that for this consumer one should produce something as far as possible from this point in the sensory space. A property of the circular model is that the two sensory dimensions are equally important for determining preferences. It is likely that this property is not the usual case, which is the reason why the elliptical model must be considered.

Elliptical model

The elliptical model is the circular one but allowing for a different quadratic coefficient for each dimension:

$$\mathbf{Y}_i = \mathbf{a} + \mathbf{b}_1\mathbf{X}_1 + \mathbf{b}_2\mathbf{X}_2 + \mathbf{c}_1\mathbf{X}_1^2 + \mathbf{c}_2\mathbf{X}_2^2 \quad [3]$$

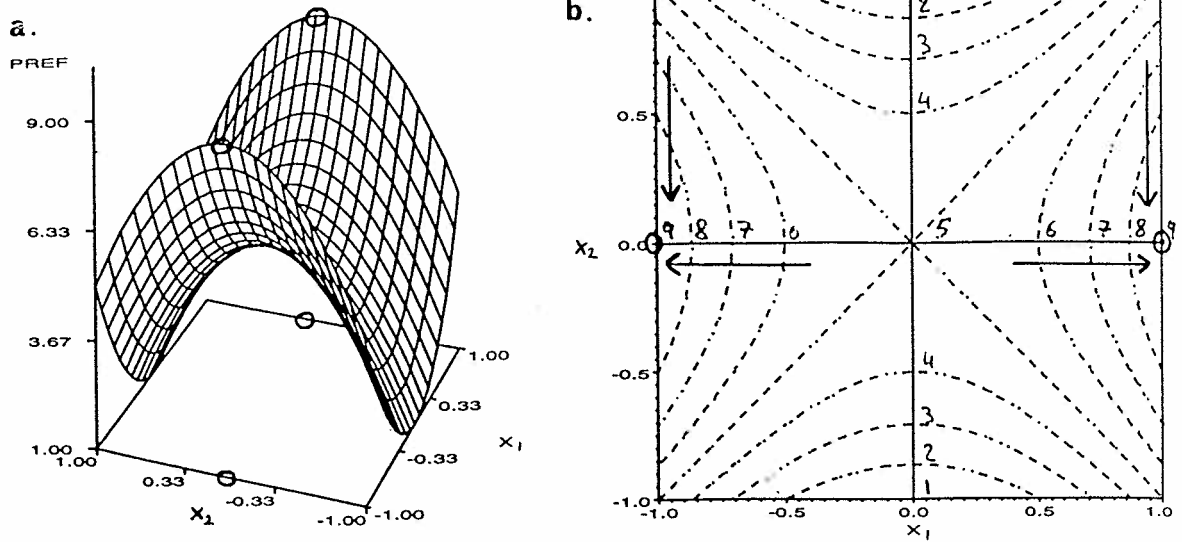


Figure 4 : Elliptical model with saddle point $PREF = 5 + 4X_1^2 - 4X_2^2$
a. response surface, b. isocontours

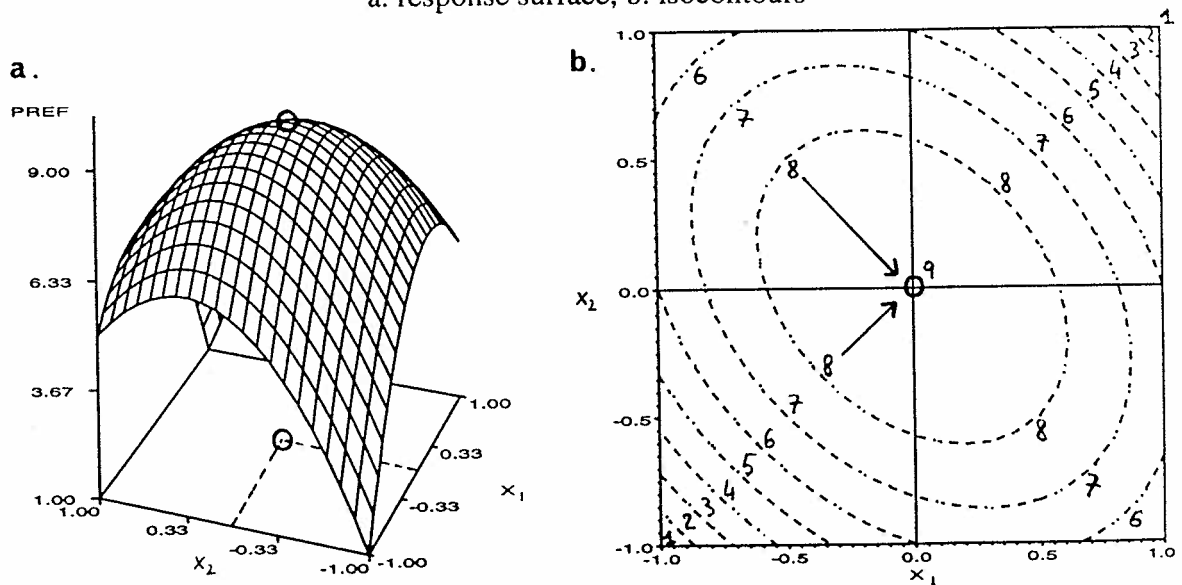


Figure 5 : Quadratic model $PREF = 9 - 3X_1^2 - 3X_2^2 - 2X_1X_2$
a. response surface, b. isocontours

In the first example of elliptical model, the weight of sensory dimension X_1 is 3 times the weight of X_2 leading to a response surface (fig. 3.a) which has a positive ideal point at the origin of the sensory space, but with elliptical isocontours around it (fig. 3.b). Therefore the preferences of this consumer are more sensitive along X_1 rather than along X_2 . As for the circular model, the ideal point is called negative when both estimations of the quadratic coefficients are negative. In this case, it is more important to avoid that point on the axis having the larger absolute value of quadratic coefficient rather than on the other.

But allowing the quadratic coefficients to be different, it is possible to get them with opposite signs, leading to the so-called saddle point. In the example (fig. 4.a), the response surface seems like a horse saddle with two different optimal points circled. The origin of the sensory map appears as a positive ideal point along X_2 and as a negative ideal point along X_1 (fig. 4.b). Therefore, this consumer seems to like the products having medium sensory properties related to X_2 and either low or high sensory properties related to X_1 .

Quadratic model

The full quadratic model is the elliptical model plus a cross-product term :

$$Y_i = a + b_1X_1 + b_2X_2 + c_1X_1^2 + c_2X_2^2 + dX_1X_2 \quad [4]$$

The response surface of an example of such a model (fig. 5.a) appears as an elliptical response surface being non parallel to the sensory axes. The most important sensory direction for preference is neither X_1 or X_2 , but it is the first bisector of the sensory map. Although this consumer prefers products having X_1 and X_2 at medium levels, if one of the basic dimension X_1 or X_2 decreases the other should compensate by increasing in the same magnitude in order to keep preference at the highest level possible (fig. 5.b). This kind of preference behaviour is perfectly possible in practice.

Model selection : AUTOFIT strategy

To the knowledge of the author, there are no published examples using elliptical or quadratic models. This lack is certainly due to the difficulty of interpreting elliptical ideal point which can even become a saddle point. As in Schlich and McEwan [7], it is most often either the vectorial or the circular model which is fitted to the whole group of consumers. Now the author thinks that any of the 4 models above is *a priori* acceptable and that a single model should be chosen *a posteriori* for each consumer on a statistical basis. In this paper, the AUTOFIT strategy is proposed to solve this problem. It is a sequence of classical statistical tests of the linear model, which seem to have been overlooked by the users of preference mapping. In a first step, the nested structure of the 4 models enables an individual model selection to be performed by means of a sequence of Fisher-ratio tests. In a second step, another statistical test enables to decide whether the selected model for a given consumer fits correctly enough his preferences or not.

For a given consumer, let SST and SSE_M be the total sum of squares and the residual sum of squares of model M and df_M be the number of degrees of freedom of SSE_M. The values of M can be V, C, E or Q standing for vectorial, circular, elliptical or quadratic model. For testing whether the sub-model M could replace the model N, where M is before N in the sequence of letters V, C, E and Q, it is known that the following statistic :

$$F_{M/N} = \frac{df_N(SSE_M - SSE_N)}{(df_M - df_N)SSE_N} \quad [5]$$

is distributed as a Fisher statistic with df_M-df_N and df_N degrees of freedom if the model M is acceptable in place of the model N. The difference sum of squares in the numerator of this statistic can be understood as the amount of error added when replacing the model N by the model M, whereas the denominator sum of squares is the error already included in the model N. The higher F_{M/N}, the less acceptable the model M is. The autofit strategy consists of testing first the elliptical model against the quadratic model, then, depending on the result of the first test, testing the circular model against either the elliptical or the quadratic model and then testing the vectorial model against the selected model from the second test. This process includes 3 out of the 6 possible tests and selects a single model for each consumer. We are used to setting the level of significance of these tests at 0.10 and this p-value is called the level of choice. Increasing this p-value would complicate the models, whereas decreasing it would simplify the models.

Assuming that the model M has been selected by the previous sequence of tests, the second step of the AUTOFIT strategy will now concentrate on whether this model fits the data or not. The proportion of variation explained by the model M, called R_M², is given by :

$$R_M^2 = 1 - SSE_M / SST \quad [6]$$

In order to test whether this R_M^2 is significant, that is whether the model M is acceptable or not, it is known that the statistic below :

$$F_X = \frac{df_M R_M^2}{(n-df_M)(1-R_M^2)} \quad [7]$$

is distributed as a Fisher statistic with df_M and $n-df_M$ degrees of freedom if the model M does not fit the data enough. We are used to setting the level of significance of these tests at 0.25 and this p-value is called the level of selection. The rationale of such a high p-value is to avoid false rejection of consumers, but we are aware that on the average one consumer out of four may just be selected by chance. At the end of the process, each consumer can be either selected with a single model or rejected.

Producing external maps

In Schlich and McEwan [7] an external map was produced for the vectorial model and another one for the circular model. The vectorial map consisted of the sensory map of the products on which each consumer was represented as an arrow pointing towards his direction of increasing preference. The length of the arrow was defined proportional to the R^2 of the regression, that is proportional to the quality of the fit. The circular map locates on the sensory map the individual ideal points as sign + or -, depending on whether these ideal points were positive or negative. Each of these signs was drawn with a size proportional to the quality of the fit.

With the AUTOFIT strategy a single map can be produced, because each consumer is fitted by a single model, or no model at all when this consumer is rejected. The location of a saddle point is represented by an "x", if the location is simultaneously a positive ideal on horizontal axis and a negative ideal on vertical axis, or by a "y" for the opposite situation, that is negative ideal on x-axis and positive ideal on y-axis. The location of a saddle point from quadratic model cannot be interpreted without the information of the orientation of the surface. Therefore the location of such a point is represented by a "z", and if a concentration of several "z" signs occurs one should look at the response surface of the corresponding consumers.

In general, looking at individual response surfaces is almost impossible, because of the large number of consumers. But one can average consumers fitted by the same model and presenting an ideal or saddle point located in the same area of the sensory map.

Sequence of analyses

To summarize the methodological part of this paper, it is worth recalling the sequence of analyses which can be performed :

- Internal preference mapping to get a picture of the preference space
- External preference mapping of consumers with the AUTOFIT selection strategy to decide whether a consumer fits to sensory and to select a model for this relation. A single external map of ideal and saddle point is produced. Visual clustering of consumers can be done on this map and the vectors of product mean scores by clusters can be investigated thanks to their isocontour plots.
- Preference clustering of consumers to identify the different patterns of preference. The vectors of product mean scores by clusters can become the input of the external analysis, making it possible to investigate the response surface of each cluster by drawing its corresponding isocontour plot.

RESULTS

Internal preference mapping of crisps data

The simplest analysis of the consumer scores is a two-way analysis of variance including the product and the consumer effects [7]. The Fisher statistic of the product effect was highly significant, but the R^2 was not very high, denoting that the comparison of product means is certainly not a complete description of these data. Nevertheless, the samples D and C were significantly less preferred by the group, whereas the sample H seems to be the best on the average over the panel.

The internal preference map (figure 6) is a biplot, on which the stars represent the products and the arrows the consumers. A consumer arrow points towards the direction of increasing preferences of this consumer. For instance, the arrow of consumer 79 (bottom left of figure 6) is apart from the other arrows, because of the particular preferences of this consumer. The orthogonal projection of samples onto his arrow demonstrates indeed that this person preferred sample D and C to the others. The sample coordinates on the first axis are roughly a graphical display of the product mean scores over all the consumers. But the second axis splits the population into 2 very different groups : at the top those who like crisps F and B and at the bottom those who do not like these 2 products. This last information was not given by the traditional analysis of variance. In fact, internal preference mapping can be seen as a way to investigate the product by consumer interaction when no replication is available.

But as the variance explained by figure 6 represents only 42 % of the total information, subsequent principal components should be considered. Most often conjoint interpretation of several principal plots is rather cumbersome, compared to the output of a clustering technique of consumers, as it will be demonstrated in the following example.

Internal preference mapping and clustering of the 6 competitive products

Internal preference mapping

Figure 7 is the internal map where the locations of the 6 products are given by numbers from 1 to 6, representing the rank of the product means, and where the consumer arrows have been replaced by stars to make the graph easier to read. Nevertheless, it is still impossible to interpret such a plot where any kind of preference seems to occur. Moreover and once again, this plot accounts for less than half of the total information, requiring at least two principal plots to be interpreted. As the total number of axes is 5, such an analysis would not be a dimensional reduction of the data, losing consequently its interest.

Internal preference clustering

The OCFA technique, described in the method section of this chapter, was applied to these data and several partitions of the consumers were investigated. It turned out that the partition in 5 clusters (table 1) would be sensible. The cluster sizes are quite balanced (Freq column in table 1) and the product effect (F column) is much more important within each cluster compared to the whole consumer panel (All line). Although no preferred, nor rejected, products clearly exist for the whole panel of consumers, it is almost possible to find a winner (in bold in table 1) and a loser (underlined in table 1) product for each cluster, and, of course, these products are different among clusters. Identifying such patterns of preferences in a simple table is definitely the strength of preference clustering compared to internal preference mapping.

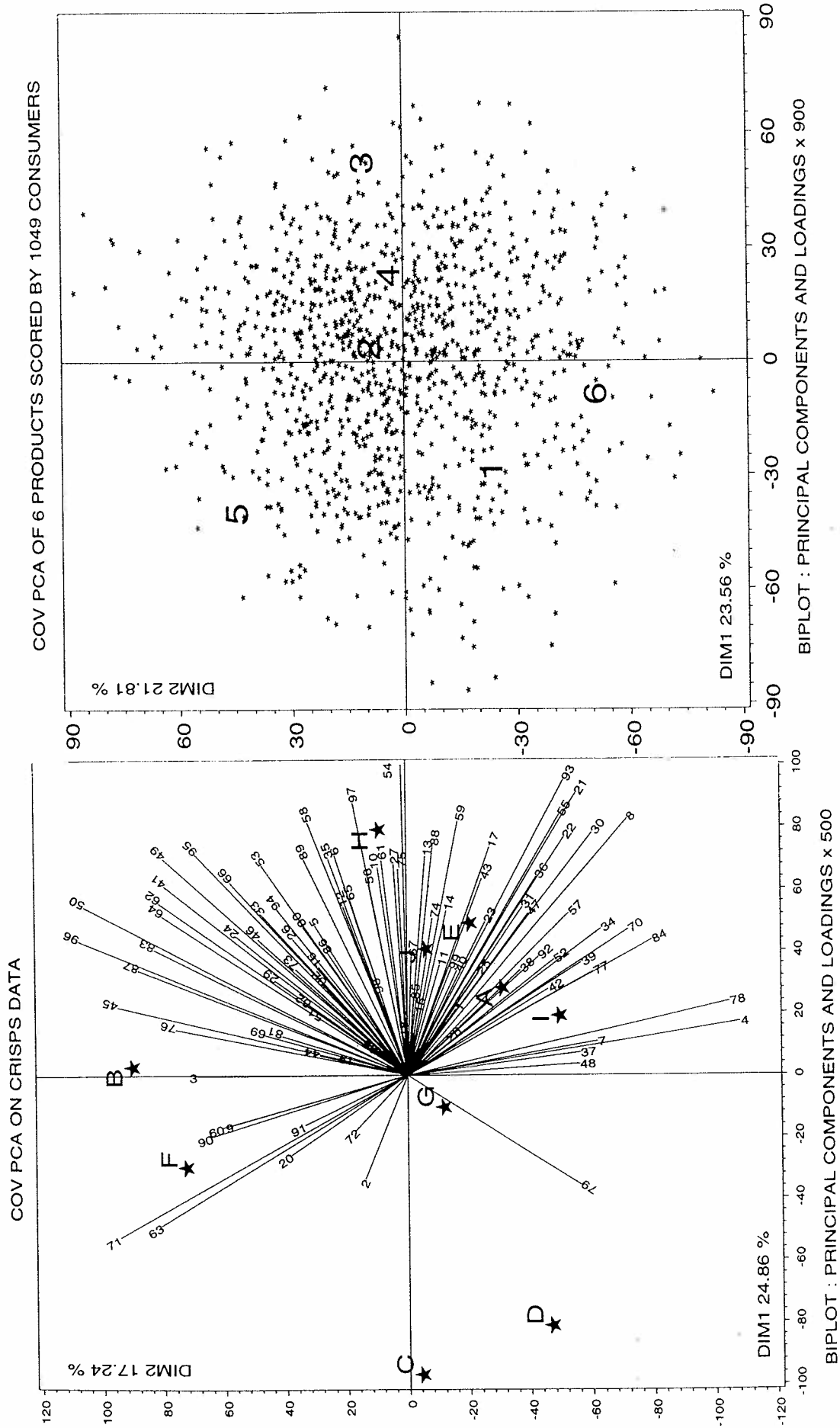


Figure 6 : Internal preference mapping of crisps data

Figure 7 : Internal preference mapping of 6 competitive products

Table 1. Mean scores of the 6 competitive products by clusters of consumers

Cluster	Freq	F ¹	Product Mean Scores ²					
			1	2	3	4	5	6
1	232	148	6.47^d	5.40 ^c	<u>2.83^a</u>	4.72 ^b	6.62^d	5.73 ^c
2	201	165	6.26 ^c	5.68 ^b	6.01 ^{bc}	7.18^d	5.82 ^{bc}	<u>2.87^a</u>
3	227	116	<u>3.08^a</u>	5.62 ^c	5.37 ^c	6.39^d	4.38 ^b	5.77 ^c
4	172	145	7.18^d	4.97 ^b	6.23 ^c	5.39 ^b	<u>2.81^a</u>	6.35 ^c
5	217	118	5.08 ^b	5.76 ^c	7.05^d	<u>3.44^a</u>	6.10 ^c	4.59 ^b
All	1049	7	5.53 ^b	5.50 ^b	5.42 ^b	5.40 ^b	5.25 ^{ab}	5.06 ^a

¹ : Fisher ratio test of PRODUCT effect in ANOVA : CONSUMER + PRODUCT

² : Two mean score products having the same letter are not significantly different (LSD, 5%).

For each cluster (line), the "winner" products are in bold, whereas the "loser" products are underlined

The point stressed by this clustering is that there is definitely room on the market for any of these 6 products. A traditional analysis of the whole group (All line) would have missed that important point. To define this room, one should characterize the clusters by demographic and sociological variables. But it might also be the sensory properties of these products which drove preference. These questions should be addressed in a subsequent paper.

External preference mapping of crisps data

Sensory map of the products

Figure 8 is the sensory map of the products obtained by GPA, on which the attributes were superimposed by means of the correlations between the attribute scores averaged over assessors and the product coordinates.

Crisps H on one side, and crisps F and B on another side are fairly different from the others. H was described as more artificial and to a lower extent more crisp. F and B were assessed as more dissolved and, to a lower extent, more fatty and stale, and less overcooked and teethclogging. The sweet attribute gathers these 3 samples as being sweeter than the others.

AUTOFIT and external preference mapping

Table 2 indicates that 59 % of the consumers were rejected due to lack of fit. This is an important conclusion missed by Schlich and McEwan [7]. One half of the selected consumers are fitted by a vectorial model and the others are more often fitted by elliptical or quadratic models rather than by circular model. This confirms that elliptical and quadratic models, most often neglected in published applications, must definitely be considered.

Table 2. AUTOFIT results of crisps data

Model	Total		Selected	
	Freq	Mean R ²	Freq	Mean R ²
Vectorial	76	0.306	19	0.539
Circular	9	0.704	9	0.704
Elliptical	10	0.869	9	0.893
Quadratic	4	0.901	4	0.901
All Models	99	0.423	41	0.688

Level of choice = 0.10 - Level of selection = 0.25

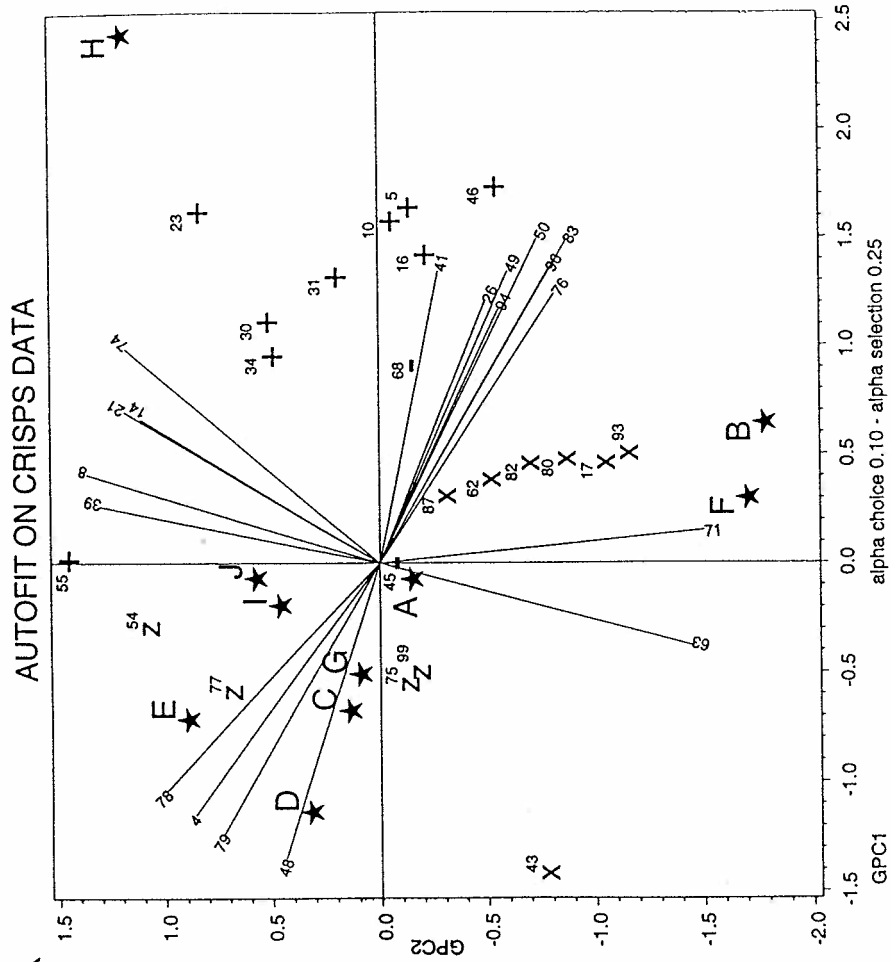


Figure 9 : External preference mapping of crisps data

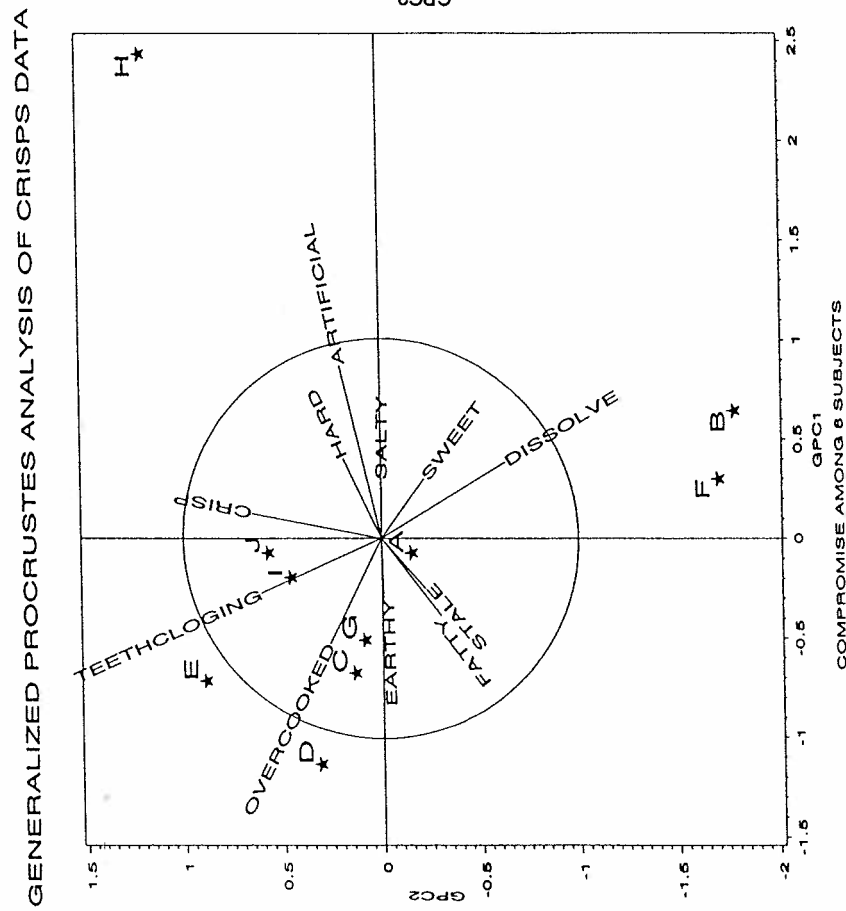


Figure 8 : Sensory map of crisps data

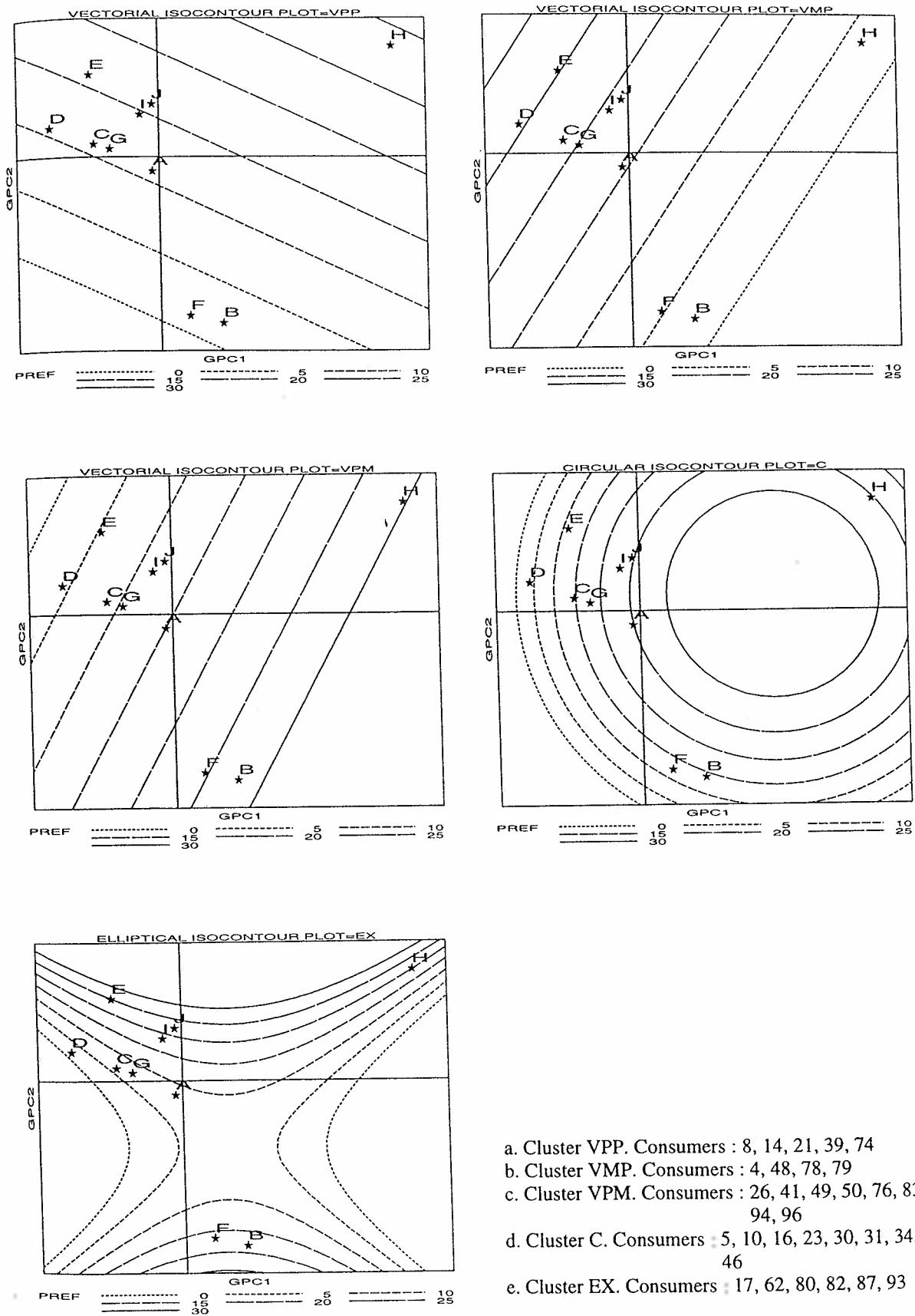


Figure 10 : Isocontour plots of the crisps data by cluster of consumers

Figure 9 is the external map of the selected consumers. Five clearly distinct groups of consumers can be identified on this map :

- a group of 5 vectorial consumers (8, 14, 21, 39, 74), called VPP, preferring crisps H. Consumer 55 could certainly be added to this group
- a group of 4 vectorial consumers (4, 48, 78, 79), called VMP, preferring crisps D and E
- a group of 8 vectorial consumers (26, 41, 49, 50, 76, 83, 94, 96), called VPM, preferring crisps H and F and B
- a group of 8 circular or elliptical consumers (5, 10, 16, 23, 30, 31, 34, 46), called C, having a positive ideal point between H and the origin of the sensory map, which can be seen as the medium crisps
- a group of 6 elliptical consumers (17, 62, 80, 82, 87, 93), called EX, having a saddle point of type x, which means that the location of product F and B is ideal for these people on the x-axis but anti-ideal on the y-axis. According to figure 8, for these people crisps F and B are sufficiently artificial but should be either much more dissolved or much less dissolved.

Because there were only four saddle points from quadratic consumers (z-point), which moreover were not located at the same place, we did not try to interpret them. Consumers 43, 45 and 68, who fit to the sensory space, have specific models which are different from the five groups identified. Consumers 45 and 68, who present a negative ideal point close to the origin, would prefer any sample as far as possible from the medium crisps.

Isocontour plots by cluster

Each of the previous visual clusters of consumers was summarized by its vector of mean score products, which was fitted to the model being common to consumers from this cluster. The 5 response surfaces are described by their isocontour plots given in figure 10 :

- The isocontours of group VPP (figure 10.a) show that a product close to H in the sensory map would insure full acceptability (maximal score of 30) for this group of consumers. The advice targeted to group VPP is therefore to copy the crisps H or even to make a crisps more artificial than H ("the more the better")
- The preference of people from vectorial group VMP (figure 10.b) would be maximized with products similar to D and E or even with products more teethclogging and overcooked than D and E
- People from vectorial group VPM (figure 10.c) would expect crisps as far as possible from the origin and on the median of the segment linking H to F and B. These consumers look for both artificial property of H and dissolved property of F and B
- For people from the circular group C (figure 10.d), the ideal crisps should be more artificial than the average but slightly less artificial than H. It is noticeable that as no crisps from the study occupied this location on the sensory map, there is room for a new formulation here
- People from the elliptical group EX (figure 10.e) require either crisps at the top, which means very crispy and definitely not dissolved, or crisps at the bottom similar to F and B. Although it is difficult in such a case to make a choice between these two alternatives, it remains that both require a product which should be similar to F and B at least on the first sensory dimension.

DISCUSSION

About the crisps data

The current analysis of this data is much more informative than that done in Schlich and McEwan [7]. Nevertheless, the sizes of the defined consumer clusters are really too small

to be the basis of new product development. A classical rule of thumb, coming from some central limit argument, says that 30 people are necessary to define a group of consumers. Therefore, assuming that one out two consumers will not fit to sensory and that k different patterns of preference are expected, one should enroll at least 60 times k consumers in the trial. Expecting 5 clusters in the crisps trial, more than 300 consumers would be required, whereas we got only 99.

Order of presentation of the samples to the consumers

For many reasons, the order of presentation of the products to the consumers must be balanced in order to get each product presented equally often in each position and followed equally often by each product or even by each sequence of products. This can be achieved thanks to William latin squares [18] or sets of mutually orthogonal latin squares first used by Cochran *et al.* [19] and defined, for instance, in John [20], Street and Street [21] or Jones and Kenward [22]. The use of these designs was recommended in the sensory field by Gacula [23], MacFie *et al.* [24], Schlich [25] and Wakeling and MacFie [26].

Number of products and number of consumers

As the quadratic model includes 5 parameters, at least six products are needed. But some degrees of freedom should be left for error estimation, therefore 8 products is really a strict minimum number of samples for preference mapping. But preference mapping techniques would be much more efficient with a total number of products close to 16. In such a case, would it be still possible to ask a consumer to score all the products ? If not, would it be possible to generalize preference mapping to incomplete designs ? If yes, depending on the total number of products in the study, would it be possible to advise as to the number of products that a consumer should score to insure validity of the results ? Would the loss of information involved by incomplete design be compensated for by increasing the number of consumers enrolled in the trial ? Research on these topics is currently being undertaken in the laboratory of the author.

Lack of fit, dimensionality and validity of the sensory space

Allowing for rejection of consumer, as AUTOFIT does, is necessary because a consumer doing a blind test may, on one hand, have no preference for the products, if he is just unable to perceive sensory differences between them, or, on the other hand, he may be unable to affect scores to his preferences. On the crisps data, more than one half of the consumers were rejected. Our experience of several data sets seems to indicate that about one half of the consumers never fit to the sensory space. Getting these people in a blind preference trial is useless. An important improvement would be to propose a strategy of consumer sampling to avoid choosing these people.

External preference mapping is most often presented when the sensory space is bidimensional. What if more than two sensory dimensions are necessary ? In such a case, the computations are still possible, but the presentation and the interpretation of the graphics are not straightforward. How could a valid number of dimensions be derived ? This is a knotty question in multivariate data analysis, which cannot be properly addressed without resampling techniques, which are not widely available in the statistical softwares yet.

What about external preference mapping when the sensory space is not a valid compromise among the trained assessors ? Detecting such a problem is one of the roles of multivariate analysis of sensory profiles, where recent progress has been made by the use of permutation tests [27-28]. When no multivariate compromise among the assessors can be defined, how could the consumer preferences still be related to the sensory attributes ? Research is in progress in the laboratory of the author on that topic.

CONCLUSION

Preference mapping methodology is in progress. The usefulness of clustering techniques and individual model selection by the AUTOFIT strategy has been demonstrated in this paper. The discussion section demonstrates that many other interesting problems for research are still open with preference mapping. This research is definitely a matter of statistics, which should interest any kind of industry where human sensory perception of products is involved.

ACKNOWLEDGEMENTS

The author thanks Jean McEwan (CFDRA, UK) for the crisps data and an industrial partner for the data set of 6 anonym competitive products. Help of Patricia Collin with english was greatly appreciated.

REFERENCES

- [1] CHANG J.J., CARROLL J.D., 1968. *Unpublished report*, Bell Telephone Laboratories.
- [2] CARROLL J.D., 1972. In R.N. Shepard, A.K. Romney, S.B. Nerlove (eds) : *Multidimensional Scaling : Theory and Applications in the behavioral Sciences (Vol1.)*. Seminar Press, New-York.
- [3] SCHIFFMANN S.S., REYNOLDS M.L., YOUNG F.W., 1981. *Introduction to multidimensional scaling. Theory, methods and applications*. Academic Press, London.
- [4] NUTE G.R., MACFIE H.J.H., GREENHOFF K., 1989. In D.M.H. Thomson. (Ed) : *Food Acceptability*. Elsevier Science Publishers Ltd, London.
- [5] JONES P.N., MACFIE H.J.H., BEILKEN S.L., 1989. *J. Sci. Food Agric.*, 47, 113-123.
- [6] GREENHOFF K., MACFIE H.J.H., 1994. In H.J.H. MacFie, D.M.H. Thomson (Eds) : *Measurement of food preferences*. Blackie Academic and Professional, Glasgow, 301p.
- [7] SCHLICH P., MCEWAN J.A., 1992. *Sci. Aliments*, 12, 267-355.
- [8] JOLLIFFE I.T., 1986. *Principal Component Analysis*. Springer-Verlag, New York.
- [9] GOWER J. C., 1975. *Psychometrika*, 40, 33-51.
- [10] DIJKSTERHUIS G.B., PUNTER P., 1990. *Food Quality and preference*, 2, 255-265.
- [11] DIJKSTERHUIS G.B., GOWER J.C., 1991/2. *Food Quality and preference*, 3, 67-87.
- [12] LAVIT CH., 1988. *Analyse conjointe de tableaux quantitatifs*. Masson, Paris.
- [13] LAVIT CH., ESCOUFIER Y., SABATIER R., TRAISSAC P., 1994. *Computational Statistics and Data Analysis*, in press.
- [14] GABRIEL K.R., 1971. *Biometrika*, 58, 453-467.
- [15] HARMAN H.H., 1976. *Modern Factor Analysis*. University of Chicago Press, Chicago.
- [16] SAS Institute Inc., 1989. *SAS/STAT® User's Guide, Version6, Fourth Edition, Volume 1 and 2*, SAS Institute Inc., Cary, NC, USA.
- [17] DRAPER N., SMITH H., 1981. *Applied Regression Analysis*. John Wiley & Sons, Inc., New York.
- [18] WILLIAMS E.J., 1949. *Aust. J. Sci. Res. Ser. A*, 2, 149-168.
- [19] COCHRAN W.G., AUTREY K.M., CANNON C.Y., 1941. *J. Dairy Sci.*, 24, 937-951.
- [20] JOHN P.W.M., 1971. *Statistical Design and Analysis of Experiments*. Macmillan, New York.
- [21] STREET A.P., STREET D.J., 1987. *Combinatorics of Experimental Design*. Oxford University Press, New York.
- [22] JONES B., KENWARD M.G., 1989. *Design and Analysis of Cross-Over Trials*. Chapman and Hall, London.
- [23] GACULA M.C., SINGH J., 1984. Academic Press, Inc., Orlando.
- [24] MACFIE H.J., BRATCHELL N., GREENHOFF K., VALLIS, L.V. 1989. *Journal of Sensory Studies*, 4, 129-148.
- [25] SCHLICH, P., 1993. *Food Quality and Preference*, 4, 223-235.
- [26] WAKELING I.N., MACFIE H.J.H., 1995. *Food Quality and Preference*, 6, in press.
- [27] WAKELING I.N., RAATS M.M., MACFIE H.J.H., 1992. *J. of Sensory Studies*, 7, 91-96.
- [28] SCHLICH P., 1995. In : T. Naes, E. Risvik. (Eds) : *Multivariate Statistics for Sensory Data*. Elsevier, in press.

Article 3

3. Schlich P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. *Multivariate Analysis of Data in Sensory Science*. T. Naes and E. Risvik. Amsterdam, Elsevier Science: 259-306.

DEFINING AND VALIDATING ASSESSOR COMPROMISES ABOUT PRODUCT DISTANCES AND ATTRIBUTE CORRELATIONS

Pascal Schlich

INRA, Laboratoire de Recherches sur les Arômes,
17 rue sully, BV 1540, 21034 Dijon cedex, France

1. INTRODUCTION

Sensory profiling of food is the process by which assessors give scores to a number of products for several attributes. The statistical analysis of profiling data requires multivariate techniques in which the attributes are the different variables. Principal Component Analysis (PCA; Jolliffe, 1986), the most basic multivariate technique, is widely used by sensory scientists to describe the data set composed of the product mean scores as the observations and the attributes as the variables. This is a reduced view of the data as it does not take into account the variance of the product mean scores due to individual differences. In the univariate framework, it is fairly accepted that no mean should be computed without its standard deviation. We argue that the same should be required on the multivariate side. Moreover, many people are used to performing this PCA on the basis of the correlation matrix, that is to perform the so-called normalized PCA. With this practice, an attribute having product means not significantly different, which is stated by means of an analysis of variance (ANOVA; Scheffe, 1959), has the same weight as a discriminant attribute among the products. To overcome this problem, one can include in PCA only the attributes being significant for the product effect. Although this practice protects to some extent against the previous problem, it suggests that the required method would be the Canonical Discriminant Analysis (CDA; Mardia *et al.*, 1979) of the product effect. CDA is indeed the natural multivariate extension of the one-way ANOVA. The input of this CDA is the full data set, where the assessors stand as replicates. But due to psychological and/or physiological reasons, the assessors may use the scale in different ways, and should therefore be considered as a block effect. Consequently, we would recommend centering each attribute to a zero mean by assessor as a pre-treatment of CDA. Although less important than centering, one could also consider standardization of each attribute to a unit variance by assessor. Unfortunately, CDA and these pre-treatments seem to have been overlooked by sensory scientists.

But the above mentioned flaws are even less important than the basic problem of sensory profiling, which can be summarized in a single statement : the attributes are not statistical variables. A statistical variable is a vector of n measurements of a random variate obtained on n experimental units characterized by their values for different controlled factors. Our point is to say that the meaning of the random variate is the same over the experimental units. In agronomy for instance, the yield or the plant size has the same meaning for every experimental

unit, whatever the levels of treatment and block factors are. In sensory profiling on the contrary, it is likely that the measurements of a given attribute from two different assessors measure two different sensory concepts. In such a case, computing for a given attribute a mean over the assessors can make no sense. Although good, but expensive, training of the assessors can substantially reduce this kind of problem, it is almost impossible to completely prevent it. Free-choice profiling (FCP; Williams & Langron, 1984), in which each assessor scores his own list of attributes, was a definite contribution to solving this basic problem. The variables being different among the assessors, neither PCA nor CDA can be applied to FCP, unless a separate analysis is performed by assessor. Although some individual analyses can be useful to elucidate particular points during an analysis of data, they are not practical enough to manage to be recommended as a basic analysis. Therefore, Williams and Langron (1984) proposed Generalised Procrustes Analysis (GPA; Gower, 1975; Arnold & Williams, 1986) to analyse FCP. The FCP-GPA coupling became popular in the sensory field and people realized that GPA could also be applied to conventional profiling data. Less well known is the fact that GPA makes possible a posteriori individual attribute selection leading to what can be called a "simulated free-choice profiling" (Schlich, 1993). Recent improvements in GPA were proposed thanks to applications in sensory analysis (Dijksterhuis & Punter, 1990; Dijksterhuis & Gower, 1991/2). Although a chapter in this book (4.2) fully describes GPA, one can say that GPA defines a consensus among assessors on the product differences by means of an iterative algorithm. Wakeling *et al.* (1992) define a test of the significance of the consensus based on permutations of the product labels by assessors. Although GPA is a significant improvement over PCA of the mean scores, a number of questions are still open concerning the analysis of sensory profiles:

- How can the dimensionality of individual sample space be estimated?
- How can the similarity between two individual sample spaces be measured?
- How can the agreement between two individual sample spaces be statistically tested?
- How can a consensus sample space be derived by an analytical solution?
- How can the significance of this compromise be tested without computing numerous permutations?
- How can several panels profiling the same products be compared?
- How can exchangeability of assessors among panels be tested?
- How can assessors be compared on the basis of attribute correlations instead of sample distances?

The present chapter aims to introduce in the sensory field a French statistical framework allowing firstly to deal with FCP and secondly to answer these questions. The basis of this framework is the RV coefficient (Escoufier, 1973; Robert & Escoufier, 1976), which is a generalised correlation coefficient between two sets of variables recorded from the same samples. It gives a way to quantify the agreement between two assessors about the sample differences. The RV coefficient is also useful in sensory science for relating sensory to instrumental measurements (Schlich *et al.*, 1987) and for analysing gas-chromatography data (Schlich & Guichard, 1989). The French acronym STATIS stands for "Structuration des Tableaux A Trois Indices de la Statistique", which could be translated into "structure of 3-way data sets in statistics". The technique was originated by L'Hermier des Plantes (1976) and was

fully described in Lavit (1988) and Lavit *et al.* (1994). It is a non-iterative 3-way multivariate analysis, based on the RV coefficient, which defines a compromise among assessors about the sample differences. Kazi-Aoual (1993) and Kazi-Aoual *et al.* (1995) defines an exact permutation test for RV coefficient which does not require numerous permutations to be actually performed. Another exact permutation test makes it possible to state whether the assessors from several panels profiling the same products can be considered as exchangeable among panels (Kazi-Aoual, 1993; 1992).

In the present chapter, these techniques are applied to an interlaboratory sensory analysis of 16 coffees evaluated by 11 different trained panels located in 8 different European countries. This experiment was organized and conducted by the European Sensory Network (ESN). ESN is currently completing a book presenting the results on that experiment, which also included consumer trials in 8 countries with half of the same coffees. Only a part of the profiling data is used in this chapter. The ESN book will make the whole set of coffee data available to the reader. More information about ESN can be obtained from the author.

2. METHODS

2.1 Raw data, sample weights and metrics

Let k be the total number of assessors and assume that assessor i scored p_i attributes ($i = 1, \dots, k$) for n samples. Let X_i be the matrix containing the scores of assessor i . The n rows of X_i are the samples, whereas the p_i columns of X_i are the attributes. Throughout this chapter, it is assumed that the attributes are centered within each X_i , that is per assessor. With the exception of Dual STATIS, every technique described in this chapter allows the attributes to be different among assessors, making it possible to handle FCP or any kind of individual selection of attributes.

Although in most applications every sample has the same weight ($1/n$), the techniques described in this chapter work with unequal weights. In this case, these weights are arranged into a diagonal matrix D of size n .

The question of choosing the attribute weight system, also called metrics, is more important. Metrics is a way of computing distances between samples. Generally speaking, it is a positive symmetric matrix Q of size p (the number of attributes) :

$$Q = (q_{lm})_{l,m=1,p} \quad (1)$$

As Q can be different among the assessors, sometimes it will be called Q_i . The squared distance between two samples $x = (x_l)$ and $y = (y_l)$, ($l = 1, \dots, p$), is given by :

$$d^2(x, y) = \sum_{l=1}^p \sum_{m=1}^p (x_l - y_l) q_{lm} (x_m - y_m) \quad (2)$$

For instance, in the framework of PCA two different metrics are commonly used . For the first one, Q is the identity matrix containing 1 as diagonal elements and 0 elsewhere, which

corresponds to the PCA of the covariance matrix and to the usual Euclidean distance :

$$d^2(x, y) = \sum_{l=1}^p (x_l - y_l)^2 \quad (3)$$

For the second one, Q is a diagonal matrix composed of the inverse of the attribute variances (σ_l^2 , $l = 1, \dots, p$), which corresponds to the PCA of the correlation matrix, also called standardized PCA, and to the weighted distance :

$$d^2(x, y) = \sum_{l=1}^p (x_l - y_l)^2 / \sigma_l^2 \quad (4)$$

As in the introduction section, where non standardized PCA was recommended for analysing the mean score products, we again recommend the non standardized PCA for analysing individual data sets. Many reasons can justify this choice, a few others can be found against it. It is out of the scope of this chapter to open such a discussion, because the techniques proposed in this chapter can accommodate any individual metrics. Furthermore, other metrics can be used with profiling data. For instance, we advocated (Schlich, 1993) the use of the individual Mahalanobis distance :

$$Q_i = (X_i' D X_i)^{-1} \quad (5)$$

when dealing with free-choice profiling, or even with conventional profiling as soon as different attribute correlation structures are expected among the subjects.

Finally, as Q is positive, it is possible to find a square matrix T of size (n, n) such as :

$$Q = T T' \quad (6)$$

Analysing X with metrics Q is thus equivalent to analyse :

$$Y = T X \quad (7)$$

with the classical identity metrics. Therefore, without loss of generality, it is assumed right now that the metrics Q is the identity and that the sample weights are all equal to $1/n$.

To summarize this section, the reader should remember that :

- Individual attributes are centered
- RV coefficient and STATIS work with any set of product weights (the same for each assessor) and with any individual metrics (which can be different by assessor)
- The selection of attributes and the choice of the metrics are two essential steps in the analysis of sensory profiles which are not addressed in this chapter.

2.2 Association matrix of individual sample space

For each assessor, the samples can be seen as a centered cloud of n points in a multidimensional space spanned by the attributes. The dimensionality of this individual sample space is equal at most to P_i :

$$P_i = \min(n-1, p_i) \quad (8)$$

The distance between two samples in this space measures the magnitude of the sensory differences between them. Because of correlations between attributes, it is likely that the main sample differences can be summarized by fewer dimensions than P_i . These new variables, called the principal components, are given by the successive eigenvectors of the association matrix of size (n, n) :

$$W_i = X_i X_i' \quad (9)$$

which contains the usual scalar products among the samples. The distance d_{uv} between samples u and v is linked to their scalar product W_{uv} through the formula:

$$d_{uv}^2 = w_{uu} + w_{vv} - 2w_{uv} \quad (10)$$

Inversely, because the columns of X_i are centered the formula (10) can be reversed into :

$$w_{uv} = -\frac{(d_{uv}^2 - d_{u.}^2 - d_{.v}^2 + d_{..}^2)}{2} \quad (11)$$

where :

$$d_{u.}^2 = (1/n) \sum_v d_{uv}^2 \quad \text{and} \quad d_{.v}^2 = (1/n) \sum_u d_{uv}^2 \quad (12)$$

$$d_{..}^2 = (1/n^2) \sum_{u,v} d_{uv}^2 \quad (13)$$

Finally, it should be remembered that the association matrix W_i contains the full information about the multidimensional differences among samples found by subject i . The association matrices are the basis of the subject comparisons in STATIS, which can therefore be done even if the assessors did not score the same attributes.

2.3 Estimating the dimensionality of an individual sample space

What PCA does is to decompose the total multidimensional variance of the sample space into successive non-correlated components which account for the maximum of this information. Precisely, the amount of variance given by the l -th principal component is the eigenvalue λ_l , of $(1/n)W_i$ ($l = 1, \dots, P_i$). A classical, but still difficult, question is to decide how many components should be analysed. We think that it is best to use resampling techniques such as

cross-validation or bootstrap (Efron & Tibshirani, 1993). However these techniques are not widely available within the statistical softwares yet and can be time-consuming. The β_i coefficient can actually be understood as an estimation of the dimensionality of the individual sample space from subject i :

$$\beta_i = \frac{(\text{trace}(W_i))^2}{\text{trace}(W_i^2)} \quad (14)$$

where the trace of a matrix is the sum of its diagonal elements. It is possible to write β_i as a function of the eigenvalues λ_l :

$$\beta_i = 1 + \frac{(2 \sum_{l < m} \lambda_l \lambda_m)}{\sum_l \lambda_l^2} \quad (15)$$

The following properties can be derived from equation (15):

$$1 \leq b_i \leq P_i \quad (16)$$

$$\beta_i = 1 \text{ if and only if a single eigenvalue is not null} \quad (17)$$

$$\beta_i = P_i \text{ if and only if the } P_i \text{ eigenvalues are all equal} \quad (18)$$

Property (17) says that the lowest dimensionality (a single axis) is obtained when all the attributes are fully correlated, whereas property (18) says that the highest dimensionality is obtained when no correlation at all exists among the attributes. These properties make it clear as to why β_i can be seen as a dimensionality coefficient. Although we do not trust this coefficient as being the exact truth about the number of dimensions involved in sensory evaluation of a set of products, we strongly rely on it for comparing dimensionality of several individual sample spaces. This concept of dimensionality is really important for the panel leader for choosing the number of attributes to be included in the profile. The β coefficient suggests to the panel leader a minimal number of ideal attributes which should be sufficient to span the sample differences. However it is almost impossible to define such ideal attributes, therefore it is recommended to include a number of attributes being at least about the double of β .

To summarize this section, it should be remembered that the β coefficient makes it possible to compare dimensionalities of individual sample spaces, which can be understood as individual complexities of assessment.

2.4. Comparing two individual sample spaces by means of the RV coefficient

The quantity :

$$\langle W_i, W_j \rangle = \text{trace}(W_i W_j) = \sum_{l,m} w_{lm}^i w_{lm}^j \quad (19)$$

is the natural scalar product between two matrices, where w_{lm}^i is the (l,m) -th element of matrix W_i . It is a generalised covariance coefficient between W_i and W_j matrices. The greater $\langle W_i, W_j \rangle$ is, the more similar assessors i and j are in terms of their raw product distances. The quantity :

$$\langle W_i, W_i \rangle = \text{trace}(W_i^2) = \sum_{l,m} w_{lm}^i{}^2 = \sum_l \lambda_l^2 \quad (20)$$

is consequently the norm of W_i or a generalised variance for subject i . The greater this quantity, the more different the products are for this subject. In this context, the RV coefficient is defined as :

$$RV(W_i, W_j) = \frac{\langle W_i, W_j \rangle}{\sqrt{\langle W_i, W_i \rangle \cdot \langle W_j, W_j \rangle}} \quad (21)$$

and appears as a generalised correlation coefficient between W_i and W_j matrices; it is worth pointing out that the RV coefficient is the classical Pearson correlation coefficient between the association matrices arranged into vectors of size n^2 . One can prove that $RV(W_i, W_j)$ is between 0 and 1. The closer the RV is to 1, the more similar assessors i and j are in terms of their standardized product distances.

Therefore, the comparison between two assessors can be based either on generalised covariance given by formula (19) or on generalised correlation given by formula (21). The author recommends the latter, because the former depends on the use of the scale. For instance, an assessor who tends to use a small portion of the scale for every attribute will get smaller covariance with the other assessors but not automatically a smaller RV. But the reader must understand that this matrix standardization is different from the classical attribute standardization, which is done by a PCA of correlation matrix. Contrarily to this normalized PCA, RV takes into account the differences between attribute variances for a given assessor.

To summarize this section, one should remember that the RV coefficient is a measure of the similarity between two individual sample spaces. RV is the classical correlation coefficient between the two square matrices of sample scalar products of size (n,n) having been previously arranged within two vectors of size n^2 .

2.5. Testing significance of a RV coefficient by an exact permutation test of the products

Testing the significance of a given RV value would require complicated parametric assumptions. Therefore, a non parametric alternative (Schlich, 1993) consists in permuting the product labels within X_i without permuting correspondingly the product labels within X_j and to recompute the RV coefficient. Providing that the two assessors agree to some extent about the sample differences, one can expect this "permuted RV" to be lower than the actual RV. This process is repeated a large number of times (say 100) in order to derive the 95 % quantile from the distribution of RV under permutation. If the actual RV is greater than this empirical quantile, it can be concluded that the agreement between the two assessors is better than what can be obtained by chance. The computations required by this permutation test can be too time-consuming to be easily implemented on a micro-computer. Kazi-Aoual (1993) proposed an efficient way to avoid computing a number of permutations. This author established

formulas (22) and (23) giving respectively the mean and the variance of all the $n!$ possible permuted RV coefficients:

$$E_p[RV(W_i, W_j)] = \frac{\sqrt{\beta_i \beta_j}}{n-1} \quad (22)$$

$$V_p[RV(W_i, W_j)] = \frac{(n-1-\beta_i)(n-1-\beta_j)(2n(n-1)+(n-3)c_i c_j)}{(n+1)n(n-1)^3(n-2)} \quad (23)$$

where :

$$c_i = \frac{(n-1)(n(n+1)\delta_i - (n-1)(\beta_i + 2))}{(n-3)(n-1-\beta_i)} \quad (24)$$

with :

$$\delta_i = \frac{\sum_l w_{il}^2}{\text{trace}(W_i^2)} \quad (25)$$

Therefore, a normalized deviation between the actual RV and its distribution under permutation can be computed as follows:

$$NRV[W_i, W_j] = \frac{RV(W_i, W_j) - E_p[RV(W_i, W_j)]}{\sqrt{V_p[RV(W_i, W_j)]}} \quad (26)$$

this normalized RV coefficient is a measure of the agreement between assessors i and j . Assuming a normal distribution of the permuted RV coefficients, formula (26) defines a test statistic for the null hypothesis of no better agreement between assessor i and j than what can be obtained after permutation of the label products. One can expect this value to be roughly greater than 2 if the agreement between the assessors i and j is better than what can be obtained by chance. Although the normal assumption has not been proved till now, it has been observed in practice when performing 100 permutations (Schlich, 1993). Anyway, the exact probability level of this test is not necessary, because the experimenter is most interested in detecting when two assessors do not agree more than what can be obtained by chance.

To summarize this section, one should keep in mind that because the magnitude of a RV coefficient depends on both the number of observations and the number of variables in the two data sets, there is a need for a statistical test of RV significance. An exact non parametric permutation test is available, it makes it possible to declare whether two assessors agree more than what can be just obtained by chance.

2.6 Defining and interpreting a compromise sample space among the assessors by means of the STATIS method

A natural way to define a compromise among subjects on the sample differences would be to compute:

$$\bar{W} = (1/k) \sum_i W_i \quad (27)$$

and to analyse this matrix by means of a principal co-ordinate analysis (PCO; Gower, 1966) in order to map the samples in accordance with the distances induced by $\bar{W}$. In most cases, the author would recommend this analysis instead of PCA of the mean scores, because it does not require "attribute alignment" among subjects and consequently is able to cope with FCP. Unfortunately, this simple method for analysing sensory profiles seems to be unknown to sensory scientists.

The STATIS compromise W differs from this natural compromise $\bar{W}$, because the latter is a classical mean, whereas the STATIS compromise is a weighted mean of the W_i :

$$W = \sum_i a_i W_i \quad (28)$$

where $(a_i)_{i=1..k}$ is the first eigenvector of the matrix of size (k,k) containing the RV coefficients between assessors. The components of this vector are positive and normalized to get a sum equal to one. This vector represents the "principal agreement among assessors". Thus, the greater the a_i , the more assessor i agrees with the panel on the sample differences. The strategy of STATIS is therefore to put weights on subjects proportionally to their agreement with the panel. Therefore, the weight of an outlier should be close to 0.

Replacing W_i by W in formula (14) makes it possible to estimate a dimensionality of the compromise, which can be an indication about the number of dimensions to be interpreted.

The product coordinates on the axes of the compromise are obtained by a PCO of W . The q compromise components are the q first eigenvectors of W being standardized to have a variance equal to the corresponding eigenvalues. These components are arranged as the columns of a matrix C of size (n,q) . The interpretation in terms of sensory attributes can be conducted thanks to the covariances or correlations between the individual attributes and these compromise components. It seems to us that the use of covariances is more logical when no attribute standardization was initially applied to the data as in a PCA of the covariance matrix. Being computed on an individual basis these covariance or correlation coefficients are numerous and it is sometimes necessary to summarize them by averaging scores over assessors having the same attributes before computing these coefficients. As no biplot property holds in this context, we are used to producing a covariance or a correlation plot not superimposed on the compromise plot. Conversely, it is possible to superimpose individual sample spaces on the compromise plot by a classical technique of projection of supplementary elements in multivariate data analysis. The sample coordinates from assessor i on the q compromise components are given by the columns of the following matrix C_i of size (n,q) :

$$C_i = W_i C E \quad (29)$$

where E is a diagonal matrix of size (q,q) containing on the diagonal the inverses of the square roots of the eigenvalues of the compromise.

The compromise location of a given product is the barycenter, for the STATIS weight system, of the k individual locations of this product. The smaller the dispersion of the individual locations of a product around its compromise, the better the agreement among assessors on this product is. In order to better visualize this dispersion, one can draw the convex hulls gathering individual locations of the same products. When two convex hulls do not overlap too much, a logical rule of thumb is to consider the two associated products as different. Alternatively, one can draw for a given product a 95 % confidence convex hull, which is the smallest convex set gathering at least 95 % of the individual locations of this product. Drawing and looking at a convex hull can also be simplified by drawing and looking at a confidence ellipsoïde, which requires a bi-normal assumption to be fulfilled. Although the number of assessors usually included in a trained sensory panel does not make it possible to check this assumption, we do think that drawing confidence ellipsoïdes or convex hulls on a compromise and individual plot is a powerful descriptive tool. Here, as most often in data analysis, it is not a strict and true p -value which is required but some evidence that a pattern makes or does not make sense.

For comparing assessors two ways exist to derive an assessor map. The first one is obtained with the two first eigenvectors of the RV matrix of size (k,k) . Each of these eigenvectors is normalized so that its sum of component squares is equal to the corresponding eigenvalue. The first axis of this map represents the direction of an assessor being equal to the compromise and the corresponding eigenvalue can be understood as the proportion of inter-individual variance explained by this compromise. The assessor coordinates on this first dimension are proportional to the STATIS weights a_i and are therefore positive. An assessor is represented on this plot as an arrow joining the assessor point to the origin. The angle between this arrow and the first axis is proportional to the disagreement of this assessor with the compromise. The length of this arrow is proportional to the quality of the representation of the assessor on this plot. In some applications, it can be necessary to produce the subsequent plots (1,3), (1,4) ... and, in such a case, the visual interpretation becomes difficult.

The second way to derive an assessor map is based on the first two eigenvectors of the RV matrix being previously and simultaneously centered in lines and in columns (that is subtract from each RV coefficient the means of the line and of the column from those it belongs to and add the grand mean of the RV matrix). Therefore, this assessor map is centered and is useful for showing whether different groups of assessors could exist on the basis of the sample differences. In some applications, more than two axes can be necessary for correctly describing the assessor structure. Obviously the RV matrix or the double centered RV matrix can also be taken as the input of any clustering algorithm.

To summarize this important section, it is worth recalling that :

- STATIS derives a compromise association matrix W which is a weighted mean of the individual association matrices W_i
- These weights are given by the first eigenvector of the RV matrix among assessors, which means that the weight of an assessor is proportional to its agreement with the panel
- The compromise sample plot is obtained thanks to a PCO of W

- Sensory interpretation is conducted through covariances or correlations between compromise components and individual attributes
- Assessor maps are derived from PCO of the RV matrix and from PCO of the doubly-centred RV matrix.

2.7 Testing panel homogeneity and significance of compromise

For testing panel homogeneity, Schlich (1993) compared the observed mean of the $k(k-1)/2$ RV coefficients to the 95 % quantile of the distribution of the mean RV coefficient when a permutation of the product labels is randomly and independently chosen for each assessor. In order to estimate this quantile, 100 sets of permutations were sampled independently. A faster solution for testing significance of STATIS compromise consists of computing the mean of the $k(k-1)/2$ normalized RV coefficients (obtained by formulas (22) to (26)) and checking whether it is roughly greater than 2. This faster solution relies on normality of the RV distribution under permutation. As soon as the number of products is larger or equal to 6, the number of possible permutations ($n!$) becomes very large and therefore the normality assumption, observed in practice, should hold.

This test of panel homogeneity is an average of homogeneity computed over pairs of assessors. This approach can be too demanding, because at the end the data is summarized by a compromise. Therefore, it seems sensible to test a weaker hypothesis, that is the individual agreement with the panel compromise. A normalized RV coefficient is computed between each assessor (W_i) and the panel compromise (W) and this paper proposes to average these k values to get a test for compromise significance.

As a summary, one should remember that the exact permutation test defined in section 2.5 makes it possible to test whether the assessors agree among themselves and whether they agree with the STATIS compromise defined in section 2.6. The strength of these tests is that they are exact and that they do not require any permutation to be actually performed.

2.8 Comparing two panel compromises about the same products

The similarity between two panel compromises can be evaluated by their normalized RV coefficient. If this coefficient is about 1 or less, then the panels disagree dramatically; if it is between 1 and 2, then the panels agree rather poorly; if it is greater than 2, then the panels agree and the interpretation should lead to the same conclusions about the sample differences. But the reader must be aware that the interpretation of these differences, in terms of the attributes, may not be equivalent from panel to panel, firstly because the attributes may not be the same among panels and among assessors within panels (FCP or individual selection of attributes), and, secondly because the scalar products in W are sums over attribute contributions making it possible to obtain equal sums composed of different attribute contributions.

This test can be based either on the whole compromise spaces or on the compromise subspaces spanned by the interpreted dimensions. The former is done directly by using the two compromise matrices in formulas (22) to (26), whereas the latter requires first to recompute new W matrices on the basis of the selected components. The advantage of the latter is to

provide insurance that no noise can destroy a significant agreement in the interpreted space, which is at the end the only information retained.

The idea to be retained from this section is that the exact permutation test makes it possible to test agreement between compromises drawn from two different panels having profiled the same samples. Finally, the technique could also be applied to check whether two competitive data analysis techniques lead to the same sample space or to the same sample plot.

2.9 Testing exchangeability of assessors among panels

When several panels have profiled the same samples and when the test described in the previous section is significant, one could wish to go further by testing whether the permutation of assessors among panels would provide us with the same amount of panel differences, as a null hypothesis. If this test is significant, it means that discrimination holds and therefore, it informs the panel leaders that they cannot exchange their assessors.

Assume that g panels were available and that the l -th panel includes k_l assessors ($\sum_{l=1}^g K_l = k$).

Having gathered the k assessors into a single data set, the RV matrix of size (k, k) among these assessors is computed. A PCO of the doubly centred RV matrix is performed in order to keep part or all of the assessor components. From this system of assessor coordinates it is now possible to apply a CDA of the panel factor. The panel discrimination can be measured thanks to a classical statistic in CDA :

$$H = \frac{\text{trace}(B)}{\text{trace}(T)} \quad (30)$$

in which B is the between-panel covariance matrix and T is the total covariance matrix. The closer to 1 H , the more different the panels are. The test proposed in Kazi-Aoual (1992) consists of permuting the assessor labels in the assessor coordinate table obtained from the PCO, without permuting the corresponding assessor coordinates and then computing a CDA of the new permuted partition of the k assessors into g panel of k_l assessors. If the null hypothesis of assessor exchangeability holds, one can expect the real H statistic to be not greater than the same statistic under permutation. Instead of performing numerous permutations in order to estimate the distribution of H under permutation, the following formulas from Kazi-Aoual (1992; 1993) give respectively the mathematical expectation and the variance of H under permutation :

$$E_p[H] = \frac{g-1}{k-1} \quad (31)$$

$$V_p[H] = \frac{2((k-1)/\beta - 1)(g-1)(k-g)(1 + (k-3) \cdot c \cdot f/2k(k-1))}{(k+1)(k-1)^2(k-2)} \quad (32)$$

where β and c are defined as β_i and c_i in equations (14) and (24) replacing W_i by the association matrix among the assessors computed from the assessor scores obtained by CDA.

The constant f is given by :

$$f = \frac{(k-1)(k(k+1)\sum_l (1/k_l) - (k-1)(g-1)(g+1))}{(k-3)(g-1)(k-g)} \quad (33)$$

making it possible to derive a normalized H statistic :

$$H_N = \frac{H - E_P[H]}{\sqrt{V_P[H]}} \quad (34)$$

when this H_N is greater than 2, one can decide that the assessors are not exchangeable across panels.

To conclude this section, it is worth pointing out that, contrary to the previous permutation tests, this one is based on permutations of the assessors instead of the products. It appears as a competitor of the multivariate analysis of variance tests such as the Hotelling-Lawley trace or the Wilks ratio (Mardia *et al.*, 1979), but contrary to these parametric tests it does not require the assumptions of multinormality and homogeneity of within-panel covariance matrices.

2.10 Defining a compromise about attribute correlations by means of the Dual STATIS method

Dual STATIS, proposed in Lavit (1988), is the STATIS method applied to the covariance matrix $X'X/n$ (or to the correlation matrix) instead of the association matrix XX' . The aim of Dual STATIS is to compare the assessors on the basis of their individual structure of attribute covariances or correlations, which can be understood as their own way to understand attributes and to link them together. The STATIS compromise becomes a weighted mean of the individual covariance matrices. Such an analysis can be very interesting for the panel leader when training his panel to conventional profiling. Furthermore, this analysis can be run even when the assessors scored different samples but for the same attributes, making it possible to investigate simultaneously correlation structures on the basis of different kinds of products.

Unfortunately, the analytical permutation tests described in Kazi-Aoual (1993), do not seem to work with Dual STATIS, because of the non centering of the columns of X' . Therefore, as in GPA, one should run numerous permutations to derive a test. Note that in this context, the permutations are applied on the attributes instead of the products.

To conclude this section, it can be underlined that with a conventional profiling data set both STATIS and Dual STATIS can be performed, making it possible to compare the assessors on the basis of both the sample differences (client need) and the attribute relationships (panel leader need). If on both aspects the agreement among the assessors is good enough, the panel leader can trust that the sample differences perceived by the assessors (significant STATIS) were also described in the same way by these assessors in terms of the sensory attributes (significant Dual STATIS).

3. COMPARISON WITH OTHER METHODS

3.1 Principal Component Analysis (PCA), Simple and Multiple Correspondence Analysis (CA and MCA) and Canonical Discriminant Analysis (CDA)

The limitations of the classical PCA of the mean score products have already been mentioned in the introduction section of this chapter. There are two other ways of performing PCA on a profiling data set, called TUCKER1 in chapter 10 of this book. The first TUCKER1 method consists of a PCA of a data set composed of n times k observations and p attributes; this data set gathers the individual data sets vertically. It cannot be applied to free-choice profiling and the author thinks that it can be quite non robust to outlier observations. The second

TUCKER1 method consists of a PCA of a data set composed of n observations and $\sum_{l=1}^k p_l$ attributes; this data set gathers the individual data sets horizontally. This technique is more interesting than the previous one, because it is a solution for analysing free-choice profiling when only a PCA program is available. Nevertheless, the weight of an assessor in this analysis can be strongly inflated or deflated according to his number of attributes. One can avoid this problem by dividing the column centered scores of a given assessor by the square root of the sum of squares of these scores and by doing a PCA of the covariance matrix.

The TUCKER2 and TUCKER3 methods, also described in this book (chapter 10), are more interesting but cannot be applied to free-choice profiling either.

The use of CA and MCA with profiling data, proposed by McEwan and Schlich (1991/2), considers sensory measurements at ordinal level instead of interval level, which is definitely not possible with STATIS or with any technique based on covariance or correlation computations. Whether this point improves significantly the interpretation of the sensory data is still not obvious. Another advantage of CA and MCA, linked to the previous one, is their ability to discover non linear relationship between attributes. Most of the linear techniques could accommodate non linear transformations of the data such as spline functions. But till now there has been a lack of published examples of these techniques in the sensory field. Finally and once again, the analysis of free-choice profiling with CA or MCA is not straightforward.

The introduction section of this chapter has suggested why CDA together with some pre-treatment of the data could be useful for analysing sensory profiles. Nevertheless and like the other methods evoked in this section, with the exception of the second TUCKER1 method, CDA cannot deal with free-choice profiling or with the problem of non attribute alignment in conventional profiling.

3.2 Generalised Procrustes Analysis (GPA)

The introduction section of this chapter has recalled that GPA is the historical leader technique for the analysis of sensory profiles because it was presented as the dedicated technique for free-choice profiling. The present book includes a chapter (7), which fully describes this technique. In the past, the author of the present chapter advocated the use of GPA through his SAS/IML[®] software for GPA (Schlich, 1989). But now, in the light of the validation concern of sensory profiles, the author prefers the STATIS framework for several reasons :

- It is a non iterative technique

- Its weighting system of assessors deals smoothly with outliers
- The β coefficient gives simple dimensionality estimation
- The analytical permutation tests of RV coefficient provide a straightforward panel homogeneity estimation, a compromise validation and a panel comparison
- Dual STATIS is a unique tool for investigating individual correlation structures

3.3 INDSCAL

INDSCAL (Carroll & Chang, 1970), described in this book in chapter 6, is the MDS (Schiffmann *et al.*, 1981) technique dedicated to three-way data analysis. It analyses a set of individual product dissimilarity matrices, in that respect it is a little more general than STATIS, which analyses a set of scalar product matrices. INDSCAL iteratively defines an a priori fixed number of optimal dimensions mapping the products. The strength of INDSCAL, not available in STATIS, is that each assessor can weight differently each of these dimensions. These individual vectors of weights are defined in order to minimize the STRESS which is a least square criterion between the observed individual dissimilarities and the compromise product distances in the fitted space.

Being a more general model than STATIS, INDSCAL can cope with free-choice profiling as soon as a dissimilarity function is chosen to be derived from the attributes scores. Unfortunately, this technique does not seem to be widely used in our field. Furthermore, to the knowledge of the author no analytical validation technique exists in the INDSCAL framework, certainly because of its iterative algorithm.

4. APPLICATIONS

4.1 The ESN coffee experiment

The European Sensory Network (ESN) was launched in 1989 as a basis for close collaboration between major food research centers in Europe. Nowadays, ESN gathers about 20 academic sensory scientists coming from about 10 different countries in Europe. The aim of ESN is to exchange ideas and to transfer results from basic research to the food industry. In order to check consistency of sensory profiling, ESN organized an interlaboratory study of 16 coffees evaluated by 11 different panels managed by ESN members in their own institutes.

The samples, the beverage preparation and service and the experimental design were absolutely identical across panels. For practical reasons, the same samples had to be assessed during a session by every assessor and no more than 4 products could be presented during a given session. Therefore, 4 sessions were conducted to complete a replicate. Three replicates were done. The allocation of the samples to the sessions was identical across panels and determined thanks to previous knowledge about the expected coffee differences. The strategy was to span the coffee space as much as possible within each of the 4 sessions of the first replicate. Second and third replicates were then defined according to the same rule and in order to respect a pair balance condition, which was to ask that any pair of samples must not be present in more than one session. Concerning the selection and the training of the assessors, as for the vocabulary development, no instruction was given to the panel leaders. Depending on

the panels, the number of assessors was between 8 and 12, whereas the number of attributes was between 14 and 56. As said in the introduction, a book will be published soon by ESN presenting the complete analysis of the coffee data and making it available to the reader. For the present chapter, only a small part of the profiling data is used to illustrate the application of the above proposed techniques.

The 16 coffee samples were provided by the International Coffee Organization (ICO) in London, who was an ESN member. The sensory panel from ICO was composed of 12 subjects highly trained to profile coffee for many years. ICO panel can be considered as an expert panel. On the contrary, one of the French panel, called F2, was poorly trained to profile coffee. Because of this opposition, we decided to present a STATIS analysis of the ICO data (section 4.2) and another one of the F2 data (section 4.3). For the sake of simplicity, we also decided to analyse only a subset of attributes (12 for ICO and 9 for F2), chosen from attributes being scored by most of the panels. Presenting the analysis of the "best" and of the "worst" panels, we aim to convince the reader that our statistical framework is actually able to detect such a diagnostic.

In order to illustrate the techniques of panel comparisons together with the Dual STATIS method, we defined a panel called S49 including 49 assessors from 5 different panels, called 1F for France (different panel than F2), IC for ICO (UK), No for Norway, Po for Poland and Sw for Sweden. These panels share the property of including 4 basic attributes for describing coffee : bitterness, acidity, astringency and body/mouthfeel. Therefore, we applied STATIS (section 4.4) and Dual STATIS (section 4.5) on the S49 panel data restricted to the above 4 attributes.

Each of the three data sets analysed were first averaged over the 3 replicates by product times assessor, making the number of observations being equal to the number of products (16) times the number of assessors (12, 9 or 49).

4.2 STATIS of the ICO panel

Table 1
ICO panel. RV coefficients between subjects

	A	B	C	D	E	F	G	H	I	J	K	L
A	1.00											
B	0.81	1.00										
C	0.76	0.83	1.00									
D	0.82	0.87	0.77	1.00								
E	0.91	0.83	0.84	0.81	1.00							
F	0.86	0.91	0.84	0.87	0.86	1.00						
G	0.86	0.76	0.72	0.77	0.79	0.81	1.00					
H	0.91	0.77	0.77	0.84	0.87	0.87	0.87	1.00				
I	0.88	0.82	0.88	0.81	0.91	0.87	0.80	0.88	1.00			
J	0.78	0.69	0.70	0.76	0.73	0.78	0.82	0.88	0.78	1.00		
K	0.76	0.81	0.86	0.75	0.84	0.83	0.70	0.77	0.84	0.70	1.00	
L	0.86	0.76	0.70	0.75	0.74	0.80	0.87	0.87	0.80	0.78	0.73	1.00

Table 2
ICO panel. Normalized RV coefficients between subjects

	A	B	C	D	E	F	G	H	I	J	K	L
B	8.56											
C	7.84	8.76										
D	8.60	9.38	7.93									
E	9.69	8.73	8.85	8.60								
F	9.10	9.75	8.73	9.24	9.10							
G	9.23	8.04	7.34	7.93	8.41	8.48						
H	9.77	8.09	8.02	8.87	9.35	9.27	9.29					
I	9.40	8.54	9.29	8.55	9.71	9.18	8.55	9.49				
J	8.18	7.15	7.15	7.88	7.63	8.11	8.66	9.41	8.24			
K	7.90	8.55	9.05	7.60	8.93	8.65	7.06	7.99	8.94	7.09		
L	9.26	8.08	7.30	7.86	7.87	8.50	9.18	9.28	8.63	8.20	7.63	
Mean	8.87	8.51	8.21	8.40	8.81	8.92	8.38	8.98	8.96	7.97	8.13	8.34

As we recommend most often, no attribute standardization was performed, but the assessors were compared on the basis of their RV coefficients (Table 1). These RV coefficients range from 0.69 to 0.91 denoting a very good agreement between panelists on the sample differences. The permutation test statistics of these coefficients, which are the normalized RV coefficients given in Table 2, range from 7.15 to 9.77. According to the normal distribution, we can therefore conclude that each pair of assessors strongly agrees about the sample structure. Such a high homogeneity in a trained sensory panel is rather unusual in practice.

Table 3
ICO panel. PCO of the non centered RV matrix

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	9.9181	82.65	82.65
DIM 2	0.6041	5.03	87.69
DIM 3	0.3237	2.70	90.38
DIM 4	0.2682	2.23	92.62
DIM 5	0.2418	2.02	94.63
DIM 6	0.1560	1.30	95.93
DIM 7	0.1294	1.08	97.01
DIM 8	0.1163	0.97	97.98
DIM 9	0.0879	0.73	98.71
DIM10	0.0685	0.57	99.28
DIM11	0.0517	0.43	99.72
DIM12	0.0337	0.28	100.00

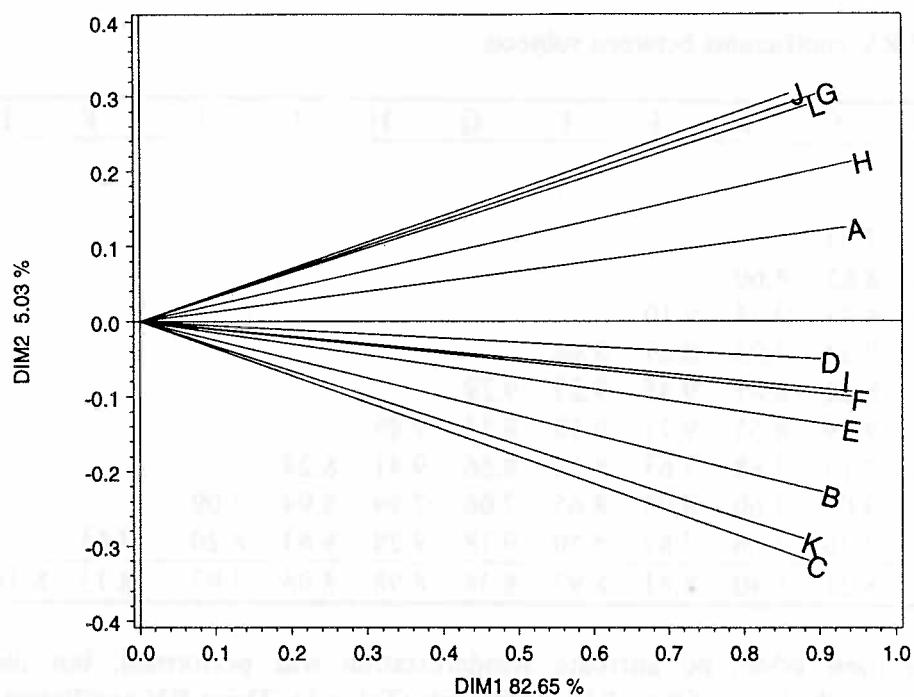


Figure 1 : ICO panel. PCO of the non centered RV matrix

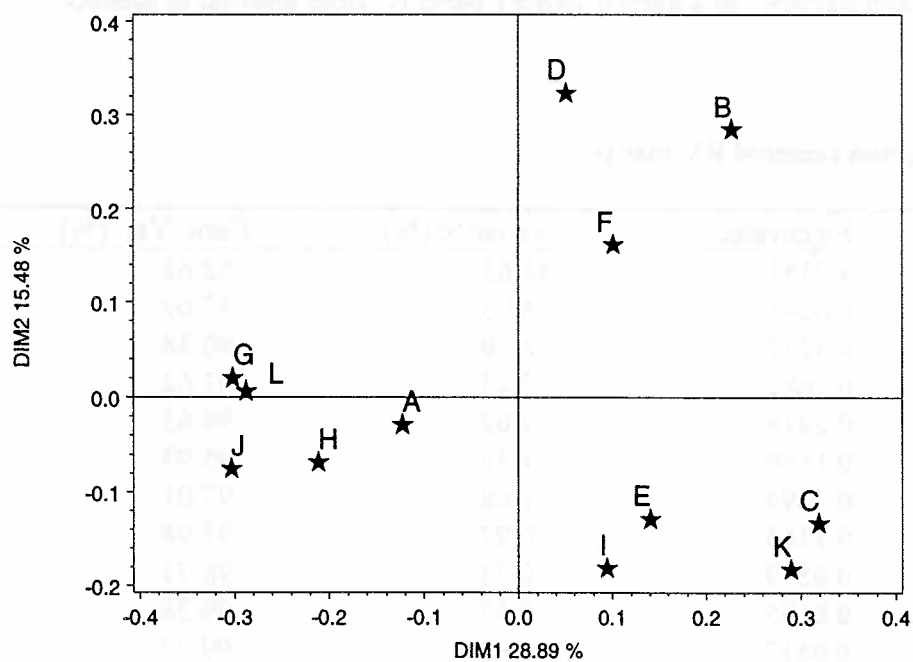


Figure 2 : ICO panel. PCO of the centered RV matrix

The first eigenvector of the RV matrix accounts for 82.65 % of the panel variation (Table 3). This can be understood as follows : the compromise association matrix W defined by STATIS will account for 82.65 % of the variation between the 12 individual association matrices W_i . Figure 1 is the first assessor plot drawn from this PCO. One can see that the assessors who might agree the least with the compromise (first axis) are assessors J, G and L at the top of the plot and assessors K and C at the bottom of the plot. One can check this point by reading the last line of Table 2, in which these assessors actually get the smallest, but still excellent, average normalized RV coefficients.

Table 4
ICO panel. PCO of the centered RV matrix

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	0.6042	28.89	28.89
DIM 2	0.3237	15.48	44.38
DIM 3	0.2715	12.98	57.36
DIM 4	0.2445	11.69	69.05
DIM 5	0.1560	7.46	76.51
DIM 6	0.1296	6.20	82.71
DIM 7	0.1172	5.61	88.31
DIM 8	0.0894	4.28	92.59
DIM 9	0.0688	3.29	95.88
DIM10	0.0517	2.47	98.35
DIM11	0.0345	1.65	100.00

Thanks to the PCO of the centered RV matrix, one can draw a centered assessor map (Figure 2), in which the compromise would be located at the origin. On this map, one can observe three groups of assessors : D, B and F at the top of the plot, C, K, E and I at the bottom of the plot and the remaining assessors on the left of the plot. This grouping of assessors might correspond to very slight differences in sample perception, as permutation tests demonstrated strong homogeneity in each pair of assessors. Moreover, as this map accounts for only 44.38 % of the total variation (Table 4), this visual grouping might also be artificial. Comparing Tables 3 and 4, one can note that the total number of dimensions in the first analysis was equal to the number of assessors, that is 12, whereas in the second analysis centering the RV matrix removed one dimension.

Because of the high homogeneity of the ICO panel, the assessor maps given by Figures 1 and 2 were not really necessary. The point is that without looking at the magnitude of the RV coefficients and without testing them thanks to the normalized RV coefficients, it would have been difficult to assess on the basis of any assessor map whether homogeneity held or not.

Table 5 is the most interesting table printed by our STATIS program. The lines of this table are the assessors. The column called 'p' gives the number of attributes. In this case, standard deviation of one or two attributes were nil for subjects A, D, G and L. The 'Scaling' column gives the coefficient by which each individual data set should be multiplied to achieve a common and equal global dispersion over both samples and attributes. This column demonstrates that subjects B and I had a clear tendency to concentrate their scores on a small

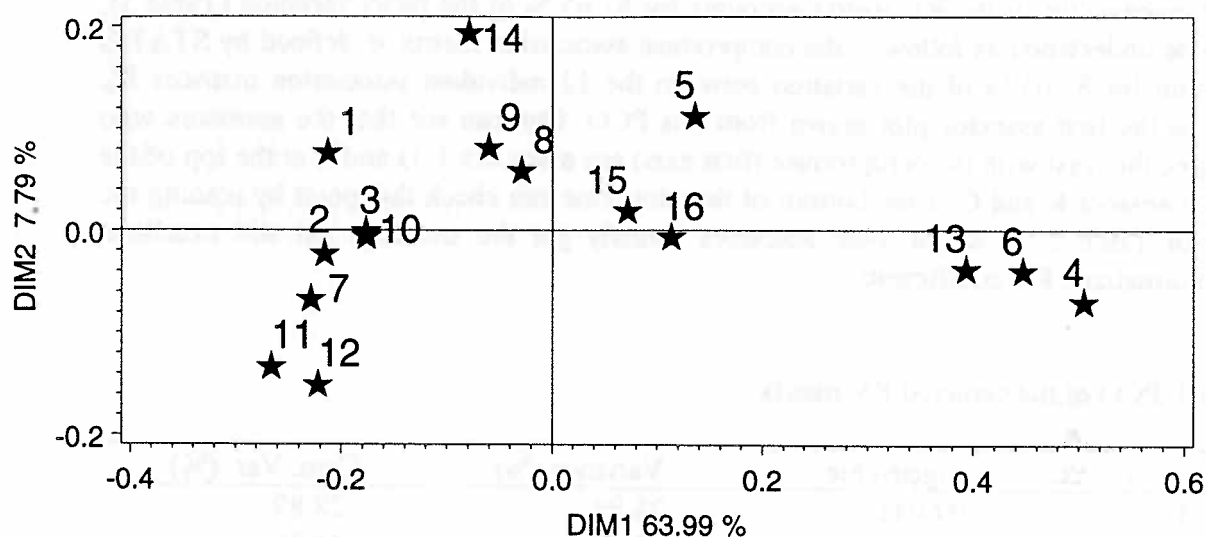


Figure 3 : ICO panel. PCO of the sample compromise among subjects

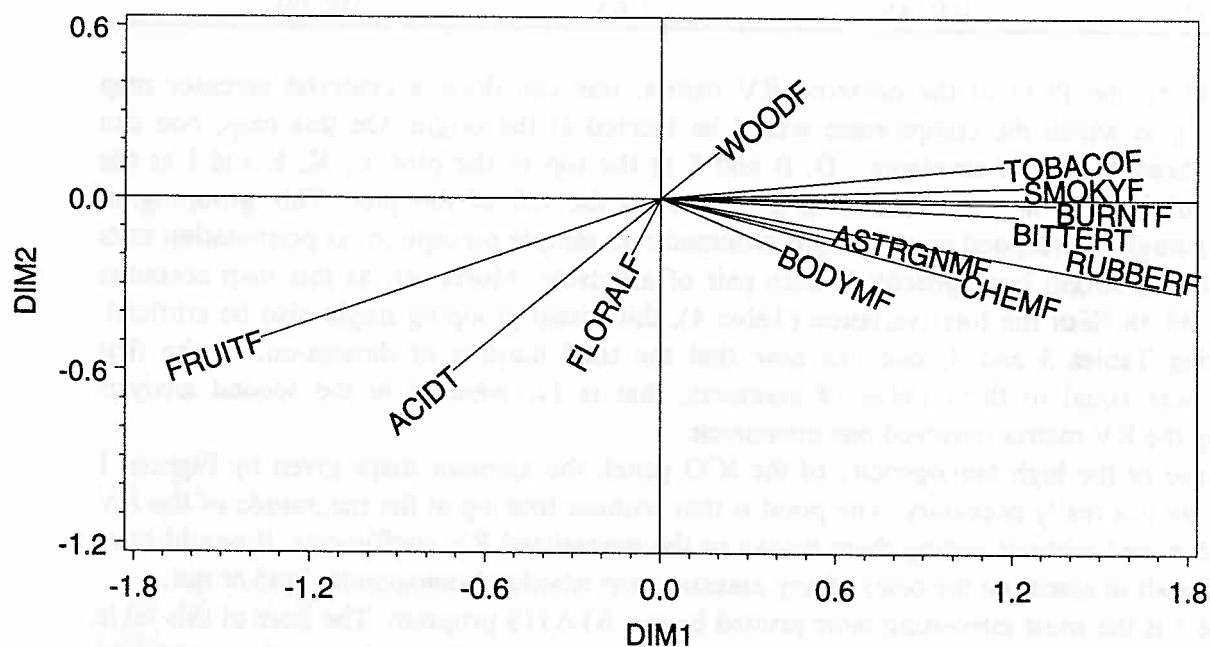


Figure 4 : ICO panel. Covariance plot of attributes

part of the scale (scaling coefficients equal to 2.131 and 2.190), whereas subjects D and G clearly behaved in the opposite way (scaling coefficients equal to 0.325 and 0.293). The next 'NRV' column is the 'Mean' line of Table 2, that is the mean of the normalized RV coefficients between the current subject and the others. Looking at this column, one can immediately detect panelists who would disagree with the rest of the panel. The 'Weight' column in Table 5 contains the STATIS weights of the assessors being normalized in such a way that their sum is equal to the number of panelists. Therefore, an individual weight greater than one corresponds to a subject who is in better agreement with the panel than the others. Because of the high level of homogeneity, the weight range is very narrow around one, denoting that there is no reason to trust some panelists more than others. The next 'BETA' column gives the dimensionality coefficient defined in section 2.3 of this chapter. This coefficient ranges from 1.467 to 2.786 with a mean value of 1.979. On this basis, one can postulate that two dimensions would be enough to span the coffee space as it is perceived by this panel. Therefore, only the two first dimensions of the compromise sample space will be interpreted below. The last column of Table 5, called 'NRVC', gives the normalized RV coefficient between each assessor and the STATIS compromise. We suggested in section 2.7 that the mean of this column can be accepted as a normal test of the null hypothesis of chance against the alternative hypothesis of compromise significance. In that respect, the ICO STATIS compromise is found highly significant : 9.752. This last column provides a check of individual agreement with compromise. In the ICO panel, NRVC is always much greater than 2, denoting that every assessor agrees with the compromise.

Table 5
ICO panel. Summary of individual STATIS statistics

Subject	p	Scaling	NRV	Weight	BETA	NRVC
A	10	0.674	8.865	1.030	1.734	10.100
B	12	2.131	8.512	0.995	1.539	9.698
C	12	1.229	8.206	0.975	2.321	9.393
D	11	0.325	8.404	0.992	2.786	9.618
E	12	1.434	8.808	1.024	1.627	10.020
F	12	0.799	8.918	1.037	2.407	10.140
G	11	0.293	8.378	0.985	2.424	9.607
H	12	0.528	8.984	1.038	1.732	10.230
I	12	2.190	8.956	1.037	1.600	10.150
J	12	0.996	7.973	0.947	1.708	9.171
K	12	0.907	8.127	0.967	2.399	9.325
L	10	0.493	8.344	0.973	1.467	9.575
Mean	11.5	1.000	8.540	1.000	1.979	9.752

The product map in Figure 3 is obtained from a PCO of the compromise association matrix, whose eigenvalue decomposition is given in Table 6. This plot is dominated by the first dimension which accounts for 63.99 % of the information. This first axis splits samples 4, 6 and 13 from the others and the second axis seems to distinguish sample 14 at the top of the plot from samples 11 and 12 at the bottom of the plot. The interpretation of this sample structure is conducted by means of an average covariance plot (Figure 4) as explained in

section 2.6. It is clear from this plot that the first axis is a gradient of coffee strength positively correlated to the bitter taste and to the tobacco, smoky, burnt and rubber flavours and negatively correlated with the acid taste and the fruity flavour. Therefore, samples 4, 6 and 13 were judged as the "strongest" coffees, whereas sample 11 and 12 were perceived as more acid and fruity and finally coffee 14 would be the weakest sample for both bitterness and acidity. The ESN book will explain why these findings make sense considering the origin of the samples.

Table 6
ICO panel. PCO of the sample compromise among subjects

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	0.9833	63.99	63.99
DIM 2	0.1197	7.79	71.78
DIM 3	0.0690	4.49	76.27
DIM 4	0.0620	4.04	80.31
DIM 5	0.0509	3.32	83.62
DIM 6	0.0473	3.08	86.70
DIM 7	0.0390	2.54	89.24
DIM 8	0.0338	2.20	91.43
DIM 9	0.0287	1.87	93.30
DIM10	0.0233	1.51	94.82
DIM11	0.0216	1.41	96.23
DIM12	0.0189	1.23	97.45
DIM13	0.0170	1.10	98.56
DIM14	0.0129	0.84	99.40
DIM15	0.0093	0.60	100.00

As suggested in section 2.6, Figure 5 locates on the previous compromise sample map (C followed by the sample numbers) the individual assessments of each coffee (sample numbers) and for one half of them (from left to right samples 12, 1, 14, 8, 16, 5, 13 and 4) draws the corresponding convex hulls. Drawing all the 16 convex hulls would have made the plot unreadable. This picture suggests that the following seven groups of coffees : (11, 12), (7, 2, 10, 3, 1), (14), (8, 9), (15, 16), (5), (13, 6, 4) may be different between groups and similar within groups.

4.3 STATIS of the F2 panel

The RV coefficients (Table 7) range from 0.13 to 0.63, showing that the best RV coefficient in F2 panel is lower than the worst in ICO panel. The normalized RV coefficients in Table 8 are often not significant (lower than 2) and quite often strongly not significant (lower than 1). From the 'Mean' line of this table, it can be concluded that only assessors B, D, E and H seem to agree with the panel.

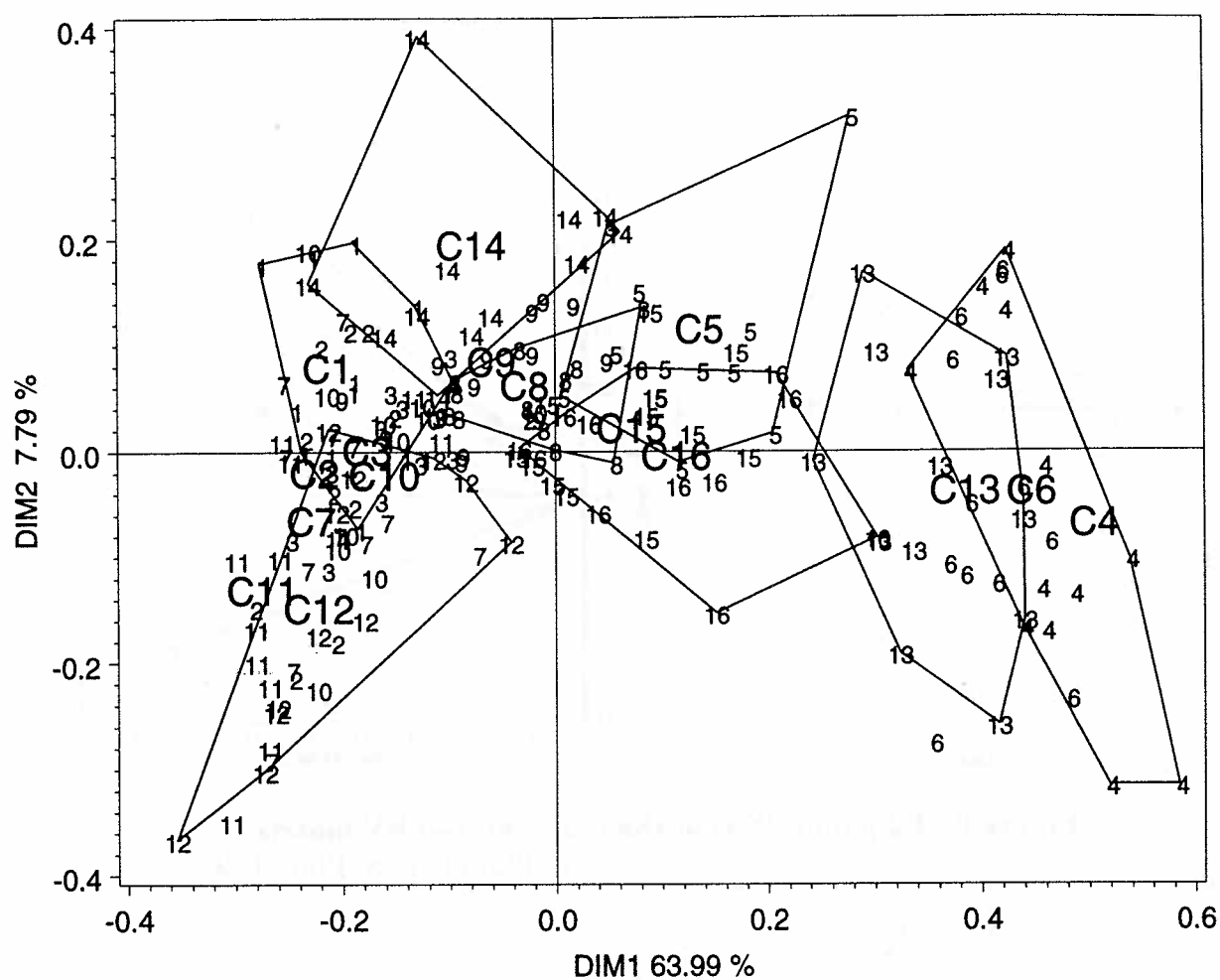


Figure 5 : ICO panel. Convex hulls of samples on the compromise plot (1, 2)

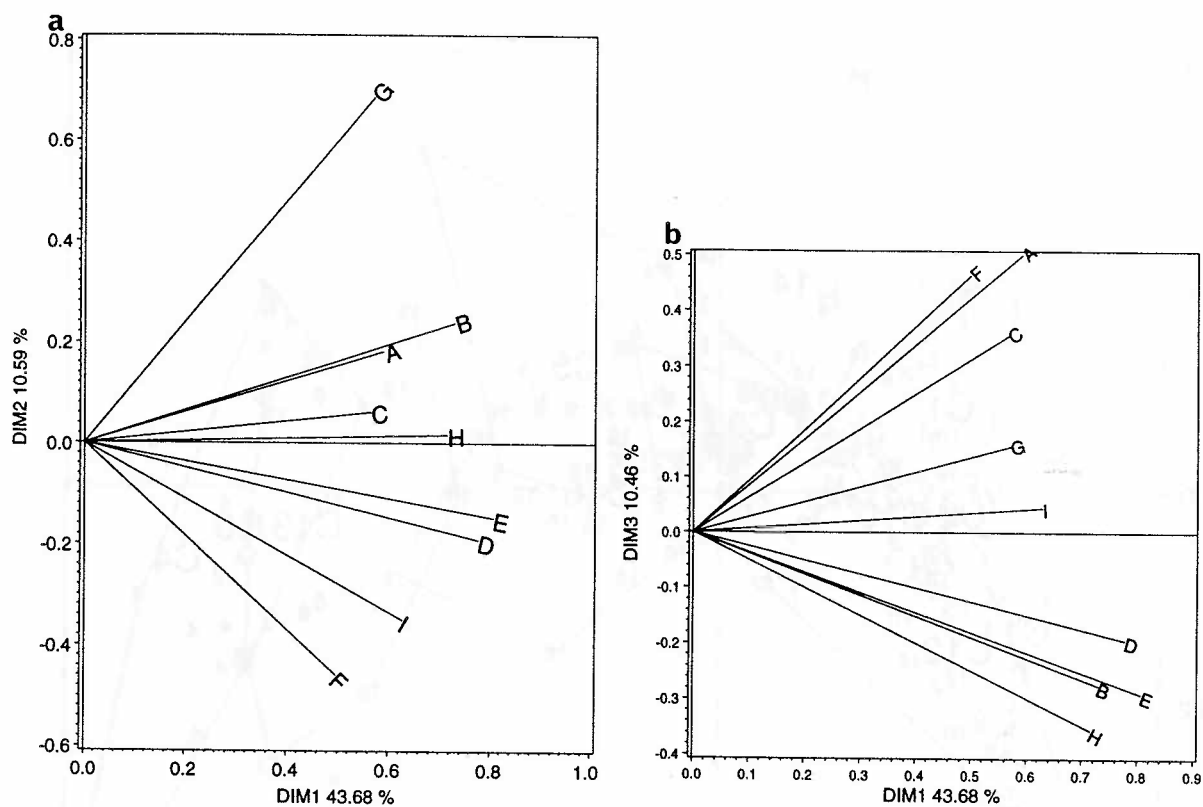


Figure 6 : F2 panel. PCO of the non centered RV matrix

a. Plot (1, 2) b. Plot (1, 3)

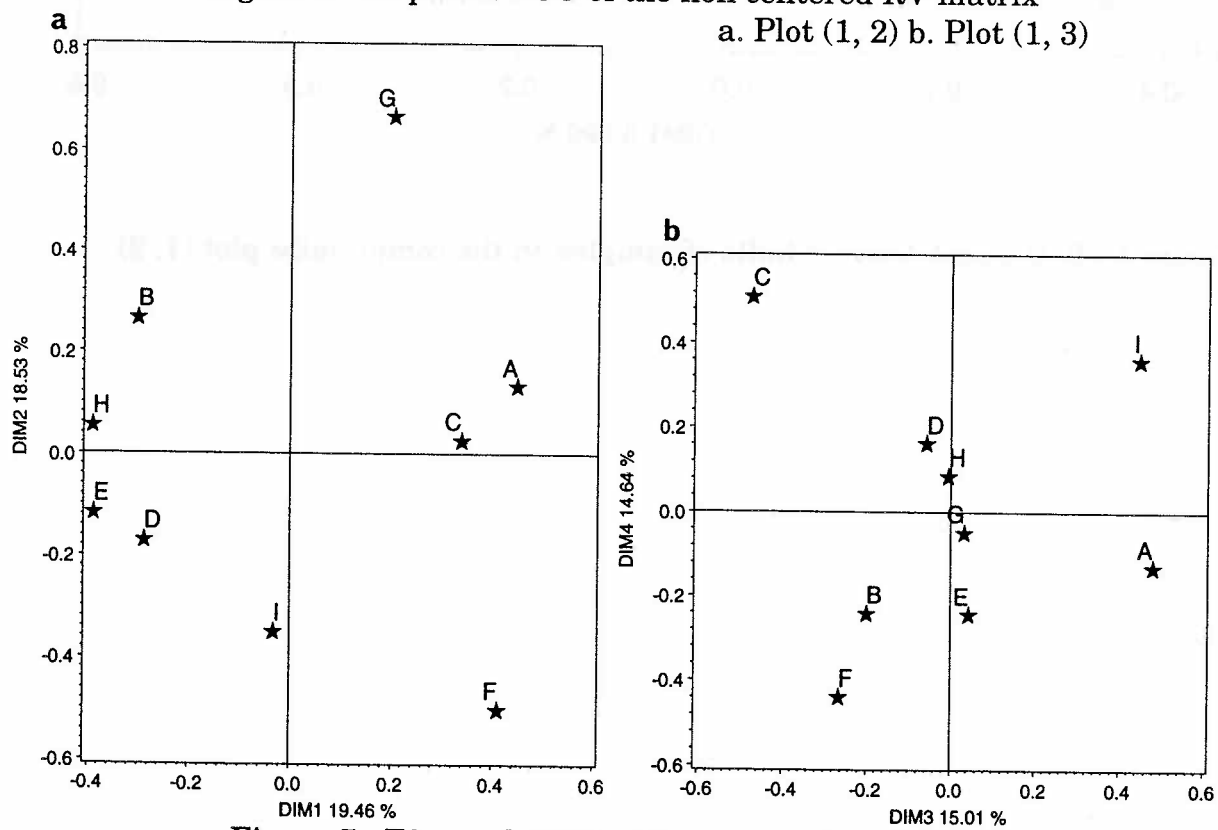


Figure 7 : F2 panel. PCO of the centered RV matrix

a. Plot (1, 2) b. Plot (3, 4)

Table 7
F2 panel. RV coefficients between subjects

	A	B	C	D	E	F	G	H	I
A	1.00								
B	0.29	1.00							
C	0.28	0.33	1.00						
D	0.29	0.47	0.40	1.00					
E	0.38	0.63	0.26	0.63	1.00				
F	0.30	0.25	0.29	0.32	0.37	1.00			
G	0.41	0.45	0.31	0.33	0.31	0.13	1.00		
H	0.29	0.48	0.32	0.52	0.58	0.20	0.33	1.00	
I	0.37	0.29	0.29	0.49	0.44	0.25	0.20	0.38	1.00

Table 8
F2 panel. Normalized RV coefficients between subjects

	A	B	C	D	E	F	G	H	I
B	1.48								
C	0.88	2.16							
D	1.17	4.29	2.94						
E	2.71	6.11	1.38	6.43					
F	1.18	1.24	1.28	1.86	2.85				
G	2.58	3.84	1.46	1.81	1.99	-0.91			
H	0.98	4.19	1.68	4.46	5.51	0.16	1.72		
I	1.66	1.38	0.88	3.91	3.62	0.43	-0.54	2.17	
Mean	1.58	3.09	1.58	3.36	3.82	1.01	1.49	2.61	1.69

Table 9
F2 panel. PCO of the non centered RV matrix

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	3.9313	43.68	43.68
DIM 2	0.9528	10.59	54.27
DIM 3	0.9413	10.46	64.73
DIM 4	0.7711	8.57	73.30
DIM 5	0.7525	8.36	81.66
DIM 6	0.5152	5.72	87.38
DIM 7	0.4617	5.13	92.51
DIM 8	0.4167	4.63	97.14
DIM 9	0.2574	2.86	100.00

As a consequence of this poor homogeneity, the first axis of the PCO of the RV matrix accounts for only 43.68 % (Table 9) of the variation among the assessors, whereas we obtained 82.65 % for the ICO panel. Because the third axis accounts for about the same amount of variation as the second one, Figure 6 presents both plots (1,2) and (1,3) of this PCO. But as these three axes account for only 64.73 % of the information (smaller arrow lengths in Figure 6 compared to Figure 1) one must interpret Figure 6 cautiously. Nevertheless,

this figure underlines assessor F as being the worst because of the large angles its arrow makes with the first axis in the two plots. This is confirmed by the lowest mean normalized RV coefficient of 1.01 (Table 8) obtained by subject F. The best assessors B, D, E and H are almost the closest to the first axis in the plot (1,2) and are gathered in the plot (1,3), suggesting that they may have directed the compromise.

Table 10

F2 panel. PCO of the centered RV matrix

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	1.0003	19.46	19.46
DIM 2	0.9527	18.53	37.99
DIM 3	0.7718	15.01	53.00
DIM 4	0.7525	14.64	67.63
DIM 5	0.5152	10.02	77.66
DIM 6	0.4658	9.06	86.71
DIM 7	0.4256	8.28	94.99
DIM 8	0.2574	5.01	100.00

This last point is clarified by the PCO of the centered RV matrix, whose distribution of eigenvalues (Table 10) suggested that four axes should be interpreted to take into account 67.63 % of the total information. These panelists B, D, E and H are clustered on the left part of plot (1,2) in Figure 7, whereas the others do not gather, and they are close to the origin on plot (3,4), whereas the others are farther to the origin. In fact, this panel is composed of four assessors (B, D, E and H) who fairly agree among themselves and of five other assessors who disagree with the group of four and among themselves. Therefore, the compromise cannot be anything else than a rough mean of assessors B, D, E and H. Interestingly this point could have been detected in Table 8, where the 6 normalized RV coefficients among these four subjects are the only ones to be above 4.

The 'Scaling' column in table 11 points out assessor G as having spanned his scores much more than the others. The 'NRV' column of Table 11 has already been interpreted above, but note that the mean of all the normalized RV coefficients is equal to 2.248, which can be considered as just significant, although it is only one fourth of the corresponding statistic in the ICO panel (8.540). The STATIS weights range from 0.760 to 1.235, but the mean weight of assessors B, D, E and H is 1.160, whereas the mean weight of the others is 0.872, meaning that in average STATIS gave to an assessor from the group (B,D,E,H) a weight being 33 % greater than the weight given to the other assessors. Interestingly, the 'BETA' column of Table 11 has a mean value of 2.876, whereas it was only equal to 1.979 for the ICO panel. Therefore, the French assessors seem to be more complex than the British assessors.... The French would perceive 3 dimensions whereas the British would perceive only 2. The reader, who may be British, should first noticed that the author is French... and secondly should not take these two last findings as the absolute truth, but rather as a useful indication of a pattern present in this data. Again more interesting is the fact that the mean b coefficient is equal to 2.306 within group (B,D,E,H), whereas it is equal to 3.331 within the other subjects, meaning that the more homogeneous assessors are at the same time the least complex subjects. The author found this kind of relation in many datasets; it is quite logical : the more dimensional,

the more chance to disagree. In our application, it is not difficult to agree about the coffee strength (first dimension common to everyone), it is much more difficult to agree about the coffee flavour (second and third dimensions not identically perceived by everyone). Finally, the 'NRVC' column in Table 11 shows that assessors from group (B,D,E,H) strongly agree with the STATIS compromise (mean normalized RV with compromise equals to 7.815) and the remaining assessors fairly agree with this compromise (mean normalized RV with compromise equals to 4.329). This is the "STATIS miracle": although in this data set some neat disagreements held between some assessors, at the end everybody significantly agreed with the compromise. The F2 panel was definitely less homogeneous than the ICO panel, but STATIS was still able to define a valid compromise. At this point of the discussion, the reader may think that the null hypothesis in these permutation tests is too weak, making these tests artificially powerful. The author agrees that a null hypothesis which is actually true would be a disaster for the panel leader. In such a case, the only thing to do is to bin the data. In some situations, which have to remain anonymous, the author actually observed mean normalized RV coefficients between assessors around 0 and between assessors and compromise around 1.

Table 11
F2 panel. Summary of individual STATIS statistics

T	SUBJEC	p	Scaling	NRV	Weight	BETA	NRVC
A		7	0.700	1.580	0.902	3.758	4.383
B		8	0.910	3.086	1.117	1.963	7.527
C		8	1.330	1.581	0.865	2.943	4.446
D		8	1.361	3.359	1.191	2.580	8.061
E		7	1.545	3.824	1.235	1.752	8.825
F		8	0.709	1.010	0.760	2.860	3.421
G		8	0.305	1.495	0.872	3.137	4.441
H		8	1.231	2.608	1.096	2.930	6.846
I		8	0.909	1.687	0.962	3.958	4.955
Mean		7.77	1.000	2.248	1.000	2.876	5.878

From the distribution of the eigenvalues of the compromise (Table 12), it was decided to analyse the first four dimensions of the compromise sample space (Figures 8 and 9). The overall structure of plot (1,2) (Figure 8) is roughly the same as ICO. Nevertheless, some differences are noticeable. This structure is more complex as exemplified by the fact that the first dimension accounts for about half of the total information than for ICO. Sample 4 is now slightly split from samples 6 and 13, which are now gathered with sample 16. Looking at the convex hulls (Figure 10), it appeared that two assessors had sent this sample 4 very far from the others. The larger surfaces of these hulls, compared to ICO, lead to a less confidence about the following visual grouping of samples : (2,11), (1,3,7,9,10,12), (5,8,15), (14), (6,13,16), (4). The covariance plot (Figure 12) confirms the interpretation made with the ICO data, namely the opposition of bitterness and acidity along the first axis and the weakness of both aspects at the top of the second axis opposed to the bottom correlated to the astringency of the coffee.

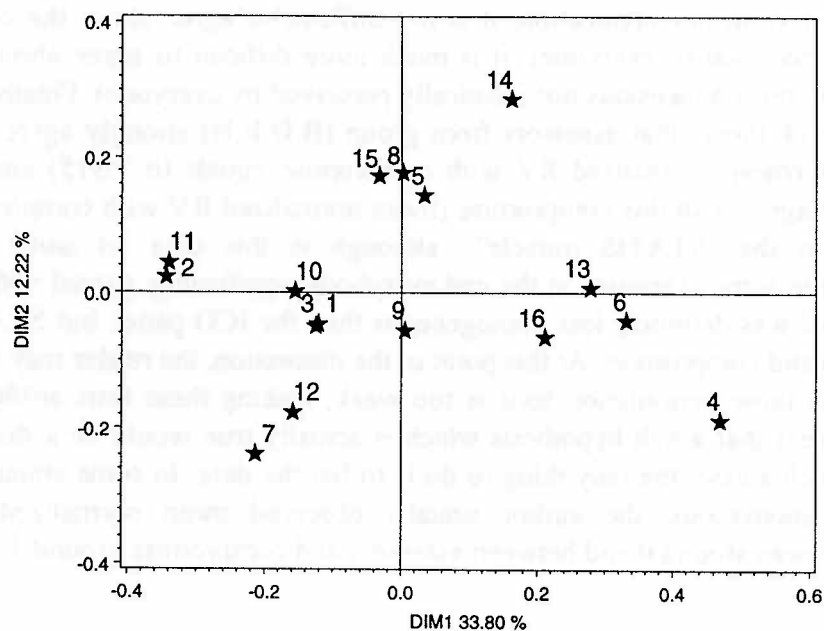


Figure 8 : F2 panel. PCO of the sample compromise among subjects. Plot (1, 2)

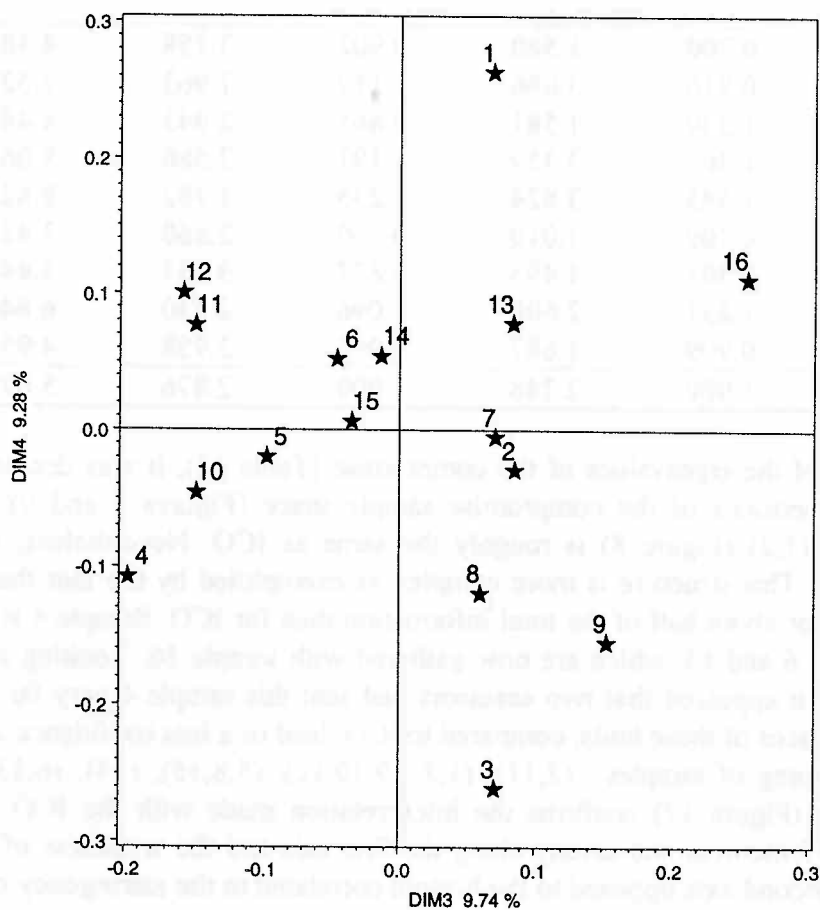


Figure 9 : F2 panel. PCO of the sample compromise among subjects. Plot (3, 4)

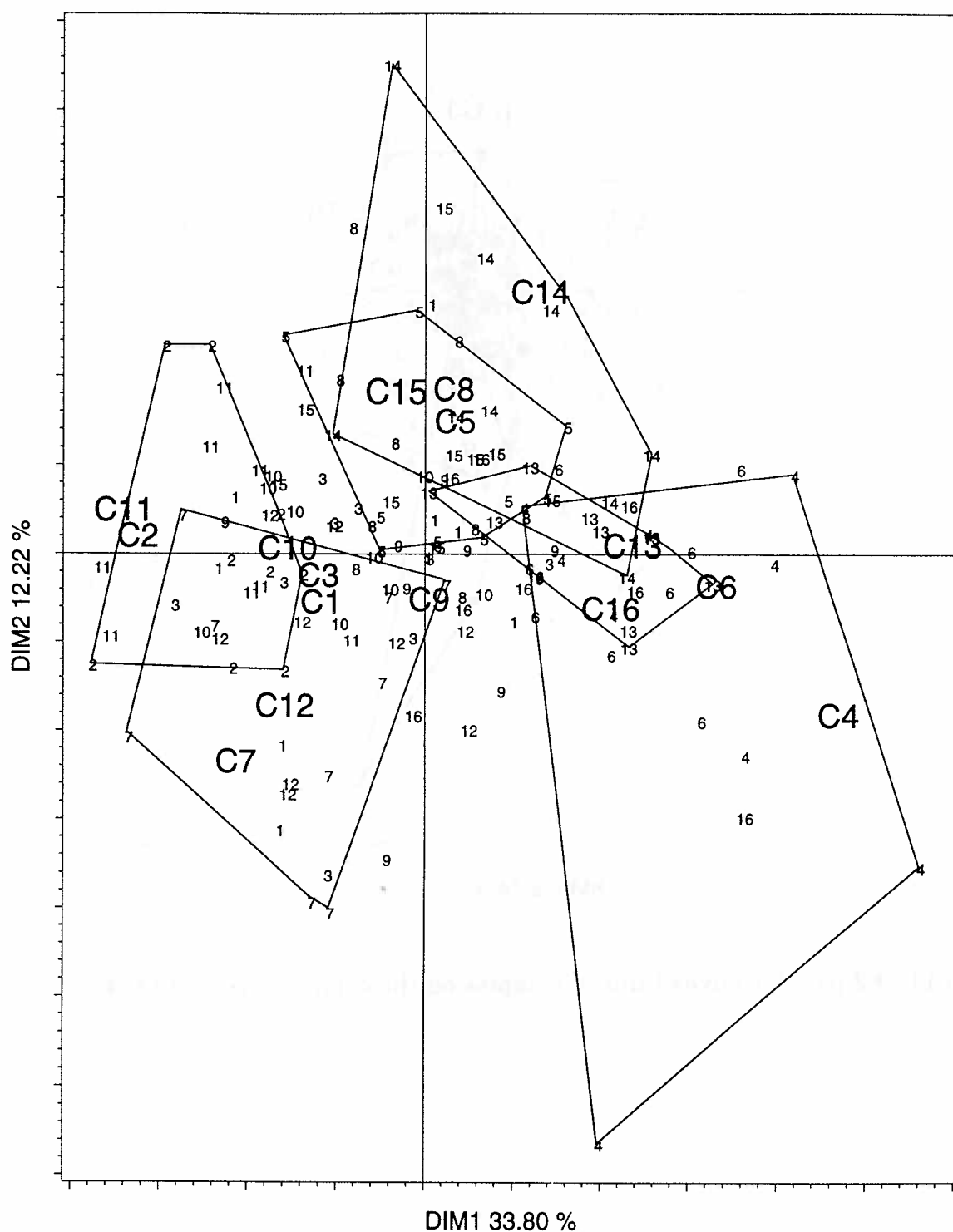


Figure 10 : F2 panel. Convex hulls of samples on the compromise plot (1, 2)

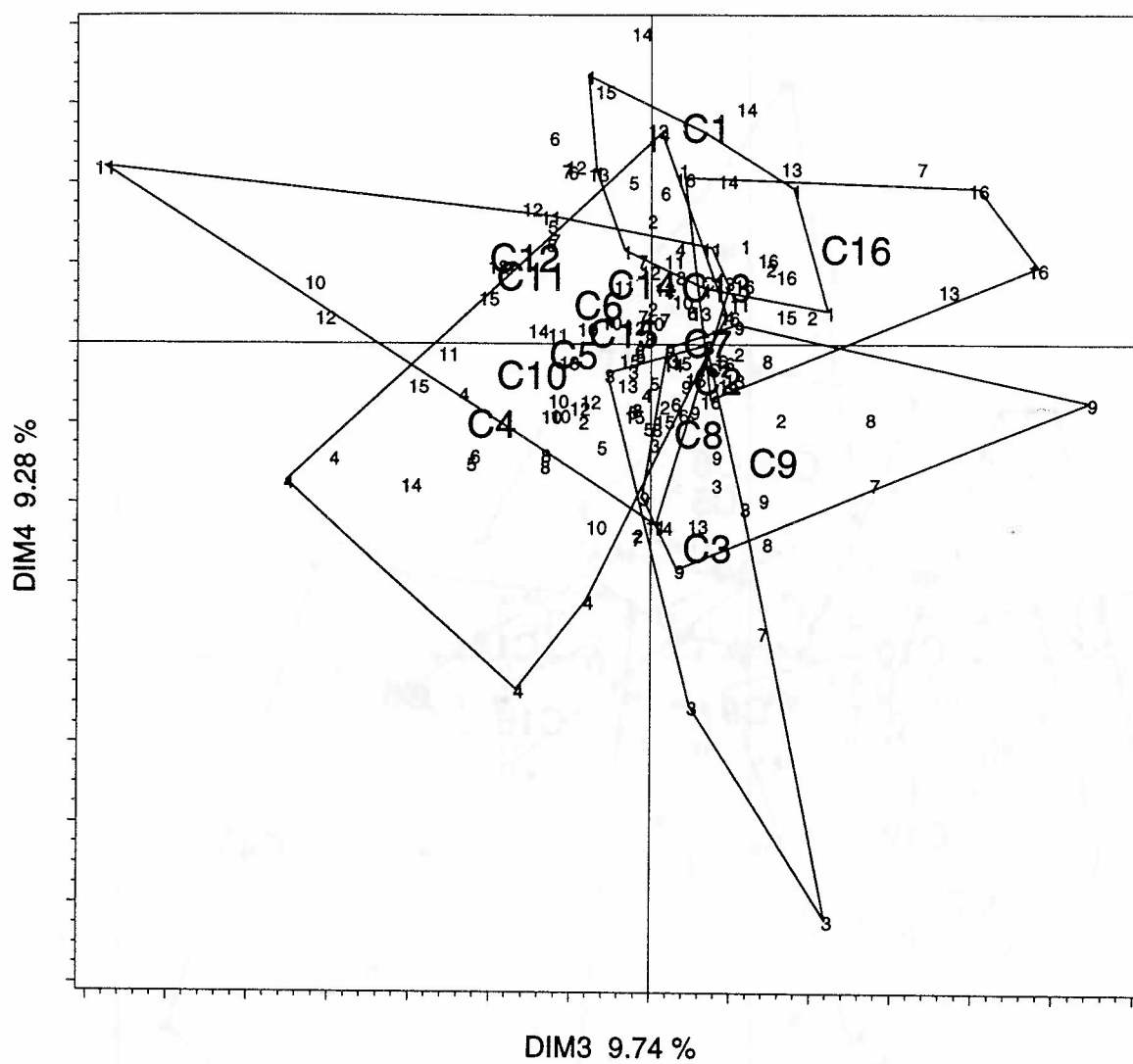


Figure 11 : F2 panel. Convex hulls of samples on the compromise plot (3, 4)

Sample and covariance plot (3,4) (Figures 9 and 13) suggest that sample 1 would have a more burnt flavour, as sample 16 which would also be more astringent and that samples 11 and 12 would have a more chemical flavour. But looking at the corresponding convex hulls in Figure 11, it is clear that the last finding about samples 11 and 12 may not be shared by most of the panel and that the previous conclusion about samples 1 and 16 is not universal.

Table 12

F2 panel. PCO of the sample compromise among subjects

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	0.8402	33.80	33.80
DIM 2	0.3038	12.22	46.02
DIM 3	0.2421	9.74	55.76
DIM 4	0.2306	9.28	65.04
DIM 5	0.1817	7.31	72.35
DIM 6	0.1239	4.98	77.33
DIM 7	0.0972	3.91	81.24
DIM 8	0.0934	3.76	85.00
DIM 9	0.0858	3.45	88.45
DIM10	0.0767	3.09	91.54
DIM11	0.0672	2.70	94.24
DIM12	0.0495	1.99	96.23
DIM13	0.0407	1.64	97.87
DIM14	0.0340	1.37	99.24
DIM15	0.0190	0.76	100.00

Figure 14 is a biplot from a covariance PCA of the mean score products. It is clear that this classical analysis leads to almost the same structure and to almost the same sample plot interpretation as in STATIS. Therefore, why should STATIS be used? Firstly because of the permutation tests provided by STATIS, secondly because of the information about assessor similarity provided by STATIS and thirdly because with less homogeneous panel than F2 the output of STATIS can be different from that obtained by means of a PCA of the mean score products. The biplot of assessor E data (Figure 15) is again very similar to the STATIS compromise illustrating why this subject had the largest normalized RV coefficient with the compromise. On the contrary, biplots from assessors G and F (Figures 16 and 17), who were the subjects who disagreed the most with the compromise, clearly show numerous discrepancies with the compromise plot.

4.4 STATIS of the S49 panel

This analysis gathers 49 assessors coming from 5 different panels having scored the same four attributes : bitterness, acidity, astringency and body/mouthfeel. The discussion will focus on panel comparisons (see sections 2.8 and 2.9 of this chapter). The mean RV coefficients within

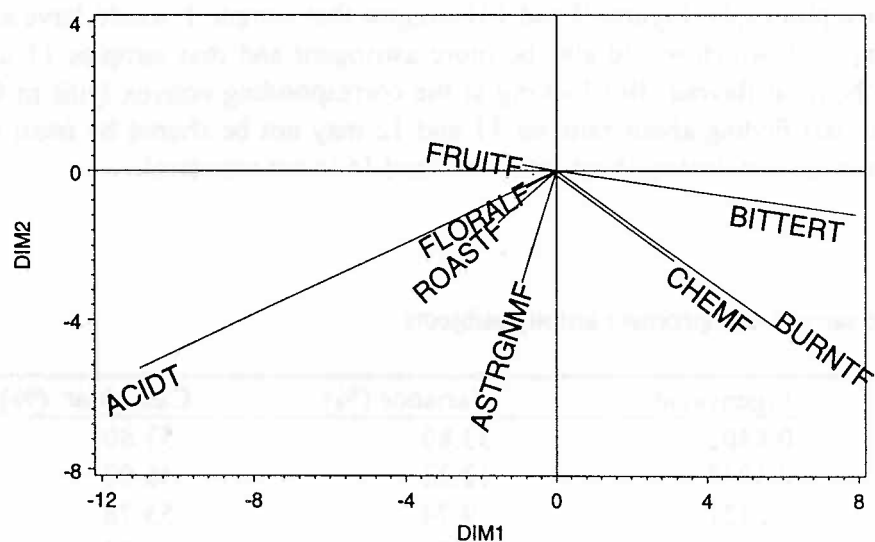


Figure 12 : F2 panel. Covariance plot (1, 2) of attributes

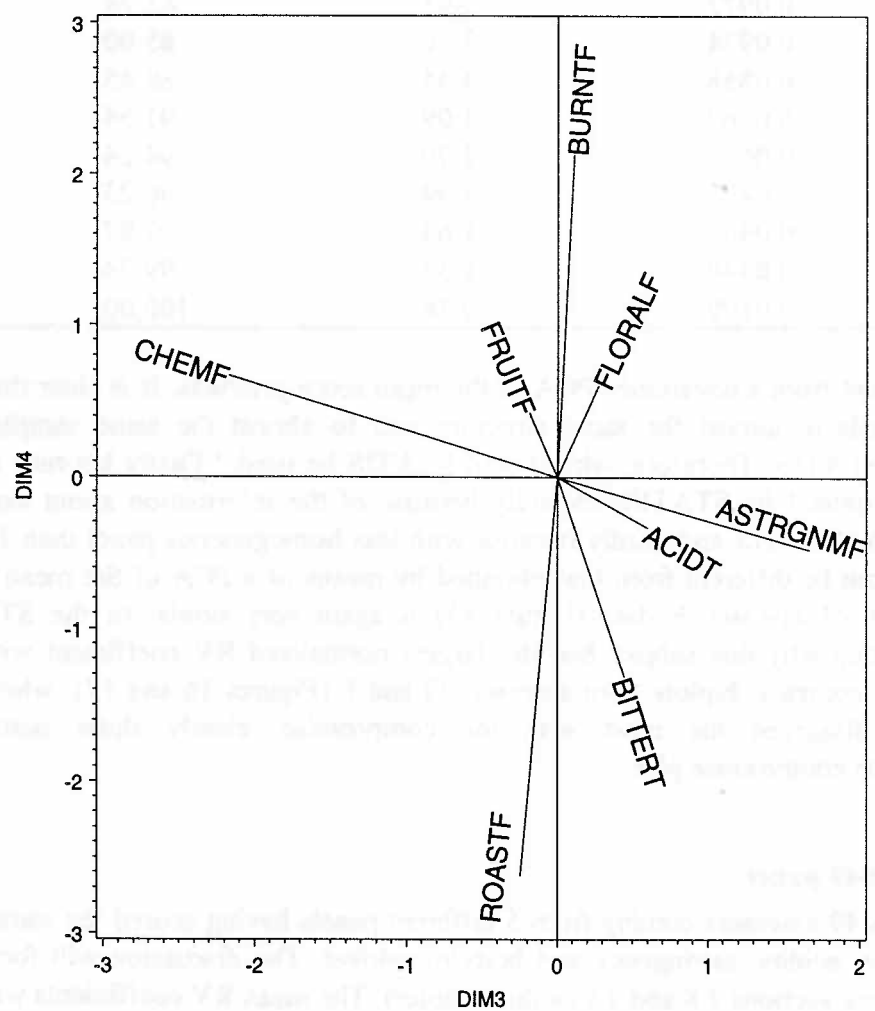


Figure 13 : F2 panel. Covariance plot (3, 4) of attributes

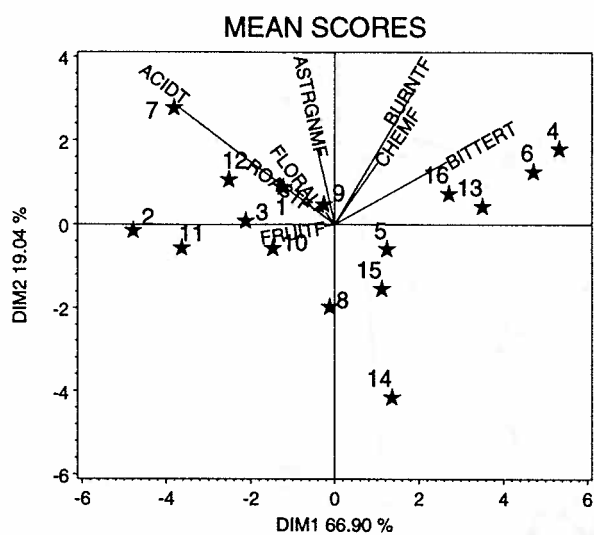


Figure 14 : F2 panel. Covariance PCA of mean score products

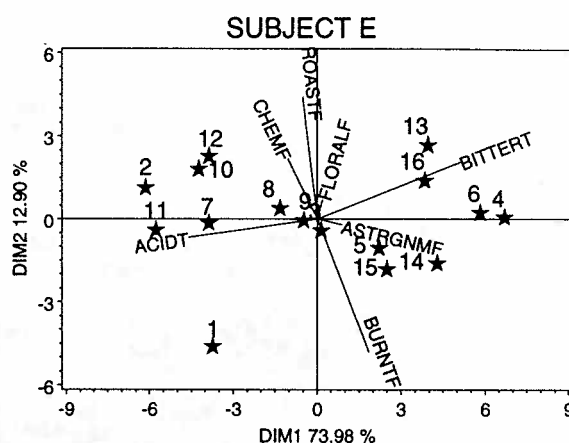


Figure 15 : F2 panel. Covariance PCA of scores from subject E

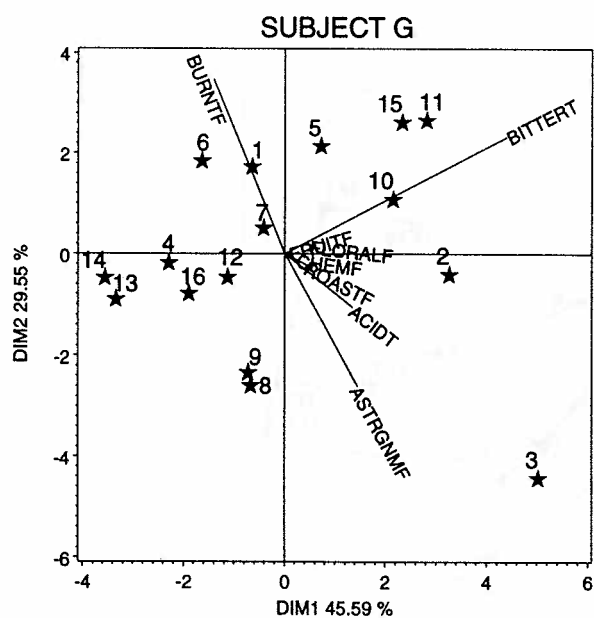


Figure 16 : F2 panel. Covariance PCA of scores from subject G

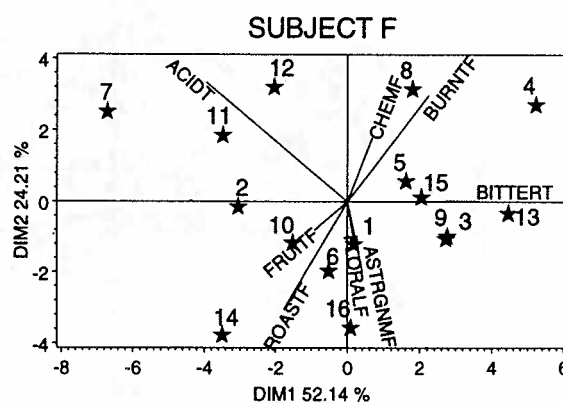


Figure 17 : F2 panel. Covariance PCA of scores from subject F

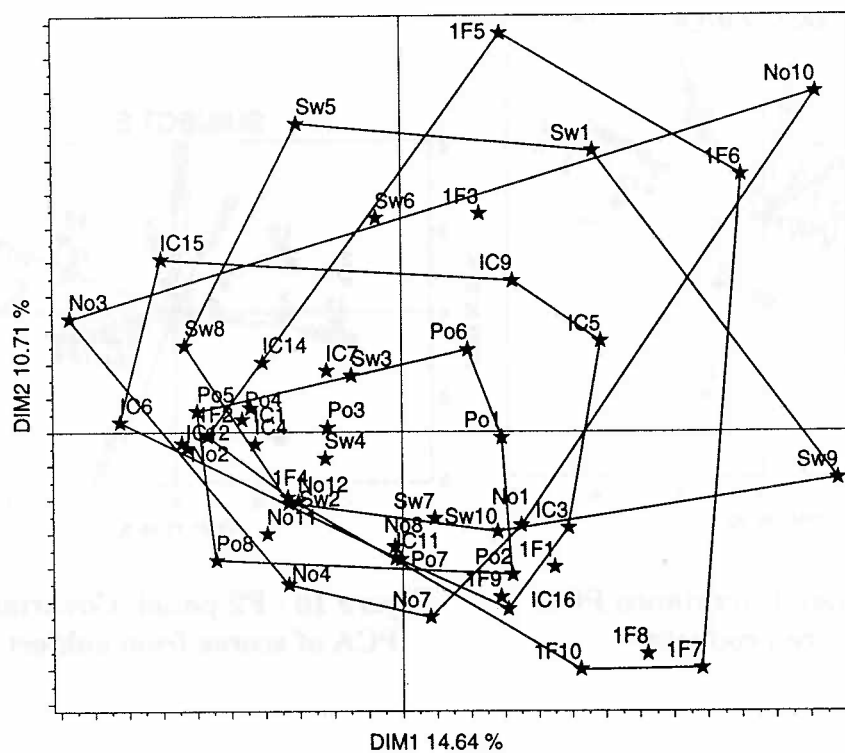


Figure 18 : S49 panel. PCO of the centered RV matrix

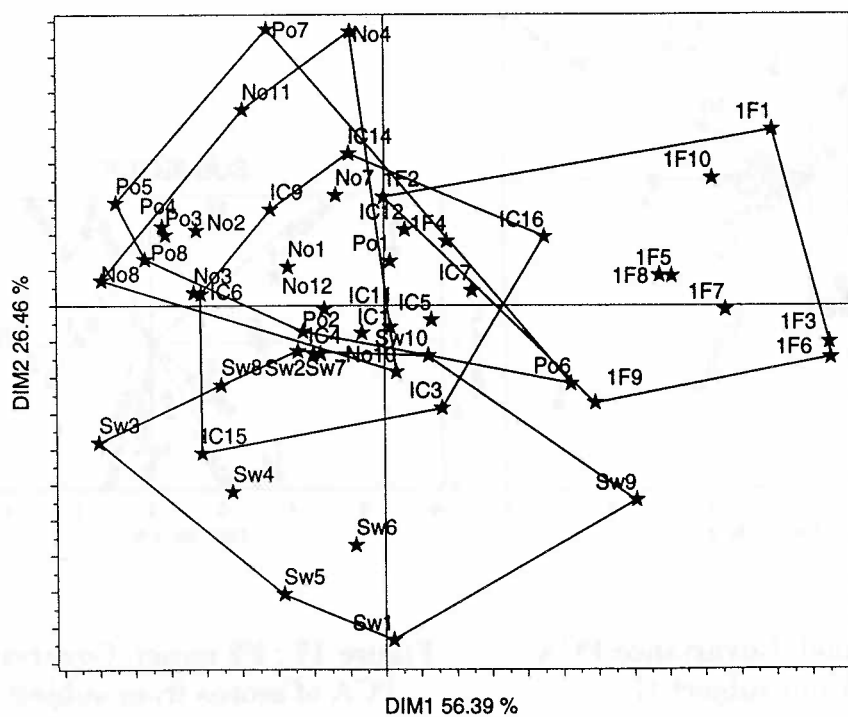


Figure 19 : S49 panel. CDA of panels 1F, IC, No, Po and Sw

Table 13
S49 panel. Mean RV coefficients within and between panels

	1F	IC	No	Po	Sw
1F	0.48				
IC	0.50	0.57			
No	0.49	0.57	0.56		
Po	0.49	0.56	0.57	0.56	
Sw	0.44	0.51	0.51	0.50	0.47

Table 14
S49 panel. Mean normalized RV coefficients within and between panels

	1F	IC	No	Po	Sw
1F	4.43				
IC	4.64	5.50			
No	4.65	5.58	5.54		
Po	4.63	5.46	5.69	5.55	
Sw	3.96	4.87	4.86	4.74	4.36

and between panels (Table 13) are rather similar and a little bit smaller when involving Swedish assessors (Sw panel). The corresponding normalized RV coefficients (Table 14) are all significant, meaning that the 5 panels are homogeneous and also agree among themselves about the sample structure, which does not automatically mean that the assessors are exchangeable among panels.

The first eigenvector of the RV matrix, which is the compromise, accounts for 54.74 % of the assessor variation (Table 15), which is a fairly good result considering that 48 axes exist in this PCO. The first plot of the PCO of the centered RV matrix does not show panel discrimination as exemplified by the panel convex hulls on Figure 18. But this plot accounts for only 25.35 % of the total information (Table 16). A good technique for mapping panel discrimination is the CDA suggested in section 2.9. For the S49 panel, the first two axes of this CDA accounts for 82.84 % of the panel discrimination (Table 17) and splits panel 1F and Sw between themselves and from the three other panels (Figure 19). The Norwegian panel (No) and the Polish panel (Po) do not seem different, meaning that the assessors might be exchangeable. Finally, the ICO panel appeared as the central panel on this map. In order to test significance of panel discrimination visually observed on this map, the permutation technique of section 2.9 was performed on the basis of the 16 first axes from the centered PCO and its output was 2.521, which is fairly significant. Therefore, one cannot exchange the assessors among the five panels, even if, on the average, they fairly agree about the sample structure. This conclusion was met by the author with several other data sets; it raises a controversial point associated to the permutation techniques described in this chapter and already mentioned above : are these tests too powerful because of a too unrealistic null hypothesis ?

Table 15

S49 panel. PCO of the non centered RV matrix

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	26.8248	54.74	54.74
DIM 2	3.1572	6.44	61.19
DIM 3	1.8522	3.78	64.97
DIM 4	1.7188	3.51	68.48
DIM 5	1.4765	3.01	71.49
DIM 6	1.3574	2.77	74.26
DIM 7	1.2311	2.51	76.77
DIM 8	1.0794	2.20	78.97
DIM 9	1.0355	2.11	81.09
DIM10	0.9631	1.97	83.05

Table 16

S49 panel. PCO of the centered RV matrix

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	3.3992	14.64	14.64
DIM 2	2.4869	10.71	25.35
DIM 3	1.7418	7.50	32.85
DIM 4	1.4779	6.36	39.21
DIM 5	1.3591	5.85	45.07
DIM 6	1.2317	5.30	50.37
DIM 7	1.0883	4.69	55.06
DIM 8	1.0385	4.47	59.53
DIM 9	0.9846	4.24	63.77
DIM10	0.8503	3.66	67.43

Table 17

S49 panel. Panel canonical discriminant analysis

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	2.4898	56.39	56.39
DIM 2	1.1684	26.46	82.84
DIM 3	0.5238	11.86	94.71
DIM 4	0.2337	5.29	100.00

Table 18
S49 panel. PCO of the sample compromise among subjects

Axis	Eigenvalue	Variance (%)	Cum. Var. (%)
DIM 1	0.7322	56.64	56.64
DIM 2	0.1416	10.95	67.59
DIM 3	0.0518	4.01	71.60
DIM 4	0.0470	3.64	75.24
DIM 5	0.0425	3.29	78.53
DIM 6	0.0412	3.19	81.72
DIM 7	0.0392	3.03	84.75
DIM 8	0.0331	2.56	87.31
DIM 9	0.0326	2.52	89.83
DIM10	0.0281	2.18	92.01
DIM11	0.0250	1.94	93.94
DIM12	0.0242	1.87	95.82
DIM13	0.0212	1.64	97.46
DIM14	0.0172	1.33	98.78
DIM15	0.0157	1.22	100.00

From the distribution of the eigenvalues of the compromise (Table 18) and from the mean b coefficient over the 49 assessors which was equal to 1.75 ('Mean' line and 'BETA' column in Table 19), it was decided to interpret the two first axes of the sample space given in Figure 20. This structure is once again similar to those already met with ICO and F2 panels, namely an opposition along the first axis of acidity and bitterness (Figure 21). Note that the astringency and the body/mouthfeel attributes are found correlated to bitterness.

The magnitude of the 'Scaling' column in Table 19, or in Table 20 which is Table 19 averaged by panel, is highly dependent on the panels simply because of the different scales used by these panels. In such a case, STATIS must be done with the global matrix standardization recommended in section 2.4. The 'NRV' and 'NRVC' columns in Table 19 point out the assessors No10 (assessor 10 from Norway) and assessor 1F6 (assessor 6 from France) as presenting a strong disagreement with the panel and with the compromise. The individual biplots of these two assessors (Figures 22 and 23) are indeed extremely different from the sample structures we have met up to now. For instance, assessor No10 found sample 8 very astringent and assessor 1F6 seems to correlate positively bitterness and acidity, therefore he did not make many differences between the acid samples 11 and 12 and the bitter samples 4, 13 and 16. One assessor for each of the three other panels (Sw1, IC9 and Po1) is also analysed by means of a biplot (Figures 24, 25 and 26). These assessors were chosen because they had the lowest agreement with the whole panel, but being still significant they do not present a very big difference with the compromise sample structure. Interestingly, all of them seem to present a null correlation between bitterness and acidity instead of a negative one. But comparison of attribute correlation structures is a matter of Dual STATIS which will be performed in the following section.

Table 19
S49 panel. Summary of individual statistics

Subject	Scaling	NRV	Weight	BETA	NRVC
1F 1	0.009	4.086	0.869	1.696	4.216
1F10	0.012	5.241	1.035	1.423	5.521
1F 2	0.015	6.129	1.197	1.729	6.475
1F 3	0.005	3.217	0.749	2.556	3.074
1F 4	0.017	6.361	1.232	1.932	6.695
1F 5	0.009	1.992	0.549	2.295	1.682
1F 6	0.011	1.227	0.442	2.609	0.769
1F 7	0.016	5.109	1.001	1.264	5.334
1F 8	0.012	5.306	1.048	1.644	5.497
1F 9	0.012	5.917	1.140	1.383	6.239
IC 1	0.017	6.264	1.211	1.951	6.568
IC11	0.011	5.993	1.178	2.163	6.249
IC12	0.057	5.835	1.153	1.579	6.209
IC14	0.016	4.304	0.907	1.464	4.543
IC15	0.037	4.808	0.997	1.825	5.024
IC16	0.023	5.672	1.111	1.843	5.939
IC 3	0.026	4.971	0.995	1.606	5.140
IC 4	0.061	5.658	1.119	1.531	5.994
IC 5	0.009	3.796	0.821	2.224	3.708
IC 6	0.037	6.031	1.179	1.309	6.477
IC 7	0.038	5.480	1.084	2.294	5.663
IC 9	0.005	3.543	0.817	3.121	3.327
No 1	0.010	5.006	1.005	1.616	5.214
No10	0.009	0.863	0.323	1.544	0.466
No11	0.037	6.312	1.219	1.462	6.717
No12	0.018	6.259	1.196	1.212	6.674
No 2	0.016	6.378	1.226	1.356	6.824
No 3	0.049	5.649	1.110	1.097	6.084
No 4	0.032	5.825	1.130	1.085	6.301
No 7	0.015	5.676	1.117	1.500	5.998
No 8	0.015	5.257	1.066	1.930	5.494
Po 1	0.651	3.787	0.845	2.458	3.785
Po 2	3.173	5.460	1.066	1.246	5.739
Po 3	0.977	5.436	1.086	1.985	5.653
Po 4	2.510	4.921	1.012	1.744	5.188
Po 5	3.605	5.605	1.110	1.575	5.955
Po 6	0.730	4.350	0.909	1.963	4.393
Po 7	2.914	5.114	1.014	1.237	5.462
Po 8	3.631	6.867	1.303	1.219	7.376
Sw 1	2.583	2.252	0.587	2.177	2.005
Sw10	2.051	5.438	1.048	1.089	5.769
Sw 2	7.994	5.672	1.116	1.237	6.045
Sw 3	1.462	4.862	0.996	2.119	5.029
Sw 4	3.983	6.143	1.190	1.651	6.462
Sw 5	1.306	3.484	0.803	2.582	3.392
Sw 6	1.392	3.239	0.771	2.749	3.123
Sw 7	4.015	5.524	1.091	1.617	5.767
Sw 8	2.270	5.400	1.069	1.341	5.764
Sw 9	3.096	3.583	0.759	1.503	3.493
Mean	1.000	4.925	1.000	1.750	5.113

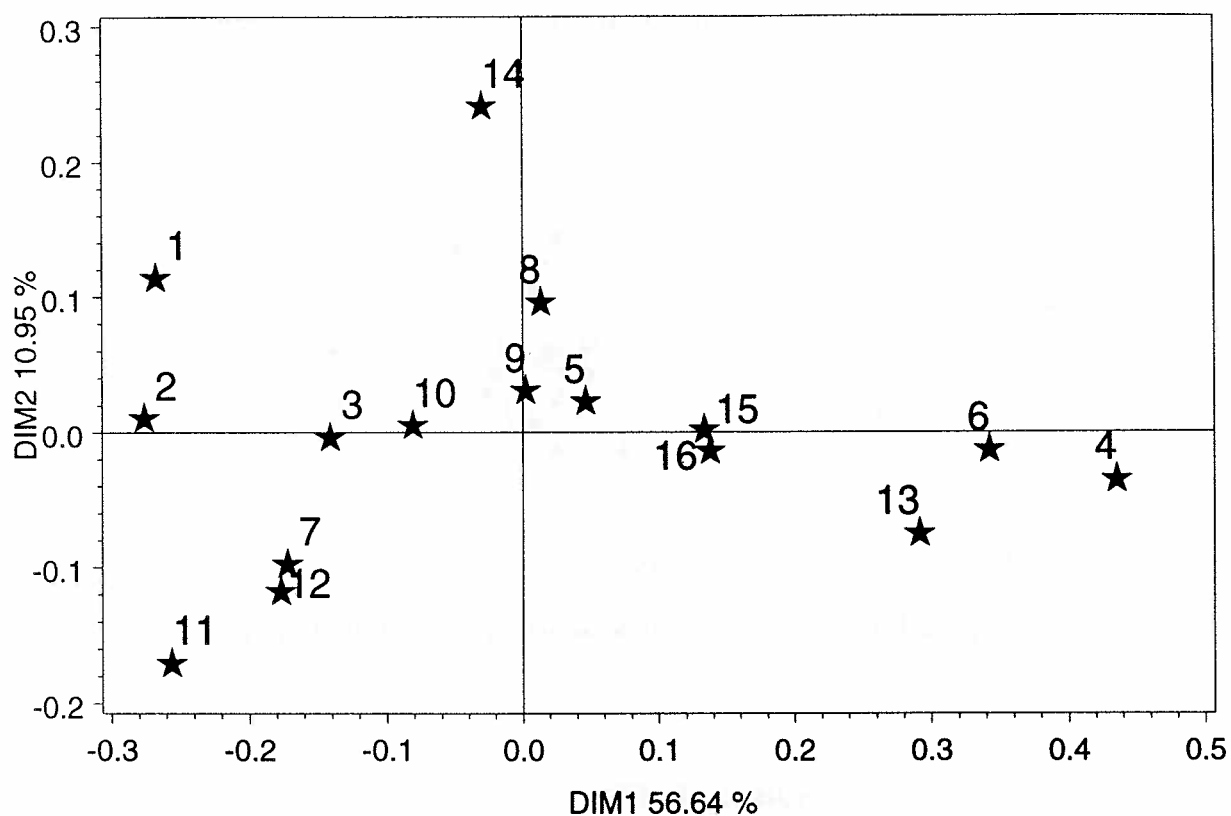


Figure 20 : S49 panel. PCO of the sample compromise among subjects

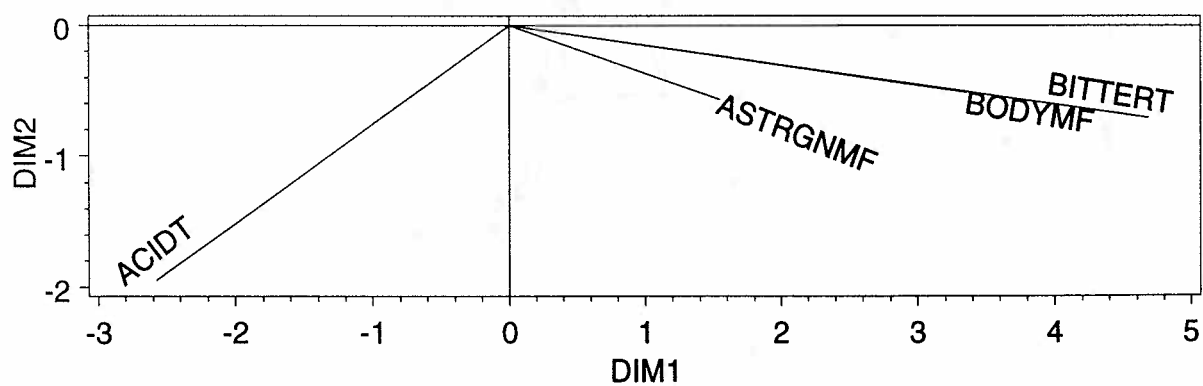


Figure 21 : S49 panel. Covariance plot of attributes

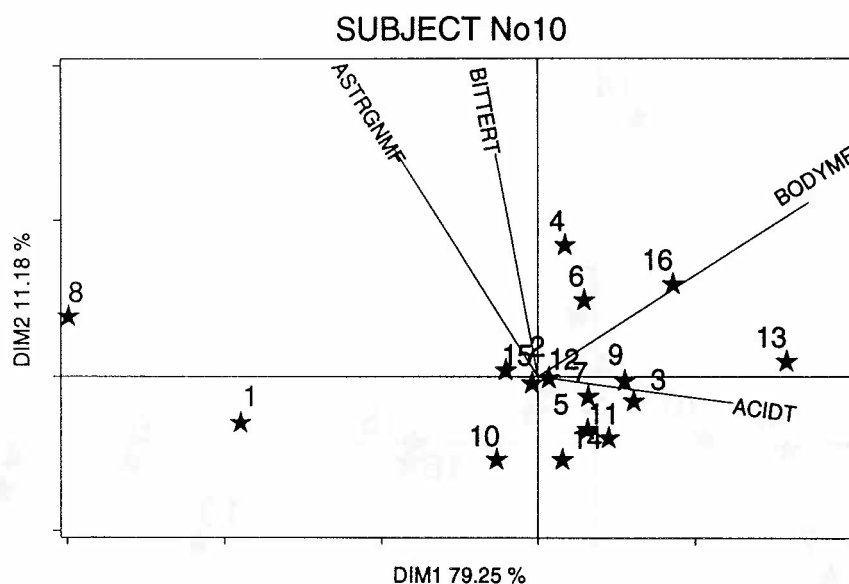


Figure 22 : S49 panel. Covariance PCA of scores from subject No10

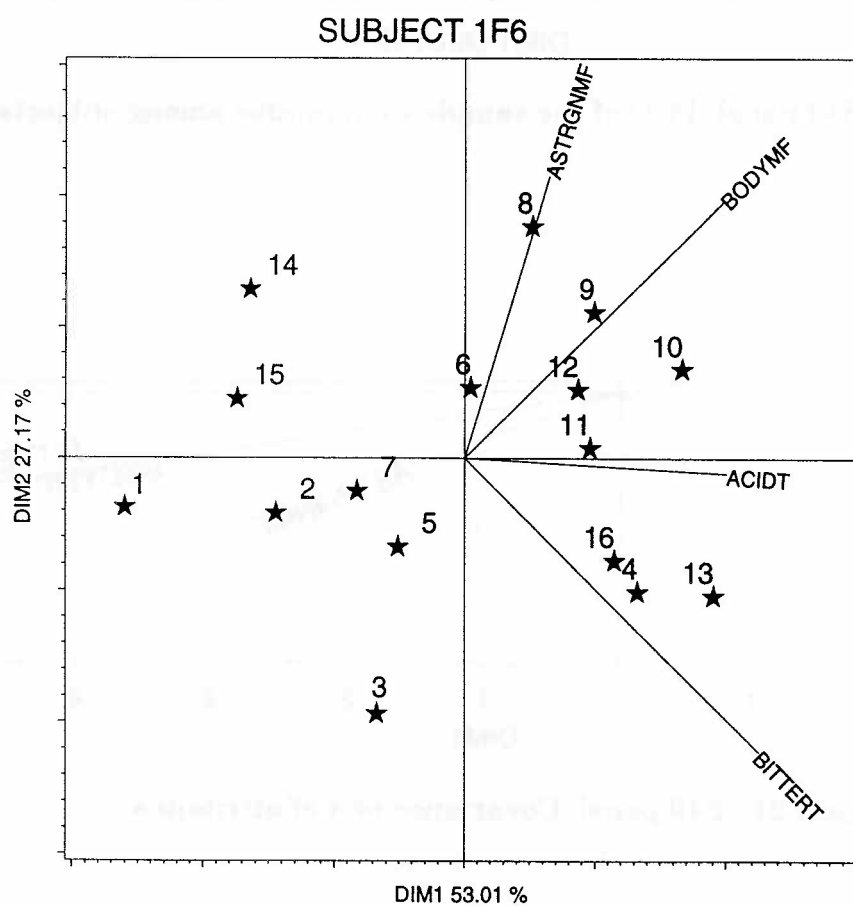


Figure 23 : S49 panel. Covariance PCA of scores from subject 1F6

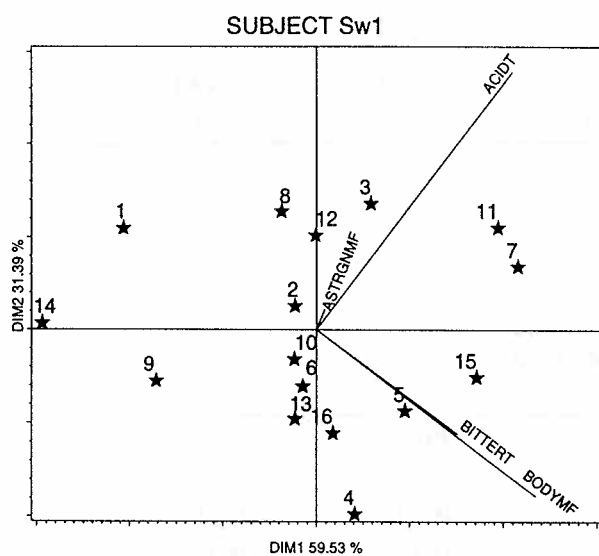


Figure 24 : S49 panel. Covariance PCA of scores from subject Sw1

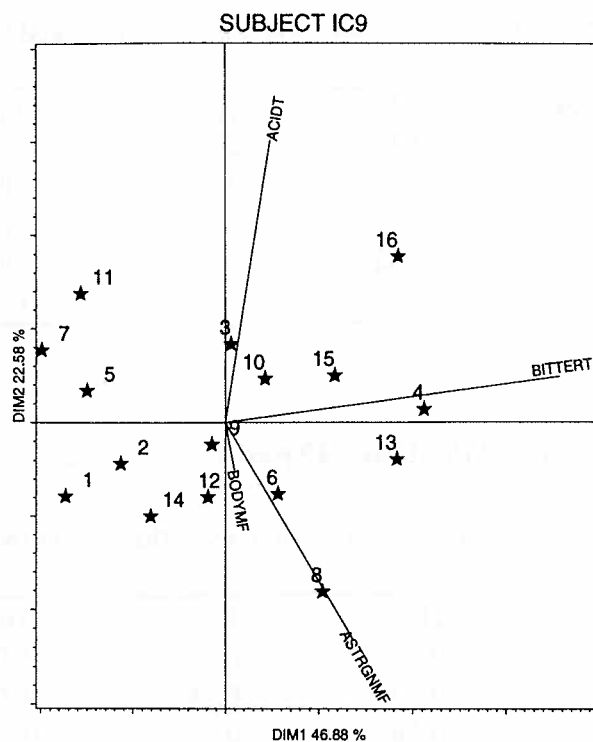


Figure 25 : S49 panel. Covariance PCA of scores from subject IC9

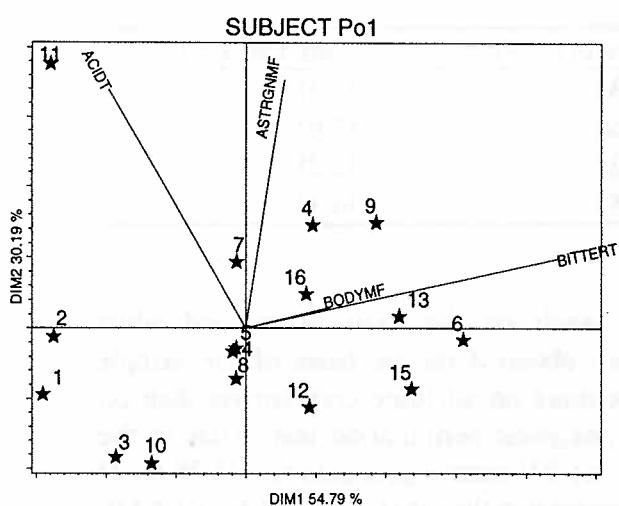


Figure 26 : S49 panel. Covariance PCA of scores from subject Po1

Table 20

S49 panel. Summary of individual statistics averaged by panel

Subject	Scaling	NRV	Weight	BETA	NRVC
IF	0.012	4.459	0.926	1.853	4.55
IC	0.028	5.196	1.048	1.909	5.403
No	0.022	5.247	1.044	1.422	5.53
Po	2.274	5.192	1.043	1.678	5.444
Sw	3.015	4.56	0.943	1.807	4.685

4.5 Dual STATIS of the S49 panel

Table 21

S49 dual panel. Mean RV coefficients within and between panels

	IF	IC	No	Po	Sw
IF	0.77	0.72	0.56	0.73	0.70
IC	0.72	0.78	0.74	0.79	0.78
No	0.56	0.74	0.75	0.73	0.75
Po	0.73	0.79	0.73	0.82	0.80
Sw	0.70	0.78	0.75	0.80	0.77

Table 22

S49 dual panel. PCO of the non centered RV matrix

Axis	Eigenvalue	Variance (%)	Cum.Var. (%)
DIM 1	36.9344	75.38	75.38
DIM 2	5.7214	11.68	87.05
DIM 3	2.5591	5.22	92.28
DIM 4	1.8882	3.85	96.13

The mean RV coefficients within and between panels on the basis of the individual correlation matrices (Table 21) are greater than those obtained on the basis of the sample spaces (Table 13), suggesting that the assessors agree more on attribute correlations than on sample differences. As explained in section 2.10, no analytical permutation test exists in the framework of Dual STATIS. The first eigenvector of the RV matrix accounts for 75.38 % of the PCO (Table 22), which again denotes a better compromise than that obtained by STATIS (54.74 %, Table 15). The first three axes of the PCO of the centered RV matrix (Table 23) explain almost all the variation, which once again was definitely not the case in classical STATIS (Table 16). But the reader should remember that this Dual STATIS compares matrices of size (4,4), whereas the classical STATIS compared matrices of size (16,16);

Table 23
S49 dual panel. PCO of the centered RV matrix

Axis	Eigenvalue	Variance (%)	Cum.Var. (%)
DIM 1	5.8397	46.54	46.54
DIM 2	2.5852	20.60	67.15
DIM 3	1.9749	15.74	82.89
DIM 4	0.8342	6.65	89.54

Table 24
S49 dual panel. Summary of individual statistics

Subject	Weight	Subject	Weight	Subject	Weight	Subject	Weight	Subject	Weight
1F 1	1.119	IC 1	1.112	No 1	0.798	Po 1	1.050	Sw 1	0.968
1F10	0.795	IC11	1.083	No10	0.611	Po 2	1.024	Sw10	1.075
1F 2	1.109	IC12	1.060	No11	1.078	Po 3	1.064	Sw 2	1.041
1F 3	0.950	IC14	1.096	No12	1.024	Po 4	0.853	Sw 3	1.041
1F 4	0.874	IC15	1.090	No 2	0.978	Po 5	1.067	Sw 4	1.090
1F 5	0.991	IC16	1.072	No 3	0.994	Po 6	1.130	Sw 5	1.107
1F 6	0.819	IC 3	0.984	No 4	0.990	Po 7	1.022	Sw 6	1.017
1F 7	0.926	IC 4	1.068	No 7	1.063	Po 8	1.138	Sw 7	1.122
1F 8	0.987	IC 5	0.708	No 8	1.119			Sw 8	1.021
1F 9	0.797	IC 6	1.000					Sw 9	0.795
		IC 7	1.084						
		IC 9	0.994						
Mean	0.937	Mean	1.029	Mean	0.962	Mean	1.044	Mean	1.028

Table 24 gives the Dual STATIS weights of the 49 assessors. One can observe that they are quite similar among assessors and also among panels. These weights allow the compromise correlation matrix (Table 25) to be derived as a weighted mean of the 49 individual correlation matrices.

Table 25
S49 dual panel. Compromise correlation matrix

	ACIDT	ASTRGNMF	BITTERT	BODYMF
ACIDT	1.00			
ASTRGNMF	-0.05	1.00		
BITTERT	-0.30	0.55	1.00	
BODYMF	-0.24	0.42	0.63	1.00

therefore the better agreement raised in Dual STATIS may be due, or partly due, to this difference. Figure 27 scatters assessors according to Dual STATIS weights and classical STATIS weights. Therefore, assessors located at the bottom of this plot disagree the most

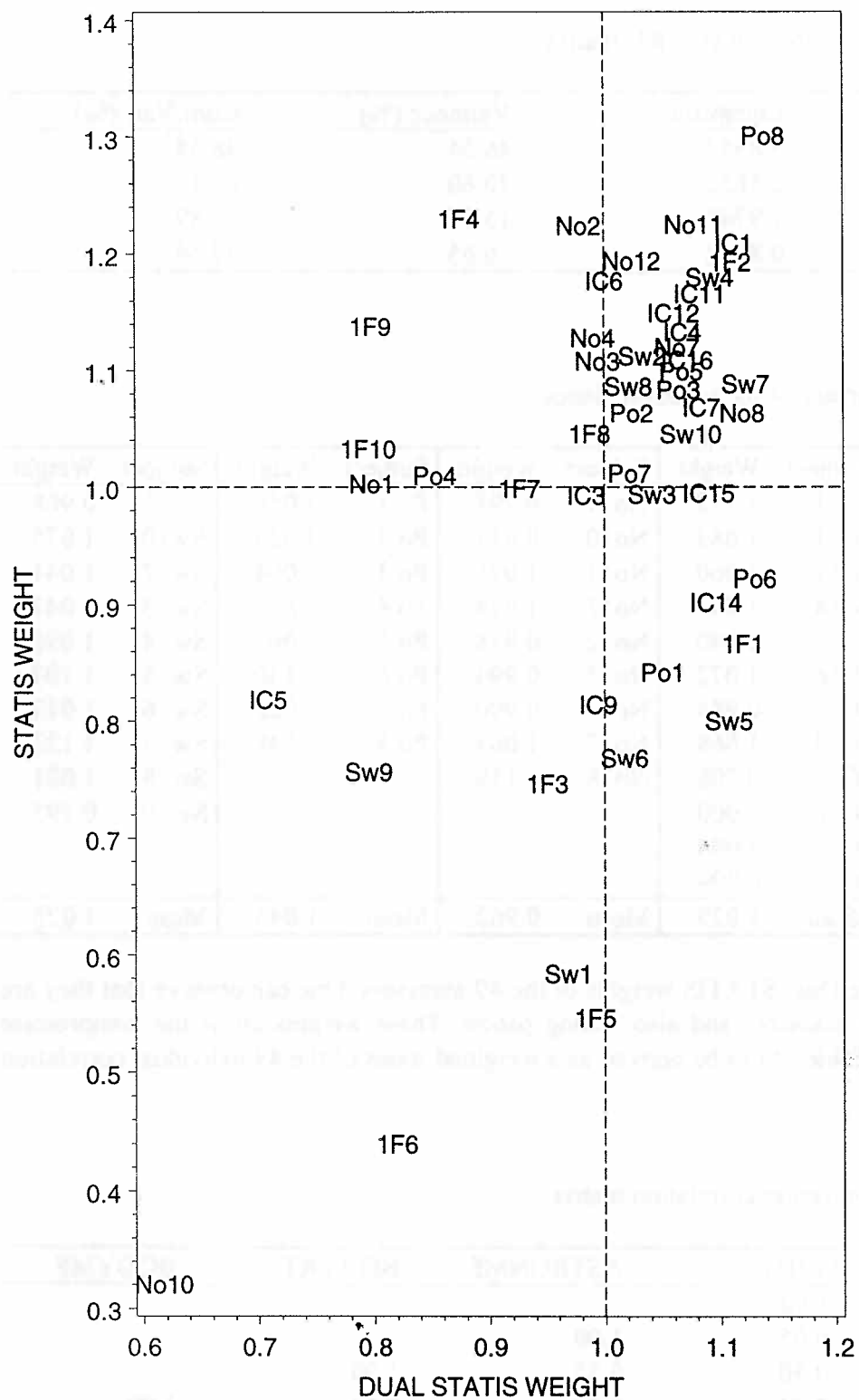


Figure 27 : S49 panel. STATIS and DUAL STATIS weights of subjects

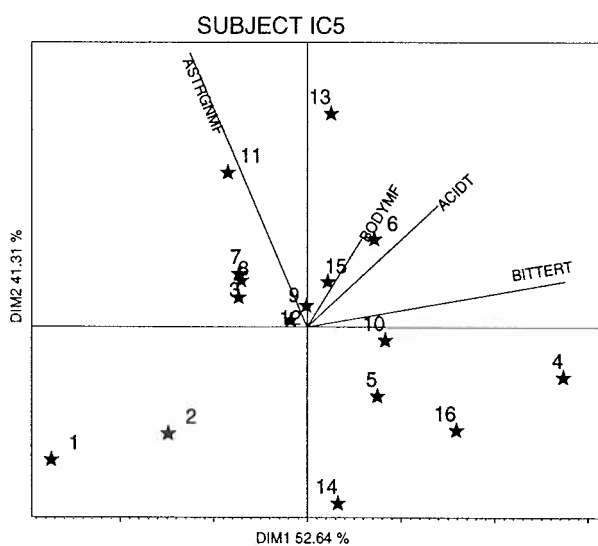


Figure 28 : S49 panel. Covariance PCA of scores from subject IC5

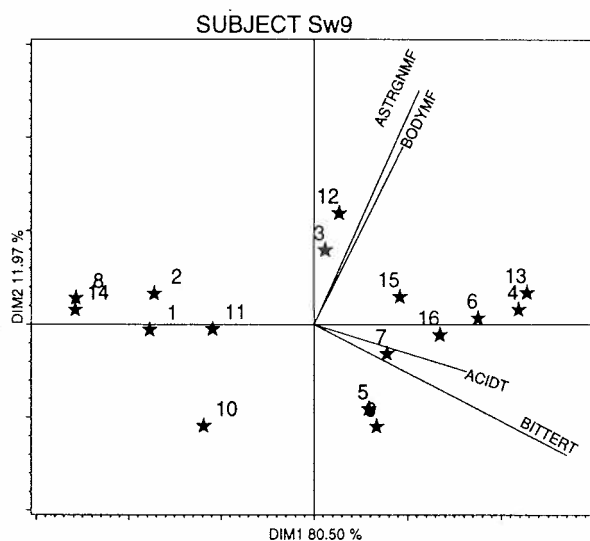


Figure 29 : S49 panel. Covariance PCA of scores from subject Sw9

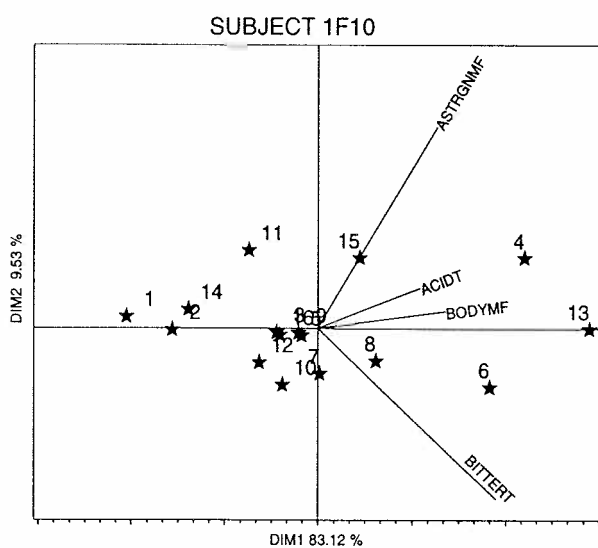


Figure 30 : S49 panel. Covariance PCA of scores from subject 1F10

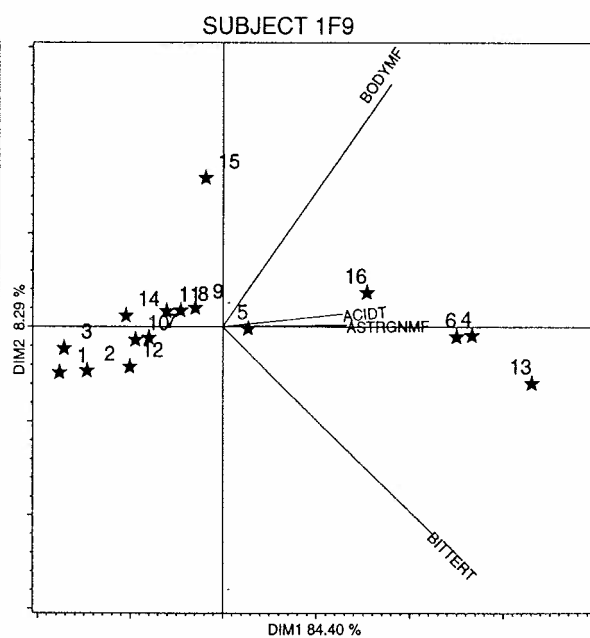


Figure 31 : S49 panel. Covariance PCA of scores from subject 1F9

with the panel on sample differences, whereas assessors located on the left of this plot disagree the most with the panel on attribute correlations. The worst assessors are obviously those located in the bottom left quadrant of this plot. The worst assessor is definitely No10, whose biplot has already been given in Figure 22. On the contrary, the best assessor seems to be Po8. The biplot of four assessors, who poorly agree with the Dual STATIS compromise, are given in Figures 28, 29, 30 and 31. IC5 and Sw9 have in common to correlate positively bitterness and acidity. The two French subjects 1F10 and 1F9 have very similar biplots, unless for acidity which is highly correlated with astringency for 1F9, whereas it is with body/mouthfeel for 1F10.

5. SOFTWARE

Every computation and graph presented in this chapter was done on a SUN workstation under the UNIX system and within the SAS[®] software, thanks to several macros developed by the author. These macros should be announced elsewhere.

6. CONCLUSION

Advantages and limitations of the techniques proposed have been discussed throughout the chapter.

The most important advantages were :

- Ability to take free-choice profiling into account
- RV coefficient for measuring similarity between two sample spaces
- Permutation tests for validation of the panel homogeneity with almost no computation
- β coefficient of individual dimensionality (complexity)
- STATIS compromise on sample distances or attribute correlations
- Compromise obtained as a mean of assessors weighted by their individual agreement with the panel
- Cross-panel comparisons
- Analytical, instead of iterative, techniques

The principal limitations were :

- Possible over-powerful tests
- No analytical permutation tests with Dual STATIS
- Individual weights only depend on agreement with the panel.

The comparison of these two lists makes it clear why the author does think that the RV related techniques have a great potential in the sensory field. Nevertheless, this cannot be recognized by the sensory community before they have been successfully applied by many sensory scientists in a wide range of situations. This could not happen without papers such as this chapter and without availability of dedicated softwares.

7. ACKNOWLEDGMENT

The author would like to thank Yves Escoufier for introducing the RV coefficient to him long time ago and Frédérique Kazi-Aoual for sharing as soon as possible her useful theoretical results with praticians. The author is also very grateful to his colleagues from ESN who allowed this paper to be illustrated by the coffee data. Finally the help of Patricia Collin with the English language and the help of Stéphanie Degoud with tables and figures were greatly appreciated.

8. REFERENCES

- Arnold, G. M., & Williams, A. A. (1986). The use of generalized procrustes techniques in sensory analysis. In J. R. Piggott (Ed.), *Statistical procedures in food research* (pp. 233-253): Elsevier Applied Science.
- Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika*, 35(3), 283-319.
- Dijksterhuis, G. B., & Gower, J. C. (1991/2). The interpretation of generalized procrustes analysis and allied methods. *Food Quality and Preference*, 3, 67-87.
- Dijksterhuis, G. B., & Punter, P. (1990). Interpreting generalized procrustes analysis 'Analysis of Variance' tables. *Food Quality and Preference*, 2, 255-265.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New-York: Chapman & Hall.
- Escoufier, Y. (1973). Le traitement des variables vectorielles. *Biometrics*, 29, 751-760.
- Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3 and 4), 325-338.
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40(1), 33-51.
- Jolliffe, I. T. (1986). *Principal component analysis*. New-York: Springer-Verlag.
- Kazi-Aoual, F. (1992). Apport de la méthode STATIS pour étudier l'homogénéité d'un panel. Paper presented at the 3èmes journées européennes "Agro-Industrie et Méthodes Statistiques", Montpellier.
- Kazi-Aoual, F. (1993). Approximations to permutation tests for data analysis (Report n° 93-06). E.N.S.A. Montpellier, unité de Biométrie.
- Kazi-Aoual, F., Hitier, S., Sabatier, R., & Lebreton, J.-D. (1995). Refined approximations to permutation tests for multivariate inference. *Computational Statistics and Data Analysis*, in press.
- L'Hermier des Plantes, H. (1976). *Structuration des tableaux à trois indices de la statistique*. Thesis, Montpellier II.
- Lavit, C. (1988). *Analyse conjointe de tableaux quantitatifs*. Paris: Masson.
- Lavit, C., Escoufier, Y., Sabatier, R., & Traissac, P. (1994). The ACT (STATIS method). *Computational Statistics and Data Analysis*, 18, 97-119.
- Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. London: Academic Press.
- McEwan, J. A., & Schlich, P. (1991/2). Correspondence analysis in sensory evaluation. *Food Quality and Preference*, 3, 23-36.

- Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods : the RV-coefficient. *Appl. Statist.*, 25(3), 257-265.
- Scheffe, H. (1959). *The Analysis of Variance*. New-York: John Wiley & Sons.
- Schiffmann, S. S., Reynolds, M. L., & Young, F. W. (1981). *Introduction to multidimensional scaling. Theory, methods and applications*. London: Academic Press.
- Schlich, P. (1989). A SAS/IML[®] program for generalised procrustes analysis. Paper presented at the "SEUGI 89" SAS European Users Group International Conference, Cologne.
- Schlich, P. (1993). *Contributions à la sensométrie*. Thesis, Paris XI Orsay.
- Schlich, P., & Guichard, E. (1989). Selection and classification of volatile compounds of apricot using the RV coefficient. *Journal of Agricultural and Food Chemistry*, 37(1), 142-150.
- Schlich, P., Issanchou, S., Guichard, E., Etievant, P., & Adda, J. (1987). RV coefficient: a new approach to select variables in PCA and to get correlations between sensory and instrumental data. Paper presented at the 5th Weurman Flavour Research Symposium, Oslo.
- Wakeling, I. N., Raats, M. M., & Macfie, H. J. H. (1992). A new significance test for consensus in generalized procrustes analysis. *J. of Sensory Studies*, 7, 91-96.
- Williams, A. A., & Langron, S. P. (1984). The use of free-choice profiling for the evaluation of commercial ports. *J. Sci. Food Agric.*, 35, 558-56

Article 4

4. Callier P. & Schlich P. (1997). La cartographie des préférences incomplètes. Validation par simulation. *Sciences des Aliments*, **17**(2), 155-172.

La cartographie des préférences incomplètes – Validation par simulation

Philippe CALLIER¹, Pascal SCHLICH^{2*}

SUMMARY Generalization of preference mapping to incomplete block designs.

Classical techniques of Preference Mapping require that each consumer gives a preference score to each product. In many situations, it is difficult to ask a consumer to taste more than, say, four or five samples, while the total number of products in a Preference Mapping study is usually greater than seven. Mutually Orthogonal Latin Squares have been recommended in these situations for designing incomplete experiments, balanced for the order and carry-over effects. The paper proposes a technique to derive homogeneous clusters of consumers from incomplete data sets. Each cluster can then be summarized by its vector of product mean scores, making it possible to apply any technique of preference mapping on these smaller, but complete, data sets. A Monte-Carlo simulation was undertaken in order to study the accuracy of this method with respect to several parameters. The most important parameters were the number of products assessed by each consumer and the complexity of the preference structure. Technically speaking, complete data sets were simulated according to a multi-cluster model and then made incomplete by removing a given number of products for each consumer without disturbing the balance of the design. Clustering from these simulated incomplete data sets, based on the proposed method, was then compared to that obtained from the complete data sets.

Key-words : *preference mapping, balanced incomplete block design, mutually orthogonal latin squares, Monte-Carlo simulation.*

RÉSUMÉ

Les techniques classiques de cartographie des préférences nécessitent que chaque consommateur évalue tous les produits de l'étude. Or, les modèles statistiques ajustés requièrent au minimum sept produits. Dans la pratique, il est parfois difficile de demander à un consommateur d'évaluer un tel nombre de produits au cours de la même session. Des carrés latins mutuellement orthogonaux sont recommandés pour la planification d'études en plans incomplets : ils assurent un équilibre parfait des effets d'ordre et de report de tous ordres. Une méthode per-

1. Groupe DANONE, 7, rue de Téhéran, 75381 Paris cedex 08, France.

2. Laboratoire de recherches sur les arômes, Institut national de la recherche agronomique, 17, rue Sully, 21000 Dijon, France.

* Correspondance.

mettant d'obtenir des classes homogènes de consommateurs à partir de données incomplètes est proposée. Chaque classe peut ensuite être résumée par son vecteur de moyennes produits établissant alors un nouveau jeu de données, réduit mais complet, sur lequel n'importe quelle technique de cartographie des préférences peut être appliquée. Une simulation de type Monte-Carlo a été entreprise dans le but d'étudier la performance de cette méthode en fonction de divers paramètres. Des données complètes, comprenant plusieurs classes de consommateurs, ont été simulées, puis rendues incomplètes, en enlevant un nombre donné de produits à chaque consommateur sans perturber l'équilibre du plan d'expériences. La classification obtenue à partir de ces données incomplètes est comparée à celle issue des données complètes.

Mots clés : cartographie des préférences, plan incomplet équilibré, carrés latins mutuellement orthogonaux, simulation de type Monte-Carlo.

1 - INTRODUCTION

L'analyse sensorielle, d'emploi fréquent dans les départements R&D de l'industrie agro-alimentaire, permet de déclarer s'il existe ou non des différences perceptibles entre des produits (tests de différences) et, si oui, de décrire et de quantifier ces différences (profils sensoriels). Les profils sensoriels des produits sont établis par un petit groupe de sujets entraînés à la dégustation, qui notent les produits selon des descripteurs sensoriels précis. Le réel souci de l'industrie agro-alimentaire est de pouvoir fabriquer des produits qui répondent le mieux aux exigences des consommateurs. Toutefois, si le consommateur peut sans trop de difficultés dire s'il aime ou non un produit et lui donner une note de préférence, il lui est beaucoup plus difficile d'exprimer les motivations de son choix. Autrement dit, il est peu vraisemblable que le consommateur moyen soit capable de décrire sans confusion les propriétés sensorielles des produits qu'il déguste et encore moins concevable qu'il puisse exhiber celles qui ont pu influencer positivement ou négativement son jugement. La cartographie des préférences (*Preference Mapping* en anglais) se propose justement d'expliquer les préférences des consommateurs à l'aide de descripteurs sensoriels établis par des sujets entraînés. Elle permet de comparer différents produits d'une même gamme et surtout d'orienter la formulation des produits en localisant sur une carte sensorielle des « zones idéales » à atteindre. Selon la position des produits sur la carte sensorielle, il peut être envisagé, soit de modifier un « produit maison » pour qu'il se rapproche de la cible, soit de développer un nouveau produit au coeur de la cible.

Les techniques de cartographie des préférences furent mises en oeuvre pour la première fois en psychométrie aux Etats-Unis (CARROL et CHANG, 1970 ; SCHIFFMAN *et al.*, 1981) et quelques années plus tard en Grande Bretagne dans le secteur agroalimentaire (NUTE *et al.*, 1989 ; JONES et KENWARD, 1989). Elles firent leur apparition en France dans le même secteur en 1992 (SCHLICH et MCEWAN, 1992).

Aujourd'hui, l'intérêt suscité par la cartographie des préférences ne cesse d'augmenter et semble motiver à la fois industriels et chercheurs. A titre d'exemple, lors des 4^e journées agro-industrie et méthodes statistiques (Dijon, 1995), environ le quart des communications ont abordé de près ou de loin ce sujet.

D'un point de vue méthodologique, la cartographie des préférences exige que des produits du même type (exemple : des yaourts brassés nature, des bières brunes...) soient dégustés à l'aveugle par un échantillon de consommateurs représentatif de la population à étudier (en général plusieurs centaines d'individus interpellés dans la rue) et que, dans le même temps, le profil sensoriel des mêmes produits soit établi par un petit groupe de sujets entraînés. Les notes de préférence de chaque consommateur (ou d'un groupe de consommateurs aux préférences homogènes) sont alors expliquées à l'aide d'un modèle de régression (choisi parmi quatre possibles) par rapport aux deux premières dimensions X_1 et X_2 de l'espace sensoriel des produits (SCHLICH, 1995 ; CALLIER et SCHLICH, 1995). Le nombre de produits exigé par les modèles de régression pour l'estimation des paramètres et l'évaluation de la qualité d'ajustement doit être au minimum égal à sept et de préférence le plus grand possible. De plus, le nombre de produits à comparer dans une étude (produits maison + produits de la concurrence) est bien souvent supérieur à huit. Le consommateur interrogé, contraint d'évaluer consécutivement un minimum de sept ou huit produits, est susceptible d'être déstabilisé et risque de l'être davantage avec 12, 13 voire 16 produits à déguster au cours de la même séance (surtout s'il s'agit de bières ou de moutardes). Une telle situation le place dans un contexte trop éloigné de la réalité et peut aussi occasionner des problèmes de saturation de certains organes sensoriels trop sollicités ou plus simplement des problèmes de temps d'expérimentation. Ces contraintes, tant sensorielles que pratiques, ont conduit à envisager la mise en place d'études en plans incomplets, où le consommateur n'évaluerait qu'une partie des produits de l'étude (soit k parmi p).

Lorsque des individus dégustent plusieurs produits consécutivement, leur jugement peut être influencé par plusieurs facteurs directement liés au plan de présentation de ces produits. Les effets déjà mis en évidence dans de nombreuses études (par exemple : PERYAM et PILGRIM, 1957 ; SCHLICH, 1993) sont :

- l'effet d'ordre : effet provenant de la position même du produit dégusté ;
- l'effet de report (du 1^{er} ordre) : effet provenant du produit précédant le produit dégusté.

La planification d'une étude en plans incomplets doit donc veiller à équilibrer au mieux tous ces effets d'ordre et de report.

De plus, dans le cadre d'une telle étude, les techniques classiques de cartographie des préférences ne sont plus applicables, puisque les modèles de régression mis en œuvre nécessitent que chaque consommateur ait évalué tous les produits de l'étude. Pour se ramener à un vecteur de notes complet comportant p notes (une note de préférence par produit), il est donc nécessaire d'établir des groupes de consommateurs aux préférences homogènes, à partir desquels la moyenne de chaque produit peut être calculée par groupe (chacun des p produits de l'étude devant être évalué par au moins un consommateur dans chaque groupe). La constitution de groupes de préférences homogènes dans un tel contexte revient à classer les individus d'un tableau (consommateurs en lignes et produits en colonnes) dont les observations ne comportent que k données renseignées sur p . Une stratégie permettant d'établir des groupes à partir d'une matrice structurellement creuse a été proposée puis validée par une simulation de type Monte-Carlo. Le principe de la validation était de comparer les groupes établis à partir de données complètes simulées (comprenant différentes classes homogènes de préférences), à ceux établis à partir des mêmes données rendues progressivement incomplètes. Les produits sont retirés un par un à chaque consom-

mateur par simple suppression de colonnes du plan initial, sans pour autant perturber l'équilibre des effets d'ordre et de report de ce dernier.

Le présent article aborde les trois points suivants :

- planification d'une expérimentation en plans incomplets équilibrant au mieux les effets d'ordre et de report ;
- élaboration d'une stratégie pour obtenir des groupes de consommateurs homogènes selon leurs préférences à partir d'une matrice structurellement creuse ;
- validation de la stratégie mise en place.

2 - PLANIFICATION

Lorsque des études sont réalisées en plans complets, les carrés latins de WILLIAMS (1949) utilisés jusqu'alors permettent d'équilibrer les effets d'ordre et de report d'ordre un. Toutefois, les effets de report d'ordre supérieur à un (effet pouvant provenir des produits placés plusieurs positions avant le produit dégusté) ne sont pas du tout pris en compte par ce genre de plans.

Des plans d'expérience, construits à partir de systèmes complets de carrés latins mutuellement orthogonaux (définis ci-après), sont désormais vivement recommandés pour la planification d'expérimentations en plans complets ou incomplets. Ils garantissent un équilibre maximal des effets d'ordre et de report.

Définitions - Propriétés

Deux carrés latins de taille p sont dits orthogonaux si, en les superposant, on obtient toutes les paires ordonnées possibles (de codes produits), soit p^2 paires ordonnées différentes (fig. 1).

	Ordre 1	Ordre 2	Ordre 3	Ordre 4		Ordre 1	Ordre 2	Ordre 3	Ordre 4	
Sujet 1	A	B	C	D		A	C	D	B	Sujet 1
Sujet 2	B	A	D	C		B	D	C	A	Sujet 2
Sujet 3	C	D	A	B		C	A	B	D	Sujet 3
Sujet 4	D	C	B	A		D	B	A	C	Sujet 4

⇓ Superposition ⇓

AA	BC	CD	DB
BB	AD	DC	CA
CC	DA	AB	BD
DD	CB	BA	AC

Figure 1

Exemple de 2 carrés latins orthogonaux de taille 4
A, B, C et D désignent les codes des produits

*Example of 2 orthogonal latin squares of order 4
A, B, C and D are the sample codes*

Des carrés latins sont dits mutuellement orthogonaux (ou MOLS pour Mutually Orthogonal Latin Squares) s'ils sont orthogonaux deux à deux. Un ensemble de $(p-1)$ MOLS de taille p est appelé un système complet de MOLS. Un tel ensemble de MOLS existe si p est un nombre premier ou une puissance d'un nombre premier ($p = 2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17$, etc.) et possède la propriété remarquable d'équilibrer les effets d'ordre et les effets de report de tous ordres (WILLIAMS, 1949 ; WAKELING et MACFIE, 1995). La figure 2 montre un système complet de MOLS de taille 4, soit trois carrés latins mutuellement orthogonaux juxtaposés verticalement. Si on regarde par exemple les colonnes O_1 et O_3 , on constate que les $4(4-1)$ couples ordonnés possibles de produits apparaissent une et une seule fois sur l'ensemble des 12 lignes. Chaque produit dégusté en 3^e position (colonne O_3) est donc précédé par chacun des autres une et une seule fois à l'ordre 2 (colonne O_1), il en serait de même en observant les colonnes O_2 et O_4 .

Cette propriété se généralise à un système complet de MOLS de taille p où chaque produit dégusté en position k (colonne O_k) est précédé par chacun des autres une et une seule fois à l'ordre 1 (colonne O_{k-1}), à l'ordre 2 (colonne O_{k-2}),... , jusqu'à l'ordre $(k-1)$ inclus (colonne O_1). En particulier, pour chaque produit dégusté en dernière position ($k = p$), l'équilibre des effets de report est réalisé jusqu'à l'ordre $(p-1)$ inclus, c'est-à-dire jusqu'au produit évalué en première position. Tout système complet de MOLS de taille p constitue un plan équilibré pour les effets d'ordre (garanti par n'importe quel carré latin) et les effets de report jusqu'à l'ordre $(p-1)$.

Une conséquence capitale de cette propriété est que l'on peut obtenir un plan incomplet équilibré à partir de tout système complet de MOLS en éliminant n'importe quel nombre de colonnes. Le plan obtenu en blocs de taille k reste alors équilibré concernant les effets de report jusqu'à l'ordre $(k-1)$ inclus.

Sujets	O_1	O_2	O_3	O_4
S_1	A	B	C	D
S_2	B	A	D	C
S_3	C	D	A	B
S_4	D	C	B	A
S_5	A	C	D	B
S_6	B	D	C	A
S_7	C	A	B	D
S_8	D	B	A	C
S_9	A	D	B	C
S_{10}	B	C	A	D
S_{11}	C	B	D	A
S_{12}	D	A	C	B

Figure 2

Système complet de MOLS de taille 4
A, B, C et D désignent les codes des produits

*Complete set of MOLS of order 4
A, B, C and D are the sample codes*

Lors d'une étude comportant au total p produits, planifiée à partir de MOLS en plans complets ou incomplets, l'équilibre des effets de report de tous ordres est optimal si le nombre n de consommateurs interrogés est un multiple de $p(p-1)$. Dans la pratique, si le nombre n de consommateurs prévu pour l'étude n'est pas exactement égal à un multiple de $p(p-1)$, l'expérimentateur construira son plan d'expériences en conservant les n premières lignes du plus petit plan en systèmes complets de MOLS (d'ordre p) dont le nombre de lignes est supérieur à n . Parmi tous les plans possibles, le plan ainsi obtenu garantit un équilibre maximum des effets de report bien que non optimal. L'équilibre des effets d'ordre, quant à lui, est optimal dès que n est un multiple de p .

Pour des compléments à propos des MOLS, le lecteur peut se référer à JOHN (1971), STREET et STREET (1987), JONES et KENWARD (1989), ou WAKELING et MACFIE (1995). Ces derniers proposent notamment des alternatives de plans lorsque p n'est ni premier ni une puissance entière d'un nombre premier.

3 - STRATÉGIE CLIP DE CLASSIFICATION SUR DONNÉES INCOMPLÈTES

3.1 Présentation de la méthode CLIP

Dans le cadre d'une expérimentation en plans incomplets où chaque consommateur déguste k produits seulement parmi p , la matrice (n,p) consommateurs $\times$ produits comporte des données manquantes, ce qui interdit toute technique de classification classique. Le pourcentage de données manquantes augmente considérablement dès que k diminue par rapport à p . Notre souhait étant de planifier des études où k est le plus petit possible, il n'a pas été envisagé d'utiliser des techniques d'estimation de données manquantes. Ces dernières, pourtant nombreuses dans la littérature, se comportent généralement très mal dès que le pourcentage de données manquantes est trop élevé. HEDDERLEY et WAKELING (1995) se sont employés à comparer certaines de ces méthodes dans le cadre d'études en cartographie interne des préférences réalisées sur des données simulées où trois différents pourcentages d'incomplétude étaient recréés : 5 %, 35 % et 65 %. Ce travail, ne s'intéressant pas directement à la stabilité d'éventuels groupes de consommateurs, mais plutôt à celle de la configuration générale des produits et des consommateurs, ne répond pas à nos préoccupations de classification sur données incomplètes. Il montre néanmoins que les méthodes comparées se comportent très bien, lorsque 5 % seulement de données manquantes existent, et que leur performance décroît sensiblement avec 35 % de données manquantes pour devenir catastrophique lorsque ce pourcentage atteint 65 %.

La technique développée ici, pour établir des groupes de consommateurs aux préférences homogènes à partir de données incomplètes, est sans rapport avec tout type de techniques d'estimation de données manquantes. Cette méthode, baptisée CLIP pour CLustering of Incomplete Preferences, doit son originalité au fait qu'elle attribue un poids différent à chacune des $n(n-1)/2$ paires de consommateurs. Le poids affecté à chaque paire d'individus est proportionnel au nombre

de produits dégustés en commun par les individus en question. Aucune technique de classification de ce genre n'a en tout cas été recensée dans la littérature. La stratégie CLIP, décrite ci-après, est très facilement accessible pour tout utilisateur du logiciel SAS® (SAS INSTITUTE, 1989). Elle se résume en trois grandes étapes :

- calcul d'une matrice S de similarités entre consommateurs ;
- analyse de S par *MultiDimensional Scaling* pondéré (Obtention de la configuration C) ;
- classification classique à partir de la configuration classique C.

Remarque

Une autre alternative à CLIP, appelée CLIP2 (CALLIER, 1996), peut être mise en œuvre par les non-utilisateurs de SAS®. En effet, à partir de la matrice de similarités S de départ, une approche inspirée d'une classification de type hiérarchique peut être entrevue. Dans ce cas, l'ultramétrie utilisée, proche du lien moyen classique (moyenne de distances), devient un « lien moyen pondéré » : moyenne pondérée de distances où le système de poids considéré est le même que celui intervenant dans la méthode CLIP. Ce système de pondération règle ainsi les problèmes de données manquantes susceptibles d'être rencontrés dans la matrice S. Cette approche, entrevue ultérieurement, n'a pu être validée par simulation aux côtés de la méthode CLIP.

3.2 Description des étapes de la méthode CLIP

3.2.1 Étape n° 1 : calcul d'une matrice S de similarités entre consommateurs

Un très grand nombre d'indices de similarités ou de dissimilarités sont proposés dans la littérature. La méthode CLIP fonctionnerait avec n'importe lequel d'entre eux. Cependant, parmi tous les choix possibles, un indice proposé par GOWER (1971) a été retenu parce qu'il semblait particulièrement adapté aux données incomplètes. Cet indice de similarité $S_{ii'}$ est calculé entre chaque paire de consommateurs (i, i') ayant dégusté $k_{ii'}$ produits en commun. Par conséquent, il n'est pas défini si les consommateurs en question n'ont évalué aucun produit en commun (cas où $k_{ii'} = 0$). Cet indice, qui a été établi sur les données centrées par consommateur, est défini par :

$$S_{ii'} = \frac{\sum_{j \in P_{ii'}} 1 - \frac{|x_{ij} - x_{i'j}|}{R(j)}}{k_{ii'}} \quad (\text{si } k_{ii'} \neq 0)$$

avec x_{ij} = note donnée par le consommateur i au produit j,

$P_{ii'}$ = ensemble des produits vus en commun par les consommateurs i et i',

$R(j)$ = (Max-Min) du produit j (calculé sur l'ensemble des consommateurs).

L'indice en question prend la valeur maximale de 1, si les consommateurs comparés ont donné les mêmes notes de préférence (après centrage ici) aux produits qu'ils ont dégustés en commun. Il vaut au contraire 0 si, pour chaque produit évalué en commun, les notes comparées correspondent respectivement aux notes minimale et maximale obtenues par le produit en question sur l'ensemble du panel interrogé.

Cette étape n° 1 conduit à l'obtention d'une matrice de similarités S ayant des valeurs manquantes dès que $k < (p + 1)/2$.

3.2.2 Étape n° 2 : analyse de S par *Multi Dimensional Scaling* pondéré

La matrice carrée S de similarités est analysée par *Multi-Dimensional Scaling* (SCHIFFMAN *et al.*, 1981 ; MEAD, 1992) avec pondération de chaque cellule S_{ij} par un poids w_{ij} proportionnel au nombre de produits k_{ij} dégustés en commun par les consommateurs i et j : $w_{ij} = k_{ij}/k$. Ce poids est nul lorsque les consommateurs n'ont dégusté aucun produit en commun. Cette étape conduit à l'obtention d'axes factoriels donnant une représentation spatiale des consommateurs.

Dans le cadre de la simulation, la procédure MDS, assortie de l'ordre WEIGHT, du logiciel SAS® a été appliquée (SAS INSTITUTE, 1989). Elle correspond à une approche non métrique itérative dont l'algorithme est très voisin de la procédure ALSCAL (TAKANE *et al.*, 1977). Trois axes pour la dimension de la configuration spatiale des consommateurs ont été retenus et le nombre maximal d'itérations a été limité à 5 pour des raisons d'économie de temps. Des tests préalables au lancement des simulations ont révélé que ces choix avaient une très faible incidence sur les résultats obtenus et permettaient un gain de temps considérable.

3.2.3 Étape n° 3 : classification classique

Toute technique de classification peut être appliquée sur la matrice complète obtenue (matrice dont les lignes sont les consommateurs et les colonnes leurs coordonnées sur les axes factoriels issus de l'étape n° 2).

Dans le cadre de la simulation, une classification de type nuées dynamiques a été établie sur les lignes brutes de la configuration tridimensionnelle des consommateurs issue de l'étape 2, à l'aide de la procédure FASTCLUS du logiciel SAS® (HARTIGAN, 1975 ; SAS INSTITUTE, 1989). Ce genre d'approche consiste à affecter chaque observation à la classe dont le noyau (centre de gravité) est le plus proche pour former des classes provisoires. La distance retenue est la distance euclidienne classique. Les noyaux sont alors remplacés par les moyennes de ces groupes temporaires (centres d'inertie mobiles) et le processus itératif d'affectation des observations à la classe la plus proche est répété jusqu'à ce que la stabilité des groupes soit atteinte. La procédure FASTCLUS se distingue des techniques traditionnelles par son algorithme de sélection des noyaux initiaux. Elle est particulièrement recommandée pour de grands jeux de données et, comme son nom le suggère, pour sa rapidité d'exécution, argument essentiel dans un contexte de simulation.

4 - VALIDATION PAR SIMULATION

4.1 Simulation de données consommateurs complètes

Une expérimentation contenant n consommateurs évaluant p produits selon un plan complet élaboré à partir de systèmes complets de MOLS est simulée, en recréant g groupes de préférences homogènes au sein de la population virtuellement sondée, selon le modèle suivant.

Tout d'abord, pour chacun des g groupes, p moyennes produits sont générées selon une loi uniforme dans un intervalle compris entre 1 et 10, dont les bornes sont des variables aléatoires dépendantes de g . Cette génération de moyennes a été établie après avoir examiné de nombreux jeux de données réelles partitionnés en plusieurs classes homogènes de mêmes préférences. La démarche choisie (non révélée ici pour des raisons de confidentialité), est censée recréer la plupart des situations susceptibles d'être rencontrées en réalité. Ensuite, les p notes de chaque consommateur sont générées aléatoirement à partir des p moyennes du groupe auquel il appartient, transformées de la façon suivante :

- dilatées ou contractées, pour restituer les différentes amplitudes d'échelle utilisées, (coefficient de dilatation/contraction généré selon une loi triangulaire) ;
- translatées, pour restituer l'utilisation d'une partie droite ou gauche de l'échelle, (coefficient de translation généré selon une loi uniforme) ;
- perturbées par un bruit gaussien, pour perturber les classements des produits, (bruit généré par une loi normale centrée d'écart-type σ) ;
- arrondies à la plus proche valeur entière, pour être en accord avec les échelles utilisées.

Après toutes ces étapes aléatoires, chaque consommateur a donné une note entière entre 1 et 10 à chacun des p produits.

Remarque

En ce qui concerne les effectifs des groupes de consommateurs, ceux des $(g-1)$ premiers groupes sont tirés au sort selon une loi binomiale $B(n, 1/g)$, alors que l'effectif du dernier est calculé de telle sorte que la somme totale des effectifs soit égale à n .

4.2 Classification sur données complètes

Bien qu'une classification traditionnelle puisse être réalisée, la technique CLIP a été appliquée en exigeant g groupes au niveau de l'étape n° 3, dans le but de servir de base de comparaison pour les classifications obtenues sur données incomplètes.

4.3 Classification sur les données rendues incomplètes

Des données incomplètes sont générées par simple suppression de colonnes du plan d'expérience. Ainsi, le jeu de données recréé correspond à une étude où les consommateurs auraient dégusté k produits seulement parmi p . Ces colonnes sont retirées une à une et, à chaque pas, le numéro de la colonne supplémentaire supprimée est tiré au sort. Une classification CLIP en g groupes est alors effectuée.

Remarque

Le problème du choix du nombre de groupes, largement discuté en statistique, demeure un sujet à polémique qui dépasse le cadre de cet article. La validation a essayé de limiter toute incertitude liée au choix du nombre de groupes, en créant artificiellement g classes sur les données complètes et en exigeant de retrouver ces g groupes lors de l'étape n° 3. Cette démarche a permis d'étudier la robustesse de la méthode CLIP indépendamment de paramètres non maîtrisés par l'utilisateur.

4.5 Comparaison COMPLET/INCOMPLET

4.5.1 Stabilité au niveau de l'ensemble de la population

La stabilité des moyennes générales des produits obtenues sur l'ensemble de la population des consommateurs a été étudiée. Ce genre de résultats est bien souvent celui qui préoccupe le plus les services marketing lors de toute étude de marché. Il était donc important de savoir si le fait de réaliser une étude en plans incomplets perturbe ou non les moyennes générales des produits. Pour s'en rendre compte, le coefficient de corrélation de Pearson entre le vecteur des moyennes générales des produits établi dans le cas complet et son homologue calculé dans le cas incomplet, a été choisi comme critère (appelé Gcorr).

4.5.2 Stabilité au niveau des groupes

Les g groupes obtenus sur données incomplètes (pour $k = p-1, p-2, \dots, 4, 3$) sont comparés à ceux obtenus sur données complètes (correspond à $k = p$). La comparaison a été établie selon deux critères tentant de répondre aux questions suivantes :

a) Les groupes comparés sont-ils les mêmes en termes d'individus qui les composent ?

b) Les groupes comparés sont-ils les mêmes en termes de préférences indépendamment des individus qui les composent ?

a) *Stabilité de la composition des groupes* (critère retenu : indice ajusté de Rand « R_a »)

Tableau 1

Table de contingence établie à partir des $n(n-1)/2$ paires de consommateurs

Table 1

Contingency table computed over the $n(n-1)/2$ consumer pairs

		Classification testée (données incomplètes)	
		Paires dont les consommateurs sont dans la même classe	Paires dont les consommateurs sont dans 2 classes différentes
Classification de référence (données complètes)	Paires dont les consommateurs sont dans la même classe	a (accord)	b (désaccord)
	Paires dont les consommateurs sont dans 2 classes différentes	c (désaccord)	d (accord)

L'indice ajusté de Rand proposé par HUBERT et ARABIE (1985) et recommandé par MILLIGAN et COOPER (1986, 1988) et JOBSON (1991), est défini à partir du tableau 1 par :

$$R_a = \frac{[a + d] - A_c}{[n(n-1)/2] - A_c} \leq 1$$

(où A_c = terme correctif pour les accords trouvés par hasard)

Plus l'indice R_a est proche de 1, plus les classifications comparées se ressemblent en termes de composition des groupes. A l'opposé, une valeur de R_a

proche de 0 correspond à la valeur que l'on obtiendrait en comparant deux classifications qui auraient été établies au hasard.

b) *Stabilité des moyennes produits des groupes* (critère retenu : moyenne des corrélations entre groupes associés « MCGA »)

Les vecteurs contenant les moyennes des produits de chaque groupe dans le cas complet et le cas incomplet ont été comparés. La comparaison de deux tableaux dont les p lignes sont les produits et les g colonnes les groupes constitués est une question délicate puisqu'il n'y a aucune association logique entre les groupes de chaque tableau. Par exemple, le groupe G_1 établi dans le cas complet n'a aucune raison *a priori* d'être associé et comparé au groupe G'_1 établi dans le cas incomplet. Le coefficient RV (ESCOUFIER, 1973) permet une comparaison globale de deux tableaux en termes de distances entre les lignes (ici les produits). Toutefois, l'orientation des vecteurs (ici le classement des produits), qui est très importante pour étudier les différentes tendances de préférence dans la population des consommateurs, n'est pas prise en compte par le coefficient RV.

Un autre critère devait donc être proposé. Parmi toutes les associations possibles entre groupes (soit $g!$), celle qui maximisait la moyenne des g corrélations (de Pearson) calculées entre vecteurs associés a été retenue. La valeur maximale ainsi obtenue correspond au critère retenu MCGA pour juger de la stabilité des moyennes des groupes dans le cas incomplet par rapport au cas complet :

$$MCGA = \max_{\sigma \in P(g)} \left\{ \frac{1}{g} \sum_{i=1}^g r[G_i, G_{\sigma(i)}] \right\} \in [-1, 1]$$

(où $P(g)$ représente l'ensemble des $g!$ permutations de $\{1, 2, \dots, g\}$)

Il est bon de noter qu'avec un tel critère, d'éventuelles corrélations négatives entre vecteurs associés jouent un rôle négatif, ce qui était un des buts recherchés. Là aussi, plus l'indice MCGA est proche de 1, plus les classifications comparées se ressemblent en termes de moyennes produits des groupes constitués.

4.6 Domaine de variation des paramètres de la simulation (tabl. 2)

En raison de leur vraisemblance pratique, les chiffres de 8, 11, 13 et 16 ont été choisis pour le nombre total p de produits et le multiple de $p(p-1)$ le plus proche de 250 a été retenu pour le nombre n de consommateurs. Le nombre g de groupes homogènes simulés correspond au nombre de tendances le plus souvent observées en réalité à savoir : 2, 3, 4 ou 5. Trois différents degrés de perturbation des données ($\sigma \in \{0,5 ; 1 ; 2\}$) correspondant à différents degrés d'homogénéité des groupes simulés, ont été testés. Toutes les combinaisons de ces différents paramètres ont été répétées 10 fois.

Au total, $4 \times 4 \times 3 \times 10 = 480$ études complètes ont été simulées donnant lieu à 4 800 classifications établies selon la stratégie proposée. Le temps moyen d'une telle classification et des calculs annexes étant de 30 min environ (sur station de travail SUN 4/65, terminal d'un serveur SUN 4/330), le coût total de ces simulations en temps de calcul s'élève environ à 2 400 h soit 100 jours. Ce temps assez conséquent ne doit pas pour autant décourager un utilisateur potentiel de la méthode. En effet, la mise en œuvre de la classification CLIP dénudée de toute étape de simulation annexe, s'effectue en une vingtaine de minutes maximum.

Tableau 2
Modalités des paramètres de la simulation

Table 2
Simulation parameter modalities

p	n	g	s	k	rep	Nombre de classifications CLIP
8	224	2, 3, 4, 5	0,5, 1, 2	8, 7, 6, 5, 4, 3	10	720
11	220	2, 3, 4, 5	0,5, 1, 2	11, 10, 9, 8, 7, 6, 5, 4, 3	10	720
13	312	2, 3, 4, 5	0,5, 1, 2	13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3	10	720
16	240	2, 3, 4, 5	0,5, 1, 2	16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3	10	720

5 - RÉSULTATS

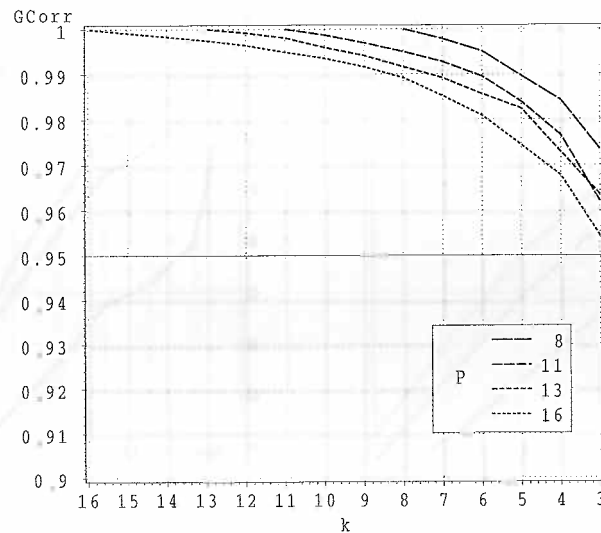
5.1 Stabilité au niveau de l'ensemble de la population

Les résultats (*fig. 3*) n'ont été donnés que dans le cas où le niveau de bruit est maximal ($\text{sigbruit} = 2$), sachant qu'ils sont encore meilleurs dans les autres cas. Pour chaque nombre p de produits, la moyenne des G_{corr} , calculée sur le nombre g de groupes et sur le nombre rep d'études répétées ($4 \times 10 = 40$ études complètes simulées), est donnée en ordonnée en fonction du nombre k de produits dégustés par consommateur.

Les résultats sont très encourageants. En effet, la perte d'information est très faible (faible pente des courbes) et la moyenne des corrélations reste toujours au dessus de 0,95 (même pour trois produits dégustés par consommateur). Les moyennes générales des produits semblent donc assez peu perturbées par une diminution du nombre de produits dégustés par consommateur, même lorsque le niveau de bruit simulé est grand.

5.2 Stabilité au niveau des groupes

Les résultats obtenus sont donnés sous forme graphique (*fig. 4* pour le critère R_a et *fig. 5* pour le critère MCGA). Pour chaque nombre p de produits, la moyenne de R_a (ou de MCGA) calculée sur le nombre de répétitions et sur le nombre de groupes ($4 \times 10 = 40$ études complètes simulées) est donnée en ordonnée en fonction du nombre k de produits dégustés par consommateur, donné en abscisse. La décroissance des courbes indique la perte progressive d'information, due au fait que les classifications sont établies à partir d'une matrice de plus en plus incomplète. Le bruit perturbant l'homogénéité des groupes simulés accentue sensiblement la vitesse de décroissance des courbes ($\text{courbe}[\text{sigbruit} = 2] < \text{courbe}[\text{sigbruit} = 1] < \text{courbe}[\text{sigbruit} = 0,5]$).

**Figure 3**

Étude de la stabilité au niveau de l'ensemble de la population

Stability study for the whole population

A titre indicatif, si on se fixe un seuil pour chaque critère (par exemple 0,5 pour Ra et 0,9 pour MCGA), on peut indiquer (tabl. 3) le nombre minimal de produits que doit déguster chaque consommateur pour que :

- ① ($Ra > 0,5$ (stabilité « jugée suffisante » de la composition des groupes) ;
- ② ($MCGA > 0,9$ (stabilité « jugée suffisante » des moyennes produits des groupes).

6 - DISCUSSION

Le choix des valeurs seuil pour les critères Ra et MCGA n'a pas été établi en fonction de tests statistiques. Les seuls tests disponibles, recensés dans la littérature, permettent de rejeter une hypothèse d'indépendance maximale et de quantifier en termes statistiques le degré de non ressemblance, alors que l'intérêt ici se porte plutôt sur le degré de ressemblance entre groupes. Des tests de permutation pourraient être envisagés et sont en cours d'étude.

L'indice ajusté de Rand est souvent recommandé pour comparer des classifications. Dans le contexte étudié, il est extrêmement exigeant de vouloir imposer qu'aucun consommateur ou presque ne change de groupe. En effet si, parmi les

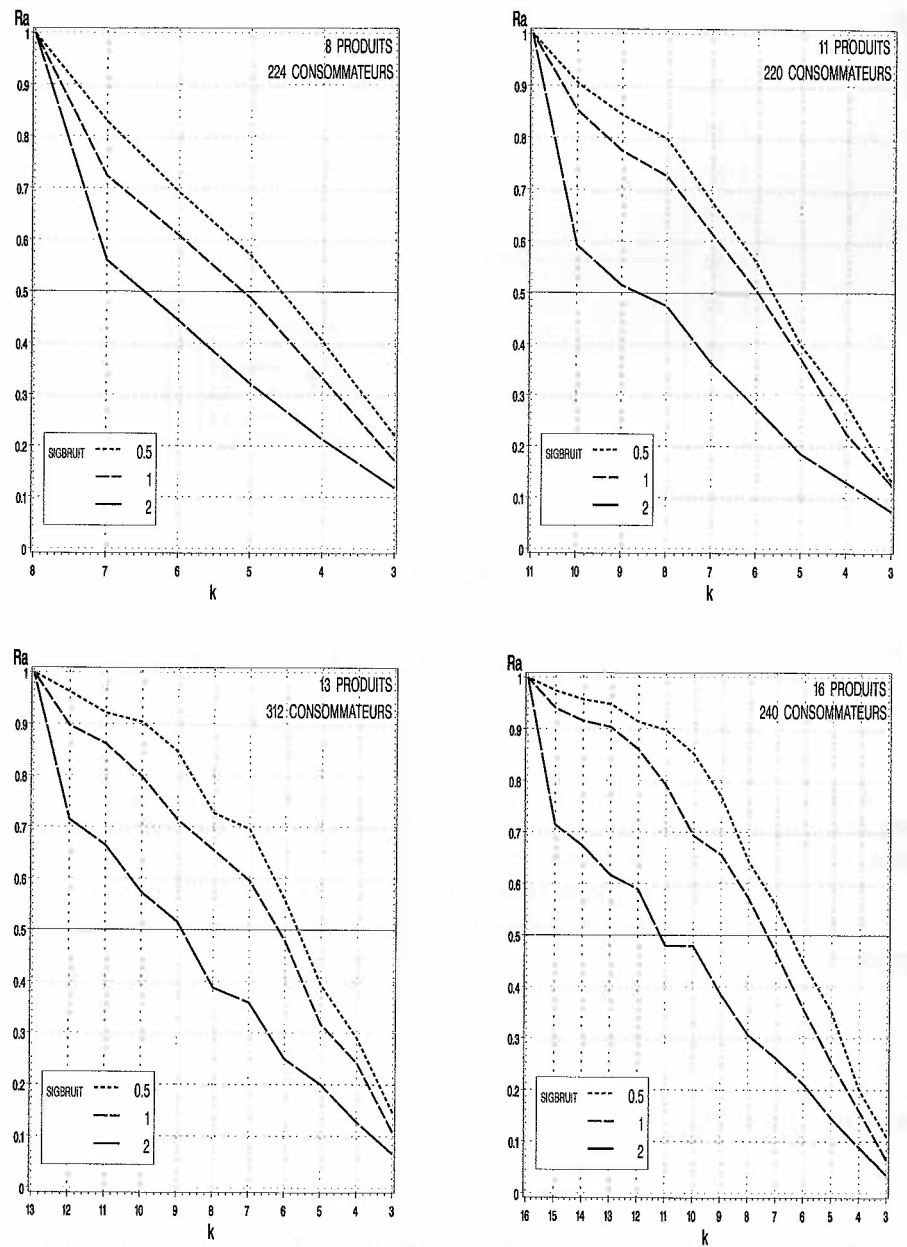
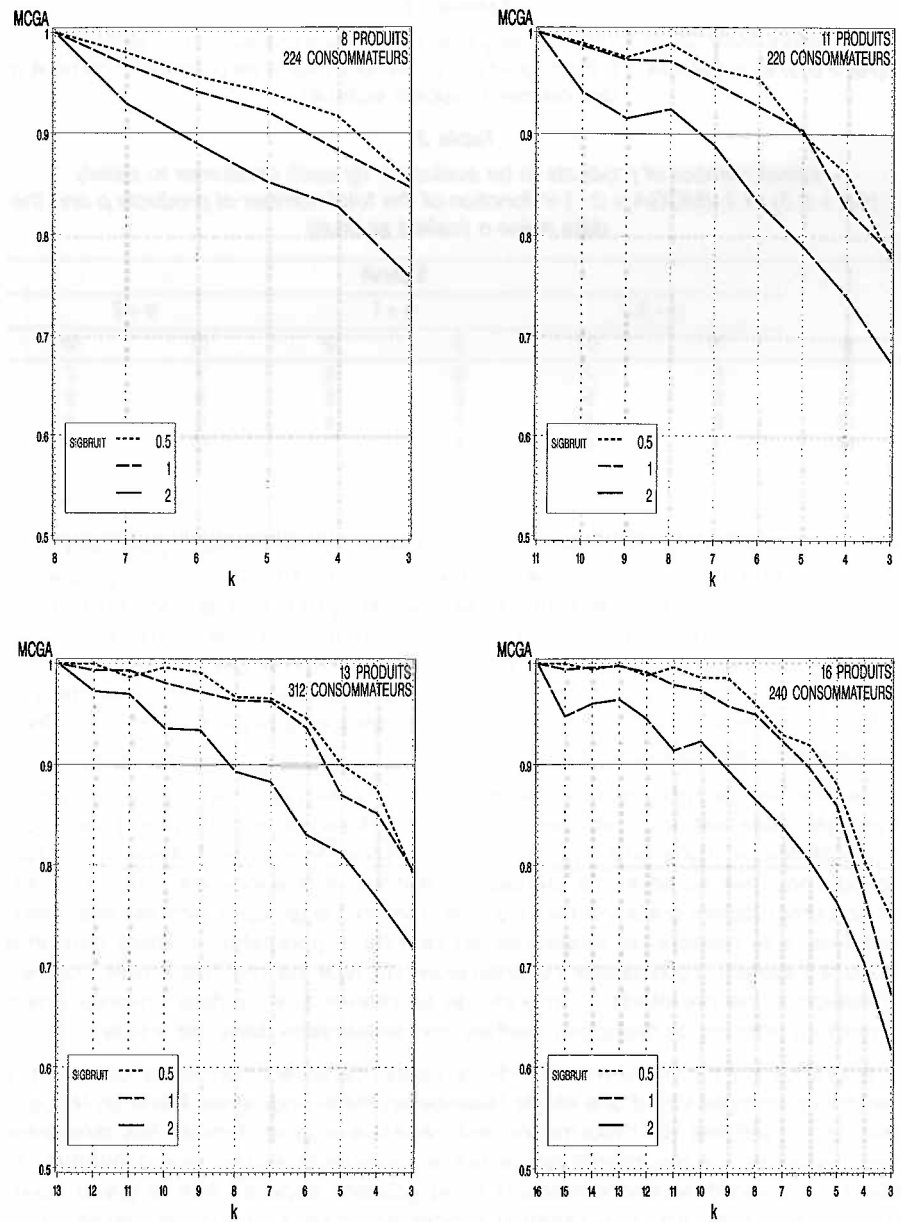


Figure 4

Étude de la stabilité de la composition des groupes

Stability study of the group composition

**Figure 5**

Étude de la stabilité des moyennes produits des groupes

Stability study of the group product means

Tableau 3

Nombre minimal de produits à déguster par consommateur pour satisfaire
① ($Ra > 0,5$) et ② ($MCGA > 0,9$) en fonction du nombre total p de produits et du bruit σ
des données (appelé sigbruit)

Table 3

Minimal number of products to be evaluated by each consumer to satisfy
① ($Ra > 0.5$) et ② ($MCGA > 0.9$) in fonction of the total number of products p and the
data noise σ (called sigbruit)

p	Sigbruit					
	$\sigma = 0,5$		$\sigma = 1$		$\sigma = 2$	
	①	②	①	②	①	②
8	5	4	6	5	7	7
11	6	5	6	5	9	8
13	6	5	7	6	9	9
16	7	6	8	7	12	10

produits dégustés par un consommateur donné, ne se trouvent pas ceux qu'il préfère ou déteste le plus, il est peu probable qu'il se retrouve dans le groupe qui aurait été le sien s'il avait dégusté l'ensemble des produits. Dans le cadre d'une étude réelle, la stabilité du profil moyen des groupes au vu des variables socio-démographiques des individus qui les composent, importe plus d'être conservée que celle de la composition même des groupes en termes d'individus. Ici, dans le cadre d'une simulation, aucune variable socio-démographique n'étant disponible, le critère Ra a été retenu.

Le nombre de tendances homogènes observées parmi les consommateurs interrogés varie selon les études. Il est donc difficile de le connaître à l'avance. Pour s'affranchir d'une telle incertitude, la moyenne de tous les critères a été établie pour tous les nombres de groupes confondus. A titre indicatif, tous ces critères se comportaient d'autant mieux que le nombre de groupes simulés était petit. Toutefois, si le nombre de tendances au sein de la population étudiée peut être connu à l'avance, (connaissance antérieure du type de marché étudié ou bien réalisation d'une pré-étude), il importe de se référer aux courbes correspondant au nombre attendu de groupes (courbes non présentées dans cet article).

A propos du choix du nombre de groupes, certaines considérations sont à prendre en compte lors d'une étude réalisée en milieu industriel. Dans un tel contexte, une synthèse de l'information est nécessaire pour donner des directives opérationnelles, ce qui interdit de ce fait la considération d'un grand nombre de groupes au sein des consommateurs (cinq groupes apparaît être le grand maximum). Le fait de répartir la population sondée en un petit nombre de classes augmente leurs effectifs potentiels. Par conséquent, le risque de constituer un groupe au sein duquel un produit pourrait ne pas avoir été évalué du tout (c'est-à-dire dégusté par aucun des individus du groupe en question) est extrêmement réduit. Si ce cas très peu probable venait toutefois à se produire, le vecteur de $(p-1)$ notes obtenu, certes non complet, permettrait tout de même l'application des méthodes de cartographie externe des préférences en ne perdant qu'un seul degré de liberté avec les modèles de régression utilisés. La moyenne générale de ce produit (non calculable directement) pourrait alors être estimée en considérant la surface de réponse des autres moyennes du groupe.

Au niveau des différents degrés de perturbation simulés, il semble là aussi bien difficile de prévoir lequel correspondrait le mieux à une situation réelle. Pour tenter de répondre à cette question cruciale, l'apport de données réelles recueillies dans le même contexte semble nécessaire. Cet aspect a été traité très récemment. Les résultats des travaux engagés seront communiqués ultérieurement (CALLIER, 1996).

REMERCIEMENTS

Cet article fait suite à une conférence présentée à Dijon le 8 décembre 1995 lors du congrès « Quatrièmes journées agro-industrie et méthodes statistiques ». Les travaux décrits ont été réalisés dans le cadre d'une thèse CIFRE effectuée pour le groupe DANONE en collaboration avec l'Inra de Dijon (Laboratoire de recherches sur les arômes).

Reçu le 29 mars 1996, accepté le 14 novembre 1996.

RÉFÉRENCES BIBLIOGRAPHIQUES

- CALLIER P., SCHLICH P., 1995. Preference Mapping in Progress. In : ASU (ed.), 4^e Journées agro-industrie et méthodes statistiques, Dijon, 7-8 décembre 1995, 283-287, ENE-SAD, Dijon.
- CALLIER P., 1996. La cartographie des préférences. Son application en milieu industriel et son extension aux plans incomplets. Thèse de doctorat de 3^e cycle, Université des sciences et techniques du Languedoc, Montpellier.
- CAROLL J.D., CHANG J.J., 1970. Analysis of individual differences in multidimensional scaling via a N-way generalization of « Eckart-Young » decomposition. *Psychometrika*, **35**, 283-319.
- ESCOUFIER Y., 1973. Le traitement des variables vectorielles. *Biometrics*, **29**, 751-760.
- GOWER J.C., 1971. A general coefficient of similarity and some of its properties. *Biometrics*, **27** (December), 857-874.
- HEDDERLEY D., WAKELING I., 1995. A comparison of imputation techniques for internal preference mapping, using Monte-Carlo simulation. *Food Quality and Preference*, **6**, 281-297.
- HARTIGAN J.A., 1975. *Clustering algorithms*. John Wiley & Sons Inc, New York.
- HUBERT L., ARABIE, P., 1985. Comparing partitions. *J. Classif.*, **2**, 193-218.
- JOBSON J.D., 1991. *Applied multivariate data analysis. Volume II. Categorical and multivariate methods*. Springer-Verlag, New York.
- JOHN P.W.M., 1971. *Statistical designs and analysis of experiments*. Macmillan, New York.
- JONES B., KENWARD, M.G., 1989. *Design and analysis of cross-over trials*. Chapman and Hall, London.
- MEAD A., 1992. Review of the development of multidimensional scaling methods. *The Statistician*, **41**, 27-39.

- MILLIGAN G.W., COOPER M.C., 1986. A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate Behavioral Research*, **21**, 441-458.
- MILLIGAN G.W., COOPER M.C., 1988. A study of standardization of variables in cluster analysis. *J. Classif.*, **5**, 181-204.
- NUTE GR., MACFIE H.J.H., GREENHOFF K., 1989. Practical application of preference mapping. In: THOMSON D.M.H. (ed.): *Food Acceptability*, 377-386, Elsevier Science Publishers Ltd, London.
- PERYAM D.R., PILGRIM F.J., 1957. Hedonic scale method of measuring food preferences. *Food Technol.*, **11**, 9-14.
- SAS INSTITUTE Inc., 1989. *SAS/STAT User's Guide, Version 6*, (Vol. 1 and 2), Fourth Edition, SAS INSTITUTE Inc. Cary, NC.
- SCHIFFMAN S.S., REYNOLDS M.L., YOUNG F.W., 1981. *Introduction to multidimensional scaling. Theory, methods and applications*. Academic Press, Orlando.
- SCHLICH P., 1993. Uses of change-over designs and repeated measurements in sensory and consumer studies. *Food Qual. Prefer.*, **4**, 223-235.
- SCHLICH P., 1995. Preference mapping relating consumer preferences to sensory or instrumental measurements. In: ETIEVANT P., SCHREIER P. (ed.), *Bioflavour 95*, 135-150, Inra Editions, Paris.
- SCHLICH P., MCEWAN J.A., 1992. Cartographie des préférences. Un outil statistique pour l'industrie agro-alimentaire. *Sci. Aliments*, **12**, 339-355.
- STREET A.P., STREET D.J., 1987. *Combinatorics of experimental design*. Clarendon Press, Oxford.
- TAKANE Y., YOUNG F.W., LEEUW J.D., 1977. Nonmetric individual differences scaling: an alternative least squares method with optimal scaling features. *Psychometrika*, **42**, 3-27.
- WAKELING I.N., MACFIE H.J.H., 1995. Designing consumer trials for first and higher orders of carry-over effect when only a subset of k samples from p may be tested. *Food Qual. Prefer.*, **6**, 299-308.
- WILLIAMS E.J., 1949. Experimental designs balanced for the estimation of residual effects of treatments. *Aust. J. Sci. Res., Ser. A*, **2**, 149-168.

Article 5

5. De La Presa-Owens C., Schlich P., Davies H., Noble A.C. (1998). Effect of "Méthode Champenoise" process on aroma of four *V. vinifera* varieties. *American Journal of Enology and Viticulture*, **49**, 289-294.

Effect of *Méthode Champenoise* Process on Aroma of Four *V. vinifera* Varieties

C. DE LA PRESA-OWENS¹, P. SCHLICH², H. D. DAVIES³, and A. C. NOBLE^{4*}

To characterize the changes in aroma resulting from the *Méthode Champenoise* process, 28 base wines and their corresponding sparkling wines were profiled by descriptive analysis. The base wines were made from Chardonnay, Pinot blanc, Pinot Meunier, and Pinot noir grapes grown in Anderson Valley, Carneros District, Napa Valley, and Russian River Valley. Intensities of seven aroma attributes were rated in base wines and sparkling wines separately. Principal Component Analyses (PCA) of the base wine data separated the white from red varietal wines, whereas the PCA of the sparkling wines did not cluster wines by region or variety. When differences in intensity ratings between each sparkling wine and its corresponding base wine were analyzed by PCA, the first dimension separated the white from red varietal wines. This pattern resulted primarily from the larger increase in citrus and apple intensity in red varietals and in vanilla/butter aroma in whites during the 18 months on yeast following the secondary fermentation.

KEY WORDS: base wine, sparkling wine, flavor, sensory, descriptive analysis

The character of bottle-fermented sparkling wines (*Méthode Champenoise*) is primarily influenced by grape variety, vineyard location, and yeast autolysis during aging. The effect on wine composition of yeast autolysis during aging [12] has been studied, as have factors influencing yeast growth such as the fermentation condition of the yeast inoculum [15] as well as bottle pressure, which affects the growth rate of yeast [6].

With the exception of two studies, sparkling sensory evaluations have only examined quality or preference ratings of base wines [2,14] and sparkling wines [4,5,9,11,13,16]. Using analytical sensory procedures, the sensory effects of varietal composition [7] and geographical parameters [10] were monitored in base wines later used in sparkling wine production. However, no investigations have profiled the changes which occur during secondary fermentation and subsequent aging on the yeast in sparkling wines made by the *Méthode Champenoise* process. The objective of this research was to quantitatively monitor the effect of the secondary fermentation process on aroma of the four major varietal wines used in sparkling wine production.

Materials and Methods

Wines: Pinot noir, Chardonnay, Pinot Meunier, and Pinot blanc base wines were made by the participating wineries from vineyards in five different regions in California (Table 1). Each wine was made from a single *Vitis vinifera* cultivar using "free run" juice with

minimal contact with grape skins or seeds. Wines were fermented in stainless steel and had not undergone malolactic fermentation. After being transferred to the university at the end of fermentation, wines were cold-stabilized, and sulfur dioxide was added, prior to bottling. For preparation of the sparkling wines, 25 mL/L inverted sugar was added to provide 6.25 Atm pressure

Table 1. Vineyard and winery of original and variety of wines.

Variety	Winery	Vineyard location
Chardonnay	Mumm	Anderson Valley
	Roederer Estate	Anderson Valley
	Scharfenberger	Anderson Valley
	Domaine Carneros	Carneros-Napa
	Gloria Ferrer	Carneros-Sonoma
	Mumm	Napa Valley
	Mumm	Napa Valley
	Schramsberg	Napa Valley
	Schramsberg	Napa Valley
	Iron Horse	Russian River
Pinot blanc	Jordon	Russian River
	Korbel	Russian River
	Piper Sonoma	Alexander Valley
	Domaine Carneros	Carneros-Napa
	Schramsberg	Napa Valley
Pinot Meunier	Mumm	Carneros-Napa
	Schramsberg	Napa Valley
	Jordon	Russian River
Pinot noir	Mumm	Anderson Valley
	Roederer Estate	Anderson Valley
	Domaine Carneros	Carneros-Napa
	Gloria Ferrer	Carneros-Sonoma
	Schramsberg	Napa Valley
	Korbel	Russian River
	Mumm	Santa Maria
	Iron Horse	Russian River

¹Dept. Viticulture & Enology, University of California, Davis, CA 95616 (FAX 530 752-0382; acnoble@ucdavis.edu, phone 530 752 0387); ²Inst. Nat. Rech.Agro., 21000 Dijon, France; ³Schramsberg Winery, Calistoga, CA.

*Corresponding author: [Fax: 530-752-3308; e-mail: <acnoble@ucdavis.edu>].

Acknowledgments: This research was funded by the participating wineries, and grants from the Spanish Government (CPO), the Adolf Heck foundation (HDD) and NATO (PS).

Manuscript submitted for publication 26 September 1997.

Copyright © 1998 by the American Society for Enology and Viticulture. All rights reserved.

with Red Star Premier Cuvée yeast (Universal Foods Corporation, Milwaukee, WI), and 0.7 mL/L bentonite solution (Clarificant S, Pechiney Cork & Seal Co., Vallejo, CA). Wines were stored at 17°C until the secondary fermentation was completed, then held at 10°C. After 18 months bottle-aging, the sparkling wines were riddled, disgorged, and 20 ppm of SO₂ added.

Sensory protocol: Preliminary evaluation of the wines indicated that few differences among the wines were found in taste, while there were large differences in aroma. Consequently, descriptive analyses of wine aroma were performed. The base wines were evaluated in February 1995 and the sparkling wines in November of 1996 by two different panels. Twenty-one judges (12 males and 9 females, 18 to 40 years of age) participated in the base wine evaluation, while 17 (13 males and 4 females, 22 to 47 years of age) were on the sparkling wine panel. Base wine judges were trained for two weeks prior to formal testing, while the sparkling wine judges were trained for four weeks. At each training session, aroma standards were presented with selected wines from the experiment.

For the last week of training, the seven aroma attributes which had been selected by the individual panels were rated in two practice sessions, using the same procedures as in the formal sessions. The composition of the reference standards which defined these aroma notes are given in Table 2.

Formal sessions: Formal evaluations were conducted in individual booths at 20°C to 22°C under red light to mask color differences. Reference standards which were evaluated prior to each session were available for subsequent evaluation as necessary. Distilled water was provided for sniffing between wines. Twenty-five milliliters of base wine or 40 mL sparkling wines were presented in coded, 170-mL, tulip-shaped

wine glasses covered with plastic lids. Base wines were evaluated at room temperature (21°C ± 2°C). Sparkling wines were cooled to 10°C before being poured. The sparkling wines were held uncovered for five minutes after being poured to stabilize the level of CO₂, then covered and served.

Six or seven base wines were presented per session in a different random order to each judge for each of the duplicate evaluations. For the triplicate evaluation of the sparkling wines, the presentation order for each judge and each replicate was balanced for position and first order carry over effect using William's Latin Squares design [17]. Seven sparkling wines per session were evaluated monadically (one at a time) to standardize evolution of carbon dioxide. For both wine types, at each session judges evaluated the standards and then scored the intensity of each term on a category scale where 0 = no, 1 = "low", and 9 = "high".

Data analysis: All descriptive data was analyzed using SAS® System for Windows, Version 6.1 (SAS Institute Inc., Cary, NC 1994). For base and sparkling wines separately, intensity ratings were analyzed by mixed model analysis of variance (ANOVA) with judges considered a random effect. A mixed model analysis of variance (ANOVA), where both grape variety and wine type (base/sparkling) were treated as fixed effects, where wines nested within variety were considered as a random effect was also performed. Principal component analysis (PCA) of the covariance matrix of judge average ratings for each wine was performed for each wine type separately and on base and sparkling wines together.

To illustrate the effect of the secondary fermentation on the four varieties, a PCA was run on the difference in intensity ratings between each base and sparkling wine. To compare the base and sparkling aroma

Table 2. Composition of aroma reference standards for base and sparkling wine descriptive analysis.

Attribute	Experiment	Composition in 500 mL Thompson Seedless wine (base) or 30 mL Chardonnay wine (sparkling)
Citrus	Base	5 g grated orange + 5 g grated lemon peels; 30 mL fresh orange, 15 mL fresh lemon, 5 mL fresh grapefruit, 5 mL fresh lime juice
	Sparkling	1.5 mL lemon juice + 2 mL orange juice + 11 mL pineapple juice
Tropical fruit	Base	5 mL imitation banana flavor, 5 mL imitation coconut flavor + 34 mL canned pineapple juice
	Sparkling	8 g coconut, 3 g Koolaid tropical fruit mix, 1.5 g banana + 5 mL frozen pear concentrate
Apple/Pear	Base	1 chopped green apple, 1 chopped Bartlett pear + 15 mL frozen pear concentrate
	Sparkling	5 g chopped green apple + 6 g chopped red apple
Floral	Base	0.5 mL phenyl ethyl alcohol, 10 g rose petals
	Sparkling	0.1 mL linalool, 0.5 mL phenethyl alcohol
Vanilla/Butter	Base	2 chopped caramel cubes, 2 drops butter extract, 5 g vanillin
	Sparkling	1 g sweet butter + 2 g caramelized sugar, 1 mL pure vanilla extract
Peach/Melon	Base	30 mL frozen peach concentrate, 1 chopped canned peach, 30 g fresh cantaloupe
	Sparkling	7 g peach pulp, 7 g chopped melon, 6 g chopped pear
Berry	Base	6 fresh strawberries, 6 frozen raspberries + 6 canned black cherries
	Sparkling	7 g blackberry, 5 g strawberry
Honey	Sparkling	3.5 g honey
Toasty	Sparkling	1 (3 mm x 3 mm) piece American white oak toasted chip, 0.1 g toasted almonds
Tea/Tobacco	Sparkling	3 g English tea, 0.1 g cigarette tobacco

configurations, the RV coefficient, a generalized correlation coefficient, and the analytical generalized permutation test of the RV value, were calculated. The value of this coefficient is expressed as normal deviation from the null hypothesis permitting a statistical estimate of the fit [18].

Results and Discussion

Base wines: All terms but tropical fruit ($p < 0.01$) varied highly significantly across the 28 base wines ($p < 0.001$). From a principal component analysis (PCA) of these mean ratings, the first PC separated base wines high in vanilla/butter versus those high in floral, citrus and apple (Fig. 1). Base wines made from white grape varieties, Chardonnay and Pinot blanc, were clustered together and were generally higher in floral, citrus, and apple. The red varietal wines, Pinot noir and Pinot Meunier, were higher in berry and vanilla/butter terms and, with the exception of two wines, were separated from the white base wines. Visual inspection of Figure 1 data failed to show clustering by region or even by winery.

Sparkling wines: The sparkling wines varied significantly in five terms: vanilla/butter ($p < 0.001$), peach/melon and apple ($p < 0.01$), tropical fruit and citrus ($p < 0.05$). PCA of the seven terms rated in base wines separated wines high in tropical fruit, vanilla/butter, floral, and peach/melon from those low in these terms on the first PC (Fig. 2), while the second PC contrasted wines high in apple and citrus from those low in those attributes. In contrast to the configuration observed in Figure 1 for the base wines, the sparkling wines were not clustered by variety, region of origin, winery, or by color of the grapes.

Comparison of base and sparkling wine aromas: Each of the seven terms rated in the base and sparkling wines varied highly significantly between the two wine types ($p < 0.001$) and, with the exception of floral (NS), all terms varied significantly across the 56 wines ($p < 0.001$ for all except peach/melon $p < 0.05$) (data not shown). Visual inspection of the differences between

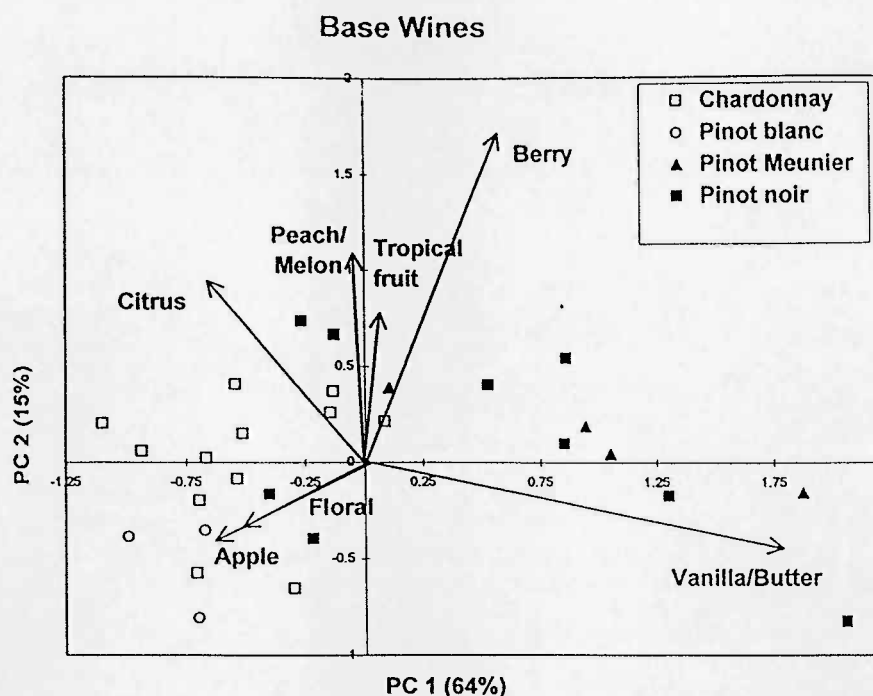


Fig. 1. Projection of the attributes on principal components I and II from principal component analysis of the judge averages for the 28 base wines. Aroma loadings (vectors) and wine factor scores (points).

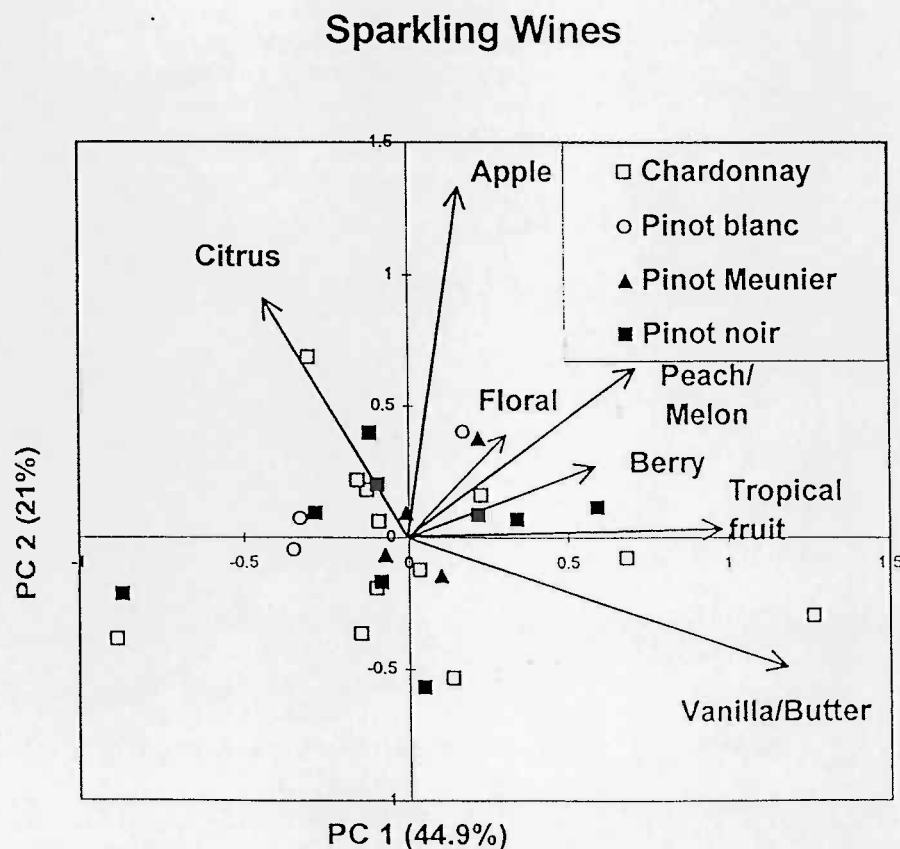


Fig. 2. Projection of the attributes on principal components I and II from principal component analysis of the judge averages for the 28 sparkling wines. Aroma loadings (vectors) and wine factor scores (points).

Table 3. Summary of analyses of variance of main effects: variety, type of wine and wines (nested withing variety), F ratios and degrees of freedom (df).

Attributes	Variety	Wine	Type	Variety × type	Wine (var.) × type
Berry	6.60**	1.63 ns	15.34***	7.90***	1.48 ns
Peach/Melon	0.30 ns	1.22 ns	61.70***	1.31 ns	1.66*
Green apple	3.82*	1.51 ns	25.37***	4.61*1	.39 ns
Tropical	0.31 ns	2.25*	62.12***	0.65 ns	1.12 ns
Citrus	2.40 ns	1.78 ns	48.33***	2.66 ns	1.49 ns
Floral	4.54*	1.90 ns	33.63***	8.03***	0.60 ns
Vanilla/Butter	5.83**	0.94 ns	14.12***	5.59**	3.55***
df	3	24	1	3	24

*, **, *** indicate significance at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

the configurations for base and sparkling wines shown in Figures 1 and 2, suggests that the aromas of the varietal base wines were modified in different ways during the secondary fermentation. Consistent with this, significant Variety × Type interactions were found for berry, apple, floral and vanilla/butter ($p < 0.05$ to $p < 0.001$) (Table 3).

To test the fit of the two wine configurations, the RV coefficient was calculated. A normalized RV value (NRV) of 0.32 was found in the permutation test, confirming the lack of correspondence between the base and sparkling wine configurations. (An NRV value greater than 2.0 is required for a significant fit). These results confirm that the changes in aroma occurring

during production of sparkling wine varied differently for either individual wines or different varieties. This means that the profiles of the base wines do not permit prediction of the sensory properties of the sparkling wines after 18 months of lees aging.

A PCA of the difference in attribute ratings between sparkling and base wines was performed to determine patterns in the changes in aroma occurring during the secondary fermentation. Similar to the PCA of the base wine data, the first PC separated the wines made from red and white varietal grapes (Fig. 3). The first PC was highly correlated with vanilla/butter and berry and negatively correlated with citrus, apple and floral. From the PCA it can be inferred that there was an increase in vanilla/butter from base to sparkling wine for both white varieties (Chardonnay and Pinot blanc) while there was an increase in citrus, apple, and floral for the red variety wines during the second fermentation in the bottle. No effect of region or winery was apparent in visual inspection of the PCA space (Fig. 3).

These inferences were confirmed by the ANOVA of main effects of variety and wine type (base/sparkling) (Table 3) and are illustrated in Figure 4 where the varietal mean ratings are plotted for base and sparkling wines. Four aroma terms varied significantly across variety: berry and vanilla/butter, apple and floral. All attributes were significantly more intense in the sparkling wines, than the corresponding base wines with the exception of berry which showed the opposite pattern. The secondary fermentation increased the intensity of vanilla/butter in the white varietal wines (Chardonnay and Pinot blanc), while no significant difference occurred in the red varieties (Pinot noir and Pinot Meunier). Conversely, the berry aroma decreased significantly in the red varieties, while no significant difference was observed for the whites. Citrus, tropical fruit, and peach/melon aromas were significantly higher in sparkling wines for all varieties, while only for the red varieties was a significant increase in floral and apple produced during the secondary fermentation.

For the first time, the effect of the secondary fermentation has been shown to be different for the four major varieties used in sparkling wine production. Due to the extensive training of the judges in both studies, it is unlikely that this is the result of the use of different judges for evaluation of the base and sparkling wines. Although the standards for the terms were prepared slightly differently for the studies, the concept of each attribute may have been quite similar. However, there is no way in which to prove this. Inclusion

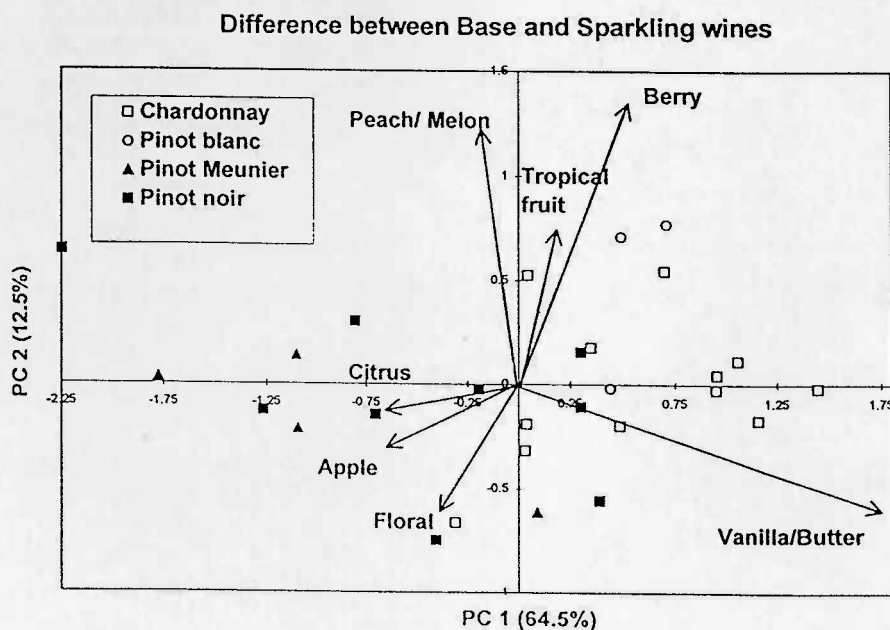


Fig. 3. Projection of the attributes on principal components I and II from principal component analysis of the differences in term ratings between each sparkling and base wine. Aroma loadings (vectors) and wine aroma difference factor scores (points).

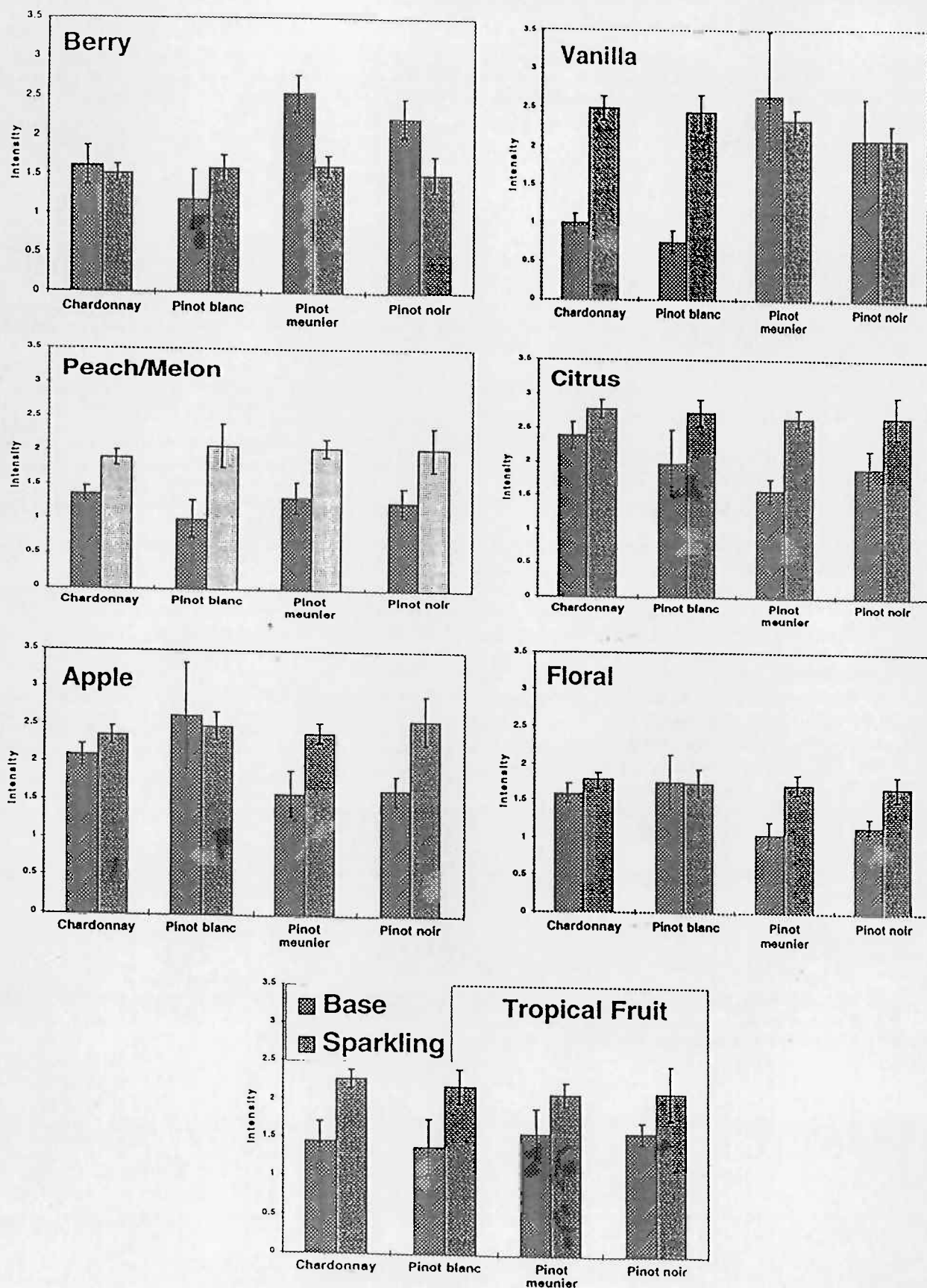


Fig. 4. Mean values for the intensity of base and sparkling wines for each variety for each aroma term. The barbells denote the 95% confidence interval. (Where there were 12 Chardonnays, 3 Pinot blancs, 4 Pinot Meuniers and 9 Pinot noirs, for base wines $n = 21$ judges $\times$ 2 reps $\times$ number of wines per variety; for sparkling wines $n = 17$ judges $\times$ 3 reps $\times$ number of wines per variety).

of one of the base wines with the sparkling wines descriptive analysis would not have standardized the analysis for two reasons: (1) the base wine would have been 18 months older, (2) the contrast between the base wine and the sparkling wines would have overshadowed other differences among the wines.

The secondary fermentation in sparkling wines produces the carbon dioxide bubbles associated with this beverage. However the primary changes in flavor are thought to arise from autolysis of yeast which occurs during aging on the lees which are reported to slow after two years [12]. Informal reports from both French and Californian sparkling winemakers suggest that changes due to yeast autolysis develop first in white varieties and may require up to three years for development in red varieties. In this study, after 18 months on the lees, the absolute difference in intensity ratings was larger for the red wines than the white (Fig. 4). However, the size of the overall change in aroma cannot be estimated from this type of study since no overall judgements of aroma intensity were made.

Conclusions

Base wines for sparkling wine production are chosen for their lack of distinctive aromas. However, a separation among base wines according to color of the cultivar was observed. The major differences in aroma between red and white base wine varieties were the characterization of Chardonnay and Pinot blanc wines as higher in intensity of floral, citrus and apple notes, while Pinot Meunier and Pinot noir base were higher in berry and vanilla/butter. The secondary fermentation of these base wines by *Méthode Champenoise* produced wines which had higher intensity ratings for all attributes except berry. The sparkling wines were no longer differentiated by variety or color of the grapes as the result of the different patterns in evolution of aroma suggesting that the sensory properties of the base wines do not permit prediction of those of the sparkling wines.

Literature Cited

1. Amerine, M. A., and C. S. Ough. Analysis of Musts and Wines. John Wiley and Sons, New York (1980).
2. Bertuccioli, M., and G. F. Montedoro. Use of Trebbiano as a base wine for the production of sparkling wines: analysis of some constituents connected with sensory characteristics. *Industrie delle Bevande* 55:368-374 (1981).
3. Boulton, R. B., V. L. Singleton, *et al.* Principles and Practices of Winemaking. Chapman and Hall (1996).
4. Carroll, D. E. Evaluation of sparkling wines produced from muscadine cultivars. *Am. J. Enol. Vitic.* 35:72-75 (1984).
5. Castino, M., and M. Ubligui. Comparison of base wines for preparation of bottled-fermented sparkling wines. *Industrie delle Bevande* 19:315-317 (1990).
6. Chen, G. N., and F. Gutamis. Carbon dioxide inhibition of yeast growth in biomass production. *Biotechnol. Bioeng.* 18:1455-1462 (1976).
7. de la Presa-Owens, C., and A. C. Noble. Descriptive analysis of three white wine varieties from Penedès. *Am. J. Enol. Vitic.* 46:5-9 (1995).
8. Dozon, N. M., and A. C. Noble. Investigation of the effect of fluorescent light on the aroma of sparkling wine and its base wine. *Am. J. Enol. Vitic.* 40:265-271 (1989).
9. Ewart, A. J. W., and J. H. Sitters. Wine assessment of Pinot noir, Chardonnay and Riesling clones. In: Proceedings of the Second International Symposium for Cool Climate Viticulture and Oenology. Auckland. R. Thornton, R. E. Smart, S. B. Rodriguez, and J. E. Young (Eds.). pp 201-205. New Zealand Soc. Vitic. & Oenol. (1988).
10. Iacono, F., M. Bertamini, A. Dalla Serra, *et al.* Use of chemical and sensory analysis for characterization of Chardonnay sparkling base wines produced in different environments in the Trentino district of Italy. *Riv. Vitic. Enol.* 43:3-14 (1990).
11. Johnson, H. The World Atlas of Wine. Simon & Schuster. New York, NY (1985).
12. Leroy, M. J., M. Charpentier, B. Duteurtre, *et al.* Yeast autolysis during Champagne aging. *Am. J. Enol. Vitic.* 41:21-28 (1990).
13. Lovino, R., A. Piracci, and M. Scazzariello. Study of the suitability of southern Italian vines for sparkling wine manufacture. *Enotecnico* 30: 85-90 (1994).
14. Markides, A. J., P. R. Dry, and A. J. W. Ewart. Australian experiences in the quality assessment of varietal and regional factors important to sparkling wine production. In: Proceedings of the Seventh Australian Wine Industry Technical Conference, Adelaide. pp 93-95 (1990).
15. Monk, P. R., and R. J. Storer. The kinetics of yeast growth and sugar utilization in tirage. The influence of different methods of starter culture preparation and inoculum levels. *Am. J. Enol. Vitic.* 37: 72-76 (1986).
16. Moulin, J. P. Champagne: the method of production and origin of the quality of this French wine. In: Proceedings of the Sixth Australian Wine Industry Technical Conference, Adelaide. pp 218-223 (1987).
17. Schlich, P. Uses of change-over designs and repeated measurements in sensory and consumer studies. *Food Qual. Pref.* 4:223-235 (1993).
18. Schlich, P. Defining and validating assessor compromises about product distances and attribute correlations. In: Multivariate Analysis of Data in Sensory Science. T. Naes and E. Risvik (Eds.). pp 259-306. Elsevier Science (1996).

Article 6

6. Peleg H., Gacon K., Schlich P. & Noble A.C. (1999). Bitterness and astringency of flavan-3-ol monomers, dimers and trimers. *Journal of the Science of Food and Agriculture*, **79**, 1123-1128.

Bitterness and astringency of flavan-3-ol monomers, dimers and trimers[†]

Hanna Peleg,¹ Karine Gacon,² Pascal Schlich³ and Ann C Noble^{4*}

¹PO Box 690, Hod-Hasharon, Israel

²13 rue Jehan de Marville, 21000 Dijon, France

³INRA, rue 17 Sully, 21034 Dijon, France

⁴Department of Viticulture and Enology, University of California, One Shields Ave, Davis, CA 95616, USA

Abstract: Intensity of astringency and bitterness of seven flavonoid compounds was evaluated by a time-intensity (TI) procedure. Eighteen trained judges rated intensity continuously from ingestion, through expectoration at 10 s until extinction of the sensation. The seven stimuli included two flavan-3-ol monomers, (+)-catechin and (–)-epicatechin, three dimers and two trimers synthesised from catechin or epicatechin by condensation with (+)-dihydroquercetin. As the degree of polymerisation increased, maximum bitterness intensity ($I_{\max}$) and total duration (T_{tot}) decreased whereas astringency $I_{\max}$ increased. The monomers were significantly higher in bitterness at $I_{\max}$ than the dimers, which were significantly higher than the trimers. Astringency $I_{\max}$ of the monomers was lower than the dimers or trimers, although no significant difference was found in T_{tot} among the polymer classes. The bond linking the monomeric units had an influence on both sensory properties. The catechin-catechin dimer linked by a 4→6 bond was more bitter than both catechin-(4→8)-catechin and catechin-(4→8)-epicatechin. Astringency was affected by both the specific linkage and the identity of the monomeric units with the dimer, catechin-(4→8)-catechin, being lower in astringency than either catechin-(4→6)-catechin or catechin-(4→8)-epicatechin.

© 1999 Society of Chemical Industry

Keywords: bitterness; astringency; time-intensity; taste; tactile; flavonoid

INTRODUCTION

Flavan-3-ols and their polymers, which are distributed widely in nature in many fruits,¹ elicit persistent bitterness and astringency. Bitter taste is perceived by taste receptors on the tongue. Although several transduction mechanisms for the perception of bitterness have been identified for individual bitterants,^{2–4} no study has addressed bitterness of flavan-3-ols. Several models have been proposed to predict bitterness from structural features or molecular properties, however, no model fits all bitter compounds. In contrast to bitterness, astringency is a tactile sensation perceived as dryness, puckering and roughness throughout the oral cavity.⁵ Perception of astringency has a slow onset and is characterised by a long persistence; it is not instantaneous but requires time for development.^{6–8} Although many compounds elicit an astringent sensation, astringency is chemically defined as the ability to precipitate proteins.⁹

Catechin and epicatechin are monomeric flavonoid phenols (tannins) which are found in fruits such as grapes and apples and in wine, tea and cider.

Comparison of equal weights of catechin and epicatechin, which are chiral isomers, revealed that epicatechin had a higher maximum intensity and longer persistence of bitterness and astringency than catechin.¹⁰ Whereas the sole difference between catechin and epicatechin is in the configuration of one hydroxyl group, polymers of these flavan-3-ols also vary in the location of bonds linking subunits, in molecular weight and identity of the subunits. As a result, comparison of monomers and polymeric fractions reflects variability other than polymer size. Overall, evaluations of polymeric fractions of tannic acid and tannins has demonstrated that larger molecules tend to be less bitter and more astringent.^{11,12} Few sensory studies, however, have evaluated fractionated polymers^{13,14} while none has tested individual dimers and trimers.

In this study, astringency and bitterness elicited by aqueous solutions of monomers, dimers and trimers of procyanidins have been studied by time-intensity (TI) methodology to characterise the time course of intensity of these persistent attributes.

* Correspondence to: Ann C Noble, Department of Viticulture and Enology, University of California, One Shields Avenue, Davis, CA 95616, USA

E-mail: acnoble@ucdavis.edu

[†] Presented in part as a poster at Second Pangborn Sensory Science Symposium, August 1995, Davis, CA, USA

(Received 17 April 1998; revised version 27 October 1998; accepted 15 January 1999)

Stimulus	Chemical structure	Source
(+)-catechin	(+)-cyanidol-3	SIGMA
(-)-epicatechin	cis-2-[3,4-dihydroxyphenyl]-3,4-dihydro-2H-1-benzopyran-3,5,7-triol	SIGMA
dimer B3	catechin (4→8) catechin	^a
dimer B6	catechin (4→6) catechin	^a
dimer B4	catechin (4→8) epicatechin	^a
trimer C2	catechin (4→8) catechin (4→8) catechin	^a
trimer C	catechin (4→8) catechin (4→8) epicatechin	^a

Table 1. Identity of the experimental stimuli

^a Synthesised as described in the text.

MATERIALS AND METHODS

Materials and sample preparation

The monomers, 3*R*(-)-epicatechin and 3*S*(+)-catechin, were purchased from SIGMA, USA. Three dimers and two trimers of procyanidins (Table 1) were synthesised by reacting catechin or epicatechin with dihydroquercetin by modification of methods of Eastmond¹⁵ and Delcour *et al.*¹⁴

Each compound was evaluated at a concentration of 0.9 g litre⁻¹ in aqueous ethanol (1% v/v). Solutions were prepared under nitrogen to avoid oxidation.

Sensory protocol

A total of 18 subjects, 4 women and 14 men, students or staff from the University of California, Davis, between 20 and 36 years old, participated in nine sessions over the course of two weeks. Astringency was rated during the first week and bitterness during the second. For each attribute, the two formal sessions were preceded by training sessions to familiarise the panelists with the concept of astringency or bitterness and the computerised time-intensity (TI) system.

Samples were tested at room temperature in individual booths illuminated with red light. For both training and formal sessions, low and high reference standards for bitterness and astringency were first evaluated to standardise use of the terms and the scale. The low and high standards in 1% (v/v) aqueous ethanol, respectively, for bitterness were 0 and 0.015 g litre⁻¹ quinine and for astringency were 0 and 1.0 g litre⁻¹ alum. The seven coded stimuli were presented in a randomised order in each formal session. The judges took the 15 ml sample in their mouth and rated the intensity continuously through expectoration at 10 s to extinction of the sensation using the computerised system described previously.¹⁶ Subjects rinsed with deionised water and waited 1 min between samples.

Determination of salivary flow

Judges ingested 15 ml of an aqueous solution of citric acid (4 g litre⁻¹) which was expectorated at 10 s. The subjects then spat stimulated saliva into a weighed container for 1 min. The saliva collected in duplicate was weighed on an analytical balance.

Statistical analysis

Weight of saliva was analysed by a two-way analysis of

variance (ANOVA). Judges varied highly significantly in their stimulated saliva weights ($F_{17,17}=24.90$, $p<0.01$) (LSD=0.51). Six judges were assigned to each of three flow groups for which the means ($\pm$ standard deviations) were low-flow=1.91 g (± 0.51), medium-flow=3.08 g (± 0.14) and high-flow=3.81 g (± 0.30). Some overlap in flow occurred between individuals in different flow groups, but all subjects were retained to increase the power of the analysis.

Parameters extracted from the TI curves included time to maximum ($T_{\max}$), maximum intensity ($I_{\max}$), total duration (T_{tot}), time (from ingestion) to decay to 60% of maximum intensity (T_{60}) and 30% (T_{30}). An initial ANOVA was performed to evaluate the main effects of subjects, compounds and replications; replications were not significant so they were not included in the subsequent ANOVA models, but were treated as two observations per cell. Judges and compounds were treated as random effects in all analyses. Nesting of samples within the monomer, dimer or trimer polymer group (sample(polymer-size)) and of individual judges within salivary flow groups (subject(saliva-group)) was performed since the same compounds or subjects did not belong to each group. This permitted comparison of the different polymer size (or flow) groups while testing for the variation of the compounds (or subjects) in each group. To compare the different polymeric groups and the effect of salivary flow status, four-way ANOVAs were run for both attributes testing the main effects: subject(saliva-group), flow-group, sample(polymer-size); polymer-size. All statistical analyses were performed using the SAS[®] System for Windows[™], Version 6.10.¹⁷

RESULTS

Effect of molecular size

The polymer class (size) and the individual compounds nested within polymer class (sample) varied significantly for several temporal parameters for astringency and bitterness as shown in Table 2. In all analyses subjects were a highly significant source of variation, but there were no significant interactions of subjects with polymer size or sample. Significant differences among polymer classes were found for bitterness $I_{\max}$ and T_{tot} ($p<0.001$), whereas for astringency, the only TI parameter approaching significance among polymer classes was $I_{\max}$

		Astringency			Bitterness		
Source of variation	Df numbers	Df error	F	Sig	Df error	F	Sig
<i>T_{max}</i>							
Polymer size (P)	2	4.19	0.29	0.766	1.71	0.69	0.609
Sample (P)	4	8	2.94	0.091	8	1.34	0.336
Saliva group	2	14.6	1.68	0.222	13.9	5.04	0.023
<i>I_{max}</i>							
Polymer size (P)	2	3.75	3.42	0.143	4	7.73	0.039
Sample (P)	4	8	3.43	0.065	8	6.05	0.015
Saliva group	2	7.62	13.18	0.003	10.9	2.09	0.017
<i>T_{tot}</i>							
Polymer size (P)	2	3.69	0.44	0.685	4.46	4.22	0.101
Sample (P)	4	8	4.61	0.032	8	5.27	0.022
Saliva group	2	13.93	0.86	0.446	13.82	3.26	0.068
<i>T₆₀</i>							
Polymer size (P)	2	2.65	0.11	0.901	4.09	3.11	0.150
Sample (P)	4	8	3.35	0.069	8	1.63	0.260
Saliva group	2	13.98	1.44	0.270	15.51	3.54	0.056
<i>T₃₀</i>							
Polymer size (P)	2	3.36	0.34	0.734	55	3.60	0.116
Sample (P)	4	8	3.75	0.053	8	2.86	0.096
Saliva group	2	14.12	1.17	0.339	14.63	3.76	0.048

Table 2. Summary of analyses of variance of time-intensity parameters. Degrees of freedom (df), *F* ratios and significance levels (Sig) for main effect of polymer size, individual compound (sample) nested within size, and salivary flow group

($p=0.143$). Contrast tests showed the monomers have significantly larger values than the trimers for both maximum intensity ($F_{1,4}=15.6$, $p<0.07$) and total duration of bitterness ($F_{1,4}=8.06$, $p<0.047$), whereas maximum intensity of astringency was higher in the dimers and trimers than the monomers ($F_{1,4}=5.20$, $p<0.08$).

The time courses for astringency and bitterness of the monomer, dimer and trimer classes are plotted in Fig 1 using averaged values extracted from individual curves. The dimer and trimer flavan-3-ols were slightly higher in astringency than the monomer class throughout the onset and decay of the sensation. In contrast, as the size of the flavan-3-ols increased, the intensity and persistence of bitterness decreased dramatically.

As shown in Fig 2, monomers were 150% more bitter than astringent at maximum intensity, conversely the trimers were 150% more intensely astringent than bitter, while bitterness and astringency were equally intense for the dimers.

Comparison of individual compounds: effect of molecular size, bond site and monomer identity

Individual compounds varied in both attributes for all TI parameters at levels of significance higher than $p<0.10$, except for time to maximum intensity (bitter) and time to decrease to 60% of I_{max} (T_{60}) (astringency) (Table 2). As typically observed in TI studies, maximum intensity was significantly correlated with duration; for astringency, $r=0.762$, $df=5$ ($p<0.05$) and bitterness, $r=0.9581$, $df=5$ ($p<0.001$), while neither parameter was correlated with time to maximum intensity. To compare the individual stimuli, maximum intensities of astringency and bitterness of

the seven compounds are plotted in Fig 3. Inspection of these I_{max} ratings shows that despite an overall trend for increased astringency and decreased bitterness with increase in molecular size, structural differences among compounds of the same size also have a significant effect. Examination of differences within each polymer class reveals that although epicatechin (EPI) was more bitter and more astringent than

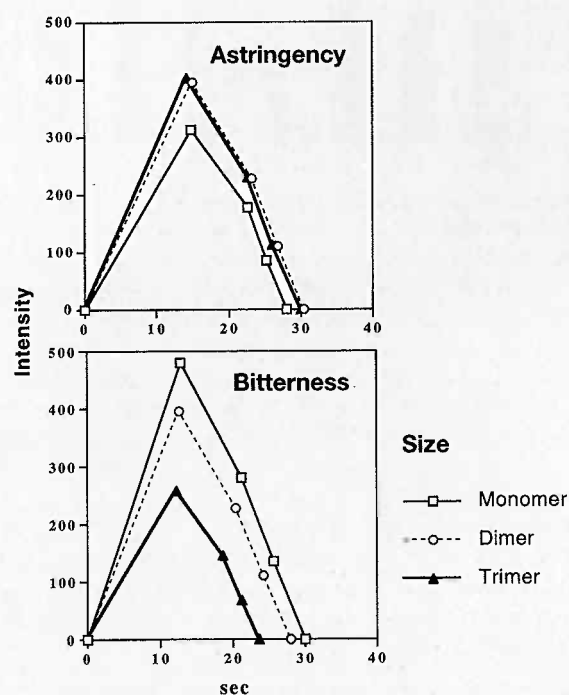


Figure 1. Mean values of intensity over time for astringency (top) and bitterness (bottom) ($n=18$ judges $\times$ 2 reps).

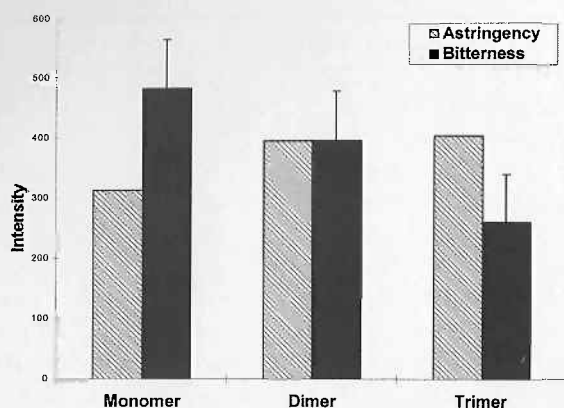


Figure 2. Mean maximum intensities of astringency (hatched) and bitterness (solid) by molecular size ($n=18$ judges $\times 2$ reps $\times n$ compounds, where $n=2$ for monomer and trimer and $n=3$ for dimer). $LSD_{(0.05)}$ for bitterness=83.1, as shown by bar bell.

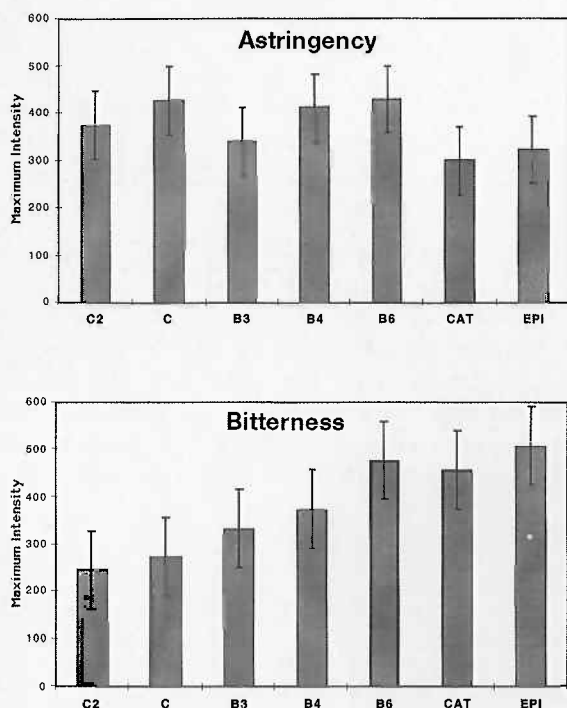


Figure 3. Mean maximum intensity of astringency (top) and bitterness (bottom) of the experimental samples. Definition of the sample codes is provided in Table 1 ($n=18$ judges $\times 2$ reps). $LSD_{(0.05)}$ for astringency=71.3 and for bitterness=80.29. LSD values are shown as bar bells.

catechin (CAT), the differences were not significant in the present study. The same was observed for the trimers; C was insignificantly higher than C2 in both astringency and bitterness. Within the dimers, B6 (CAT-(4 $\rightarrow$ 6)-CAT) was significantly higher in bitterness than B4 (CAT-(4 $\rightarrow$ 8)-EPI) and B3 (CAT-(4 $\rightarrow$ 8)-CAT). The dimer B6 was highest in astringency followed by B4 and the two trimers from which it did not differ significantly.

Effect of salivary flow status on perception of astringency and bitterness

The effect of salivary flow status was significant for

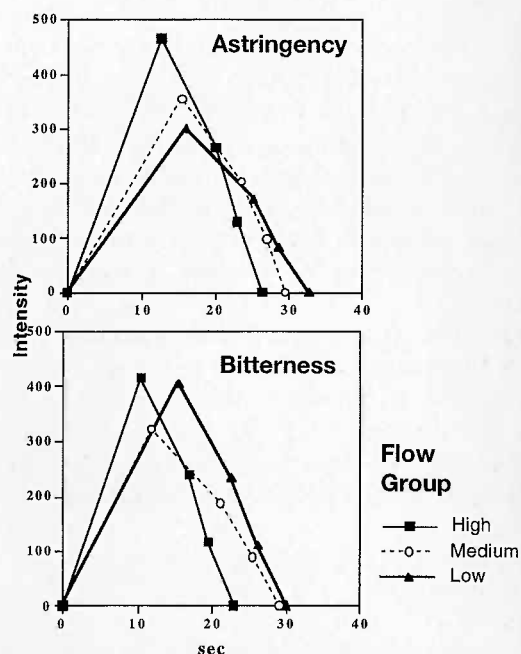


Figure 4. Mean values of intensity over time for astringency (top) and bitterness (bottom) for each salivary flow group ($n=6$ judges $\times 7$ compounds $\times 2$ reps).

selected parameters for each attribute (Table 2). Whereas all five TI parameters varied significantly across flow groups for bitterness, only maximum intensity varied significantly among flow groups for astringency ($p < 0.002$).

As illustrated in Fig 4, the high-flow subjects rated I_{max} of astringency higher than the low-flow subjects, in contrast to previous reports where high-flow subjects rated astringency lower.^{18,19} Consistent with these previous studies, however, low-flow assessors took 5.1 s longer to reach maximum bitterness than the high-flow group and bitterness persisted significantly longer for the medium- and low-flow subjects. Time to maximum and total duration of astringency showed the same trends but no significant difference among flow groups was found.

DISCUSSION

In this study, variation in intensity and duration of astringency and bitterness of procyanidin oligomers can be attributed to three structural attributes: chain length (monomer, dimer, trimer); stereochemistry of subunits (catechin versus epicatechin); and site of bond between subunits (C4 $\rightarrow$ C6 versus C4 $\rightarrow$ C8). Bitterness was highest in the monomers and lowest in the trimers, whereas monomers were rated lower in astringency than the dimers or trimers. This is consistent with previous research in which astringency increased and bitterness decreased with the degree of polymerisation.^{7,11,20}

In a study comparing different complex fractions of grape seed tannin,¹³ the polymeric fraction composed of hexameric or larger molecules was more intensely

bitter and astringent on a per weight basis than the smaller fractions. This is in contrast to the present results in which equal weights of monomers elicited higher $I_{\max}$ and longer T_{tot} for bitterness than dimers or trimers. The unique difference between the present study and previous research is the use of single molecules of known configuration, rather than employing fractions which contained mixtures of compounds.

Whether bitterness is elicited by interaction with a specific bitter membrane-bound receptor²¹ or through surface membrane interactions²² increasing the size of these procyanidins decreases bitterness. This may result from increased steric interference which reduces the strength of interactions between the flavonoid and the receptor or the receptor membrane causing the trimers to be perceived as least bitter.

Neither catechin nor epicatechin is chemically defined as an astringent, primarily since polyphenols with molecular weights (MW) below 500 typically do not precipitate proteins.²³ Substituted benzoic acids (MW from 122 to 170) have been found to elicit astringency.²⁴ Astringency of these small phenols has been hypothesised to result from precipitation, of, or strong binding with, proteins due the presence of one 1,2-dihydroxy or 1,2,3-trihydroxy groups,²⁵ which may explain the astringency of the monomeric flavan-3-ols. Possibly other alterations of the protein occur since both monomeric flavonoid phenols were found to induce astringency in the present experiment and in previous studies.^{10,26}

The increase in perceived astringency with degree of polymerisation may be due to more extensive formation of phenol-protein complexes via hydrogen bonding between hydroxyl groups of the phenolic compounds and the carbonyl groups of peptide linkages in the proteins. This more effective ability of trimers to bind proteins may be the result of the greater number of hydroxyl groups per molecule. This speculation is indeed consistent with preliminary research measuring the relative chemical astringency of these compounds using a modification of the haemanalysis of Bate-Smith²³ and Porter and Woodruffe.²⁷ The monomers had a lower maximum binding capacities and affinities than the dimers, which in turn were lower than trimer C2. Trimer C exhibited protein binding affinity equivalent to that of dimer B6 (Peleg and Noble, in preparation).

In the present sensory study, differences in astringency were observed among the dimers, with B6 (CAT-(4→6)-CAT) more astringent than B3 (CAT-(4→8)-CAT). These data are in accordance with the results for chemical astringency of dimers.²⁸ Dimeric procyanidins with a 4→6 bond (which was presumed to have the greatest chemical astringency) had more effective antiviral activity than dimers with a 4→8 bond. Antiviral activity was measured as the amount of tannin required to reduce the number of plaques to 50% of those found in control viral cultures.²⁸

The more planar conformation of the C ring of epicatechin than catechin²⁹ may further contribute to the differences in both chemical and sensory properties observed between B4 and B3, and between C and C2, as speculated previously.¹⁰ Although no significant differences were found in astringency, trimer C (CAT-(4→8)-CAT-(4→8)-EPI) was more astringent than C2 (CAT-(4→8)-CAT-(4→8)-CAT); dimer B4 (CAT-(4→8)-EPI) was more astringent than B3 (CAT-(4→8)-CAT); epicatechin was higher in astringency than catechin. For these seven compounds, the stereochemical difference also affects bitterness and astringency but to a lesser degree than differences in molecular size.

CONCLUSION

For the seven flavonoids examined, molecular size was the major factor influencing the sensory properties of bitterness and astringency. As molecular size increased, bitterness decreased and astringency increased. The monomers were more intensely bitter than astringent. Conversely trimers were more astringent than bitter, although the most astringent compound was a dimer, the 4→6 linked catechin-catechin. Both the stereochemistry of the monomeric flavan-3-ol subunits and bonds linking them affects bitterness and astringency, however, the relative importance of the bond versus the subunit identity in influencing either sensory property cannot be estimated from this small number of samples.

REFERENCES

- 1 Kuhnau J, The flavonoids. A class of semi-essential food components: their role in human nutrition. *Wld Rev Nutr Diet* 24:117-191 (1976).
- 2 Kinnamon SC, Taste transduction: Linkage between molecular mechanisms and psychophysics. *Food Qual Pref* 7:153-159 (1996).
- 3 Spielman AI, Huque T, Whitney G and Brand JG, The diversity of bitter taste signal, in *Sensory Transduction*, Ed by Corey DP and Roper SD, Rockefeller University Press, NY, pp 307-324 (1992).
- 4 Spielman A, Nagal H, Sunavala G, Dasso M, Breer H, Boekhoff I, Huque T, Whitney G and Brand J, Rapid kinetics of second messenger production in bitter taste. *Am J Physiol* 270:C926-C927 (1996).
- 5 Lawless HT and Corrigan CJ Semantics of astringency, in *Olfaction and Taste*, Ed by Kurihara K, Suzuki N and Ogawa H, Springer-Verlag, Tokyo, pp 288-292 (1994).
- 6 Joslyn MA and Goldstein JL, Astringency of fruits and fruit products in relation to phenolic content, in *Advances in Food Research*, Ed by Chichester CO, Mrak EM and Stewart GF Academic Press, New York, pp 179-209 (1964).
- 7 Fischer U, The influence of ethanol, pH, and phenolic composition on the temporal perception of bitterness and astringency, and parotid salivation. MSc Thesis, University of California, Davis, CA (1990).
- 8 Noble AC, Bitterness and astringency in wine, in *Bitterness in Foods and Beverages*, Ed by Rousseff R Elsevier Scientific, New York, pp 145-158 (1989).
- 9 Bate-Smith EC and Swain T, Flavonoid compounds, in *Comparative Biochemistry*, Ed by Mason HS Florkin AM, Academic Press, New York, pp 755-809 (1962).

- 10 Thorngate JH III and Noble AC, Sensory evaluation of bitterness and astringency of 3R (–)-epicatechin and 3S(+)-catechin. *J Sci Food Agric* 67:531–535 (1995).
- 11 Robichaud JL and Noble AC, Astringency and bitterness of selected phenolics in wine. *J Sci Food Agric* 53:343–353 (1990).
- 12 Lea AGH and Arnold GM, The phenolics of ciders: bitterness and astringency. *J Sci Food Agric* 29:478–483 (1978).
- 13 Arnold RA, Noble AC and Singleton VL, Bitterness and astringency of phenolic fractions in wine. *J Agric Food Chem* 28:675–678 (1980).
- 14 Delcour JA, Vanderbergh MM, Corten PF and Dondeyne P, Flavor threshold of polyphenolics in water. *Am J Enol Vitic* 35:134–136 (1984).
- 15 Eastmond R, The separation and identification of a dimer of catechin occurring in beer. *J Inst Brew* 80:188–192 (1974).
- 16 Matysiak NL and Noble AC, Comparison of temporal perception of fruitiness in model systems sweetened with aspartame, acesulfameK blend or sucrose. *J Food Sci* 56:822–826 (1991).
- 17 SAS Institute, SAS System for Windows, Version 6.10, SAS Institute Inc, Cary, NC (1994).
- 18 Fischer U, Boulton RB and Noble AC, Physiological factors contributing to the variability of sensory assessments: relationship between salivary flow rate and temporal perception of gustatory stimuli. *Food Qual Pref* 5:55–64 (1994).
- 19 Ishikawa T and Noble AC, Temporal perception of astringency and sweetness in red wine. *Food Qual Pref* 6:27–33 (1995).
- 20 Noble AC, Bitterness in wine. *Physiol Behav* 56:1251–1255 (1994).
- 21 Kinnamon SC, Taste transduction: A diversity of mechanisms. *Trends Neurosci* 11:491–496 (1988).
- 22 Kumazawa T, Nomura T and Kurihara K, Liposomes as a model for taste cells: Receptor sites for bitter substances including N-C=S substances and mechanism of membrane potential changes. *Biochemistry* 27:1239–1244 (1988).
- 23 Bate-Smith E, Haemanalysis: The concept of relative astringency. *Phytochem* 11:907–912 (1973).
- 24 Peleg H and Noble AC, Perceptual properties of benzoic acid derivatives. *Chem Senses* 20:393–400 (1995).
- 25 McManus JP, Davis KG, Lilley TH and Haslam E, The association of proteins with polyphenols. *JCS Chem Comm.* 309–311 (1981).
- 26 Kallithraka S, Bakker J and Clifford MN, Evaluation of bitterness and astringency of (+)-catechin and (–)-epicatechin in red wine and in model solution. *J Sensory Studies* 12:25–37 (1997).
- 27 Porter LJ and Woodruffe J, Haemanalysis: The relative astringency of proanthocyanidin polymers. *Phytochemistry* 23:1255–1256 (1984).
- 28 Hagerman AE, Chemistry of tannin-protein complexation, in *Chemistry and Significance of Condensed Tannins*, Ed by Hemingway RW and Karchesy JJ, Plenum Press, New York, pp 323–333 (1989).
- 29 Haslam E, Proanthocyanidins, in *The Flavonoids: Advances in Research*, Ed by Harborne JB and Mabry TJ, Chapman and Hall, London, pp 417–447 (1982).
- 30 Arnold RA and Noble AC, Bitterness and astringency of grape seed phenolics in a model wine solution. *Am J Enol Vitic* 29:150–152 (1978).
- 31 Bartoshuk LM, The biological basis of food perception and acceptance. *Food Qual Pref* 4:21–32 (1993).
- 32 Hyde RJ and Pangborn RM, Parotid salivation in response to tasting wine. *Am J Enol Vitic* 29:87–91 (1978).
- 33 Noble AC, Application of time-intensity procedures for the evaluation of taste and mouthfeel. *Am J Enol Vitic* 46:128–133 (1995).
- 34 Spielman AI, Interaction of saliva and taste. *J Dent Res* 69:838–843 (1990).
- 35 Thorngate JH III, Flavan-3-ols and their polymers in grapes and wines: chemical and sensory properties. MSc Thesis, University California, Davis, CA (1992).

Article 7

7. Schlich P. & Chabanon C. (2000). L'analyse sensorielle appliquée au confort de freinage. *Revue Française de Marketing*, **179/180**, 129-142.

L'ANALYSE SENSORIELLE APPLIQUEE AU CONFORT DE FREINAGE¹

Pascal SCHLICH,
Inra, Laboratoire de Recherches sur les Arômes

Christian CHABANON
Renault/DIV/DSPV

INTRODUCTION ET RAPPEL DES PRINCIPES DE L'ANALYSE SENSORIELLE

Quels sont les aspects sensoriels qui génèrent les préférences des consommateurs et comment établir une recette sensorielle conduisant à un produit idéal pour un groupe ciblé de consommateurs ? L'analyse sensorielle de produits alimentaires postule qu'en situation de dégustation à l'aveugle d'un ensemble de produits concurrents, un consommateur peut dire ce qu'il préfère, mais rarement pourquoi. Nous postulons que la sensation de confort de freinage peut rentrer dans ce schéma. Je sais que j'aime ou que je n'aime pas le freinage d'une voiture que j'essaye, mais je serais bien incapable de décrire et de quantifier pourquoi.

D'où la démarche de l'analyse sensorielle qui consiste à sélectionner un groupe restreint de sujets et à l'entraîner à décrire ces produits de manière quantitative à l'aide d'un bon nombre de " descripteurs sensoriels " (par exemple une vingtaine), la qualité de la mesure ayant fait l'objet de tests appropriés [Sch94a]. Les méthodes de cartographie des préférences ([Sch92], [Sch95a]) modélisent alors les préférences subjectives ou " hédoniques " d'un échantillon de consommateurs appelés parfois " naïfs ", représentatif d'une population cible, à l'aide des profils sensoriels objectifs établis par les seuls sujets entraînés appelés parfois " experts ".

A côté de la description des propriétés sensorielles des produits, l'intérêt primordial de l'analyse sensorielle est d'expliquer les préférences des

consommateurs. Cette problématique intéresse aujourd'hui vivement l'industrie agro-alimentaire, mais il est clair que toute industrie de produits personnels, voire de services, est concernée. Une application au confort de freinage en essai automobile est présentée dans cet article. Son objectif est de déterminer dans quelle mesure nous pouvons utiliser les descriptions des " experts " essayeurs pour modéliser des données de préférence de consommateurs " naïfs ".

HISTORIQUE

Dans une première phase, de 1994 à 1996, Renault et l'Inra ont collaboré dans un projet de transfert des techniques de l'analyse sensorielle des aliments vers l'essai automobile. Le modèle retenu fut le confort de freinage, sachant que tout autre aspect de l'essai automobile mettant en jeu la perception sensorielle de l'être humain aurait pu être choisi.

L'essai automobile par des pilotes mécaniciens est, depuis toujours, une activité classique chez les constructeurs automobiles. Toutefois, cette activité relève davantage de l'expertise que de la métrologie sensorielle [Law98], dans la mesure où il n'y a pas

(1) —

Cet article a déjà fait l'objet d'un exposé [Sch98] aux *journées Modulab* organisées par Renault et le CISIA en novembre 1998 sous le titre " *La cartographie des préférences ou l'optimisation du plaisir subjectif à l'aide d'un réglage sensoriel objectif*". Quelques modifications ont cependant été faites pour en faciliter la compréhension.

constitution et entraînement d'un groupe à noter des intensités sensorielles plutôt que des préférences subjectives. L'analyse sensorielle propose donc une " revalorisation " du métier d'essayeur, car en codifiant la pratique de ces essais au moyen des concepts de plans d'expérience, d'étalonnage et de répétabilité, elle confère aux activités de ce métier une lisibilité et une traçabilité favorisant sa transmissibilité.

Les plans d'expérience de l'analyse sensorielle ([Mac89], [Sch93], [Wak95], [Cal97]), qui permettent d'équilibrer l'ordre de dégustation des produits afin que chaque produit soit dégusté un même nombre de fois dans chaque ordre de service et un même nombre de fois après chaque autre produit, ont été utilisés afin d'estimer la présence et l'importance de ces effets d'ordre en essai automobile. Un vocabulaire de dégustation décrivant les sensations de freinage a été établi sous la forme d'une liste de 18 " descripteurs " notés sur une échelle non structurée (segment libellé aux deux extrémités, sur lequel le sujet porte une croix qui donnera lieu à une note de 0 à 100). La qualité de la mesure sensorielle a été contrôlée au moyen de répétitions et à l'aide de techniques statistiques développées en analyse sensorielle pour ce problème [Sch94a]. Nos travaux se sont achevés, dans le cadre de cette première phase, avec l'établissement et la validation [Sch96] d'une carte sensorielle du freinage dans laquelle 32 véhicules du marché français ont pu être positionnés. Cette carte nous a permis de comprendre les dimensions sensorielles principales mises en jeu dans la perception de la sensation de freinage. Pour des raisons de confidentialité, cette carte ainsi que la liste des descripteurs du freinage ne peuvent être dévoilées. A titre d'exemple, on peut tout de même indiquer que ces descripteurs traitaient de notions physiques telles que force et course, mais aussi de sensations telles que mordant et sécurité et, enfin, d'équilibre entre intensité physique et niveau de sensation.

Dans une deuxième phase, en 1997, le volet consommateur de cette étude a été réalisé avec 112 personnes (personnel Renault), qui ont essayé 8 véhicules pour leur donner une note " hédonique " : Mégane 2,0 - Laguna 2,0s - Laguna V6 - Scénic 1,6^e - Audi A4 - Safrane 2,2dt - Safrane 2,5dt - Espace. De plus, 9 pilotes ont jugé les mêmes véhicules en utilisant les 18 " descripteurs sensoriels " obtenus dans la phase précédente.

Nous allons maintenant décrire cette deuxième phase, en précisant les conditions de cette expérimentation, puis les méthodes utilisées et enfin les résultats obtenus.

EXPÉRIMENTATIONS

Les 112 personnes ont donc roulé successivement (2 séances de 4 véhicules par personne) dans 8 véhicules selon un plan d'expérience en deux systèmes de 7 carrés latins 8x8 mutuellement orthogonaux. Ce plan assure que chaque véhicule sera testé 14 fois en 1^{ère}, 2nd, ... et 8^{ème} position et que chaque véhicule précédera 14 fois chaque autre véhicule, 12 fois chaque autre véhicule à l'ordre 2 (c'est-à-dire avec un autre véhicule entre les eux), 10 fois à l'ordre 3 (2 véhicules entre eux), ..., 2 fois à l'ordre 7 (6 véhicules entre eux). Un tel plan équilibre tout effet de report (*carry-over effect* dans la littérature des plans d'expérience) qui pourraient exister entre les notes successives données par un même sujet, par exemple une sous estimation, par effet de contraste, de l'intensité d'un véhicule essayé juste après un véhicule très intense.

A l'issue de chaque essai, un certain nombre de questions étaient posées aux sujets dont l'une portait sur le confort de freinage qui devait être évalué sur une échelle non structurée de 0 à 10 allant de " Je n'aime pas du tout " (0) à " J'aime beaucoup " (10). Ce travail utilise uniquement cette note comme note de préférence.

D'autre part, les consommateurs ont aussi rempli un questionnaire général leur demandant leur sexe, leur taille, leur poids, la fréquence d'utilisation de leur voiture en général, en ville, sur route et sur autoroute et enfin le type de conduite (sportive, rapide ou familiale) qu'ils pensent avoir.

En parallèle, 9 pilotes ont réalisé le profil sensoriel de ces huit véhicules à l'aide de la liste des 18 descripteurs. En général, les pilotes ont testé une seule fois chaque véhicule, mais en raison de la disponibilité des véhicules, certains pilotes n'ont pas vu certains véhicules, alors que d'autres ont pu tester deux fois le même véhicule. Au total, 69 profils sensoriels ont été réunis.

MÉTHODES

Dans un premier temps, les données de profil sensoriel seront analysées par analyse de la variance et les résultats présentés sous la forme d'une table Flash (tableau 1) [Sch95b]. L'analyse multidimensionnelle de ces données sera conduite par une analyse canonique discriminante de l'effet véhicule dans le modèle additif de Manova [Mar79]: véhicule+sujet, où les variables explicatives sont les 18 descripteurs de freinage, modèle calculé à partir des évaluations moyennes de chaque sujet pour chaque véhicule. L'intérêt de cette analyse est d'une part d'établir une carte sensorielle des véhicules (Figure 1), mais d'autre part de tester statistiquement le nombre d'axes significatifs de cette carte (option Canonical de l'ordre Manova de la procédure GLM de SAS/STAT®). Enfin, on superposera sur cette carte les ellipses de confiance au seuil 90 % de la position moyenne de chaque véhicule. Ce procédé permet d'évaluer visuellement le niveau de significativité des distances entre véhicules sur la carte sensorielle. On effectuera également, sur les profils sensoriels, l'analyse la plus courante, à savoir une analyse en composante principale (ACP) non normée du tableau des moyennes des véhicules pour les 18 descripteurs. La carte sensorielle obtenue (figure 2) est traditionnellement l'entrée sensorielle dans les techniques de cartographie des préférences.

L'analyse des données de préférence sera réalisée par cartographie interne des préférences (figure 3), qui est essentiellement une ACP non normée du tableau *véhicule x consommateur* des notes de préférences. Une classification des consommateurs sera conduite sur la base de leurs seules notes de préférence par la procédure Varclus de SAS/STAT® (option Centroid). Il s'agit d'une méthode de classification de variables en ACP, où le noyau d'une classe est la moyenne des consommateurs de la classe, l'allocation

des consommateurs aux classes étant réalisée de telle sorte que la somme des variances des noyaux soit maximale. L'algorithme de partition est hiérarchique descendant, avec une possibilité d'échange d'individus qui fait qu'aucune relation d'inclusion d'une partition fille dans la partition mère n'est imposée. Les partitions en 2, 3 et 4 segments (tableau 2) seront examinées afin de découvrir quels sont les véhicules qui segmentent le plus facilement les préférences de notre population de consommateurs.

Le lien entre le profil sensoriel et les préférences des consommateurs sera étudié à l'aide des deux modèles les plus classiques de la cartographie externe des préférences : le modèle vectoriel (figure 4) et le modèle circulaire (figure 5). Le premier détermine, sur la carte sensorielle de la figure 2, la direction de préférence croissante de chaque consommateur, alors que le second localise sur cette carte un point correspondant à un véhicule idéal pour chaque consommateur. Ces modèles souffrent de quelques limites assez claires dans cet exemple, ainsi nous présenterons l'application d'une technique nouvelle (en cours de publication), dénommée PrefMaX, qui conduit à une carte des optimums sensoriels des consommateurs et à une segmentation de ces consommateurs sur cette base. La technique permet alors d'établir un " réglage sensoriel " optimal des descripteurs, soit pour le groupe de consommateurs tout entier, soit pour chacun des groupes identifiés dans l'étape de segmentation (tableau 3).

Enfin, une caractérisation des groupes de consommateurs issus du PrefMaX à l'aide des variables du questionnaire consommateur a été tentée, mais on verra qu'elle n'a pas donné de résultats intéressants.

Le tableau 0 ci-après devrait aider le lecteur à naviguer dans ce papier.

Tableau 0			
Synoptique du papier			
Données en entrée	Problème	Analyse	Résultats en sortie
X_{69,18} 69 essais de 8 véhicules par 9 pilotes notant l'intensité de 18 descripteurs du freinage	Structure de corrélation des descripteurs et différences entre véhicules	Flash : classification Varclus des descripteurs, Anova et tests de contraste contre la moyenne	Tableau 1
	Carte sensorielle des véhicules avec zones de confiance et test du nombre d'axes significatifs	Analyse canonique de l'effet véhicule (en MANOVA véhicule + pilote) dans l'espace des 18 descripteurs	Figure 1
	Carte sensorielle des véhicules (1)	ACP non normée du tableau des 8 moyennes des véhicules pour les 18 descripteurs	Figure 2
Y_{8,112} Notes de préférence pour chacun des 8 véhicules de 112 consommateurs	Carte des préférences des véhicules	Cartographie interne des préférences. ACP non normée du tableau Y	Figure 3
	Segmentation des consommateurs sur leurs préférences	Classification Varclus (option Centroid) de SAS®, en 2, 3 ou 4 groupes de consommateurs	Tableau 2
Y_{8,112} et X_{69,18} Les colonnes de Y jouent le rôle de variables à expliquer séparément par les colonnes de X ou par les dimensions sensorielles principales de X	Directions de préférence optimales dans la carte sensorielle (1)	Cartographie externe vectorielle des préférences. Régression linéaire multiple par consommateur	Figure 4
	Localisation de points idéals dans la carte sensorielle (1)	Cartographie externe circulaire des préférences. Régression quadratique par consommateur	Figure 5
	Segmentation des consommateurs sur la base de leur véhicule idéal	PrefMaX	Figure 6
	Niveaux optimaux des descripteurs par groupe de la segmentation précédente	Réglage sensoriel du PrefMaX	Tableau 3

RÉSULTATS

L'espace sensoriel du freinage

Analyse Flash de la variance

Les notes du bas du tableau 1 (voir page suivante) expliquent la technique. Elle a été mise au point pour permettre une présentation claire et synthétique des résultats d'une batterie d'analyses de la variance et des comparaisons de moyennes associées. Bien entendu, nous simplifions le problème en ne comparant chaque moyenne qu'à la moyenne générale, mais c'est le prix à payer pour gagner en clarté.

L'examen de ce tableau révèle une importante première dimension de différenciation du freinage des véhicules qui oppose l'Audi A4, la Mégane 2.0 et la Laguna 2.0s, d'une part, à l'Espace et à la Safrane 2.2dt, d'autre part. Cette dimension reçoit une interprétation sensorielle claire (du moins lorsque l'identité des descripteurs n'est pas cachée) en terme de freinage : opposition des descripteurs 18, 1, 14, 12, 4, 5 et 8 aux descripteurs 13, 3, 2 et 6, d'une part, mais aussi aux descripteurs 10 et 11, d'autre part. Le descripteur 14 est particulièrement intéressant car il induit de plus une opposition de la Scenic 1.6e à la Laguna V6.

On note que les descripteurs 15, 7 et 16 ne présentent aucune différence significative entre nos

véhicules. Le lecteur doit se souvenir que la liste de véhicules a été générée sur la base de 32 véhicules et non pas seulement des huit de cette étude. Nous aurions pu retirer ces descripteurs des analyses suivantes sans que cela change quoi que ce soit aux résultats. On notera aussi que ces trois descripteurs sont ceux ayant une petite moyenne (à peine de l'ordre de 10 sur 100), ils correspondent à un aspect sensoriel dont l'intensité est trop faible pour être bien perçu par les sujets.

Présenter une table Flash avec les noms des descripteurs cachés est une tâche assez ingrate, mais lors de la phase d'étalonnage avec 32 véhicules cette technique s'est avérée très utile pour intégrer et synthétiser cette volumineuse information.

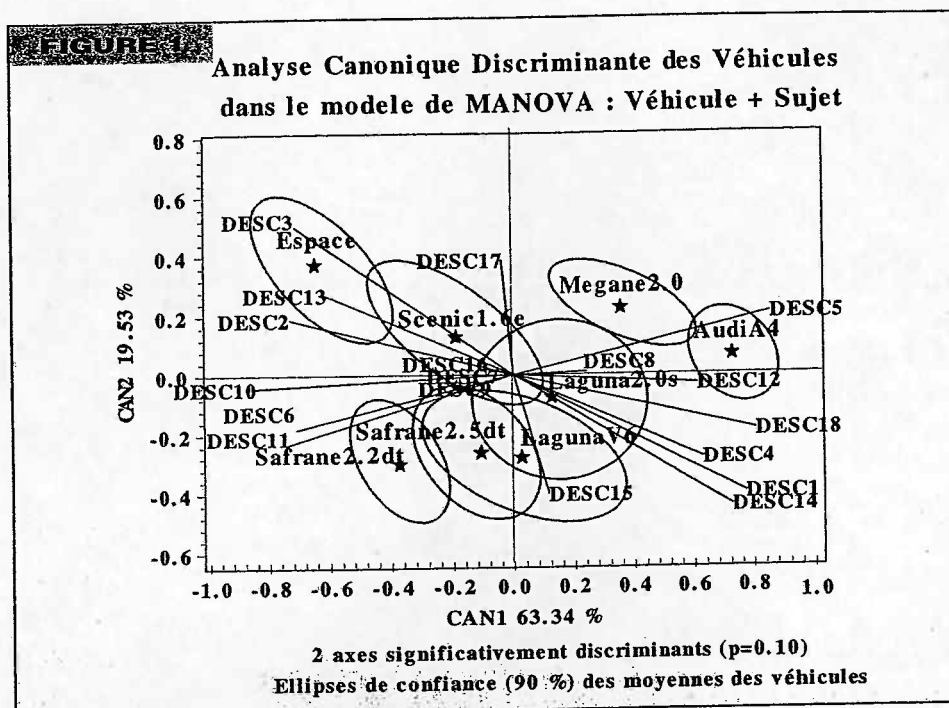
Analyse canonique discriminante

Cette analyse (figure 1, page suivante) est la contrepartie multidimensionnelle de nos analyses de la variance. Elle diffère de l'analyse factorielle discriminante (AFD) classique dans le sens où ce sont les mêmes sujets qui ont évalué les différents véhicules. Toutefois, les axes qu'elle nous donne sont les mêmes que ceux que l'on obtiendrait après centrage des descripteurs par sujet et application de l'AFD. Mais les tests de MANOVA que donneraient ces logiciels seraient alors incorrects au niveau des degrés de liberté.

La figure 1 nous apprend que les différences significatives de perception du freinage entre les véhicules " vivent " dans un espace bidimensionnel.

Analyse Flash des profils sensoriels des véhicules				Espace	Safrane 2.2 dt	Safrane 2.5 dt	Scenic 1.6e	Laguna V6	Laguna 2.0s	Megane 2.0	Audi A4
Descripteur	F	Prob	Moyenne								
DESC18	18,5	0,0000	69	49-	58-	63-			77+	77+	86+
DESC1	14,1	0,0000	68	46-	58-				76+	75+	82+
DESC14	15,4	0,0000	61	37-	53-		48-	70+	72+	68+	76+
DESC12	5,54	0,0000	61	47-	50-					74+	75+
DESC4	6,68	0,0000	63	42-					74+		77+
DESC5	22,9	0,0000	29	10-	6-	11-				53+	69+
DESC8	1,82	0,1070	17							30+	
DESC15	0,72	0,6590	6								
DESC7	0,85	0,5520	9								
DESC9	1,36	0,2460	10				19+				
DESC17	4,99	0,0000	50	62+	43-			37-	43-		
DESC13	10,2	0,0000	41	63+	50+			21-	24-		29-
DESC3	13,3	0,0000	37	71+			49+		27-		21-
DESC2	12,3	0,0000	28	50+	40+			19-		19-	12-
DESC6	13,7	0,0000	37	54+	67+	57+			24-	13-	14-
DESC10	19,4	0,0000	47	66+	60+			62+		30-	20-
DESC11	13,8	0,0000	42	54+	52+			59+		29-	22-
DESC16	0,39	0,9020	12								

Les descripteurs sont classés en groupes homogènes par la procédure Varclus option Centroid de SAS
Le nombre de groupes est choisi afin que les composantes des groupes expliquent au moins 70 % de la variance totale
Le premier groupe est le plus important en terme de variance globale restituée
Le second est le mieux corrélé positivement au premier, le troisième au second, et ainsi de suite...
Au sein d'un groupe, les descripteurs sont triés du meilleur au moins bon représentant du groupe
Les véhicules sont triés par moyenne croissante du premier descripteur du tableau
F et Prob sont le test F de l'effet Véhicule et son risque associé, dans l'Anova Véhicule + Sujet
La moyenne d'un véhicule est imprimée seulement si elle est significativement ($p=0.05$) supérieure (+) ou inférieure (-) à la moyenne générale



Cette carte restitue donc l'essentiel des différences entre véhicules. L'axe horizontal pèse pratiquement pour deux tiers du total des différences et oppose exactement, et heureusement, les mêmes descripteurs que ceux soulignés par la table Flash. Notons que les descripteurs sont représentés sur cette carte à l'aide de leur coefficient de corrélation totale (calculée sur la base des 69 observations) avec les axes de l'analyse. La flèche d'un descripteur indique juste sa direction d'intensité croissante, sans pour autant récupérer les propriétés de *biplot* comme en ACP.

L'ellipse de confiance d'un véhicule nous indique la zone où la moyenne de ce véhicule avait 90 % de chance d'être observée. Cette information graphique est précieuse pour apprécier visuellement la significativité des distances entre véhicules. Par exemple, les deux Safrane semblent légèrement différentes dans l'espace du freinage alors que la Laguna V6 et la Safrane 2.5dt semblent similaires. La surface des ellipses est inversement proportionnelle à l'accord des sujets sur le véhicule ; en ce sens, les deux Laguna semblent induire le plus de désaccord.

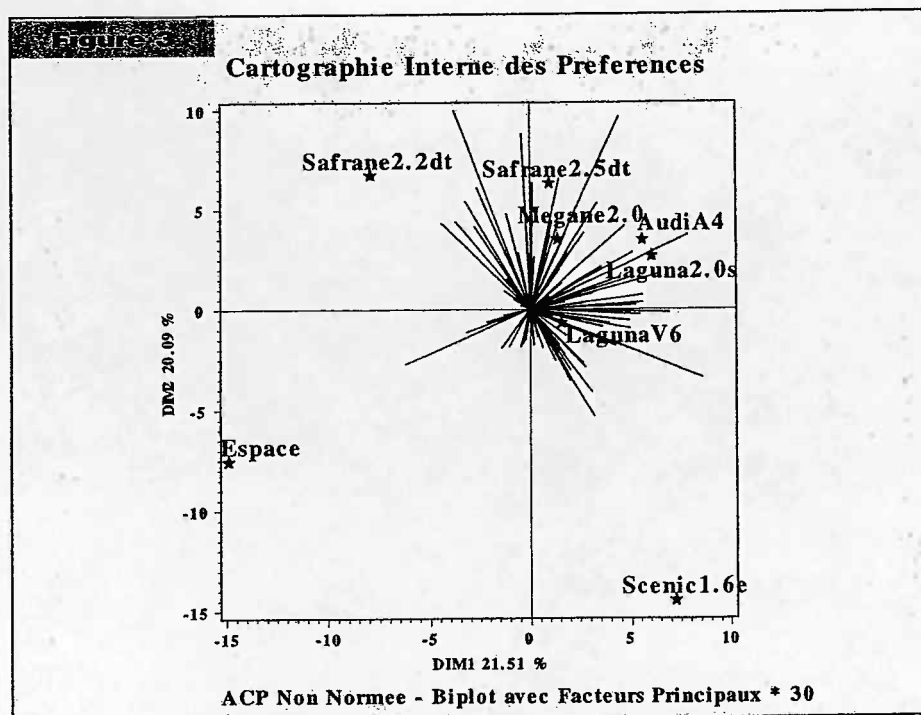
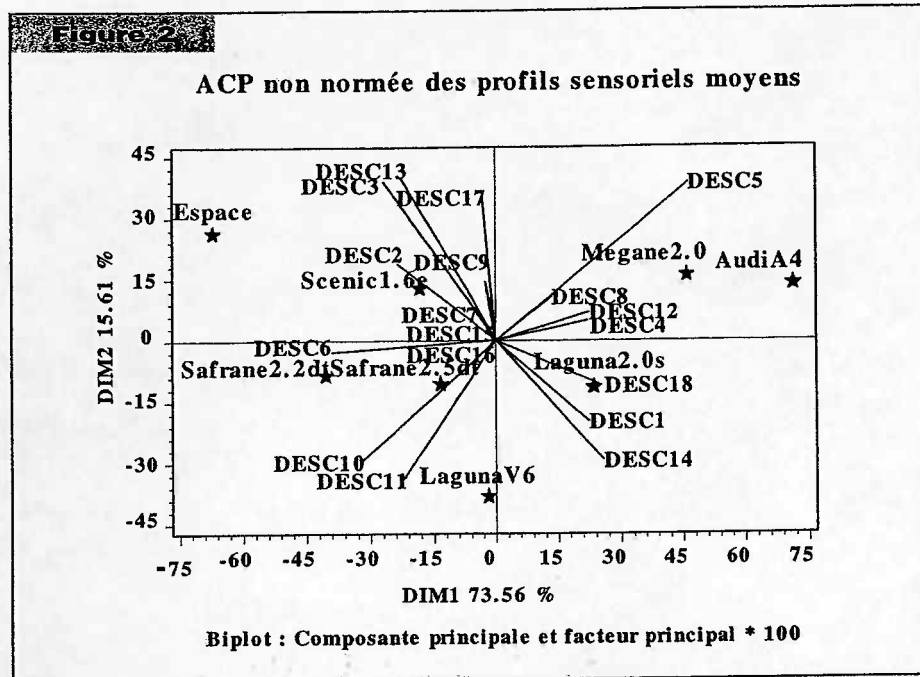
Cette figure distingue de manière claire certains véhicules. D'abord l'Espace, qui s'oppose aux autres véhicules sur le gradient allant des descripteurs 14, 1, 4 et 18 aux descripteurs 3, 13, et 2 et ensuite l'Audi A4 qui reste la plus marquée positivement par le gra-

dient de l'axe 1. Il convient de noter que de tous ces véhicules Renault, c'est la Mégane 2.0 qui a un freinage le plus similaire à celui de l'Audi A4.

Analyse en composantes principales

Dans la pratique courante des laboratoires d'analyse sensorielle, l'analyse canonique discriminante n'est pas encore très répandue et c'est l'ACP du tableau des moyennes des véhicules qui aurait été mise en œuvre dans 95 % des cas pour dresser la carte sensorielle des véhicules. Nous recommandons fortement d'effectuer cette ACP à partir de la matrice de covariance et non à partir de celle de corrélation (ACP non normée plutôt que normée), sinon les descripteurs non significatifs prendraient une importance qu'ils ne méritent pas.

La figure 2 (voir page suivante) est le *biplot* résultant de cette analyse. Il est clair que sa structure est très similaire à celle de la figure 1. Toutefois, quelques différences légères apparaissent, notamment sur le positionnement de la Laguna V6 qui semble maintenant bien différenciée dans le bas de la figure, alors que les deux Safrane se sont quelque peu éloignées des Laguna. L'interprétation en terme de freinage reste globalement la même que celle de la figure 1.



Lorsque les données sensorielles sont de bonne qualité, on constate généralement que ces deux analyses conduisent aux mêmes conclusions. Dans le cas contraire, l'analyse canonique peut donner des répartitions moins uniformes des véhicules, car elle pénalise ceux dont la note moyenne est entachée d'une plus grande variance et, au contraire, elle surpondère ceux remportant un fort accord entre les sujets.

LES PRÉFÉRENCES EN CONFORT DE FREINAGE

Cartographie interne des préférences

La figure 3 est la carte des préférences établie à partir de la note d'appréciation du freinage des 112 consommateurs. En réalité, seuls 110 consommateurs ont participé à cette carte, car deux d'entre eux avaient des notes constantes. Dans ce *biplot*, chaque consommateur est représenté par une flèche dirigée dans sa direction de préférence croissante. Ainsi, deux véhicules proches sont appréciés en général par les mêmes personnes. La longueur d'une flèche

mesure l'importance des différences de préférence exprimée par le consommateur correspondant. On note que cet aspect est hautement variable d'un consommateur à l'autre. Il peut s'agir d'un artefact d'utilisation de l'échelle de notation (auquel cas une ACP normée aurait été préférable), comme d'une absence de préférence. Nous observons plus souvent la seconde situation, c'est pourquoi nous appliquons systématiquement une ACP non normée.

Un effet de taille apparaît clairement le long de la première bissectrice distinguant l'Espace comme le véhicule dont le freinage est de loin le moins apprécié de tous. La Scenic a une position segmentante : les consommateurs pointant vers le haut de la figure n'apprécient pas le freinage de ce véhicule.

Il convient de noter que ce plan ne restitue que 42 % de l'information de l'espace des préférences qui comporte 7 axes au total. Il reste pourtant la meilleure carte des préférences et, comme l'interprétation de plusieurs cartes simultanément est toujours délicate, nous allons utiliser un algorithme de classification des consommateurs pour mieux saisir les traits principaux des préférences.

Tableau 2				Classification des consommateurs sur la seule base de leurs notes de préférence							
Classe	N	F	Moyenne	Espace	Safrane2.2dt	AudiA4	Scenic1.6e	LagunaV6	Megane2.0	Safrane2.5dt	Laguna2.0s
P1C1	110	8,6	7,88	6,95-						8,25+	8,35+
P2C1	90	11	7,93	7,02-	7,61-		8,57+				8,29+
P2C2	20	14	7,67	6,66-	8,43+		4,72-			8,69+	8,60+
P3C1	31	15	7,51	5,52-		8,11+	8,45+	6,85-		8,43+	
P3C2	19	14	7,61	6,62-	8,35+		4,56-			8,62+	8,52+
P3C3	60	7,8	8,16	7,80-	7,73-	7,64-	8,62+	8,54+			8,79+
P4C1	30	8,1	7,93	7,24-			8,40+	7,14-	8,36+	8,71+	7,21-
P4C2	17	13	7,59	6,67-		8,44+	4,44-			8,64+	8,37+
P4C3	45	21	7,88	6,50-	7,06-	8,36+	8,51+	8,50+	7,53-		8,82+
P4C4	18	16	8,09			5,37-	8,64+		8,91+		9,02+

PiCj désigne la classe j de la partition en i classes. N est l'effectif de chaque classe
 F est le test de Fisher de l'effet véhicule dans l'analyse de la variance Véhicule + Consommateur
 Une moyenne d'un véhicule est imprimée si elle est significativement ($p=0.05$) supérieure (+)
 ou inférieure (-) à la moyenne générale.
 Pseudo F des classifications : P2=18.23 P3=15.72 P4=16.34

La première ligne du tableau 2 montre que sur l'ensemble de la population les véhicules Laguna 2.0s et Safrane 2.5dt ont des freinages légèrement préférés, alors que celui de l'Espace est clairement moins apprécié que la moyenne.

La segmentation la plus patente de la population est obtenue en isolant un groupe de 20 personnes (P2C2) qui n'aiment pas du tout le freinage de la Scenic, alors qu'en moyenne les 90 autres personnes (P2C1) classent ce véhicule premier en terme de freinage. Ce petit groupe critique vis-à-vis de la Scenic est toujours présent à l'étape suivante de classification en 3 groupes (alors que l'algorithme permet tout type de transfert). Par contre, nous distinguons maintenant un groupe de 60 personnes (P3C3) qui apprécie moins l'Audi A4 et davantage la Laguna V6 et un groupe de 30 personnes (P3C1) dont les préférences sont opposés au précédent. Avec 4 classes, on conserve toujours un groupe de " rejeteurs " de la Scenic et l'on voit apparaître un groupe de personnes aimant moins le freinage des deux Laguna (P4C1).

Choisir le nombre de groupes est toujours délicat. Ici, le Pseudo F nous conduirait à retenir 4 classes, mais les effectifs sont vraiment trop faibles pour avoir confiance en des groupes si petits. On admet en général que la constitution de groupes de préférence basés sur moins de 50 individus en moyenne par groupe est dangereuse. Ici, avec 112 consommateurs, une segmentation en plus de deux groupes peut donc être dangereuse. De plus, nous désirons vraiment effectuer cette classification sur la base des corrélations aux aspects sensoriels plutôt que sur la base des seules notes de préférence. C'est l'objet du chapitre suivant.

L'OPTIMISATION SENSORIELLE DU CONFORT DE FREINAGE

Cartographie externe vectorielle des préférences

La carte des véhicules de la figure 4 (page suivante) est exactement la carte sensorielle de la figure 2 dans laquelle les consommateurs sont projetés sous la forme d'une flèche indiquant leur préférence croissante. Mais, par rapport à la cartographie inter-

ne du paragraphe précédent, il faut bien comprendre que si cette projection est bonne (longueur de la flèche suffisante), il existe une corrélation entre préférences et différences sensorielles. Autrement dit, une longue flèche représente un consommateur pour qui le véhicule idéal serait un véhicule le plus éloigné possible dans la direction sensorielle indiquée par la flèche. Au vu de la figure 4, la tendance générale des préférences serait donc dans la direction des deux Laguna et les meilleurs " drivers " sensoriels de la préférence seraient ainsi les descripteurs 14, 1, 11, 10 et 18.

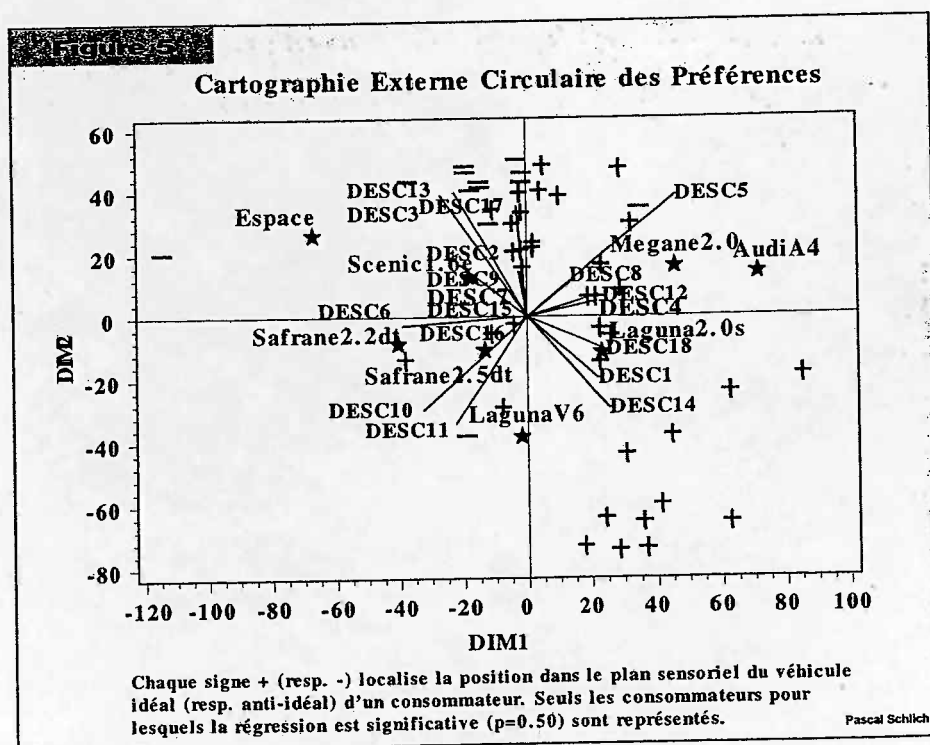
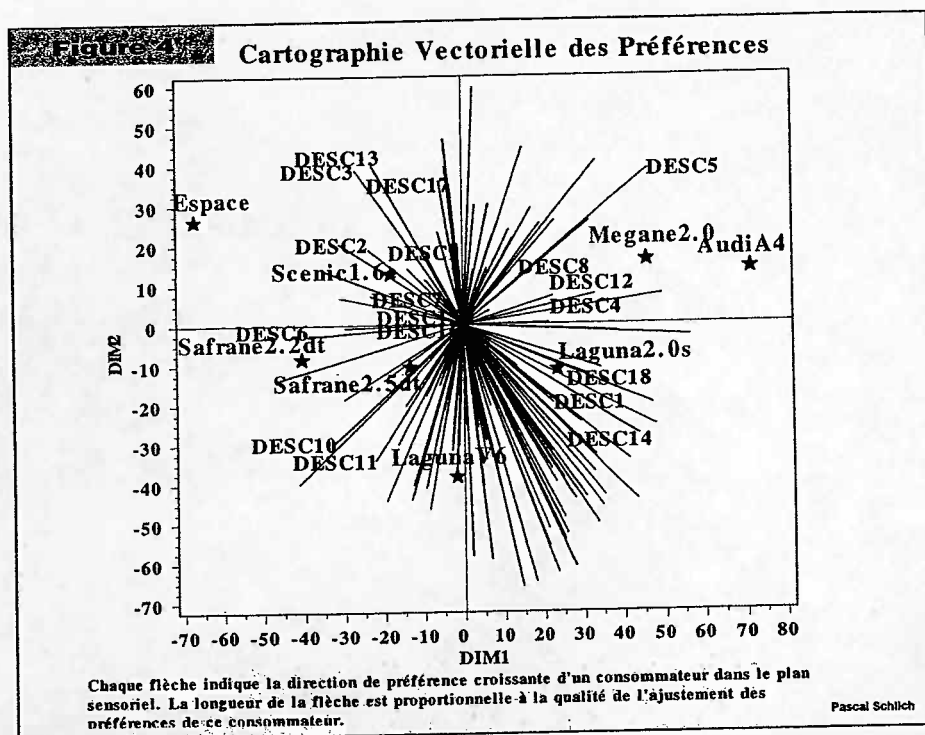
Toutefois, la qualité de ces ajustements est médiocre comme en témoigne la moyenne des R^2 (rapport de la variance du modèle à la variance totale) qui s'établit à 0,297, avec certes un maximum à 0,854 mais aussi avec un minimum à 0,002. Une fois de plus, la variabilité interindividuelle est très forte dans cet aspect.

Le modèle vectoriel, surnommé '*the more, the better*', est limité, dans la mesure où il suppose qu'il faut maximiser les meilleurs descripteurs et minimiser les moins bons pour réaliser le meilleur produit. La réalité n'est pas si simple et souvent, des intensités moyennes peuvent s'avérer être les meilleures. Ceci constitue la justification du modèle circulaire.

Cartographie externe circulaire des préférences

La carte sensorielle des véhicules de la figure 5 (page suivante) est par construction la même que celle de la figure 4. Mais, comme le modèle d'ajustement des notes de préférence de chaque consommateur inclut en plus un terme quadratique, la surface de réponse associée est un paraboloïde situé au-dessus du plan sensoriel. Sa forme peut être concave ou convexe et donc mettre en évidence dans la carte sensorielle, soit un point idéal, soit un point anti-idéal.

Seuls les consommateurs pour lesquels la qualité de l'ajustement semblait commencer à battre le hasard ($p=0,50$) ont été retenus. Il est troublant de constater que, même avec une règle de sélection aussi peu sévère, seule une petite moitié des consommateurs sont sélectionnés. De plus, ces



consommateurs ont, pour la plupart, une allure vectorielle, c'est-à-dire soit un idéal positif assez loin dans la direction indiquée par le descripteur 14, soit des idéaux négatifs loin dans la direction opposée (ce qui revient au même). Seul, un petit groupe de 7 consommateurs ayant un idéal positif plus ou moins sur le triangle Laguna 2.0s, Megane 2.0 et Audi A4, apportent une information neuve.

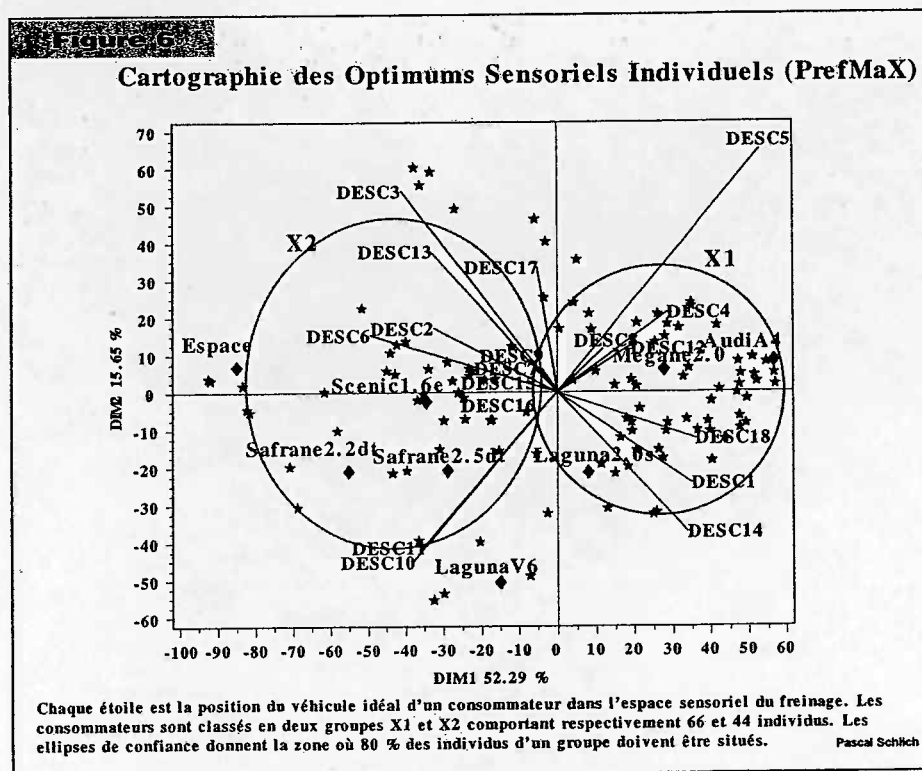
Ces limites des techniques classiques de cartographie des préférences ont été discutées (Schlich 1995). D'autres modèles (elliptique et quadratique) existent mais ne donnent pas de meilleurs résultats sur ces données. Le petit nombre de véhicules est un facteur limitant pour la mise en œuvre de modèles trop compliqués, car à la fin on ne sait plus si le bon ajustement obtenu relève de la chance ou d'une véritable corrélation.

PrefMaX

Cette technique (en cours de publication) est une réponse aux limitations des techniques classiques que nous venons d'expérimenter sur nos don-

nées de freinage. Sans vouloir rentrer ici dans les détails techniques de cette méthode, disons qu'elle localise et factorise directement les idéaux de chaque consommateur (forcément positifs). Le poids de chaque consommateur dans les calculs est proportionnel à l'ajustement des préférences de ce consommateur à l'espace sensoriel, ce qui règle de manière douce le problème des consommateurs mal ajustés (pondération au lieu de sélection). Le résultat de cette méthode est une carte sensorielle (figure 6), dans laquelle chaque consommateur est localisé par un point exprimant le niveau optimal des descripteurs pour ce consommateur. Les véhicules peuvent alors être localisés dans cette carte sur la base de leurs notes moyennes d'intensité pour les descripteurs. Ainsi, la position d'un véhicule correspond à un consommateur fictif pour lequel cette voiture aurait un freinage absolument idéal.

La méthode inclut une phase de segmentation des consommateurs qui, dans notre cas, a recommandé une partition en deux groupes de 44 et 66 consommateurs respectivement. On note que l'on obtient, par cette méthode, deux groupes d'effectifs raisonnables et dont l'interprétation en terme de



véhicule est claire. Le groupe X1 de 66 consommateurs est bien compact et rassemble les gens pour qui l'idéal sensoriel du freinage est localisé sur le triangle Audi A4, Mégane 2.0 et Laguna 2.0s. A l'opposé, les 44 personnes du groupe X2 préfèrent dans leur majorité le freinage de la Scenic. Toutefois, ce groupe est moins nombreux et moins compact que le précédent, il ramasse aussi les très rares personnes préférant le freinage de l'Espace. La Laguna V6, qui se distinguait dans le modèle vectoriel, ne ressort pas dans cette analyse comme un pôle de forte attraction des préférences. Il convient de noter que l'origine de cette carte correspondrait à un véhicule fictif maximisant les préférences sur l'ensemble du panel (le tableau 3 donnera les coordonnées exactes

de ce point en fonction des descripteurs). Il est remarquable que très peu de consommateurs soient en réalité proches de l'origine et donc que ce véhicule optimal pour tous remporterait finalement peu de préférence absolue au niveau individuel.

Par rapport à la segmentation des consommateurs effectuée sur la seule base de leurs notes de préférence, cette technique exhibe des groupes qui, s'ils ont un sens, doivent pouvoir s'interpréter d'un point de vue sensoriel. Précisément, la technique permet d'établir, pour la population toute entière et pour chacun des groupes de consommateurs, un réglage sensoriel optimal pour chacun des descripteurs (tableau 3)

Statistique de qualité du PrefMax et réglage optimal des descripteurs									
Descripteur	R ² Moyen	Sensibilité	F inter segment	110 sujets		X1 : 66 sujets		X2 : 44 sujets	
				Moyenne	Err. Std.	Moyenne	Err. Std.	Moyenne	Err. Std.
DESC1	0,344	1,2	210,2	72,64	0,92	78,32	0,66	61,89	0,93
DESC18	0,336	1,2	190,5	73,60	0,93	79,41	0,69	63,26	0,95
DESC2	0,347	1,2	163,5	23,39	0,93	17,79	0,77	33,55	0,92
DESC14	0,339	1,1	133,9	65,92	1,03	71,29	0,87	54,04	1,16
DESC4	0,315	1,1	112,7	67,04	0,98	72,79	0,80	57,76	1,24
DESC5	0,263	0,8	108,5	41,02	1,67	52,38	1,56	27,56	1,84
DESC3	0,347	0,9	104,0	33,56	1,42	25,87	1,31	47,60	1,62
DESC10	0,294	1,0	91,00	39,72	1,19	31,62	1,18	48,48	1,34
DESC13	0,357	1,0	77,07	33,36	1,09	29,33	0,96	46,79	1,73
DESC11	0,263	1,1	76,13	37,53	1,23	28,14	1,44	44,73	1,24
DESC12	0,292	1,5	75,14	64,42	0,71	68,04	0,73	58,13	0,83
DESC6	0,295	0,8	64,21	34,19	1,80	26,57	1,97	51,49	1,88
DESC8	0,271	1,8	50,23	20,77	0,62	23,54	0,59	16,01	0,94
DESC9	0,295	2,9	8,89	9,09	0,56	7,65	0,68	10,89	0,89
DESC17	0,317	1,8	3,48	45,56	0,66	44,68	0,70	47,26	1,32

Dans le tableau 3, le R² moyen est une mesure de confiance dans la détermination du réglage optimal de chaque descripteur au niveau du groupe tout entier. Dans cet exemple, ces valeurs ne sont de fait pas très élevées et l'on n'observe pas de grosses variations d'un descripteur à l'autre. La sensibilité d'un descripteur est un taux d'accroissement des préférences imputable à une unité de variation d'in-

tensité du descripteur (rapportée à la même échelle que celle des préférences). Ce critère varie davantage et fournit de précieuses indications sur les descripteurs qu'il faudra contrôler en première priorité. Le F entre les segments mesure l'importance du descripteur pour discriminer les groupes X1 et X2 au niveau du réglage sensoriel. Dans ce tableau, les descripteurs sont triés par valeurs décroissantes de

ce critère. Les colonnes suivantes donnent ce réglage sensoriel optimal (Moyenne) et son erreur standard associée (Err. Std.) qui peut permettre de bâtir des intervalles de confiance ou des tests de comparaison de moyennes. Les descripteurs du haut du tableau indiquent des différences de réglage considérables entre les deux groupes. La confidentialité de l'application nous interdit d'aller plus loin dans l'interprétation.

La caractérisation des 2 segments de préférence

Si, comme le paragraphe précédent nous l'a montré, il existe deux tendances bien distinctes pour l'optimisation du confort de freinage, l'étape ultérieure consiste à caractériser qui sont les consommateurs de chacun de ces deux groupes. Nous avons appliqué différentes techniques pour croiser nos 8 variables du questionnaire consommateur avec la composition des groupes X1 et X2, mais toutes ont abouti à la même conclusion : aucun recoupement solide entre les deux aspects. Le seul point à la limite de la significativité fut d'observer proportionnellement davantage d'usagers fréquents de l'autoroute dans le groupe X1 que dans le groupe X2. Mais, ceci est bien maigre pour envisager une stratégie marketing tournée vers le freinage et dédiée à un groupe donné de consommateur... Nous pensons sincèrement que ce questionnaire est bien trop succinct pour permettre d'expliquer les préférences. Dans le domaine de l'agro-alimentaire, nous sommes habitués à croiser des questionnaires comportant plus d'une centaine de questions permettant de comprendre les attitudes et les habitudes de chacun face au type de produit considéré. Nous sommes persuadés que cette information est également recueillie dans le domaine de l'industrie automobile et qu'il s'agirait peut-être seulement de coordonner les efforts de différents services pour aboutir à une riche cartographie des préférences.

DISCUSSION ET CONCLUSION

Le but de cet article est de convaincre le monde de l'automobile de l'intérêt des techniques de l'analyse sensorielle pour une meilleure compréhension des attentes de ses clients. Le lecteur doit comprendre que ce que nous présentons aujourd'hui sur

le confort de freinage peut s'étendre à tous les aspects subjectifs mis en jeu dans le plaisir de conduire. L'intégration de ces différents aspects du plaisir subjectif dans une stratégie de qualité totale vise une seule chose : satisfaire le client, maintenant et demain.

Cette étude est, à notre connaissance, la première dans laquelle un relativement grand nombre de consommateurs essayent successivement plusieurs véhicules selon un plan d'expérience. Elle montre que les préférences subjectives des consommateurs en matière de freinage sont variées. Elle suggère que ces préférences peuvent s'expliquer par des descripteurs sensoriels objectifs, qu'un panel de pilotes entraîné à cet exercice est à même de mesurer. Enfin, elle segmente les consommateurs en deux groupes de préférence, et pour chacun d'eux, elle propose une cible chiffrée pour ces descripteurs.

Une différence de taille entre les produits alimentaires et les automobiles est qu'il est difficile, voire impossible, d'essayer un véhicule à l'aveugle... ainsi, rien ne dit que la note de préférence du freinage n'est pas en fait une note de préférence globale pour ce véhicule. Mais, la plus grosse limite de cette étude réside, à notre avis, dans l'échantillonnage des consommateurs restreint au personnel Renault. Le personnel Renault n'est pas représentatif de la cible des clients, ou futurs clients de Renault.

Toujours pour rester dans le domaine de la comparaison avec l'industrie agro-alimentaire, le nombre de 112 consommateurs est souvent trop faible, dès lors que les préférences sont complexes et que l'on s'achemine vers 3, 4 voire 5 segments de préférence. Une règle empirique consiste à dire qu'un groupe de moins de 30 sujets est souvent instable. De plus, avec les mesures subjectives de préférence, on récupère irrémédiablement des réponses au hasard ou, au moins, beaucoup de hasard dans les réponses. Ainsi, il convient de doubler ce seuil de 30 à 60. Si l'on suppose l'existence de k segments de préférence, nous conseillons de réunir $60 \times k$ consommateurs.

Enfin, on réalise une cartographie des préférences pour comprendre le marché et, pour cela, il convient d'inclure les différents éléments du marché dans l'étude, ce qui conduit en général à plus de huit produits. Avec un plus grand nombre de produits,

l'expérimentation se complexifie, mais les statistiques gagnent en robustesse. Des solutions ont été proposées pour étendre la cartographie des préférences aux plans incomplets ([Cal96], [Cal97]), où chaque consommateur ne déguste pas l'ensemble des produits de l'étude.

Ces travaux, ainsi que d'autres non publiés à ce jour, ouvrent un nouveau champ d'investigation pour l'industrie automobile : la mesure directe de la satisfaction du client. Les traditionnels questionnaires de satisfaction ou d'intention d'achat sont relayés par l'essai automobile par des consommateurs. D'un point de vue méthodologique, on passe de l'enquête au plan d'expérience dans l'objectif avoué de

dépasser l'observation des préférences pour atteindre leur modélisation.

Pour des raisons évidentes, le coût d'une cartographie des préférences dans le domaine automobile est bien supérieur à celui de ce type de cartographie dans le domaine alimentaire. Par contre, nous pensons que son potentiel est encore plus fort dans le domaine automobile. En effet, après l'alignement des différents constructeurs d'un point de vue des performances mécaniques et des gages de sécurité, la différence se fera sur le plaisir subjectif qui résulte d'une intégration des différentes perceptions sensorielles mises en œuvre dans la conduite d'une automobile.

BIBLIOGRAPHIE

- [Cal96] Callier, P. (1996). *La cartographie des préférences. Son application en milieu industriel et son extension aux plans incomplets*. Thèse de l'Université de Montpellier II.
- [Cal97] Callier, P. and Schlich, P. (1997). "La cartographie des préférences incomplètes. Validation par simulation", *Sciences des Aliments*, 17(2), pp 155-172.
- [Law98] Lawless, H. T. and Heymann, H. (1998). *Sensory evaluation of food. Principles and practices*. New York, Chapman & Hall.
- [Mac89] MacFie, H. J., Bratchell, N., Greenhoff, K. and Vallis, L.V. (1989). "Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests", *Journal of Sensory Studies*, 4, pp 129-148.
- [Mar79] Mardia, K.V., Kent, J.T. and Bibby, J.M. (1979). *Multivariate analysis*. London, Academic Press.
- [Sch93] Schlich, P. (1993). "Uses of change-over designs and repeated measurements in sensory and consumer studies", *Food Quality and Preference*, 4, pp 223-235.
- [Sch94a] Schlich, P. (1994). "GRAPES : a method and a SAS program for graphical representations of assessor performances", *Journal of Sensory Studies*, 9, pp 157-169.
- [Sch94b] Schlich, P. (1994). "L'évaluation sensorielle : une technologie statistique ?" *XXVIème journées de statistique de l'ASU*, Université de Neuchâtel Ed.
- [Sch95a] Schlich, P. (1995). "Preference mapping : relating consumer preferences to sensory or instrumental measurements", *Bioflavour 95*, Dijon, INRA Ed.
- [Sch95b] Schlich, P. and Degoud, S. (1995). "FLASH Analysis of Profiling Data", *4èmes Journées Agro-industrie et Méthodes Statistiques*, Dijon, ASU Ed.
- [Sch96] Schlich, P. (1996). "Defining and validating assessor compromises about product distances and attribute correlations", in *Multivariate analysis of data in sensory science*, T. Naes and E. Risvik Eds., Amsterdam, Elsevier Science, pp 259-306.
- [Sch92] Schlich, P. and McEwan, J.A. (1992). "Cartographie des préférences. Un outil statistique pour l'industrie agro-alimentaire", *Sciences des Aliments*, 12, pp 339-355.
- [Sch98] Schlich, P. (1998). "La cartographie des préférences ou l'optimisation du plaisir subjectif à l'aide d'un réglage sensoriel objectif", in *Exploitation statistique des mesures industrielles objectives & subjectives, Applications industrielles de l'analyse des données, 4èmes journées Modulad*, Guyancourt, 17-18 novembre 1998, Renault et CISIA Eds.
- [Wak95] Wakeling, I. N. and MacFie, H. J. H. (1995). "Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of k samples from t may be tested." *Food Quality and Preference*, 6(4), pp 299-308.

Article 8

8. Guinard J.X., Uotani B. & Schlich P. (2001). Internal and external mapping of preferences for commercial lager beers: comparison of hedonic rating by consumers blind versus with knowledge of brand and price. *Food Quality and Preference*, **12**(4), 243-255.

Internal and external mapping of preferences for commercial lager beers: comparison of hedonic ratings by consumers blind versus with knowledge of brand and price[☆]

Jean-Xavier Guinard^{a,*}, Bunsaku Uotani^a, Pascal Schlich^{b,1}

^a*Department of Food Science and Technology, University of California,
Davis, Davis, CA 95616, USA*

^b*INRA, 21034 Dijon, France*

Received 5 December 1999; received in revised form 17 October 2000; accepted 18 January 2001

Abstract

The individual preferences of 170 consumers in six categories of age (20s, 30s, 40s) and gender (men, women) for 24 domestic, imported or specialty lager beers, tasted first blind and then with knowledge of brand and price, were investigated by preference mapping techniques. Internal preference mapping revealed differences in the preferences of consumers, with some consumers preferring domestic or ice beers, and others preferring specialty or imported beers. Hedonic ratings changed significantly from the blind to the informed tasting condition, particularly for consumers in their twenties, thereby documenting the significant role of non-sensory variables in the formulation of a hedonic judgement by the consumer. In an external preference map relating the consumers' hedonic ratings to the first two principal components of a principal component analysis of descriptive ratings for the 24 beers, 75% of the consumers were fitted by the vectorial, circular, elliptical (with maximum or saddle point) or quadratic AUTO-FIT models, at the required minimum level of significance ($P \leq 0.25$ for wrongly not simplifying the model, and $P \leq 0.25$ for wrongly selecting a consumer). This improvement over previous studies is credited to the high number of samples (24) in the design, and to the large differences in sensory properties among samples. © 2001 Elsevier Science Ltd. All rights reserved.

Keywords: Preference mapping; Beer; Consumer testing

1. Introduction

Understanding which sensory attributes drive consumer acceptance of food and beverage products is critical to the food and beverage industries. Until recently, most techniques relating consumer (hedonic) and analytical (sensory descriptive and/or instrumental) data sets would regress averaged hedonic ratings onto mean analytical ratings (Response Surface Methodology, Regression Method, Stepwise and Multiple regression

techniques; Giovanni, 1983; Schutz, 1983). Such approaches, however, failed to account for inter-individual differences among consumers and made predictions based on an “average consumer”. The consumer population for a given product often is heterogeneous in its likes and dislikes, however, and a variety of techniques have recently been developed to examine the preferences of each consumer and in some cases to regress the hedonic ratings of each consumer onto a set of analytical (sensory or instrumental) variables. These techniques include internal and external preference mapping (Arditti, 1997; Dalliant-Spinnler, MacFie, Beyts, & Hedderley, 1996; Greenhoff & MacFie, 1994; Hough & Sanchez, 1998; Jaeger, Andani, Wakeling, & MacFie, 1998; Mc Ewan, 1996; Monteleone, Frewer, Wakeling, & Mela, 1998; Murray & Delahunty, 2000; Schlich, 1995), multiple factor analysis (Belin-Batard, Huon de Keradec, & Barthélémy, 1996; Escofier & Pagès, 1990)

[☆] Presented in part at the Annual Meeting of the Institute of Food Technologists, Orlando, Florida, USA, June 1997.

* Corresponding author. Tel.: +1-530-754-8659; fax: +1-530-752-4759.

E-mail address: jxguinard@ucdavis.edu (J.-X. Guinard).

¹ Author Schlich was on sabbatical leave at the University of California, Davis, at the time of the study.

and partial least square regression (Huon de Kermadec, Durand, & Sabatier, 1997; Murray & Delahunty, 2000). Internal preference mapping refers to the analysis of preference data only, and provides a summary of the main preference directions and the associated consumer segments (Greenhoff & MacFie, 1994). Recent developments in internal preference mapping include the significance testing of individual consumers fitted by the preference model and whether differences among product preferences are significant (Dalliant-Spinnler et al., 1996; Monteleone et al., 1998). Using a number of regression models (from linear to quadratic ones), external preference mapping regresses the preferences of each consumer onto the first two principal components of a principal component analysis of the products' sensory characteristics (typically derived from descriptive analysis; Schlich, 1995).

A number of questions or pitfalls have been raised regarding preference mapping methodology. First, it is necessary to have consumers sample a large number of products (for the preference maps to be meaningful), and this goes against common practices in market research (where consumers usually are asked many questions about one or two products only). Other limitations are specific to the type of preference mapping technique used. For internal preference mapping: (1) The first two principal components which are typically plotted to represent consumer preferences (the preference map) often account for a limited amount of the variance in the data (sometimes less than 50%); as a result, experimenters are left wondering whether they are missing valuable information (and preference dimensions) by overlooking the other principal components. (2) There is no convenient way to translate the preference dimensions in terms of product characteristics, since the consumer data is not regressed onto a set of descriptive or instrumental data. Information about the sensory properties driving preference may be obtained however, by "projecting" sensory attributes onto the sample map spanned by the first internal preference dimensions (MacFie & Hedderley, 1993). For external preference mapping: (1) The number of consumers who may be accounted for (in a significant manner) with the various regression models can be low (sometimes less than 50% of the consumers). This means that a significant portion of the consumer population may not be accounted for in ensuing formulation efforts. (2) It can be difficult to translate the preference directions or clusters in terms of product characteristics (and to reformulate the product accordingly) since the consumer data is regressed onto principal components (which are linear combinations of the original sensory variables), not sensory attributes. The present paper addresses these potential limitations.

There is clear evidence and extensive knowledge from market research that many non-sensory variables may affect the sensory acceptability of a food product. The

importance of variables such as nutritional information, price, package, context, expectations and attitudes have been documented in the sensory literature (Dransfield, Zamora, & Bayle, 1998; Guinard & Marty, 1997; Murray & Delahunty, 2000; Solheim, 1992; Solheim & Lawless, 1996; Tuorila, Andersson, Martikainen, & Salovaara, 1998; Vickers, 1993). In these studies, however, mean ratings across a consumer population have usually been compared among conditions, not the ratings of each individual consumer. In the present research, we wanted to apply the preference mapping methodology to investigate the effects of two of these non-sensory variables — brand and price, to get a sense for how individual consumers would be affected by these variables (not just the consumer population as a whole).

The objectives of this study were: (1) to investigate individual consumer preferences for lager beers using internal preference mapping; (2) to relate these individual preferences to descriptive attributes in the beers using external preference mapping; and (3) to assess the effect of knowledge of brand and price on those individual preferences.

2. Materials and methods

2.1. Beers

Twenty-four lager beers divided in three categories (e.g. 10 domestic/ice lagers, eight imported lagers, and six specialty lagers) were evaluated in this study (Table 1). Beers were selected based on volume of sales in the US, and to cover the range of sensory properties typically found in lager beers. The number of beers in the design was chosen based on the number of beers consumers could reasonably be expected to taste over 3 sessions (e.g. 2 sets of 4 beers = 8 beers per session).

2.2. Consumer tests

2.2.1. Consumers

One-hundred and seventy consumers, users and likers of beer, between 21 and 59 years of age, participated in the consumer tests. They represented six categories based on gender (men and women) and age (20s, 30s, and 40s). The distribution of the subjects among the various categories is shown in Table 2. Power analysis using data from previous studies measuring attitudes and liking established that about 30 consumers per gender and age category was required to achieve adequate statistical power (Guinard, Sechevich, Meaker, Jonnalagadda, & Kris-Etherton, 1999). That target number was reached for four of the six categories in the study. Women in their 30s and 40s were more difficult to recruit, and the number of subjects in these two categories was lower (22 and 23, respectively).

2.2.2. Consumer recruitment and screening

The study was advertised in the local and student newspapers, by posting and handing out fliers in local restaurants, bars, supermarkets, and setting up a sign-up desk at the Davis biweekly Farmer's Market. Campus bulletin boards and email ads were also used. Consumers expressing an interest in participating were screened by phone interview for usage and liking of beer using a Product Attitude Survey instrument (Stone & Sidel, 1993). Potential participants were selected if they answered yes to the question "do you like beer?" and reported that they consumed between 2 and 20 bottles or cans of beer (or equivalent) weekly.

2.2.3. Site of study

Consumer tests were carried out in the Sensory Science Laboratory of the Department of Food Science and Technology at the University of California, Davis. The laboratory consists of a preparation area, five booths, and a data processing, analysis and reporting

area. This set up (rather than mall intercepts or drinking-establishment- or home-use tests) was selected to afford better control of the experimental conditions.

2.2.4. Experimental design and protocol

Hedonic (attribute diagnostic questionnaire) and purchase intent ratings were obtained from the consumers during three blind-tasting and three informed-tasting sessions held over two consecutive weeks, with three sessions per week, held on three different days. In each session, consumers tasted two sets of four beers, with a 10-min break between sets.

The 9-point hedonic scale (Peryam & Pilgrim, 1957) was used for hedonic scaling and a 5-point scale (Meilgaard, Civille, & Carr 1991) was used for purchase intent scaling. A special set of questions was also designed to ask the consumers in which situation or on which occasion they would drink the beers.

Samples (50 mL) were served monadically according to one of 10 different randomized orders of presentation. The appropriateness of the number of samples and sessions' format was established in preliminary trials with an in-house panel of consumers.

In the blind condition, samples were served monadically in wine glasses topped with a Petri dish to maintain their flavor for as long as possible. In the informed condition, samples were served as in the blind condition, except that an empty bottle showing the brand name and label, with the price per 6-pack tagged around its neck, was placed alongside the glass for the consumer to examine.

Consumers were recruited in batches of 20–30 consumers at a time, for each 2-week period in the study. Testing was held between 16:00 and 19:45 pm. Consumers were scheduled every 45 min. Upon arrival into the testing area they were invited to sit down at a booth and to review the scorecard. The instructions on the scorecard were also reviewed verbally by the experimenter to ensure compliance. The consumer then received the first beer and rated the various parameters on the scorecard. When done with the first beer, the judge flipped a light switch and was given the next beer to evaluate. The first glass was removed from the booth (monadic presentation). The judge repeated the procedure for the four beers in the set. The second set was tested in the same manner after a 10-min break. Consumers took between 30 and 45 min to complete their evaluation of the two sets.

Samples were brought up from a cold room to a refrigerator in the laboratory the evening before the day of testing. Bottles were opened immediately before serving and recapped right after pouring. A bottle could be reused for other judges within the same 45-min session. This procedure allowed for a stable temperature across samples and it maintained the carbonation at, or close to its original level. Beers were poured carefully into 6-oz

Table 1

Commercial beers included in the study, listed by category with their price per 6-pack (average US retail price at the time of the study)

Category	Brand identification	Price per 6-pack in US dollars
Domestic/ice lagers	Budweiser	4.29
	Bud Ice	3.99
	Coors Original	4.29
	Ice House	4.29
	Michelob	4.29
	Miller Genuine Draft	4.29
	Miller High Life	4.29
	Molson Ice	5.29
	Red Dog	4.39
	Rolling Rock	4.99
Imported lagers	Beck's	7.29
	Corona	6.49
	Foster's	5.99
	Heineken	7.29
	Kirin	5.99
	Labatt's	5.39
	Molson Gold	5.29
	Pilsener Urquell	8.99
Specialty Lagers	Berghoof Lager	3.99
	George Killian's	4.49
	Henry Weinhard's Lager	4.49
	Leinenkugel Red	5.49
	Red Wolf	4.29
	Samuel Adams Lager	5.99

Table 2

Composition of the consumer population sampled

Gender	Age		
	21–30 years	31–40 years	41–50 years
Men	33	32	29
Women	31	22	23

wine glasses to allow for some foam formation (but not excessive foaming).

2.3. Descriptive analysis

Descriptive analysis was carried out by a panel of 17 trained judges with expertise in beer and brewing, using a method that combined elements of the Quantitative Descriptive Analysis® and Spectrum® methods. The panel received 25 h of training over a month. Depending on the beer category, the panel rated the intensity of 26–36 attributes on a 15-point category scale. The beers were evaluated in quintuplicate by category, in the following sequence: domestic/ice beers first, imported beers second, and specialty lagers third. Two sets of five beers were evaluated per session, with a 10-min break between sets. The order of presentation of the samples was randomized across sessions. Details of the descriptive analysis procedures (i.e. term generation, training, protocol, etc. . .) are given elsewhere (Guinard, Yip, Cubero, & Mazzucchelli, 1999).

2.4. Data analysis

Analysis of variance was applied to the hedonic ratings in the blind and informed conditions, with beers, consumers, gender and age (nested) and their interactions as sources of variation. Internal preference mapping was carried out on the hedonic ratings by the 170 consumers (for the blind and informed conditions). Additional analyses were carried out on the combined ratings per segment (e.g. men in their 20s vs. women in their 40s, etc. . .), and on the difference in hedonic ratings between the blind and informed conditions. The matrix of mean ratings for the 23 attributes across the 24 beers was analyzed by principal component analysis (PCA) of the covariance matrix.

External preference mapping (AUTOFIT procedure — Schlich, 1995) was used to regress the preferences of the consumers (in the blind and informed conditions) onto the principal component descriptive space derived from the PCA of descriptive ratings. The different models used to regress the hedonic ratings onto the first two PCs in the AUTOFIT procedure are the vectorial, circular, elliptical (with maximum or saddle point) and quadratic models. The equation relating DOL (Y) for a consumer to PC 1 (X_1) and PC2 (X_2) may therefore range from a simple, linear one, e.g:

$$Y = a + bX_1 + cX_2$$

to a complex, second-order one with quadratic and cross-product effects, e.g.

$$Y = a + bX_1 + cX_2 + dX_1^2 + eX_2^2 + fX_1X_2$$

All statistical analyses were carried out with PC SAS (SAS, 1991).

3. Results

The first step in the analysis of the data was to assess the sources of variation in the hedonic ratings for each condition (blind and informed) with analysis of variance procedures. F -ratios for the sources of variation and their significance are shown in Table 3. In both conditions, degree of liking differed significantly across the beers. Differences in mean hedonic ratings were also observed among age groups, but not between genders. Interactions between beer and gender, and between beer and age group were significant, however, indicative of individual differences among consumers within a gender or an age group. These typical findings exemplify the need for examining individual data by preference mapping or related techniques.

3.1. Internal preference mapping

Internal preference mapping of the hedonic ratings for each condition (e.g. blind and informed) revealed the existence of significant differences in preferences among consumers, and showed an effect of the tasting condition on expressed preferences. In the blind condition (Fig. 1), a half-circle spread of consumers was observed, which revealed the presence of sub-groups of consumers with preferences for ice and Canadian beers in the upper portion of the figure, specialty lagers in the right-hand portion of the figure, and specialty lagers/true imports in the lower portion of the figure. This

Table 3

F -ratios from the ANOVA procedure applied to the hedonic ratings in the blind and informed conditions

Sources of variation	d.f.	F-ratios	
		Blind	Informed
Beer	23	22.06***	21.36***
Gender	1	0.88	2.59
Age group	2	43.47***	37.00***
Gender×Age	2	15.47***	9.66***
Beer×Gender	23	2.62***	2.09**
Beer×Age	46	1.65**	2.06***
Beer×Gender×Age	46	0.78	0.87
Consumers (Gender×Age)	164	7.69***	9.41***
<i>Based on MS Consumer (gender×age)</i>			
Gender	1	0.11	0.28
Age group	2	5.65**	3.93*
Gender×Age	2	2.01	1.03

* $P < 0.05$.

** $P < 0.01$

*** $P < 0.001$

clearly indicates that averaged hedonic ratings do not provide a complete and accurate account of consumer preferences. Fig. 2 shows the results of the same statistical procedure applied to the informed-condition ratings. A slight, but significant shift (rotation) of the beers was observed, with some domestic beers moving out of the range covered by consumer vectors and some imported beers moving into that range. This means that

upon finding out beers were domestic, some consumers lowered their ratings, and the opposite upon finding out some beers were imported (consistent with the results of the comparisons between blind and informed ratings averaged across consumers — Guinard, Uotani, Tagushi, Masuoka, & Fujino, 2000).

Fig. 3 shows an internal preference map of the difference between informed and blind ratings. A long vector

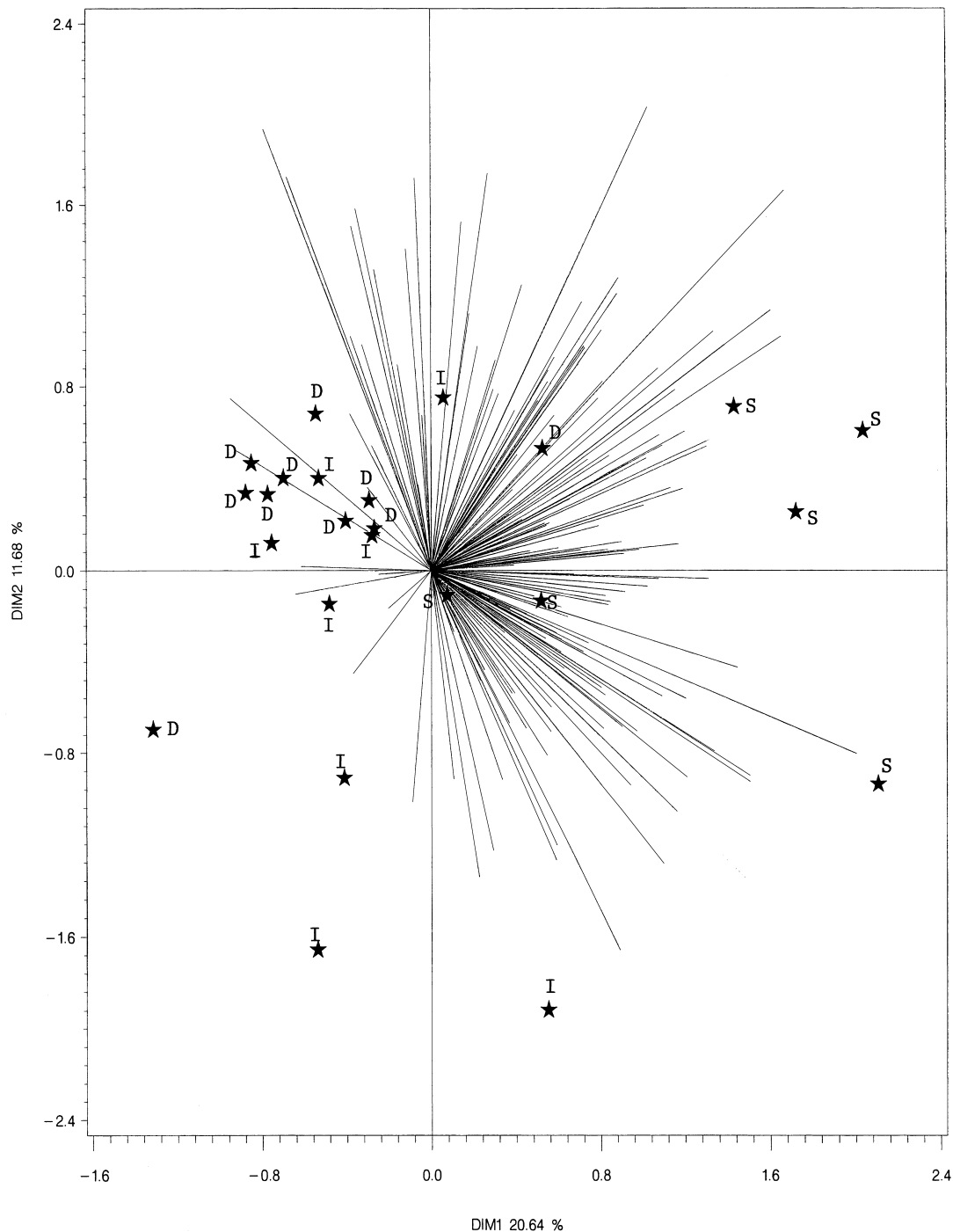


Fig. 1. Internal preference mapping of hedonic ratings in the blind tasting condition. Each vector represents a consumer. D, domestic/ice beers; I, imported lagers; S, specialty lagers.

indicates the preferences of that consumer changed significantly from the blind to the informed condition. It is clear that for a large number of consumers, the information provided in the informed condition affected their hedonic ratings of the beers. It is also apparent that some consumers were influenced by the information given more than others (based on the range of vector lengths seen in Fig. 3). The higher density of long consumer vectors in the proximity of the imported beers

(and opposite the domestic beers) suggests that consumers who changed their ratings did it mostly for beers from these two categories. This is consistent with our observations for the mean data (averaged across age and gender groups; Guinard, Uotani et al., 2000).

The results of the same analysis, this time on mean ratings per consumer group, are shown in Fig. 4. The length of the vectors for the men in their 20s (M20) and the women in their 20s (F20) suggests their preferences

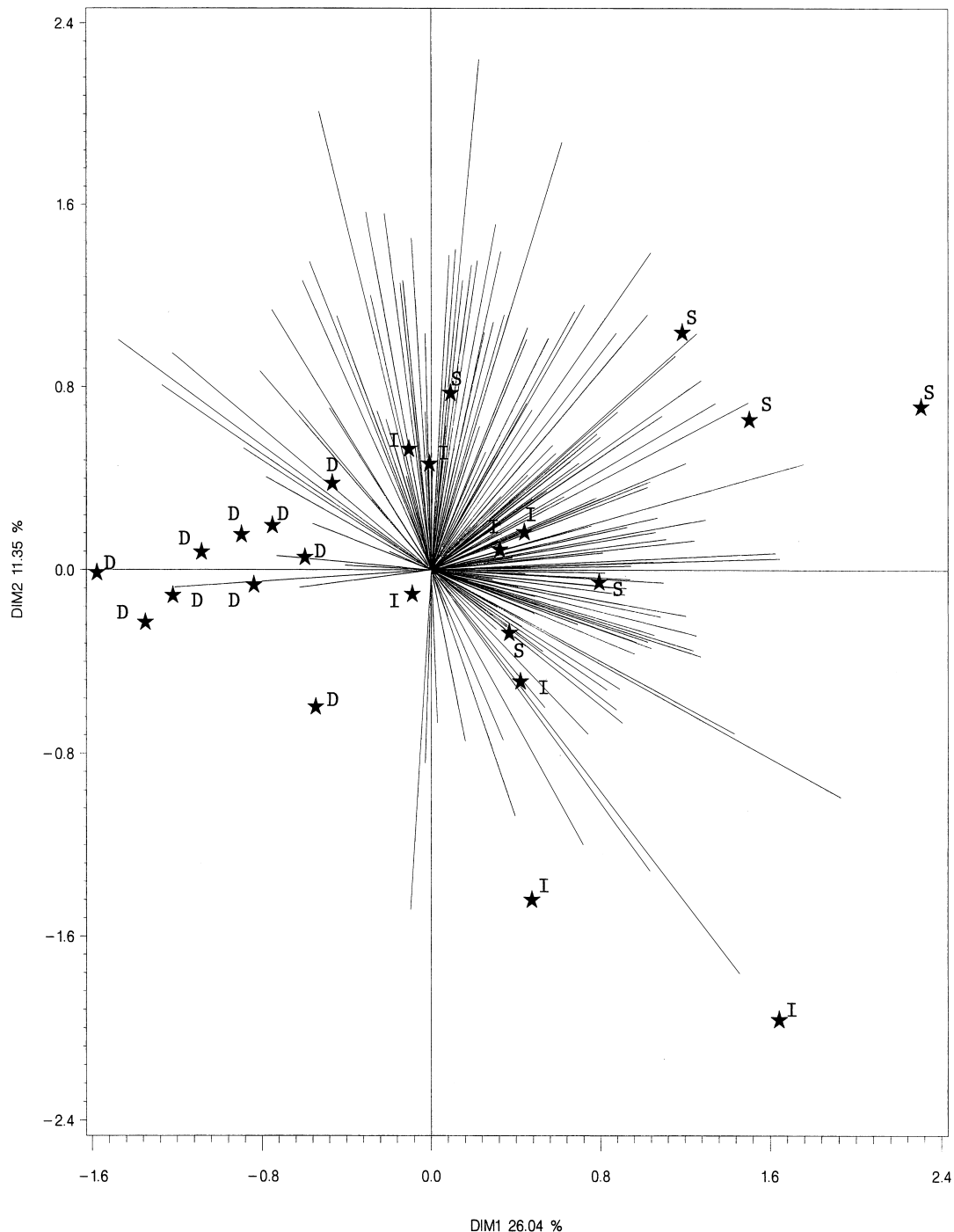


Fig. 2. Internal preference mapping of hedonic ratings in the informed tasting condition. Each vector represents a consumer. D, domestic/ice beers; I, imported lagers; S, specialty lagers.

changed more than those of older consumers upon finding out the brand of beer they were tasting. This is an important finding which suggests younger beer consumers are more likely to be influenced in their preferences by non-sensory variables.

The mean hedonic ratings per group and per condition were then analyzed by internal preference mapping (Fig. 5). Preferences clearly were affected by (1) gender, (2) age, and (3) condition (blind vs. informed), consistent with the results of the analyses of variance presented above (Table 3).

3.2. External preference mapping

External preference mapping using the AUTOFIT procedure (Schlich, 1995) was then used to regress individual consumer preferences onto the principal components of the covariance matrix of descriptive ratings across beers. The PCA of the descriptive attributes across beers based on the covariance matrix is shown in Fig. 6. The use of a covariance matrix allows for a more accurate representation of the importance of each of the descriptive attributes (e.g. a longer vector generally

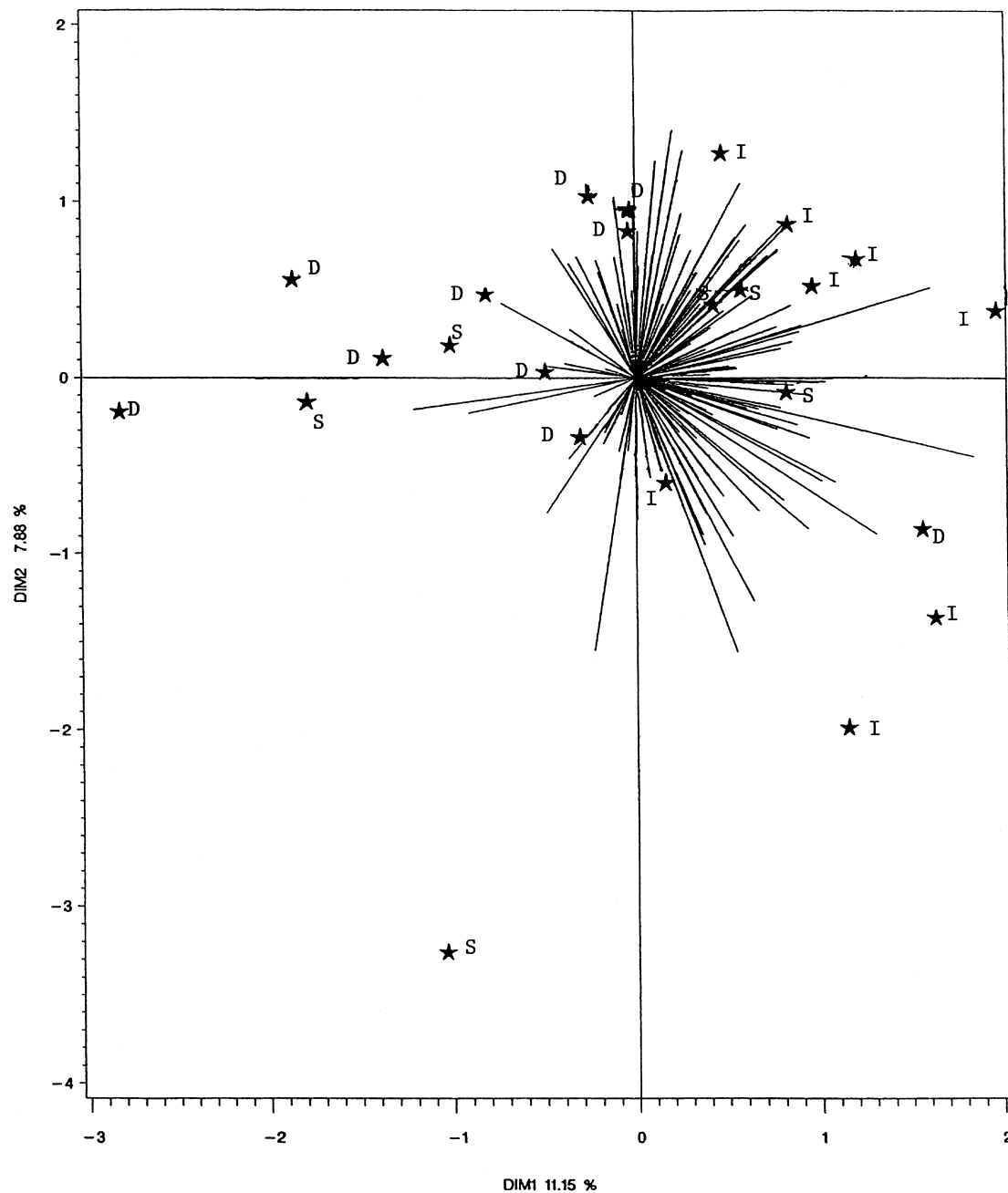


Fig. 3. Internal preference mapping of the difference in hedonic ratings between the informed and the blind tasting conditions. Each vector represents a consumer. D, domestic/ice beers; I, imported lagers; S, specialty lagers.

indicates use of a wider range of the scale, typically, but not exclusively, because of greater discrimination among the samples). The first principal component, which accounted for 57.7% of the variance, contrasted desirable aromatic notes with carbonation. The second principal component (29.7% of the variance) accounted for the presence or absence of mostly undesirable sensory attributes such as oxidized, bitterness, astringent, worty, grainy, sulfury and dimethyl sulfide (DMS). We deem these attributes undesirable based on their negative correlation with expert quality ratings for the same beers (Guinard, Yip et al., 1999), or on research showing that bitterness is an innately unpleasant sensation, even in beer (Guinard et al., 1996). It is remarkable that the beers form three separate clusters corresponding to the three categories in the design — domestic/ice, imported or specialty lagers. This means that ratings of flavor and mouthfeel properties alone, could be used to classify lager beers.

The external preference map for the blind tasting condition is shown in Fig. 7. The beers, shown as stars, are not labeled so as to avoid overcrowding of the figure with symbols. The main purpose of this figure is to show the incidence of the various models used in the AUTO-FIT procedure. For some consumers a vectorial model gave the best fit, and for others the elliptical model with saddle point did. The numbers of consumers fitted by

the various models, along with mean coefficients of determination of the regressions, are shown in Table 4. The models that fitted the most consumers were the vectorial model (shown as straight lines on the biplot), followed by the circular and quadratic models (Table 4). The circular and elliptical models are displayed on Fig. 7 with positive (+) or negative (–) ideal points. In the case of a positive ideal point, the symbol on the plot (+) shows the location of the ideal beer for that consumer. In the case of a negative ideal point, the symbol (–) shows the location of the least-liked beer for that consumer. Symbols x and y represent consumers best fitted by a quadratic model with saddle point.

Table 4

Numbers of consumers fitted (*n*) out of 170 consumers (best model shown), and coefficients of determination (R^2) for the various models in the external preference map. (*P*-values for wrongly not simplifying the model, or wrongly selecting a consumer ≤ 0.25)

Model	<i>n</i>	R^2			
		Mean	S.D.	Minimum	Maximum
Vectorial	55	0.351	0.133	0.184	0.743
Circular	26	0.410	0.138	0.233	0.719
Elliptical	16	0.487	0.149	0.292	0.817
Quadratic	26	0.525	0.105	0.343	0.709
All	123	0.418	0.148	0.184	0.817

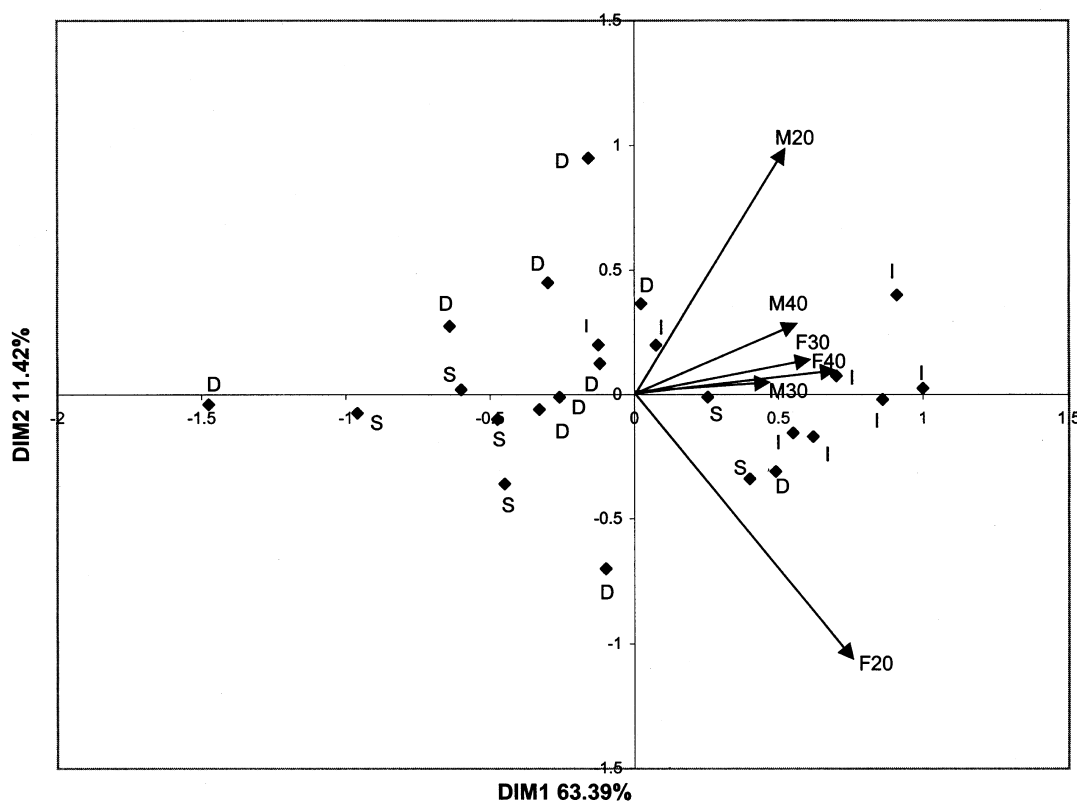


Fig. 4. Internal preference mapping of the difference in mean hedonic ratings per consumer group, between the informed and the blind tasting conditions. Each vector represents a consumer group. M, men; F, women; 20, 21–30 years old; 30, 31–40 years old; 40, 41–50 years old; D, domestic/ice beers; I, imported lagers; S, specialty lagers.

4. Discussion

Internal preference mapping clearly showed a range of consumer preferences in both the blind and informed conditions. This stresses the importance of examining individual consumers, not just averaged hedonic ratings. Given the number of original dimensions in the matrix of hedonic ratings (170 consumers), it is reassuring that the first two PCs still accounted for 32 and 37% of the variance in the blind and informed conditions, respectively.

To our knowledge, this is the first time preference mapping techniques are used to compare blind and informed hedonic ratings by a consumer population. The relevance of preference mapping for investigating how the tasting conditions affect ratings, is the same as what it is for examining hedonic ratings per se. Consumers respond differently to the sensory properties of a product because they have heterogeneous likes and dislikes. Similarly, they respond differently to non-sensory variables, and it is important to examine the behavior of each consumer, rather than that of a population as a whole. As far as which beers were affected most by the transition from blind to informed tasting, more consumers lowered their ratings for domestic beers than for others, and more consumers increased their ratings for imported beers (consistent with the results of the comparisons

between blind and informed ratings averaged across consumers — Guinard, Uotani et al., 2000). The wide range of vector sizes in Fig. 3, where the difference between blind and informed ratings was mapped, shows how some consumers changed their ratings significantly, whereas other did not. When we examined these patterns for each age and gender group (Figs. 4 and 5), it became evident that younger beer consumers were more likely to be influenced in their preferences by non-sensory variables, and to change their ratings from a blind to an informed tasting condition, yet not in the same way (the M20 and F20 vectors are opposite each other on PC2 in Fig. 4). This indicates that they are influenced more by brand and/or price than their older counterparts. It would be interesting to sort out the respective weights of these two influences, but we can speculate that price might have had a greater influence on younger consumers, given that their financial income is more limited.

Our findings regarding the effects of brand and price reinforce those of other investigators. Non-sensory variables of food products are major determinants of consumer behavior (Dransfield et al., 1998; Guinard & Marty, 1997; Murray & Delahunty, 2000; Solheim, 1992; Tuorila et al., 1998; Vickers, 1993). Furthermore, consumer variables such as attitudes and expectations, interact with product sensory and non-sensory variables

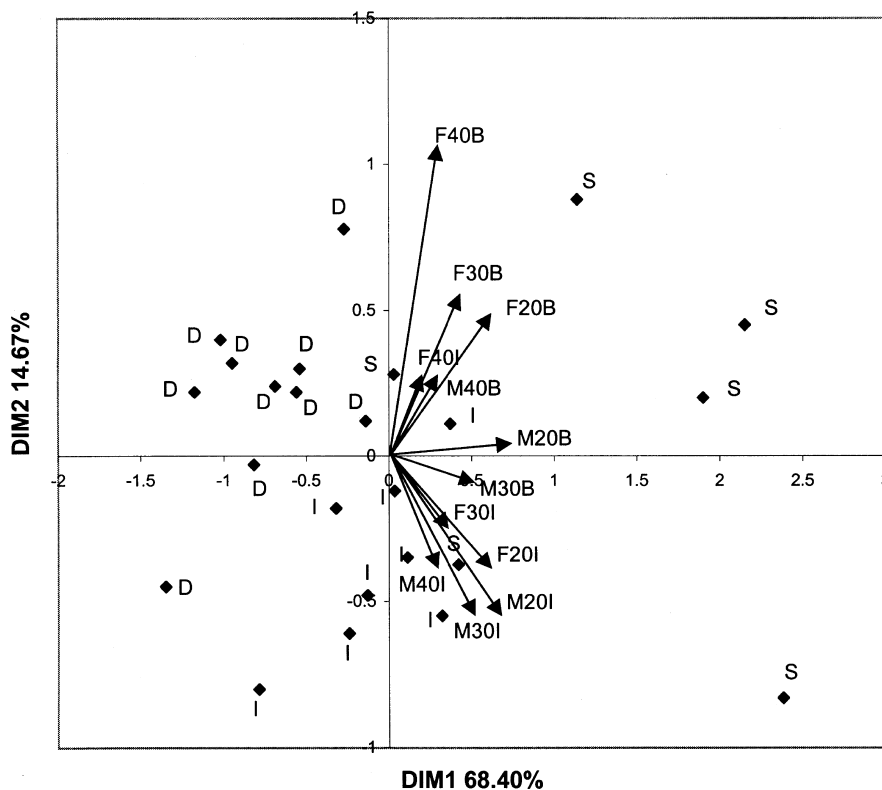


Fig. 5. Internal preference mapping of the mean hedonic ratings per consumer group, in each tasting condition. M, men; F, women; 20, 21–30 years old; 30, 31–40 years old; 40, 41–50 years old; B, blind tasting condition; I, informed tasting condition; D, domestic/ice beers; I, imported lagers; S, specialty lagers.

to shape consumer responses (Deliza & MacFie, 1996; Solheim & Lawless, 1996). Integration of all these variables in measures of consumer behavior is critical to successful consumer and market research.

The findings from the internal preference maps raise the issue of whether a central location test where consumers sample small amounts of several beers can accurately predict actual consumer preferences and behavior. Indeed, the half-circle of consumer vectors in Figs. 1 or 2 generally were directed toward specialty and to a lesser extent imported beers. Yet, domestic lager beers clearly dominate actual volumes of sales in the USA. This could mean that home-use tests where consumers sample the beers in conditions more like actual conditions of consumption, may be needed to get accurate

measures of consumer behavior. For preference mapping, however, consumers must sample many different beers, which would raise the time and cost of a home-use test considerably. Additional research is warranted in this area of predictability of consumer behavior based on consumer test responses.

To assess the quality of an external preference map, it is important to consider: (1) what proportion of the variance is accounted for by the first two dimensions (as for any other PCA), and (2) what percentage of the consumers are fitted by the various models (not an issue with internal preference mapping, which fits all the consumers on the biplot). With regards to the first issue, the first 2 PCs of the descriptive data's PCA accounted for 87.4% of the variance, and as a result, we regressed

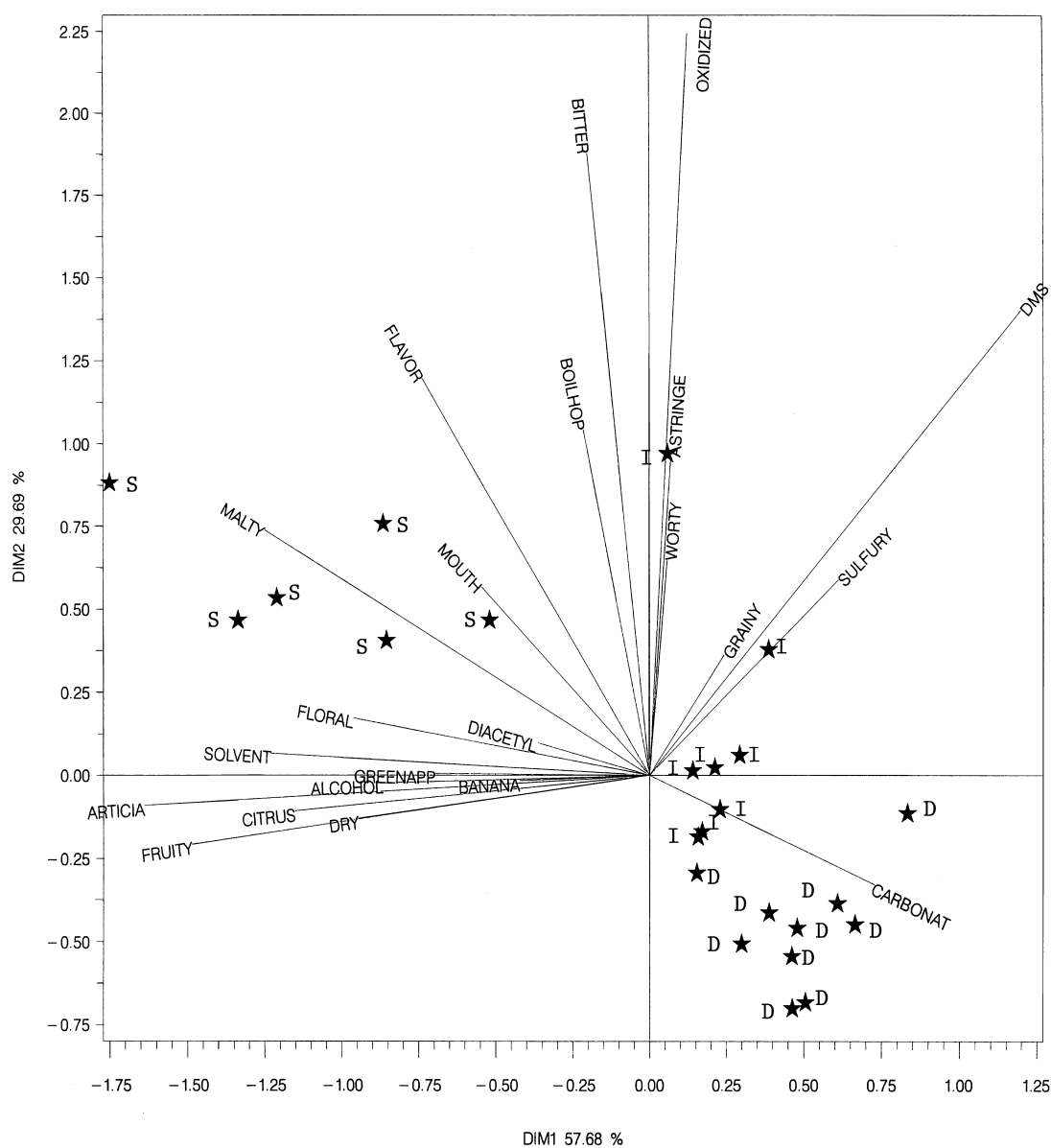


Fig. 6. Principal component analysis of the matrix of mean descriptive analysis ratings from the expert panel across the beers, showing the sensory attributes and the beers (represented as stars on the biplot) in the first two principal components' space.

consumer preferences onto these 2 PCs only. In cases where PC3 or even PC4 account for significant amounts of the variance in the descriptive data (as per an eigen value higher than 1, or per the Scree test), additional external preference maps should be developed for PC1 vs. PC3 and, if necessary, PC1 vs. PC4. As for the second issue, the best model for a given consumer also has to fit the data for that consumer (as per the significance level of the regression equation) or else that consumer is excluded from the external preference map. So far, the percentage of consumers who could be fitted by one of the models had been the main limitation of external preference mapping. Indeed, if only 50% of the consumers are represented in the preference map biplot, the technique may not be that useful to product developers and marketers. This problem is apparent in the actual examples used by Schlich (1995). In this study, 75% of the consumers were fitted by the external preference mapping models at the required minimum level of significance

($P \leq 0.25$ for wrongly not simplifying the model, and $P \leq 0.25$ for wrongly selecting a consumer). This improvement over previous studies can probably be credited to the high number of samples (24) on which the relationships were based, and to the large range of attribute intensities found in the sample set. Indeed, higher degrees of freedom (from a large sample set), and a higher variance in the independent variables — the 2 PCs from the descriptive data, make for more powerful polynomial regressions (statistically speaking). It is also worth noting that all the models were used in fitting these consumers onto the map (Table 4). For some consumers a vectorial model gave the best fit, and for others the elliptical model with saddle point did. The models that fitted the most consumers, however, were the vectorial model (shown as straight lines on the biplot), followed by the circular and quadratic models (Table 4).

One limitation of this case study with lager beers, is that color and other appearance attributes were not

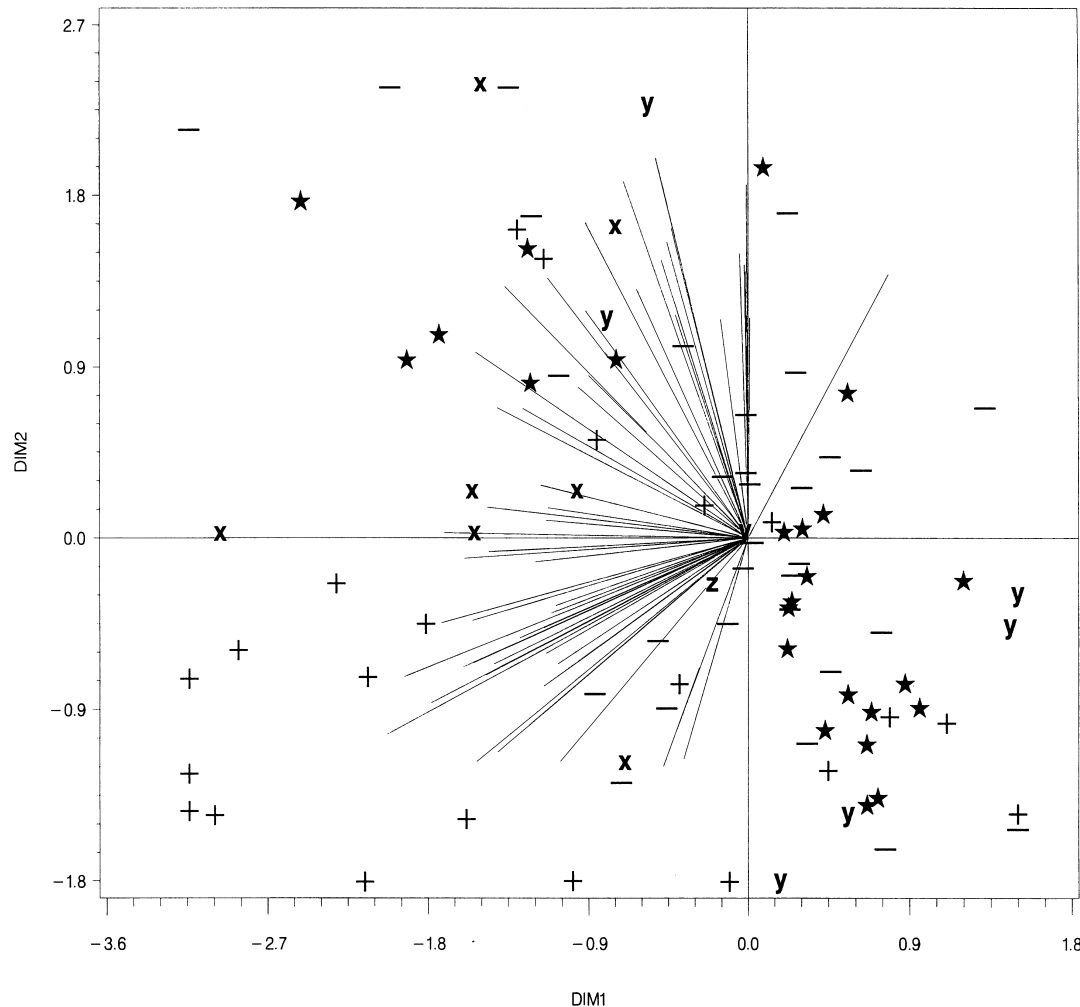


Fig. 7. External preference map of hedonic ratings in the blind tasting condition. The consumers and the various models used to fit them are identified as follows: straight line = vectorial model; + = circular or elliptical model with positive ideal point; - = circular or elliptical model with negative ideal point; x = elliptical model with saddle point, with ideals away from symbol in the direction of DIM 2; y = elliptical model with saddle point, with ideals away from symbol in the direction of DIM 2; and z = quadratic model with saddle point. DIM 1 and DIM 2 are the same as in Fig 6.

included in the descriptive analysis scorecard. We can assume that these attributes must have played a significant part in determining consumer responses to the beers, so that our external preference map is not quite complete. We can only speculate that an additional “significant” principal component would have arisen in a PCA of descriptive ratings with appearance attributes included, that this PC would have contrasted golden vs. amber colors, and that most consumer vectors (or optima) would have been found on the amber side of that PC. That is because specialty malts (used to a greater extent in imported and specialty beers) are responsible for both stronger flavors and darker colors in beer, and most consumers preferred stronger-flavored beers.

The challenge of external preference mapping (and one point of criticism) is the interpretation of the results, since preferences are expressed as a function of two or more principal components, not simple analytical variables (sensory attributes or instrumental measures) as in the Regression Method by Schutz (1983). However, one can start by examining the position of the consumers in the plot in relation to the beers. In Fig. 7, the same observations made for the internal preference maps can be made again: the preferences of most consumers were oriented towards the specialty or ice lagers in the blind condition. One striking feature of both the internal and external preference maps was the lack of beers in one of the (preferred) quadrants for all biplots (e.g. lower left quadrant in Figs. 6 and 7). That niche (which would have been occupied by a type of beer liked by many consumers) probably is that of specialty ales, which were not included in this study. We base this hypothesis on the combination of sensory attributes that would have placed a beer in that quadrant, e.g., malty and dry hop characteristics, low carbonation and absence of defects- all characteristics of a quality ale.

This case study shows the advantage of external preference mapping over internal preference mapping. Once (1) a dense area of consumer preferences has been identified, one can then (2) characterize an ideal beer based on its position on the first two PCs, (3) develop the profile of sensory attributes that correspond to that position on the PCA, and (4) formulate the beer accordingly. This multistep approach is rendered feasible by the fact that brewers nowadays understand quite well the relation between ingredients/processes and beer flavor and are reasonably able to formulate a beer with a specific set of sensory attributes.

5. Conclusions

(1) There was significant variability in the individual preferences of consumers. Such variability must be taken into consideration when analyzing the results of consumer tests; techniques such as internal and external preference

mapping provide that kind of information; (2) hedonic ratings changed significantly from a blind to an informed tasting condition, particularly for consumers in their twenties; this documents the significant role of non-sensory variables in the formulation of a hedonic judgement by a consumer; (3) external preference mapping performed on a large number of samples proved a powerful statistical technique which provides valuable tools (e.g. sensory dimensions that drive consumer liking) for the formulation of successful new products.

Acknowledgements

Supported by a grant from Kirin Brewery Company, Ltd., Yokohama, Japan.

References

- Arditti, S. (1997). Preference mapping: a case study. *Food Quality and Preference*, 8(5), 323–326.
- Belin-Batard, E., Huon de Kermadec, F., & Barthélémy, J. (1996). Méthodes statistiques multidimensionnelles d'analyses de données sensorielles. *Psychologie Française*, 41(3), 301–306.
- Dalliant-Spinnler, B., MacFie, H. J. H., Beyts, P. K., & Hedderley, D. (1996). Relationships between perceived sensory properties and major preference directions of 12 varieties of apples from the southern hemisphere. *Food Quality and Preference*, 7(2), 113–126.
- Deliza, R., & MacFie, H. J. H. (1996). The generation of sensory expectation by external cues and its effect on sensory perception and hedonic ratings: a review. *Journal of Sensory Studies*, 11(2), 103–128.
- Dransfield, E., Zamora, F., & Bayle, M.-C. (1998). Consumer selection of steaks as influenced by information and price index. *Food Quality and Preference*, 9(5), 321–326.
- Escofier, B., & Pagès, J. (1990). *Analyses factorielles simples et multiples, objectifs, méthodes et interprétation*. Paris: Bordas.
- Giovanni, M. (1983). Response surface methodology and product optimization. *Food Technology*, 37(11), 41–45, 83.
- Greenhoff, K., & MacFie, H. J. H. (1994). Preference mapping in practice. In H. J. H. MacFie, & D. M. H. Thompson, *Measurement of food preferences* (pp. 137–166). London: Blackie Academic & Professional.
- Guinard, J.-X., & Marty, C. (1997). Acceptability of fat-modified foods to children, adolescents and their parents: effect of sensory properties, nutritional information and price. *Food Quality and Preference*, 8(3), 223–231.
- Guinard, J.-X., Zoumas-Morse, C., Dietz, J., Goldberg, S., Holz, M., Heck, B., & Amoros, A. (1996). Does consumption of beer, alcohol and bitter substances affect perception of bitterness? *Physiology & Behavior*, 59(4), 625–631.
- Guinard, J.-X., Yip, D., Cubero, E., & Mazzucchelli, R. (1999). Quality ratings by experts, and relation with descriptive analysis ratings: a case study with beer. *Food Quality and Preference*, 10(1), 59–68.
- Guinard, J.-X., Sechevich, P., Meaker, K., Jonnalagadda, S. S., & Kris-Etherton, P. (1999). Sensory responses to fat are not affected by varying dietary energy intake from fat and saturated fat over ranges common in the American diet. *Journal of the American Dietetic Association*, 99(6), 690–696.
- Guinard, J.-X., Uotani, B., Tagushi, A., Masuoka, S. and Fujino, S. (2000). Consumer testing of commercial lager beers: relation with descriptive analysis and expert quality ratings. *Journal of the Institute of Brewing*, 106(1), 11–13.

- Hough, G., & Sanchez, R. (1998). Descriptive analysis and external preference mapping of powdered chocolate milk. *Food Quality and Preference*, 9(4), 197–204.
- Huon de Kermadec, F., Durand, J. F., & Sabatier, R. (1997). Comparison between linear and nonlinear PLS methods to explain overall liking from sensory characteristics. *Food Quality and Preference*, 8(5), 395–400.
- Jaeger, S. R., Andani, Z., Wakeling, I. N., & MacFie, H. J. H. (1998). Consumer preferences for fresh and aged apples: a cross-cultural comparison. *Food Quality and Preference*, 9(5), 355–366.
- MacFie, H. J. H., & Hedderley, D. (1993). Current practices in relating sensory perception to instrumental measurements. *Food Quality and Preference*, 4(1), 41–49.
- McEwan, J. A. (1996). Preference mapping for product optimization. In T. Naes, & E. Risvik, *Multivariate analysis of data in sensory science* (pp. 71–102). Amsterdam: Elsevier.
- Meilgaard, M., Civille, G. V., & Carr, B. T. (1991). *Sensory evaluation techniques* (2nd Ed). Boca Raton, FL: CRC Press.
- Monteleone, E., Frewer, L., Wakeling, I., & Mela, D. J. (1998). Individual differences in starchy food consumption: the application of preference mapping. *Food Quality and Preference*, 9(4), 211–220.
- Murray, J. M., & Delahunty, C. M. (2000). Mapping consumer preference for the sensory and packaging attributes of cheddar cheese. *Food Quality and Preference*, 11(5), 419–435.
- Peryam, D. R., & Pilgrim, F. J. (1957). Hedonic scale method of measuring food preferences. *Food Technology*, 11(9), 9–14.
- SAS (1991). *SAS® User's Guide: Statistics (Version 6.07)* (4th ed). Cary, NC: SAS Institute.
- Schlich, P. 1995. Preference mapping: relating consumer preferences to sensory or instrumental measurements. In *Bioflavour' 95. Analysis/precursor studies/biotechnology*, eds. P. Etiévant and P. Schreier (pp. 135–150), INRA Editions, Versailles.
- Schutz, H. (1983). Multiple regression approach to optimization. *Food Technology*, 37(11), 46–48, 62.
- Solheim, R. (1992). Consumer liking for sausages affected by sensory quality and information on fat content. *Appetite*, 19, 285–292.
- Solheim, R., & Lawless, H. T. (1996). Consumer purchase probability affected by attitude towards low-fat foods: liking, private body consciousness and information on fat and price. *Food Quality and Preference*, 7(2), 137–143.
- Stone, H., & Sidel, J. L. (1993). *Sensory evaluation practices* (2nd ed). San Diego, CA: Academic Press.
- Tuorila, H., Andersson, A., Martikainen, A., & Salovaara, H. (1998). Effect of product formula, information and consumer characteristics on the acceptance of a new snack food. *Food Quality and Preference*, 9(5), 313–320.
- Vickers, Z. M. (1993). Incorporating tasting into a conjoint analysis of ste, health claim, price and brand for purchasing strawberry yoghurt. *Journal of Sensory Studies*, 8(4), 341–352.

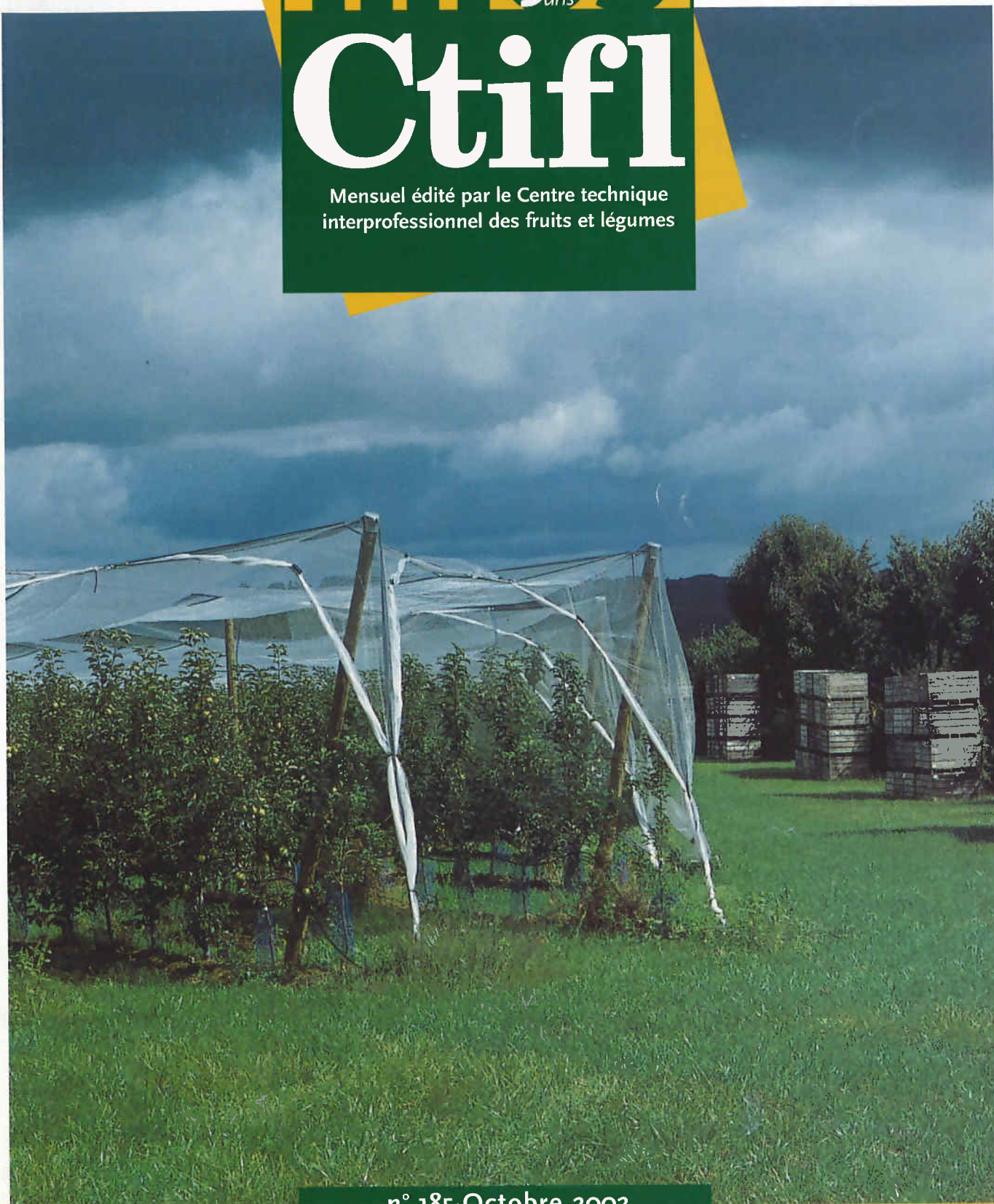
Article 9

9. Scandella D., Navez B., Lespinasse N., Vénien S., Jost M., Schlich P. (2002). La segmentation gustative. Cibler les envies du consommateur. L'exemple de la tomate. *Infos-Ctifl*, **185**, 9-13.

infos Ctifl

1952
2002
CTIFL
50
ans

Mensuel édité par le Centre technique
interprofessionnel des fruits et légumes



n° 185-Octobre 2002

- Le Mur fruitier en récolte manuelle
- Darwin et le Mur fruitier
- Le 1-MCP :
un procédé aux vertus étonnantes
- Pomme : la stratégie néo-zélandaise
- La cerise : c'est Noël au printemps 217
- L'azote du melon : la pratique
du pilotage



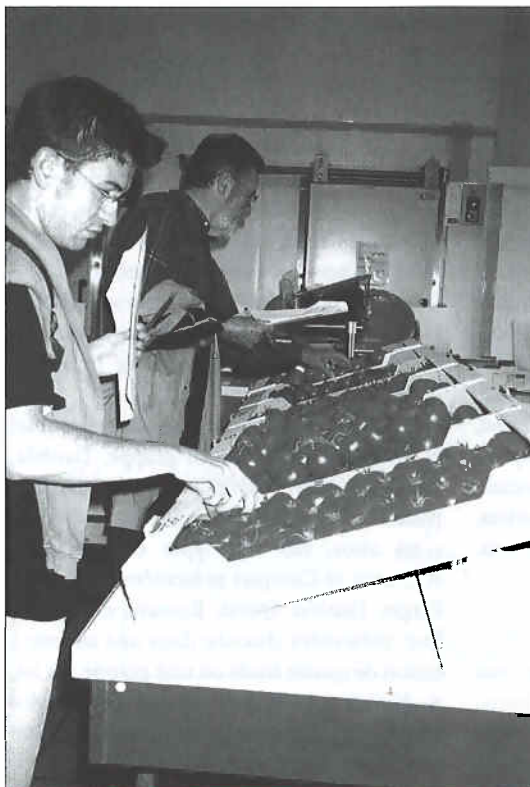


RÉSUMÉ

La cartographie des préférences est une approche méthodologique qui permet d'identifier les caractéristiques sensorielles des produits attendues par les différents segments de consommateurs. Une étude de ce type (financée par l'Acta) a été réalisée sur la tomate en 1999 et 2001, en collaboration avec l'Inra de Dijon. Face à la diversification de l'offre tomate, ce travail permet d'apporter des éléments tangibles de reconnaissance pour les consommateurs. En effet, l'application de la méthodologie de la cartographie des préférences a permis de différencier des groupes de produits bien identifiés ainsi que des segments de consommateurs. Ces résultats peuvent d'ores et déjà être utilisés sur les rayons pour mieux cibler l'offre et parallèlement communiquer auprès des consommateurs.

SEGMENTATION BY TASTE: TARGETING CONSUMER EXPECTATIONS. THE EXAMPLE OF THE TOMATO

Preference mapping is a method used to identify the sensorial characteristics of products that various consumer segments expect. A study of this type, funded by Acta, was conducted on tomato in collaboration with the Inra unit in Dijon in 1999 and 2001. In the face of an increasingly diversified tomato offering, this study provides tangible elements of recognition for consumers. For example, the use of preference mapping allowed differentiation between clearly identified product groups as well as consumer segments. These results can already be used in retail displays to improve the targeting of the offering and communication to consumers.



La cartographie des préférences cristallise les groupes de consommateurs

La segmentation gustative

Cibler les envies du consommateur

L'exemple de la tomate

Les professionnels des fruits et légumes frais se trouvent confrontés à une nécessité de diversification du marché. Pour cela, il est essentiel de bien définir les créneaux et les cibles à atteindre. L'utilisation de la méthode de la cartographie des préférences doit permettre, à partir de la mise en relation des données d'analyses sensorielles descriptives et de celles des tests consommateurs, de déterminer les différents segments d'un produit et de repérer des groupes de

consommateurs en fonction de leurs préférences. Elle permet de vérifier qu'il n'existe en général pas de produit idéal apte à satisfaire 100 % de la population mais différents groupes de produits pouvant chacun répondre à des attentes diversifiées. L'analyse plus détaillée des groupes de consommateurs identifiés permet non seulement de quantifier l'importance d'un segment mais également de qualifier le profil de ces consommateurs.

Avec l'aide du ministère de la Recherche, gérée par l'Acta.

Remerciements aux équipes tomate Balandran et Carquefou pour la préparation de l'essai, la production et l'approvisionnement en tomate, ainsi que les entreprises qui ont collaboré à la collecte des lots.

Les variétés testées

La cartographie des préférences impose de tester un nombre important de produits très diversifiés en même temps afin de couvrir au maximum l'espace sensoriel choisi. Celui-ci représentait la gamme variétale des tomates vendues en vrac ou en grappe, en excluant les types cerises. Treize lots ont été testés en 1999 et dix-sept lots en 2001, avec onze variétés communes aux deux années (en gras ci-dessous).

En 1999 : **Campari, Daniela, Durinta, Élégance, Félicia, Fergie, Labell, Quest, Romana, Romance, Tradiro, Trust, Zapata**. En 2001 : **Alpino, Cameron, Campari, Cencara, Clotilde, Cuor di Bue, Daniela, Durinta, Élégance, Fergie, Labell, Quest, Romana, Romance, Temptation, Tradiro, Trust**.

Les analyses descriptives

Afin de caractériser d'un point de vue organoleptique les différents lots de tomates, des analyses descriptives ont été réalisées au laboratoire d'analyse sensorielle de Balandran.

En 1999, un panel composé de treize juges entraînés a effectué le profil des variétés à l'aide de neuf descripteurs : **ferme, farineux, peau gênante, intensité de l'odeur, arôme, acide, juteux et sucré**.

En 2000, un travail a été réalisé pour préciser les descripteurs et mieux définir les produits. Les descripteurs communs aux deux années sont surlignés en gras.

En 2001, un panel composé de quatorze juges entraînés a effectué le profil des variétés à l'aide de quinze descripteurs : couleur extérieure, taille, caractère côtelé, **ferme, farineux, juteux**, fermeté à la coupe, croquant, fondant, **persistance de la peau, intensité de l'odeur, intensité de l'arôme, persistance de l'arôme, acidité globale, sucré**.

Les mesures physico-chimiques

Les mesures classiques de sucre (réfractométrie et HPLC), d'acidité (acidité totale par neutralisation à la soude et HPLC), de fermeté des fruits (Durofel) et pourcentage de gel (mesure par pesée), ont été réalisées sur tous les lots. La coloration de chaque lot a également été notée d'après le code de couleur OCDE. Ces différentes analyses permettent de définir au mieux la qualité des lots dégustés. Cependant, leur exploitation n'est pas exposée ici car elle nécessite une autre méthode de traitement (cas particulier des relations entre

données expérimentales et données sensorielles) qui est en cours de mise au point.

Les tests hédoniques

Afin d'enregistrer la satisfaction des consommateurs, 215 personnes réparties entre Nantes, Montpellier et Paris en 1999 et 361 personnes réparties entre Lyon, Nantes et Dijon en 2001 ont dégusté les tomates.

Chacun a répondu au protocole suivant :

1- Classement de préférence au niveau de l'aspect visuel afin d'enregistrer le niveau de satisfaction *a priori*

- en 1999, sur six types de tomates : Campari et Durinta présentées en grappe, Daniela, Romana, Quest et Trust présentées dans les plateaux d'expédition anonyme.

- en 2001, sur sept types de tomates : Romance et Campari présentées en grappe, Fergie, Daniela, Quest, Romana et Cuor di Bue, présentées chacune dans une assiette à raison de quatre fruits ou une grappe par lot.

2- Notation hédonique de satisfaction globale à la dégustation : échelle continue non structurée de 0 à 10

3- En 2001, dégustation d'une paire constituée de la variété préférée et de la variété moins aimée : chaque consommateur doit exprimer celle qu'il préfère et pourquoi (d'après une liste de termes proposés).

4- Un questionnaire d'usages et d'habitudes de consommation de 19 questions est rempli par chaque consommateur.

Traitement statistique : la cartographie externe des préférences

La cartographie externe des préférences vise à localiser sur la carte sensorielle établie par le jury expert, le point du produit optimum pour chaque consommateur. Les réponses de chaque consommateur sont projetées sur la carte sensorielle de sorte que l'espace de représentation fasse apparaître simultanément le positionnement des produits et des points consommateurs. On obtient une représentation dans laquelle apparaissent à la fois un produit optimum pour chaque consommateur et le positionnement de chaque variété caractérisée par le panel entraîné.

La carte sensorielle la plus communément utilisée est celle du plan (1,2) de ACP (Analyse en composantes principales) des moyennes des descripteurs. C'est à partir de cette carte sensorielle que la cartographie externe des préférences est réalisée : les notes de préférence de chaque consommateur sont expliquées à l'aide d'un modèle de

MÉTHODOLOGIE

La méthodologie de la cartographie des préférences repose sur la dégustation d'un grand nombre de produits de l'espace sensoriel par les consommateurs, accompagnée d'un questionnaire permettant de mieux connaître ces derniers. Parallèlement les produits sont qualifiés par un panel entraîné. L'analyse des données obtenues avec le panel entraîné permet dans un premier temps de réaliser une segmentation produit c'est-à-dire d'identifier les groupes de produits aux caractéristiques sensorielles comparables.

Dans un second temps, les données « consommateurs » sont utilisées pour relier leurs préférences aux caractéristiques sensorielles afin de réaliser une segmentation consommateurs.

régression par rapport aux deux premières dimensions de l'espace sensoriel des produits.

Plusieurs modèles pouvant être utilisés, il faut disposer d'un outil permettant de choisir le modèle ajustant au mieux les préférences de chaque consommateur. Une fois le modèle choisi, la qualité de l'ajustement est testée pour chaque consommateur. Certains consommateurs pour lesquels l'ajustement n'est pas satisfaisant peuvent être éliminés des calculs.

Les graphiques proposés montrent la répartition des consommateurs dans la carte descriptive des produits.

Les premiers enseignements

Seuls les résultats de 2001 seront présentés ici car ils valident globalement ceux de 1999 en partie présentés lors de la journée tomate à Balandran en mars 2000.

Test descriptif

Dans un premier temps on a pu observer que le panel entraîné s'est avéré plus discriminant en 2001 qu'en 1999 (cf. ellipse de confiance à 90 % contre 50 %). L'ajout de nouveaux descripteurs travaillé en 2000 y a certainement contribué.

L'analyse canonique du profil sensoriel a établi une carte des variétés dirigée selon deux gradients qui porte dans le plan 1-2, 73,5 % de l'information totale : un gradient « gustatif » et aromatique qui porte sur l'arôme



et le sucre et distingue dans l'ordre les variétés Temptation, Alpino, Campari et Elegance, qui s'opposent à Daniela, Labell et un gradient « textural » qui porte sur le ferme et croquant (Durinta et Clotilde) et s'oppose au fondant et farineux (Cuor di Bue, Trust, Cencara et Romana).

Romana et Cencara perçues plus farineuses s'opposent aux variétés de type cocktail, Alpino, Temptation et Campari caractérisées par les critères d'intensité de l'odeur, de l'arôme, sucré et juteux plus élevée. Cuor di Bue et Trust sont caractérisées par leur caractère fondant et Cuor di Bue par le critère farineux. Cuor di Bue et Trust s'opposent à Durinta, Fergie, Clotilde, acides, fermes et croquantes, et à Labell et Daniela qui sont, de plus, jugées « farineuses ».

Cameron, Tradiro, Quest, Romance sont mal représentées, mais il semble (plan 1-3) que Quest et Romance (comme Fergie, dont elles sont proches) se distinguent par leur caractère acide et une peau peu persistante. Elegance est une variété qui tend à présenter les caractéristiques des tomates « cocktails » tout en restant à des niveaux d'intensité plus faibles.

L'analyse de variance à deux facteurs (variété et répétitions sur les séances 1 et 2 de dégustation par le panel) révèle un effet produit significatif au seuil de 5 % pour tous les critères, de même qu'un effet juge pour tous les critères. Ceci signifie qu'il y a une différence de notation entre les juges qui peut s'expliquer par une utilisation différente de l'échelle par les sujets ou par une hétérogénéité des lots.

L'étude des groupes par la méthode de la ppds (plus petite différence significative) met en évidence des groupes qui se chevauchent. On peut cependant retenir :

Intensité de l'odeur : Cuor di Bue, Labell ont les notes moyennes plus faibles et ne sont pas différentes de Daniela, Cencara.

Croquant : Cuor di Bue, Trust moins croquantes que toutes les autres variétés. Clotilde la plus croquante n'est pas différente de Daniela et Durinta.

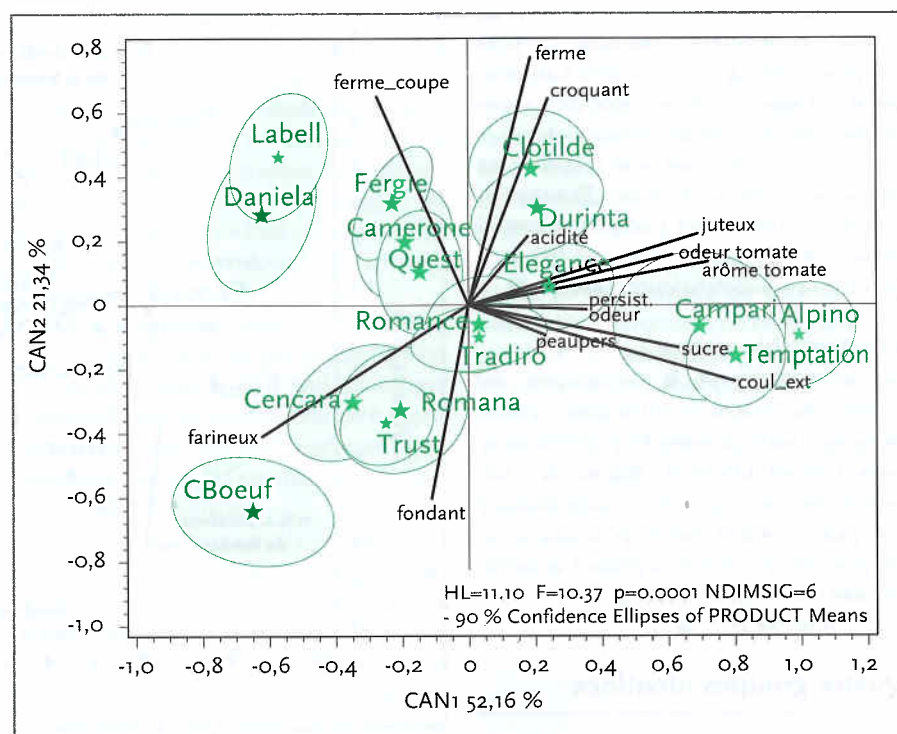
Ferme : Cuor di Bue, Trust sont moins fermes que toutes les autres.

Fondant : Cuor di Bue, Trust sont plus fondantes que toutes les autres.

Farineux : Cuor di Bue la plus farineuse est seule dans son groupe. Elle est suivie par Cencara, Daniela, Romana qui forment un groupe homogène différent de toutes les autres.

Sucré : Alpino, Temptation, Campari forment un groupe plus sucré différent de tou-

FIGURE 1-Carte sensorielle par CVA sans cotelée et taille



tes les autres.

Afin de ne pas entraîner de confusion sur la caractéristique « farineuse », il faut signaler qu'elle est dans la majorité des lots testés peu élevée (de 0,5 à 3,1) ; pour les plus farineuses, on enregistre des valeurs de 4,9 pour Cencara à 6,4 pour Cuor di Bue. Cette caractéristique de texture reste à mieux définir : elle semble correspondre plutôt à une sensation opposée au « juteux » et proche du « fondant » ou plutôt du « pâteux » qu'à une réelle « farinosité » (FIGURE 1).

Tests consommateurs

■ Classement visuel

Comme prévu, les tomates, Romance et Campari, présentées en grappe entraînent une appréciation très favorable d'après l'aspect visuel. A l'opposé, Cuor di Bue ne s'est pas révélée séduisante sur aucun des sites : il semble que son aspect « hors norme », côtelée à l'extrême voire « difforme », ait entraîné quelques suspensions de la part des

consommateurs (TABLEAU 1).

Daniela et Quest, essentiellement différenciables par leur calibre (Quest plus grosse que Daniela), ont ensuite été respectivement moins appréciées. Quant à Fergie et Romana leur appréciation varie selon les sites. Fergie est positionnée très proche de Quest dans deux sites sur trois et ne s'en différencie que par son calibre plus proche de celui de Daniela.

■ Notation hédonique

Cinq lots présentent une note moyenne inférieure à 4,5 : Tradiro, Cencara, Daniela et Cuor di Bue et Labell. Elles sont donc globalement très peu appréciées.

Six lots présentent des notes moyennes comprises entre 4,5 et 5,5 : Clotilde, Romana, Romance, Fergie et Trust : l'observation des histogrammes de répartition des notes révèle que ces positionnements moyens correspondent à une absence de consensus entre les consommateurs et non pas à une

TABLEAU 1-Classement visuel-Tomates 2001

	Romance	Campari	Daniela	Fergie	Quest	Romana	Cuor di Bue
Nantes	1	2	3	4	5	6	7
Lyon	1	3	6	2	5	4	7
Dijon	1	2	6	5	4	3	7

Les chiffres correspondent au rang attribué (1 : la mieux appréciée à 7 la moins)

appréciation « moyenne » par l'ensemble des consommateurs.

Quatre lots présentent une note moyenne comprise entre 5,4 et 5,6 : Durinta, Cameron, Quest et Elegance : ces lots sont bien appréciés par une majorité de consommateurs.

Trois lots présentent une note moyenne supérieure à 6,5 : Alpino, Campari, Temptation. Ces lots, hormis le lot Campari dégusté à Dijon, ont été appréciés par une très forte majorité de consommateurs (TABLEAU 2).

■ Cartographie externe

La méthodologie de la cartographie des préférences permet de distinguer, comme en 1999, quatre groupes de préférences situées dans les différentes régions de « l'espace tomate ». L'analyse du questionnaire d'usages de consommation pour chacun de ces groupes de préférence permet de mieux identifier les consommateurs qui les constituent (FIGURE 2).

Quatre groupes identifiés

Un premier groupe composé de 49 % des consommateurs qui s'oriente d'après la carte sensorielle, vers des tomates « goûteuses » : aromatiques, sucrées, juteuses mais également acidulées, fermes et croquantes.

L'enquête déclarative permet de confirmer ce positionnement : le caractère sucré des tomates correspond à une attente forte de ces consommateurs qui reprochent plus que les autres groupes le manque de sucre de l'offre actuelle. En dégustation, ils apprécient particulièrement Alpino, Temptation et Campari, mais aussi plus que les autres groupes, Durinta et Clotilde. Ils aiment en revanche beaucoup moins que la moyenne des autres groupes : Cuor di Bue, Romana, Cencara et Trust.

Ces consommateurs déclarent préférer les tomates cerises en premier et les cocktail en second. Ils consomment ces dernières plus que la moyenne des autres groupes. Cette déclaration est confortée par le fait que lors de la dégustation, ils ont classé Campari première en présentation visuelle. À l'opposé, ils déclarent, en fréquence plus importante que les autres groupes, ne jamais consommer de côtelées.

Les jeunes consommateurs (18 à 29 ans) sont sur représentés dans ce groupe tandis que les seniors (50-65 ans) y sont sous représentés.

Dans ce groupe, les tomates sont nettement

FIGURE 2-PrefMaX tomates 2001-379 consommateurs -17 variétés-13 descripteurs

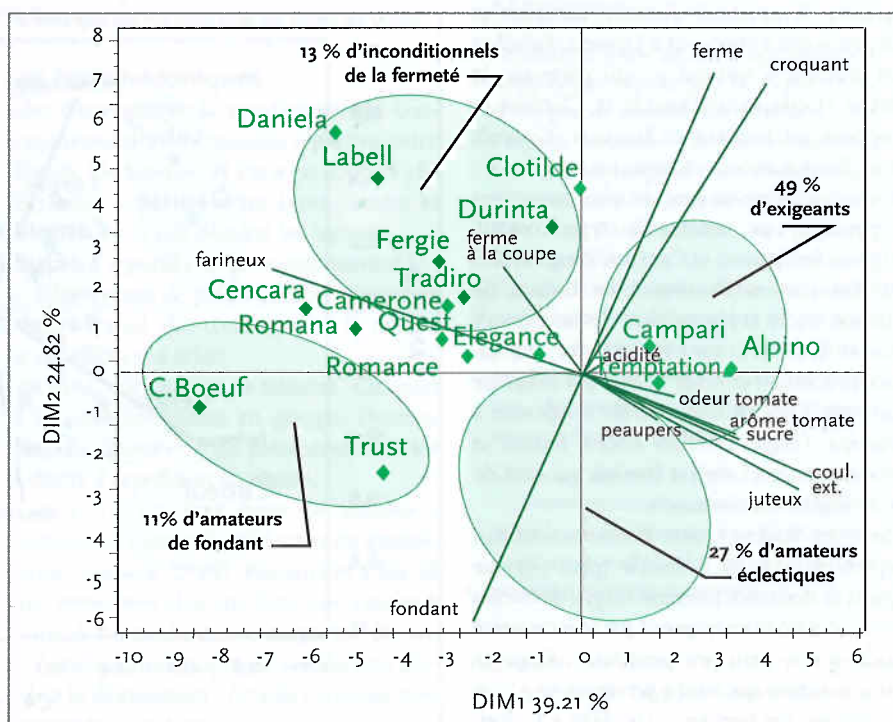


TABLEAU 2-Moyenne des notes hédoniques enregistrées en 2001

Variété	Moyenne générale	Écart type	Variété	Moyenne générale	Écart type
Alpino	7,1	2,6	Romance	5,0	2,6
Temptation	6,8	2,8	Clotilde	4,9	2,7
Campari	6,6	2,8	Romana	4,8	2,8
Elegance	5,7	2,4	Labell	4,2	2,6
Quest	5,7	2,6	Cuor di Bue	4,0	3,3
Camerone	5,5	2,5	Daniela	3,8	2,5
Durinta	5,5	2,6	Cencara	3,2	2,7
Trust	5,4	2,7	Tradiro	3,2	2,7
Fergie	5,3	2,7			

plus souvent consommées natures, telles quelles ou à la croque au sel.

Ce groupe pourrait se résumer ainsi : *Les jeunes, amateurs de tomates cerises et cocktail : gustatives, juteuses et fermes.*

Un second groupe composé de 27 % des consommateurs aimant à la fois les tomates aromatiques, sucrées et juteuses comme celles décrites ci dessus mais ne recherchant pas les mêmes caractères de croquant, ferme et acide que le groupe précédent : aucune tomate actuelle ne se situe dans cet espace.

Lors de la dégustation, les consommateurs de ce groupe ont nettement plus aimé que ceux des autres groupes, des variétés assez différentes entre elles telles que Cuor di Bue, Trust et Temptation mais également dans une moindre mesure Romana. À l'opposé, ils

ont moins aimé les variétés Labell, Daniela, Fergie et Durinta, variétés identifiées par le panel entraîné comme plus fermes ou farineuses. Cela est confirmé par l'enquête déclarative où ce groupe cite plus que la moyenne le caractère sucré comme caractéristique importante tandis que la fermeté ne l'est pas. Lors de la dégustation, l'attrait visuel pour la Cuor di Bue a été prépondérant chez ces consommateurs.

En terme d'usage de la tomate, ce groupe consomme moins que la moyenne les tomates nature ou en soupe.

Ce groupe pourrait se résumer ainsi : *Les amateurs éclectiques qui ne trouvent pas vraiment dans l'offre actuelle la tomate goûteuse et non ferme qu'ils apprécient. Ils recherchent un produit autre que le type rond aujourd'hui majo-*



ritairement proposé.

Un **troisième groupe** composé de 11 % des consommateurs particulièrement orientés vers une préférence en terme de texture : ils recherchent une tomate fondante et « farineuse » (cf. remarque sur ce caractère dans le paragraphe sur les résultats descriptifs) mais également ferme à la coupe. Le caractère fondant attendu est bien supérieur (7,19/10) à celui proposé en moyenne par les tomates présentées (4,98/10)

Ces consommateurs aiment particulièrement Cuor di Bue et Trust mais également Daniela, Cencara et Labell et dans une moindre mesure Romana. À l'opposé, ils n'aiment pas Temptation, Campari et Alpino. En notation visuelle, ils ont mieux classé Fergie et moins bien Campari que les autres groupes.

Cette attente marquée au niveau de la texture est confirmée par l'enquête déclarative où ils notent que « ferme quand on la croque » est la caractéristique la plus importante pour le goût de la tomate tandis que leur attente au niveau du parfum est inférieure à celle des autres groupes.

Ils déclarent, plus que la moyenne, être intéressés par une tomate garantissant une qualité gustative supérieure ainsi que par des informations sur les modes de production et l'origine française des tomates dans leur acte d'achat.

S'ils déclarent dans l'enquête préférer les variétés allongées, ils ne citent jamais les tomates cerises et cocktail en tomates préférées et ne les consomment que rarement.

Les seniors sont sur-représentés dans ce groupe ainsi que le nombre de consommateurs disposant de tomates issues de potager. Ce groupe pourrait se résumer ainsi : *les seniors à la recherche quasi exclusive du fondant mais attentif également au mode de production et à l'origine française des produits.*

Un **quatrième groupe** composé de 13 % des consommateurs dont les préférences s'orientent sur la carte sensorielle vers à la fois une tomate ferme, croquante et « farineuse » (cf. remarque sur ce caractère dans le paragraphe sur les résultats descriptifs).

On retrouve bien dans le déclaratif ces préférences gustatives c'est-à-dire que le caractère ferme est la caractéristique la plus importante tandis que le manque de sucre n'est pas le motif le plus important de non satisfaction. En revanche, l'odeur et le parfum ne sont pas une priorité pour ce groupe.

Ils aiment particulièrement Daniela et Labell et aiment moins que la moyenne

Temptation, Alpino, Campari et Trust.

En notation visuelle, ils ont nettement préféré Fergie par rapport à la moyenne des préférences des autres groupes.

En déclaratif, ces consommateurs révèlent, en plus grand nombre que la moyenne, ne jamais consommer de grappes, préférer nettement plus les rondes et les allongées. Ils ne citent jamais la cocktail comme variété préférée et déclarent consommer rarement les cerises et les cocktails.

Dans ce groupe, on considère nettement plus que dans les autres que la tomate n'est pas une base dans la cuisine et moins que la moyenne que la tomate participe à l'équilibre alimentaire. Ils sont également peu intéressés par des informations sur les différents types de tomates.

Enfin, ce groupe se caractérise également par le fait qu'ils font leur achat principalement en grandes et moyennes surfaces et nettement moins souvent que les autres groupes au marché ou chez des détaillants spécialisés. Ce groupe quasi-identique à celui identifié en 1999 pourrait se résumer ainsi : *les inconditionnels de la fermeté, qui ne semblent pas intéressés par d'autres types de tomates.*

Un outil perfectible

La méthodologie de la cartographie des préférences mise en œuvre pour la tomate en 2001 a permis globalement de valider les résultats de 1999 tout en précisant mieux l'attente des consommateurs grâce à l'introduction d'un plus grand nombre de variétés et de nouveaux descripteurs de texture tels que « fermeté à la coupe » et « fondant ». Les différents segments produits sont ainsi mieux définis et pourront désormais être utilisés sur le rayon pour une meilleure identification de l'offre tomate. Face à ces segments produits, quatre segments de consommateurs ont également été identifiés et devront être utilisés par les professionnels pour mieux cibler cette offre.

Il n'en demeure pas moins que pour 27 % des consommateurs la tomate idéale : « goûteuse et fondante » ne semble pas exister dans la gamme actuelle d'après les analyses précédentes et reste donc à créer par les éditeurs. Cependant, on peut s'interroger sur le fait que pour ces consommateurs aimant une gamme diversifiée de tomates, la méthode tend à « moyenner » pour ce groupe le goût de chacune d'elles

pour en tirer un produit idéal qui en fait ne correspond pas à une réelle attente. En effet, ils peuvent plutôt préférer selon les utilisations ou les occasions de consommation, soit une tomate fondante comme Trust ou Cuor di Bue soit une tomate juteuse, sucrée et parfumée.

La méthodologie de la cartographie des préférences bien adaptée pour un produit agroalimentaire, dont on peut faire varier toutes les composantes du goût dans des formulations différentes, nécessite encore certainement des adaptations pour les fruits ou légumes. En effet, pour ces derniers on peut estimer qu'il y a un produit pour chaque usage et donc des attentes potentielles multiples pour un individu donné. Enfin, il reste à approfondir encore la description et la mesure de la texture pour mieux définir des caractéristiques telle que le farineux par exemple. Un programme est en cours avec l'Inra de Nantes sur ce sujet. ■

Bibliographie

■ *L'approche sensorielle de la qualité tomate.* Infos Ctifl, Mars 2000.

■ P. SCHLICH, J.A.MC EWAN. *Cartographie des préférences, un outil statistique pour l'industrie agro-alimentaire.* Sciences des aliments, 12 (1992) 339-355.

GLOSSAIRE

Les tests descriptifs cherchent à décrire qualitativement et quantitativement les produits testés avec l'aide d'un panel entraîné à la dégustation qui évaluera l'intensité de descripteurs sensoriels appropriés : c'est l'approche analytique de l'évaluation sensorielle.

Les tests consommateurs cherchent à enregistrer le niveau de satisfaction d'un échantillon dit « naïf » de la population : c'est l'approche hédonique de l'évaluation sensorielle.

Espace sensoriel : Ensemble des produits qui représente, pour une gamme donnée, l'exhaustivité des caractéristiques sensorielles susceptibles d'être rencontrées, associées à une gamme étendue de leurs intensités respectives. En général, l'envergure de l'espace est directement liée au niveau de précision de la gamme (ex. : espace des tomates, incluant les tomates de tout calibre, l'espace des tomates cerise).

Article 10

10. Schlich P. (2003). Tests de différences en analyse sensorielle. In *Commission de réflexion sur l'enseignement des mathématiques*. Y. Escoufier. Paris: 18-25

**Composition de la seconde commission mise en place le 10 mars 2001 à la demande de
Jacques Lang, Ministre de l'Education Nationale**

Président :

Jean-Pierre KAHANE, professeur émérite à l'Université Paris-Sud, membre de l'Académie des Sciences, président de la CREM

Membres :

- Michèle ARTIGUE, professeur à l'Université Denis Diderot, directrice de l'IREM de Paris 7
- Roger BALIAN, physicien au CEA, membre de l'Académie des Sciences
- Frédéric BONNANS, maître de conférences à l'Ecole Polytechnique, directeur de recherches à l'INRIA
- Rémy COSTE, professeur au lycée Edmond Michelet à Arpajon, membre du CNP
- Claude DESCHAMPS, professeur de mathématiques spéciales au lycée Louis le Grand, membre du CNP
- Catherine DUFOSSE, professeur au lycée de Marseilleveyre à Marseille
- Jean-Claude DUPERRET, professeur à l'IUFM de Reims, responsable du centre IUFM de Troyes
- Yves ESCOUFIER, professeur à l'Université de Montpellier II
- Catherine HOUEMENT, maître de conférence à l'IUFM de Rouen
- Francis LABROUE, IA-IPR de mathématiques, Académie de Créteil
- Rémi LANGEVIN, professeur à l'Université de Bourgogne
- Michel MERLE, professeur à l'Université de Nice, membre du CNP
- Daniel PERRIN, professeur à l'Université Paris-Sud et à l'IUFM de Versailles
- Antoine PETIT, professeur à l'ENS de Cachan
- Claudine ROBERT, professeur à l'Université Joseph Fourier et à l'IUFM de Grenoble, présidente du GEPS de mathématiques
- Marc ROSSO, directeur du département de mathématiques de l'ENS, rue d'Ulm
- Claudine RUGET, doyenne de l'Inspection Générale de mathématiques.

SOMMAIRE

Introduction :Yves Escoufier.....	04
Incertitudes des mesures de grandeur : Claudine Robert, Jacques Treiner.....	06
Tests de différences en analyse sensorielle : Pascal Schlich.....	18
La notation statistique des emprunteurs ou “ scoring ” : Gilbert Saporta.....	26
Estimation de courbes de référence pour l’analyse de propriétés biophysiques : Jérôme Saracco, Ali Gannoun, Christiane Guinot.....	34
Les traitements statistiques de données textuelles : Ludovic Lebart.....	40
Sensibilisation à la Statistique : Yves Escoufier.....	44
Postface : Yves Escoufier.....	52

Introduction

Yves Escoufier
Laboratoire de Probabilités et Statistiques
Université Montpellier 2
yves.escoufier@univ-montp2.fr

Invité à coordonner un rapport sur la Statistique, j'ai proposé aux membres de la commission de réflexion sur l'enseignement des mathématiques de le construire par le rassemblement d'articles courts écrits par différents auteurs. Ce choix voudrait apporter plusieurs éclairages complémentaires sur cette discipline.

D'abord il met en évidence qu'il n'y a pas de pratiques statistiques sans données et qu'il n'y a pas de données sans des questions issues des domaines scientifiques, économiques, sociaux ou industriels qui justifient l'intérêt qu'on leur porte et le plus souvent provoque et organise leur collecte. Les textes qui suivent abordent des problèmes issus de la banque, de l'industrie des cosmétiques, de l'industrie agro – alimentaire, de l'étude des questionnaires ou de la pratique quotidienne de la mesure de différentes grandeurs physiques. On retiendra que dans les démarches qu'ils décrivent, la Statistique n'intervient pas après un raisonnement interne au domaine concerné ; elle ne s'applique pas après comme une fioriture finale ; elle est inhérente à la construction méthodologique ; elle fournit les outils de la structuration des étapes successives et ceux de l'analyse des résultats intermédiaires ou terminaux. Cette interpénétration de la Statistique et des domaines scientifiques, économiques, sociaux et industriels est une spécificité de cette discipline. Elle alimente l'intérêt qu'on peut lui porter puisqu'elle ouvre à ses spécialistes des champs multiples de l'activité humaine. Elle nourrit aussi les critiques que l'on entend à son sujet lorsque ses résultats deviennent des éléments du débat citoyen. Le statisticien a-t-il pris en compte toute la complexité du problème ? Le spécialiste du domaine concerné a-t-il su mettre en œuvre les méthodes statistiques les mieux adaptées ? Le journaliste ou l'homme politique ont-ils une compréhension suffisante des résultats qu'ils commentent ?

Une seconde information apportée par ces articles réside dans la diversité des méthodes mises en œuvre. Des descripteurs simples comme la moyenne et la variance sont utilisés mais les auteurs font aussi appel à des outils plus complexes issus de la Statistique non paramétrique, des méthodes d'analyses multidimensionnelles ou des tests. La panoplie des méthodes statistiques est très riche et elle s'enrichit constamment. Comme toute discipline, elle le fait de manière endogène en s'interrogeant sur les résultats acquis et en essayant de forger des outils qui ne souffrent pas des limitations reconnues aux outils existants. Elle le fait aussi en se confrontant à des problèmes nouveaux donnant naissance à des problématiques et des données d'un type original. Cet affrontement permanent au réel est une motivation incessante pour le statisticien : la recherche n'est jamais finie ; la pratique est toujours renouvelée. Elle est aussi une exigence car elle interdit de considérer qu'on n'a plus à apprendre. On notera aussi dans tous ces articles la place prise par l'informatique. Elle apparaît de façons variées : pour faire des calculs, pour permettre des simulations, pour gérer de grands ensembles de données ou pour la mise en œuvre de certaines méthodes. C'est là un élément incontournable de la Statistique d'aujourd'hui, évident pour les pratiques statistiques, tout aussi nécessaire dans la recherche elle-même. Et ceci a bien sûr des conséquences dans la formation des statisticiens.

Un troisième constat peut être fait : même si la plus grande réserve dans l'emploi des mathématiques leur a été recommandée, les différents auteurs font tous appel à des formules et des expressions mathématiques pour décrire ce qu'ils font. Comme j'essaie de le montrer dans le texte que j'ai écrit pour cette annexe, les concepts et le langage des probabilités et donc des mathématiques sont nécessaires à la description des objets de la Statistique et des démarches du statisticien. Certes, dans des cours de découverte de la Statistique, on peut, en particulier en faisant appel à la simulation, faire appréhender certaines notions fondamentales par l'expérimentation. Ceci est important dès lors que tout citoyen doit être préparé à lire un journal ou écouter une radio. Il rencontrera inévitablement des pourcentages, des graphiques, des prévisions. Rechercher par des expériences simples à développer son esprit critique à leur sujet est une nécessité. Mais dès que l'usage des statistiques devient plus intense et plus professionnel, l'écriture mathématique est inévitable. Comme d'autres, je crois que cette constatation justifie que l'enseignement de la Statistique soit confiée dans le secondaire aux enseignants de mathématiques. Bien sûr, il faut les aider à découvrir cette discipline et à savoir y trouver des applications particulières d'objets et de résultats mathématiques plus généraux.

Ces professeurs de mathématiques des lycées ont été désignés comme les lecteurs préférentiels des articles qui suivent. En conséquence, dans chacun des textes, une partie au moins est assez détaillée pour pouvoir être reproduite dans une situation analogue par un lecteur possédant une formation mathématique raisonnable et prêt à manipuler des données. Ce serait une grande réussite pour ce rapport si sa lecture contribuait au développement de travaux de groupes en Statistique ou alimentait la réflexion sur les laboratoires de mathématiques que la commission appelle de ses vœux dans les lycées.

Tests de différences en analyse sensorielle

Pascal Schlich
Centre Européen des Sciences du Goût, Dijon
Schlich@cesg.cnrs.fr

Introduction

L'analyse sensorielle consiste à décrire les propriétés organoleptiques (aspect, odeur, texture, saveur, arômes, ...) et hédoniques (plaisir procuré au consommateur) des aliments et des boissons à l'aide de jurys de dégustation, plutôt appelés groupes ou panels d'évaluation sensorielle. L'un des objectifs de l'analyse sensorielle est d'établir si deux produits A et B sont perçus comme différents par un groupe d'évaluation sensorielle lors d'une dégustation à l'aveugle. Il s'agit de choisir et d'appliquer un test de différence, aussi appelé test de discrimination. Par exemple, l'épreuve triangulaire consiste à goûter 3 échantillons non distinguables visuellement dont on sait que deux sont le produit A alors que le troisième est le produit B; il s'agit pour le sujet de désigner l'échantillon provenant du produit non répété B. S'il ne perçoit pas de différence de goût entre ces trois échantillons, alors le sujet répondra au hasard et le nombre de bonnes réponses du panel sera alors distribué selon une loi binomiale de paramètre n (le nombre de sujets) et $1/3$ (la probabilité de succès). Dans cette procédure, le sujet reçoit la consigne de toujours fournir une réponse, quitte à choisir un produit au hasard s'il pense ne percevoir vraiment aucune différence.

Une fois le test réalisé par les n sujets, on peut, à partir du nombre de bonnes réponses, calculer la probabilité de se tromper en déclarant les deux produits comme perçus différemment par les sujets (risque de première espèce, α). Mais l'analyse sensorielle utilise le plus souvent ce test afin de confirmer l'absence d'une différence perceptible entre les deux produits ; il convient dans ce cas de calculer une autre probabilité correspondant au risque de se tromper en déclarant les produits identiques. Le calcul de ce risque, appelé risque de seconde espèce β , nécessite de fixer a priori la grandeur d de la différence entre les produits que le test devrait être en mesure de détecter au niveau α du risque de première espèce. La probabilité $1-\beta$, appelée puissance du test, est donc celle de détecter une différence d entre les produits avec un risque inférieur à α de se tromper. Ainsi, la puissance apparaît comme une fonction de n , d et α .

Après avoir introduit la loi binomiale dans le contexte de l'épreuve triangulaire, nous montrerons comment l'utiliser pour réaliser les calculs de risque de seconde espèce et de puissance. Nous présenterons alors des tables statistiques (Schlich, 1993) qui permettent à l'analyste sensoriel de bâtir ses expérimentations, c'est-à-dire de choisir le nombre de sujets à recruter en fonction des choix raisonnés des risques α et β .

Nous discuterons ensuite de l'intérêt de faire des répétitions, c'est-à-dire de demander à chaque sujet de réaliser plusieurs fois le même test. Nous évoquerons, sans rentrer dans les détails, comment le traitement statistique de ces données peut prendre en compte les répétitions.

Enfin, nous conclurons en montrant pourquoi le cadre de la loi binomiale n'est qu'un modèle peut-être trop simpliste face à la complexité de la situation d'un test de différence.

Définition et principe d'un test de différence

La loi de Bernoulli et la loi Binomiale

Dans l'épreuve triangulaire décrite en introduction, il existe deux alternatives, ou bien le sujet perçoit une différence entre les deux répétitions et on s'attend alors à ce qu'il donne la bonne réponse, ou bien il n'en perçoit pas et il donne alors une réponse au hasard qui a de fait une chance sur trois d'être la bonne. Une telle épreuve aléatoire à deux issues, échec ou succès (codées respectivement 0 ou 1), où la probabilité d'observer un succès est connue à l'avance et nommée p (ici $p=1/3$), est dite suivre une loi de Bernoulli de paramètre p . La probabilité d'observer un échec est alors égale à $1-p$ de sorte que la somme des probabilités des deux événements possibles est bien égale à 1.

Lorsque l'on dispose d'une collection de n épreuves indépendantes chacune régie par la même loi de Bernoulli de paramètre p , la somme du nombre de succès suit par définition une loi Binomiale de paramètre n et p . Ainsi, le nombre de bonnes réponses fournies par un panel de n sujets à une même épreuve triangulaire est distribué selon une loi binomiale de paramètres n et $1/3$, du moins si l'on fait l'hypothèse que les deux produits comparés dans le test ne sont pas perçus différemment par le groupe. Il convient d'insister sur l'importance de l'indépendance des épreuves de Bernoulli qui signifie pratiquement que chaque épreuve ne doit pas être influencée par le résultat de chacune des autres épreuves ; en clair, il suffit de s'assurer que les sujets de l'épreuve triangulaire ne communiquent pas entre eux au cours de la dégustation.

En notant X_i le résultat (codé 0 ou 1) de la i -ème épreuve de Bernoulli, le nombre total de succès, qui suit la loi Binomiale, est noté : $X = \sum_{i=1}^n X_i$ et est une variable aléatoire discrète prenant ses valeurs entre 0 et n . Grâce à l'indépendance des épreuves, la probabilité de chacune de ces valeurs est donnée par :

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad (1)$$

où $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ est le nombre de façons de choisir les x épreuves à succès parmi les n épreuves.

Le principe du test statistique

Si réellement les deux produits ne sont pas distinguables au goût, alors le nombre de bonnes réponses du panel est distribué selon cette loi binomiale de paramètre n et $1/3$. Cette situation est appelée l'hypothèse H_0 , dite "hypothèse nulle". On s'attend à observer "en moyenne" un nombre de bonnes réponses proche de $n/3$. Cette valeur moyenne attendue est désignée en statistique par le terme d'espérance mathématique de la variable aléatoire X . On peut calculer l'espérance mathématique de toute variable aléatoire discrète en additionnant toutes les réalisations possibles de la variable aléatoire chacune étant multipliée (c'est-à-dire pondérée) par sa probabilité de réalisation.

Si au contraire les sujets, ou du moins certains d'entre eux, sont capables de faire la différence entre les deux produits, alors le nombre de bonnes réponses du panel est attendu supérieur à $n/3$. Le but du test statistique est de décider à partir de quel nombre de bonnes réponses observées x on peut considérer que les produits sont perçus différemment par le panel en tant qu'unité. Pour cela, on considère la probabilité suivante :

$$P(X \geq x) = \sum_{i=x}^n P(X = i) \quad (2)$$

qui est celle d'observer au moins x bonnes réponses. Cette probabilité est d'autant plus petite que x est grand. Lorsque qu'elle devient suffisamment petite, cela signifie que le nombre de bonnes réponses est devenu suffisamment grand pour remettre en cause l'hypothèse nulle. Il s'agit là du principe de base des tests statistiques : **on calcule une statistique, dont on connaît la distribution de probabilité sous l'hypothèse nulle, afin de pouvoir infirmer l'hypothèse nulle si la valeur observée de la statistique devient peu probable.**

Classiquement, on demande à la probabilité donnée par la formule (2) d'être inférieure à 0.05 pour rejeter l'hypothèse nulle et donc pour déclarer les deux produits différents ; le risque de se tromper, aussi appelé risque de première espèce α ou niveau du test, en prenant cette décision est alors inférieure à 5%. On peut aussi décider de faire un test au niveau 10%, voire 20%, selon le niveau du risque que l'on est prêt à prendre en déclarant les produits différents alors qu'en fait ils ne l'étaient pas. Il faut bien comprendre qu'en général l'hypothèse nulle devrait être l'hypothèse dangereuse, celle dont le préjudice d'un rejet à tort aurait un coût dont l'importance va nous guider dans le choix d'un risque de première espèce. Par exemple, l'hypothèse nulle " le patient est malade ", car si l'on conclut qu'il ne l'est pas, il ne s'agit pas de se tromper ; ou encore, " la sécurité de cette centrale nucléaire est déficiente ", hypothèse que l'on souhaitera même tester au risque, non pas de 5%, mais de 1%, voire 1‰. En revenant à l'épreuve triangulaire, rejeter l'hypothèse nulle, c'est rejeter que les produits ne soient pas distinguables et c'est donc prendre le risque de refuser de substituer le produit B au produit A, alors que son coût de revient est peut être bien inférieur.

Selon cette logique, l'épreuve triangulaire devrait être mise en œuvre dans des situations où l'on pense que la différence existe et où il est important de le démontrer. Par exemple, mon produit perd des parts de marché depuis l'arrivée d'un nouveau produit concurrent sur le marché. Après avoir dégusté le nouveau produit avec mon équipe, je pense qu'il a effectivement un goût différent du nôtre et que c'est peut-être la raison de notre perte de parts de marché. Toutefois, avant de me lancer dans la modification de la formule de notre produit ; je souhaiterais m'assurer que la population de consommateurs de ce type de produits fait bien la différence entre les deux produits en concurrence. Je vais donc organiser un test de différence en choisissant un risque de première espèce petit, par exemple 5%, afin de n'engager un coûteux programme de recherche visant à reproduire le nouveau produit, que si je suis quasiment certain que les deux produits étaient vraiment perçus différemment par les consommateurs.

Lorsque j'aurai mis au point notre nouveau produit, je vais être confronté au problème inverse du précédent, à savoir démontrer la similitude gustative de ce produit avec celui de nos concurrents. Je vais alors à nouveau organiser un test de différence, mais cette fois-ci dans le but de ne pas rejeter l'hypothèse nulle et surtout en contrôlant le risque β , dit risque de seconde espèce, de me tromper en déclarant les produits identiques alors qu'en fait ils ne l'étaient pas. Il s'agit là du problème du contrôle de la puissance d'un test statistique

La puissance du test

Définition

On appelle puissance d'un test sa probabilité de détecter une différence existante. C'est donc la probabilité de rejeter à juste titre l'hypothèse nulle, soit la probabilité complémentaire du risque β d'accepter à tort cette hypothèse. Ainsi la puissance d'un test est égale à $1-\beta$.

Si l'on désigne par H_1 , l'hypothèse alternative à l'hypothèse nulle H_0 , le tableau 1 résume les différentes situations possibles :

Tableau 1. Les risques en jeu lors d'un test statistique

		Décision	
		H_0	H_1
Réalité	H_0	OK	Risque de première espèce α
	H_1	Risque de seconde espèce β	Puissance $1-\beta$

Pour calculer la puissance, il faut spécifier H_1 c'est-à-dire la hauteur de la différence à mettre en évidence. Dans le cadre de l'épreuve triangulaire, il s'agit de fixer une probabilité p_1 de réussite supérieure à la probabilité $p=1/3$ de réussite sous le seul effet du hasard, c'est-à-dire en l'absence de différence perceptible entre les produits. Les hypothèses H_0 et H_1 du test statistique associé à l'épreuve triangulaire sont donc :

$H_0 : p=1/3$ $H_1 : p=p_1=1/3+d$ avec $d>0$ représente la différence entre les deux produits

En se fixant un risque α , on a vu que l'on pouvait déterminer en utilisant les formules (1) et (2) le nombre x de bonnes réponses à obtenir pour rejeter H_0 . Ce nombre x étant fixé, la donnée de p_1 permet maintenant de calculer β de la manière suivante :

$$\beta = P(X < x \mid p = p_1) = \sum_{i=0}^{x-1} \binom{n}{i} p_1^i (1-p_1)^{n-i} \quad (3)$$

C'est-à-dire en sommant les probabilités d'observer 0, 1, 2, ..., $x-1$ bonnes réponses, soit l'ensemble des événements qui conduisent à ne pas rejeter H_0 , alors qu'en fait c'est H_1 qui est vraie, justifiant l'emploi de p_1 à la place de p dans la formule binomiale.

Tables statistiques pour le contrôle des risques α et β

Les tables statistiques à la disposition des praticiens de l'analyse sensorielle ne présentaient classiquement que le nombre de réponses correctes x parmi n à obtenir pour pouvoir rejeter l'hypothèse nulle avec un risque α inférieur à 10%, 5% ou 1%. Nous avons proposé (Schlich, 1993) deux séries de tables permettant de contrôler simultanément les risques α et β dans le

cadre de l'épreuve triangulaire ($p=1/3$) mais aussi pour d'autres épreuves de différences correspondantes à $p=1/2$. Pour cela, nous avons choisi d'exprimer d , la différence entre les deux produits à mettre en évidence, sous la forme de p_c appelée la proportion corrigée du hasard des bonnes réponses et donnée par la formule suivante :

$$p_c = (p_1 - 1/3)/(2/3) \quad (4)$$

En remplaçant p_1 par la proportion observée de bonnes réponses, on comprend p_c comme la proportion de sujets qui font réellement (sans l'aide du hasard) une différence entre les produits. Dans la première série de tables, dont le tableau 2 présente un extrait, nous avons fixé 3 valeurs pour p_c qui, exprimées en pourcentage, étaient de 50%, 37,5% et 25% et que nous avons arbitrairement appelées grande, moyenne et petite différences. Pour n variant de 5 à 50, nous avons alors tabulé les risques α et β associés à chaque règle de décision définie par le rejet de l'hypothèse nulle si x bonnes réponses sont observées. Nous n'avons pas inclus les x conduisant à α supérieur à 50% et nous n'avons pas imprimé de β supérieur à 50%.

Tableau 2. Extrait de l'annexe 1 de Schlich (1993)

APPENDIX 1. Table of risks (%) for triangular test.

n	x	α	β			n	x	α	β			n	x	α	β		
			50%	37.5%	25%				50%	37.5%	25%				50%	37.5%	25%
5	3	21	21	35	50	21	8	40	0	2	9	30	11	42	0	1	5
6	3	32	10	20	34	22	9	24	1	5	19	30	12	28	0	1	10
	4	10	32	49			10	12	2	11	33		13	17	0	3	18
7	3	43	5	11	23		11	6	6	22	50		14	9	1	7	29
	4	17	17	32	50		12	2	12	37			15	4	2	13	43
	5	5	43				13	1	24				16	2	4	23	
							14	0	40				17	1	9	35	
8	4	26	9	20	36	22	8	46	0	1	7	31	11	47	0	0	4
	5	9	26	44			9	29	0	3	14		12	32	0	1	7
9	4	35	4	12	25		10	16	1	8	26		13	20	0	2	14
	5	14	14	30			11	8	3	16	42		14	12	0	5	24
	6	4	35				12	3	8	28			15	6	1	10	36
							13	1	16	44			16	3	3	17	
10	4	44	2	7	17	23	14	0	29			32	17	1	6	28	
	5	21	8	20	38		15	0	46				18	0	12	41	
	6	8	21	41			9	35	0	2	11		19	0	20		
	7	2	44				10	21	1	5	20		20	0	32		
11	5	29	4	12	27		11	11	2	11	34		21	0	47		
	6	12	12	28	50		12	5	5	21	50		12	37	0	1	6
	7	4	29				13	2	11	35			13	24	0	1	11
							14	1	21								
							15	0	35								

Les 3 dernières colonnes de ces tables donnent le risque β associé successivement aux valeurs de p_c de 50 %, 37,5 % et 25 %. Par exemple avec 30 sujets, un test réalisé au niveau α de 5 % correspondrait à une règle de décision de $x=15$ bonnes réponses, mais ce test n'aurait qu'une probabilité de 57 % (risque β de 43 %) d'être significatif si 25 % des sujets faisaient réellement une différence entre les produits, alors qu'il détecterait presque à coup sur ($\beta=2$ %) la différence si un sujet sur deux ($p_c=50$ %) percevait réellement la différence. On voit bien ici l'importance du paramètre p_c . Si le choix de n variant de 5 à 50 correspond bien à la réalité des tailles de panels d'analyse sensorielle, celui des 3 valeurs de p_c ne correspond pas toujours à la réalité économique. Il suffit parfois de 5 % de consommateurs qui ne retrouve plus le goût de leur produit usuel et donc qui menacent de changer de marque pour fragiliser la rentabilité de cette marque.

L'intérêt de cette première série de table est également pédagogique car la tabulation simultanée du risque α associé à chaque nombre de bonnes réponses avec 3 risques β correspondant à 3 hypothèses H_1 différentes est utile lorsqu'il s'agit d'expliquer le concept de test statistique.

Afin de couvrir un éventail plus large de valeurs de p_c , une seconde série de tables a été proposée dans le même article. Le tableau 3 donne cette table pour le test triangulaire. On peut se rendre compte que s'il s'agit de détecter un p_c de 5 % avec une probabilité $1-\beta$ de 95 %, alors même en tolérant un risque élevé α de 20 % il faudra rassembler un panel d'au moins 1273 sujets ! En conservant les mêmes risques α et β , il est intéressant de regarder l'évolution de n : 1273, 325, 86, 39, 25 et 16 selon que p_c augmente de 5 % à 50 %. Ceci illustre bien que la taille de la différence à mettre en évidence est vraiment le concept clef pour le contrôle de la puissance. Ces tables ont eu le mérite d'alerter des analystes sensoriels sur l'inadéquation à leur problème pratique des tests de différence qu'ils utilisaient en routine.

Tableau 3. Annexe 3 de Schlich (1993)

APPENDIX 3. Triangular test. Minimum total number of responses (n) and associated number of correct responses (x) for a range of type 1 and type 2 risks.

Risk 1	Risk 2 (1-power)									
	0.200		0.100		0.050		0.010		0.001	
	n	x	n	x	n	x	n	x	n	x
$p_c = 50\%$										
0.200	7	4	12	6	16	8	25	11	36	15
0.100	12	7	15	8	20	10	30	14	43	19
0.050	16	9	20	11	23	12	35	17	48	22
0.010	25	15	30	17	35	19	47	24	62	30
0.001	36	22	43	25	48	27	62	33	81	41
$p_c = 40\%$										
0.200	12	6	17	8	25	11	36	15	55	22
0.100	17	9	25	12	30	14	46	20	67	28
0.050	23	12	30	15	40	19	57	26	79	34
0.010	35	19	47	24	56	28	76	36	102	46
0.001	55	30	68	36	76	39	102	50	130	61
$p_c = 30\%$										
0.200	20	9	28	12	39	16	64	25	97	37
0.100	30	14	43	19	54	23	81	33	119	47
0.050	40	19	53	24	66	29	98	41	136	55
0.010	62	30	82	38	97	44	131	57	181	76
0.001	93	46	120	57	138	64	181	81	233	101
$p_c = 20\%$										
0.200	39	16	64	25	86	33	140	52	212	77
0.100	62	26	89	36	119	47	178	68	260	97
0.050	87	37	117	48	147	59	213	83	305	116
0.010	136	59	176	74	211	87	292	117	397	155
0.001	207	91	257	110	302	127	396	162	513	205
$p_c = 10\%$										
0.200	149	55	238	86	325	116	529	186	819	285
0.100	240	90	348	128	457	166	683	244	1011	357
0.050	325	123	447	166	572	210	828	299	1181	421
0.010	525	201	680	256	824	307	1132	415	1539	557
0.001	803	310	996	379	1165	439	1530	568	1992	730
$p_c = 5\%$										
0.200	593	208	931	323	1273	439	2074	710	3201	1090
0.100	927	328	1359	476	1761	613	2682	926	3943	1353
0.050	1271	452	1763	621	2225	779	3246	1127	4616	1592
0.010	2059	737	2660	944	3236	1142	4440	1554	6021	2093
0.001	3162	1137	3903	1393	4585	1628	6006	2116	7827	2739

p_c , Percentage above chance.

Compenser le manque de sujets par des répétitions

Lorsque le nombre de sujets requis dépasse les moyens disponibles en pratique, la tentation est forte de compenser le manque de sujets par des répétitions, c'est-à-dire de donner plusieurs fois le même test à chaque sujet. Si par exemple n sujets ont effectué k répétitions chacun du même test est-il correct de traiter cette situation comme si $n.k$ sujets avaient réalisé chacun une seule fois ce test ? La réponse est bien évidemment négative. Cependant, il n'existe pas encore de consensus parmi les théoriciens de l'analyse sensorielle sur la manière de traiter ce problème. Nous allons brièvement décrire l'état de la controverse.

Il convient tout d'abord de remarquer que sous l'hypothèse nulle chaque sujet a bien une même probabilité de 1/3 de donner une bonne réponse à chaque répétition, si l'on suppose que les répétitions d'un même sujet sont indépendantes entre elles, alors le nombre total de bonnes réponses suit bien une loi binomiale de paramètres $n.k$ et 1/3 et il est fort tentant de tester H_0 en utilisant cette loi, on appellera ce test le "test optimiste". Pourtant, nous pensons que ce test n'est pas correct dans la mesure où nous n'avons plus à faire à un échantillon aléatoire d'une population cible, mais à un échantillon à deux strates (les sujets et les répétitions). Poussé à l'extrême, le test optimiste pourrait conduire à n'utiliser qu'un seul sujet effectuant $n.k$ répétitions.... Ce qui serait évidemment grotesque.

A l'opposé, on peut penser à un test "pessimiste" qui consisterait à ne retenir pour chaque sujet que sa proportion de réussite sur l'ensemble des répétitions et d'utiliser la somme de ces proportions individuelles comme nombre de réussite dans un test binomial classique de paramètres $(n, 1/3)$. Mais sous l'hypothèse H_1 où les deux produits sont perçus comme différents, le test pessimiste, comme le test optimiste d'ailleurs, ne resterait valable que sous la condition que les probabilités de réussite de chacun des sujets soient toutes égales, on parle alors d'homogénéité de la population quant à sa capacité à détecter la différence. Ainsi, les calculs de puissance, qui nécessitent de spécifier une hypothèse alternative H_1 , ne seraient valables que pour une population homogène quant à sa probabilité de détecter la différence.

Brockhoff et Schlich (1998) sont partis du principe que le test optimiste était forcément la stratégie la plus puissante possible, alors que le test pessimiste est la moins puissante. Puisque l'hétérogénéité de la population quant à la probabilité de réussir le test constitue le frein principal à l'utilisation d'une stratégie binomiale permettant des calculs de puissance, ces auteurs ont proposé un test qui coïncide avec le test optimiste pour une population parfaitement homogène et avec le test pessimiste pour une population complètement hétérogène. Ce test consiste à réaliser un test binomial classique en considérant que le nombre total de réponses était égal à $n.k/\sigma^2$, ou à son nombre entier le plus proche, et que le nombre de bonnes réponses était lui égal à x/σ^2 ; où σ^2 est un coefficient de *sur-dispersion* qui consiste à estimer le facteur multiplicatif de la variance de la loi binomiale occasionné par le fait que chaque sujet n'a pas eu le même taux de réussite aux k répétitions. Ce coefficient d'hétérogénéité est compris entre 1 pour une population parfaitement homogène et k pour une population complètement hétérogène.

Le test de Brockhoff et Schlich requiert une estimation du coefficient σ^2 selon des formules fournies par les auteurs, mais qui nécessitent que des données issues d'un test avec répétitions soient déjà disponibles. Or dans la phase de planification du test, l'analyste sensoriel doit contrôler la puissance de son test par un choix raisonné du nombre de sujets et de répétitions qui nécessitera une estimation a priori de ce coefficient. Ce problème est débattu par Schlich *et al.* (2000) sur la base de 6 jeux de données de tests triangulaires intensément répétés (entre 8 et 60 répétitions) avec des groupes de sujets composés de 12 à 61 étudiants. Il ressort de ces travaux qu'une estimation correcte du véritable niveau d'hétérogénéité d'une population requiert une première expérimentation basée sur un grand nombre de répétitions (environ une dizaine), ce qui est rarement faisable.

Le test de Brockhoff et Schlich a été discuté et critiqué par Kunert et Meyners (1999). Il n'est guère possible de rentrer dans le détail de cette discussion ici, mais il faut savoir que celle-ci n'est pas close, puisque Brockhoff va prochainement publier un nouvel article sur ce sujet qui devrait relancer le débat de manière contradictoire entre statisticiens et praticiens de l'analyse sensorielle. D'autre part, Ennis et Bi (1998) proposent une autre approche basée sur la loi beta-binomiale, qui correspond assez bien à la réalité d'un test de différence avec répétitions.

Enfin, il convient d'ajouter qu'aucun de ces tests ne prend en compte que les résultats successifs d'un même sujet à un même test au fil d'une même séance sont peut-être auto-corrélés en raison de phénomène d'apprentissage ou au contraire de fatigue sensorielle.

Cette discussion montre qu'il serait prématuré de vouloir ériger aujourd'hui l'une ou l'autre de ces méthodes en une norme AFNOR ou ISO, même si les praticiens de l'analyse sensorielle réclament cette norme avec insistance. Elle montre aussi qu'un paradigme aussi simple que l'épreuve triangulaire soulève en fait une foule de questions se situant à la frontière des statistiques, de la psychophysique et de la psychologie.

Références :

- Brockhoff P. B. et Schlich P. (1998). Handling replications in discrimination tests. *Food Quality and Preference*, **9**, 303-312.
- Ennis D.M. et Bi J. (1998). The beta-binomial model : accounting for inter-trial variation in replicated difference and preference tests. *Journal of Sensory Studies*, **13**, 389-412.
- Kunert J. et Meyners M. (1999). On the triangle test with replications. *Food Quality and Preference*, **10**, 477-482.
- Schlich P. (1993). Risk tables for discrimination tests. *Food Quality and Preference*, **4**, 141-151.
- Schlich P., Dacremont C. et Brockhoff P.B. (2000). Application of replicated difference testing. *Food Quality and Preference*, **11**, 43-46

Postface

Yves Escoufier
Laboratoire de Probabilités et Statistiques
Université Montpellier 2
yves.escoufier@univ-montp2.fr

La mise en forme de cette annexe au rapport sur la Statistique s'achève alors que l'Institut international de statistique (IIS) annonce le lancement d'un grand projet " d'alphabétisation statistique ", connu sous l'appellation anglaise d' International Statistical Literacy Project (ISLP).

La première décision prise dans le cadre de ce projet a été d'inclure un plus grand nombre de termes concernant l'enseignement de la Statistique dans le dictionnaire des termes statistiques de l'Institut international. Ce dictionnaire multilingue est accessible par internet à l'adresse : <http://www.cbs.nl/isi> en cliquant sur la rubrique ISI Glossary of Statistical terms. Il suffit de repérer un mot du vocabulaire statistique dans une des dix neuf langues connues par le dictionnaire pour avoir sa traduction dans les dix huit autres.

La seconde décision concerne la création d'un site web destiné à apporter des éléments d'information à toutes les personnes concernées par la Statistique. Dans le cadre de cette postface , il n'est pas sans intérêt de relever les objectifs de quelques unes des pages de ce site :

- page B : pour les enseignants de l'école élémentaire, des supports à utiliser dans leurs classes
- page C : pour les enseignants de l'école secondaire, des supports à utiliser dans leurs classes
- page D : pour les enseignants de l'école élémentaire, des supports à utiliser pour améliorer leur connaissance et leur enseignement de la Statistique et des Probabilités
- page E : pour les enseignants de l'école secondaire, des supports à utiliser pour améliorer leur connaissance et leur enseignement de la Statistique et des Probabilités
- page F : des supports pour les enseignants intervenants dans la formation initiale et continue des enseignants.

D'autres pages visent d'autres cibles : les statisticiens des offices gouvernementaux ; les journalistes, les adultes qui veulent apprendre par eux mêmes les bases de la Statistique...

Ce site est accessible à l'adresse suivante : <http://course1.winona.edu/cblumberg/islplist.htm> .

La page C contient déjà un certain nombre de références. Il n'est pas inutile de souligner que comme l'accès au dictionnaire, l'accès à ce site est gratuit. Le but poursuivi par ses initiateurs est de faciliter l'accès de tous aux connaissances statistiques qui les concernent : la communauté internationale des statisticiens se sent une responsabilité à ce sujet et en assume le coût. Pour le moment, les pages de ce site sont exclusivement rédigées en anglais. Des contributions des statisticiens francophones seraient certainement acceptées puisque le français est l'une des langues officielles de l'IIS. Certains feront – ils l'effort nécessaire ?

Revenons à l'annexe que clôt cette postface. Son objectif n'est pas éloigné de celui des pages E et F citées plus haut : apporter des éléments de culture statistique aux enseignants du secondaire. Les programmes de secondes, premières et terminales des établissements secondaires français se sont récemment enrichis de chapitres de Statistique et Probabilités. De nombreux enseignants de mathématiques se sentent démunis devant cette nouvelle tâche à

COMMISSION DE REFLEXION SUR L'ENSEIGNEMENT DES MATHEMATIQUES

STATISTIQUE

Rapport adopté le 15 mars 2003

assurer. La commission de réflexion sur l'enseignement des mathématiques a voulu les aider d'abord par la rédaction de son rapport sur les Statistiques et les Probabilités puis par la mise en chantier de cette annexe au rapport. Le choix qui a été fait ici est double : il voudrait susciter la curiosité des enseignants pour la Statistique en montrant les interventions de cette discipline dans des domaines variés ; il voudrait également contribuer à la formation de ces enseignants en mettant en œuvre sous leurs yeux les objets statistiques qu'ils enseignent : fréquences, caractéristiques numériques, distributions... Les plus curieux pourront même trouver la possibilité d'aller plus loin en se rapportant à la bibliographie volontairement limitée donnée par chaque article ; elle permet d'entrevoir ce qu'est la recherche dans les domaines de la Statistique concernés par l'article.

Au passage, certaines de ces contributions font toucher les problèmes que l'usage de la Statistique peut poser à la société. Le score de risque bancaire qui est le résultat d'un calcul statistique, peut – il à lui seul déterminer la décision d'accorder ou non un prêt à un individu ? Non, rappelle Gilbert Saporta, mais il a fallu que la Commission nationale informatique et liberté (CNIL) se préoccupe du sujet et délibère. En cherchant à déterminer des courbes de références , Jérôme Saracco et ses co – auteurs touchent à la séparation d'individus en deux groupes : ceux qui sont “ dans la norme ” et ceux qui n'y sont pas. Concernant un problème de cosmétique, l'enjeu de cette séparation ne suscite pas de débat de société. En serait il de même s'il s'agissait de santé ou de résultats scolaires ? La décision au sujet des cosmétiques a cependant des conséquences importantes dans le domaine économique pour l'entreprise concernée. Ce même enjeu économique sous-tend l'article de Pascal Schlich sur les tests de différences en analyse sensorielle. Remarquons que l'auteur souligne que les praticiens de l'analyse sensorielle réclament une norme AFNOR ou ISO pour ces tests et que le statisticien défend l'idée qu'en l'état actuel des connaissances , il serait prématuré de vouloir la fixer.

La Statistique est une discipline impliquée dans les débats de la société ; les conclusions des méthodes qu'elle met en œuvre sont des arguments de ces débats. Contribuons à donner au plus grand nombre la capacité de comprendre les méthodes et leurs résultats : c'était le but de cette annexe .

Article 11

11. Cordelle S., Lange C. & Schlich P. (2004). On the consistency of liking scores: Insights from a study including 917 consumers from 10 to 80 years old. *Food Quality and Preference*, **15**(7-8), 831-841.

On the consistency of liking scores: insights from a study including 917 consumers from 10 to 80 years old

Sylvie Cordelle ^{*}, Christine Lange, Pascal Schlich

Centre Européen des Sciences du Goût, 15 rue Hugues Picardet, 21000 Dijon, France

Available online 2 June 2004

Abstract

Preference for 7 orange juices was recorded monadically. Three of the samples were replicates of the same juice (RJ). The 4 other samples were RJ slightly spiked with either sucrose, or citric acid, or quinine or an orange flavor. Each subject then performed a paired preference test composed of RJ and the modified juice whose score was the furthest from the RJ score. Finally, subjects were asked to choose one among 8 reasons why they preferred that sample. A number of individual statistics were computed in order to compare the consistency of liking scores within session, across ages and between genders. Women were slightly more consistent than men. However, no significant effect of age nor gender was found in the coherence between the first hedonic test and the paired preference test. Moreover, children and seniors found it more difficult than other consumers to identify the reason for their preference.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Basic tastes; Preference; Consistency of hedonic measurements; Age; Gender

1. Introduction

Consistency of liking judgments toward foods or beverages has been discussed by several authors in sensory analysis. Most of the studies have focused on the evolution of preferences for unfamiliar or complex products. Thus, in 1999, Levy and Köster conducted two experiments in order to study the relevance of hedonic judgment, when such a judgment is followed by a second judgment in the form of a choice among the products just tested. The influence of repeated exposure on the development of preference was also studied. The first study compared 3 soft drinks (one was found on the French market; the other two were reformulations of the first one and were unfamiliar), in a design based on an alternation of laboratory measurements and home exposure to one of the drinks by different groups. The second study consisted in repeated measurements at home of 3 alcoholic beverages (as previously, one was found on the French market, whereas the other two were unfamiliar). In both experiments, a considerable amount of preference change was observed within sessions, and a development of preference over sessions was

clearly shown in the choice behavior, but not in the hedonic judgment. According to the authors, variety seeking, the exploratory context of exposure, stimulus properties such as novelty and complexity and the individual physiological state of the subject are among the factors that might explain such choice changes. In 2003, Köster, Couronne, Léon, Lévy and Marcelino discussed repeatability in hedonic sensory measurement. Their argumentation was illustrated by data obtained from groups of children and adults who were exposed to the same hedonic sensory measurements a number of times. Some of the products used were popular on the French market, whereas others were unfamiliar products. The authors concluded from the results that the initial hedonic impression was a poor predictor of final liking and choices.

Although several studies have focused on the evolution of liking judgments for unfamiliar or complex products, no study, to our knowledge, has focused on the consistency of liking judgments for familiar and simple food products, apart from the specific studies using product boredom tests (Köster, 1990). These tests are designed to compare boredom caused by different varieties of the same type of product: the test generally consists of several sessions, one for each variety of the product. In each session, the subject is asked to evaluate his liking of a series of at least 15 samples. He is told

^{*} Corresponding author. Fax: +33-3-80681632.

E-mail address: cordelle@cesg.cnrs.fr (S. Cordelle).

that these samples are slightly different, but in fact all samples are exactly the same. The development of the pleasantness ratings over the monotonous series of samples is then studied and compared for the different varieties of the product.

In the present study, we focused on short term consistency—that is, consistency within one sensory session—of hedonic measurements for orange juice, which is a familiar and simple product. We collected hedonic responses for 7 orange juices from a large number of consumers belonging to several age and gender groups. Three of the samples were replicates of the same juice. The four other samples were obtained by adding either sucrose, or citric acid, or quinine or an orange flavor to the replicated juice. It was interesting to first evaluate the influence of the consumers' individual characteristics on the liking patterns. Then, we tried to determine whether age and gender can explain the differences in individual short term consistencies. Finally, by means of an additional paired preference test, we examined whether and to what extent consumers can name the reason for their preference in such a design.

2. Material and methods

2.1. Products

Preliminary tests were carried out on 10 commercial orange juices (with no pulp), in order to choose one as neutral as possible in terms of acidity, bitterness, sweetness and flavor. The selected juice was a pure juice provided by Eckes Granini Company, France. This orange juice was popular on the French market. By adding chemicals to this orange juice, we obtained 5 juices (Table 1). The concentrations of each added chemical were defined by means of a preliminary discriminative test with 16 subjects. This test was designed to determine the lowest concentration at which almost the whole panel was able to discriminate the modified juice from the basic juice. Nevertheless, the 5 juices were in the end quite similar compared to the diversity of the products on the French market.

The 5 juices were prepared each morning before the sessions. They were put into glass bottles and stored at 11 °C until further use. During the sessions, the samples

Table 1
Descriptions of the 5 orange juices

Juice	Code	Added chemical
Basic juice	RJ	No addition
Flavored juice	FJ	0.025 g/l Flavoring ^a
Acidic juice	AJ	3 g/l Citric acid
Bitter juice	BJ	15 mg/l Quinine
Sweet juice	SJ	20 g/l Sugar

^a Provided by Symrise Company, Germany.

were served in plastic transparent glasses coded with 3 digit numbers.

2.2. Subjects

Nine hundred and seventeen consumers participated in the study. Table 2 gives the distribution of the consumers by age group and gender.

The subjects were not paid for their participation, but they could attend, after the sensory session, a presentation of the laboratory, a talk about the principles of sensory analysis and an attractive presentation of their own tasting results.

2.3. General design

Each consumer took part in one session. Forty sessions were organized during the same week, with a maximum of 37 consumers at each session. All the sessions took place in a sensory room equipped with separate booths. The sessions were held at different times of the day and lasted approximately half an hour. A computerized system was used to capture the data (FIZZ, Biosystèmes, France). However, for the elderly people and for the consumers who had never used a computer, data were collected on paper questionnaires. The samples were evaluated under white lights, as there was no visual difference between juices.

Each session consisted of three parts:

2.3.1. First test: collection of liking scores

Seven samples were presented monadically and a single overall liking score was collected for each sample. Since emphasis was on comparing subjects rather than products, every consumer received the juices in the same order (Table 3). The basic juice was presented three times, in order to allow computation of consistency statistics. The amount of each sample served was 20 ml. The subjects were told to drink approximately the same quantity of each sample, with no obligation to finish the

Table 2
Distribution of the consumers by age group and gender

Age group	Men	Women	Total
[10–12[	24	42	66
[12–14[	45	55	100
[14–16[	49	67	116
[16–18[	36	92	128
[18–21[	20	81	101
[21–25[	32	51	83
[25–35[	42	46	88
[35–45[	31	47	78
[45–55[	23	38	61
[55–65[	30	29	59
[65–80[	17	20	37
Total	349	568	917

Table 3
First hedonic test: order of presentation of the 7 samples

Order of presentation	Sample
1	Basic juice (RJ1)
2	Acidic juice (AJ)
3	Flavored juice (FJ)
4	Basic juice (RJ2)
5	Sweet juice (SJ)
6	Bitter juice (BJ)
7	Basic juice (RJ3)

glass. Between two samples, subjects were asked to rinse their mouths with water. For each sample, consumers had to rate their hedonic judgment on a linear scale labeled at one anchor “I do not like the juice at all” and at the other anchor “I like the juice very much”. The hedonic ratings were then translated into scores from 0 to 100.

2.3.2. Consumer questionnaire

After the first hedonic test, each consumer had to fill in a questionnaire concerning socio-demographic information (age, gender, marital status, incomes, professional activity, etc.), health status (smoking status, height, weight), status at the session time (last cigarette, last coffee, last meal) and habits regarding orange juice consumption (consumption level, types of orange juice, brands, etc.).

2.3.3. Second test: Contrast Paired Preference Test

After having completed the above questionnaire, each consumer received two samples of orange juice. The first one was the basic juice, whereas the second one was chosen for each consumer, according to his responses in the first test. The mean (MRJ) of the scores given to the basic juice in positions 4 and 7 was computed for each consumer. Then the modified juice (i.e. the acidic, the bitter, the sweet or the flavored juice) whose hedonic score was the furthest (either positively or negatively) from MRJ was chosen to be the second sample presented in the paired test. However, in order to avoid too much imbalance between the total number of “negative juices” and the total number of “positive juices” presented in the paired test against the basic juice, the previous rule was adapted as follows:

- Let J+ be the modified juice whose score is positively the furthest from MRJ;
- Let DP be the absolute value of the difference between J+ and MRJ;
- Let J− be the modified juice whose score is negatively the furthest from MRJ;
- Let DN be the absolute value of the difference between J− and MRJ;
- If $DP \geq DN/2$ then J+ was chosen as the second sample presented in the paired test for that consumer, if not, J− was chosen.

The amount of each sample served was 40 ml. The subjects were first told to drink half of the first sample (i.e. the basic juice), then to rinse their mouth with water, to drink half of the second juice and to answer the question “Which of these two juices do you prefer?”. The subjects were then told that they could drink the rest of the two samples in order to answer a second question “Could you please choose the main reason for preferring that sample over the other one?”. Eight answers were proposed (“That sample is sweeter”, “That sample is less sweet”, “That sample is more acidic”, “That sample is less acidic”, “That sample has more flavor”, etc.) and each consumer had to click on the answer he found the most appropriate.

2.4. Data analysis

2.4.1. Effect of individual characteristics on the liking scores

In order to examine the effect of age and gender, an analysis of variance was performed on the liking scores given to the 7 samples using the following model:

$$\text{Liking score} = G + A + G * A + C(G * A) + S + G * S + A * S + G * A * S + \text{Error}$$

where C corresponds to Consumer, G to Gender, A to Age and S to Sample. Gender and Age effects and their interaction were tested against Consumer (Gender * Age), whereas the other effects were tested against the usual error term.

Comparisons of mean liking scores were made visually with confidence intervals at 5%, by age and gender, in order to illustrate the effect of these individual characteristics on average liking. Then, an internal preference mapping was performed on the liking scores in order to examine the difference in liking patterns among the consumers.

Finally, an analysis of variance was performed on the liking scores for each of the other individual characteristics noted on the consumer questionnaire.

2.4.2. Effect of the individual characteristics on the consistency of the liking scores

Five individual statistics were computed (Table 4). STD7 measures the overall dispersion of the hedonic scores given by each consumer. STD3 measures the replication dispersion for each consumer. RF was computed in order to express the consistency of the liking scores in the first test: individual one-way analysis of variance was performed on the liking scores of the 5 variants of juices (using the 3 replicates of the basic juice to estimate the residual error). In such a design, the F-product corresponds to the individual discrimination. The square-root of the individual F-product was computed to obtain RF, which is close to the overall

Table 4
Description of the 5 computed statistics

Name	Statistics	Description
STD7	Standard deviation of the 7 hedonic scores	Overall dispersion
STD3	Standard deviation of the 3 replicated juices	Replication dispersion
RF	Square root of F-product in individual one-way ANOVA	Discrimination
COHER	Coherence between the liking scores and the paired test	0/1 values
RIGHTR	Choice of the right reason for preference in the paired test	0/1 values

dispersion divided by the replication dispersion. Thus, the higher RF, the more consistent the consumer. Indeed, it means that the consumer found major differences between the 5 variants of orange juice and minor differences between the 3 replicates. COHER is the coherence between the liking scores and the results of the paired test: if the consumer chose the juice to which he gave the lower hedonic score, then COHER is equal to 0, if not, it is equal to 1. Finally, RIGHTR was defined as follows: if the consumer chose the right reason for preference in the contrast paired preference test, then RIGHTR is equal to 1, if not, it is equal to 0.

In order to examine the effect of age and gender on the consistency of the liking scores, an analysis of variance was performed on each of these 5 statistics, using the following model:

$$\text{Statistics} = \text{Gender} + \text{Age} + \text{Gender} * \text{Age} + \text{Error}$$

Comparisons of mean liking scores were made visually with confidence intervals at 5%, by age and gender.

Finally, an analysis of variance was performed on the 5 computed statistics for each of the other individual characteristics noted on the consumer questionnaire.

2.4.3. Influence of age and gender on the order effect in the first test

An analysis of variance was performed on the liking scores of the 3 replicates, with the following model:

$$\begin{aligned} \text{Liking score} = & G + A + G * A + C(G * A) + O \\ & + G * O + A * O + G * A * O + \text{Error} \end{aligned}$$

where C corresponds to Consumer, G to Gender, A to Age, O to Order.

In such a design, the sample effect corresponds to an order effect since the 3 samples are the same. Gender and age effects and their interaction were tested against Consumer(Gender * Age), whereas the other effects were tested against the usual error term. Comparisons of mean liking scores were made visually with confidence intervals at 5%, by age and gender, in order to illustrate the order effect.

3. Results and discussion

3.1. Effect of the individual characteristics on the liking scores

Concerning age and gender, the results of the analysis of variance are presented in Table 5. The effects of age and gender on the liking scores are significant, but there is no effect of the age by gender interaction. Examination of the mean scores with confidence intervals at 5% by age and gender (Fig. 1) shows that the youngest and the oldest consumers gave higher scores on the hedonic scale than the other age groups. Approximately the same pattern was revealed in a recent European study: about 900 consumers aged from 20 to 70+ and native from different countries (Denmark, France and Germany) performed a preference test on 12 samples of soup, rice and biscuits. The results showed that for soup (Gourillon et al., 2001) and to a lesser extent for biscuits (Issanchou, 2003), significantly higher scores were obtained from the elderly compared to younger people, whatever the country. However, this pattern is not as obvious when considering rice: in Germany, signifi-

Table 5
Effects of age and gender on liking scores^a

Source	DF	Mean Sq	F-value	Prob
Gender	1	16,071	11.42	0.0008
Age	10	8962	6.37	<0.0001
Gender * Age	10	522	0.37	0.9591
Consumer(Gender * Age)	895	1407		
Sample	6	67,736	119.62	<0.0001
Gender * Sample	6	399	0.7	0.6457
Age * Sample	60	1230	2.17	<0.0001
Gender * Age * Sample	60	692	1.22	0.1174
Residual	5370	566		

^a $R^2 = 0.4$. The effects in the first part of the table are tested against Consumer(Gender * Age), whereas those in the second part are tested against the usual Error term. A bold *F* indicates a significant effect ($p < 0.05$).

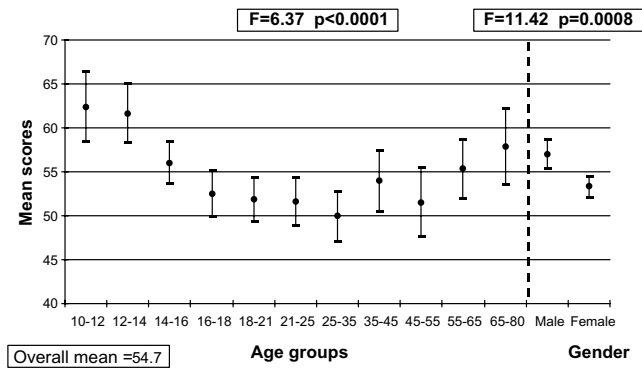


Fig. 1. Age and gender effects on mean liking scores.

cantly higher scores were obtained from the two oldest groups, whereas the opposite situation was observed in Denmark and no age effect was found in France (Simchen et al., 2003). Further analysis showed no impact of the sensory abilities of the consumers on the liking scores. The differences observed between the young and the elderly seemed partly due to different levels of familiarity with the products. However, in Germany, the oldest groups gave higher hedonic scores whatever the product and no relationship was found to their food habits, which may suggest a different way of using the hedonic scale. Concerning biscuits, the author suggested that the differences between young and elderly subjects might be due to a stronger attraction in the elderly for sweetness, or a less negative attitude towards sugar and consequently sweetness on health.

Concerning gender, Fig. 1 shows that the men gave higher scores than the women. Again, this finding is in agreement with those of other studies. Thus, in a European study conducted on coffee (ESN, 1996), hedonic scores were collected in eight different countries for 16 samples of coffee. In each country, 51–85 consumers aged from 16 to 65+ performed the test. There was a fairly large gender effect and a smaller effect on the results due to age. Examining the overall mean scores, it was apparent that on average women tended to score lower than men, and that the 65+ group scored on average significantly higher than the other age groups. Also, Lange, Martin, Chabanet, Combris, and Issanchou (2002) observed that hedonic scores collected for Champagnes in blind conditions were significantly higher for men than for women. However, there was a significant gender by sample interaction: only women undervalued the lowest priced Champagne.

Table 5 also shows that the sample effect is highly significant. This effect is underlined in the internal preference mapping (Fig. 2): most of the consumers are projected on the right side of the map; those consumers gave higher hedonic scores to the basic and the sweet juices than to the acidic, the flavored and especially the bitter juices. However, some consumers preferred the

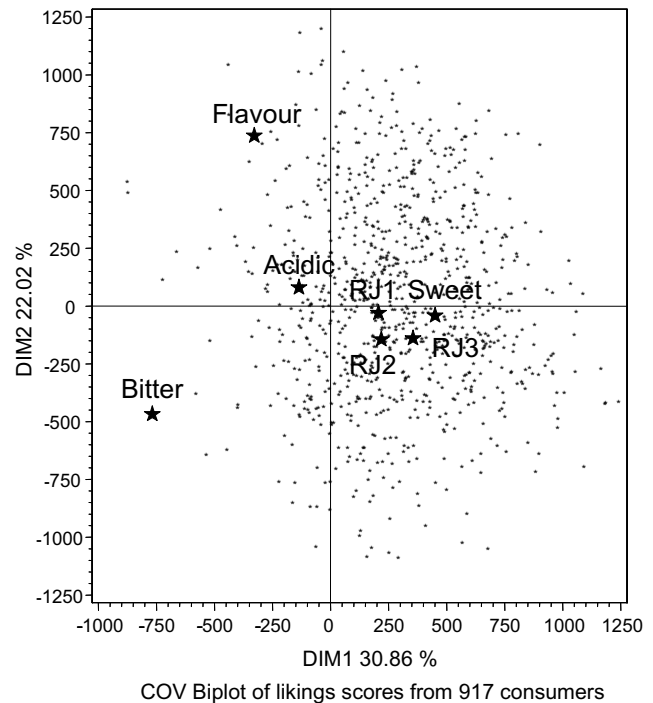


Fig. 2. Internal preference mapping of the liking scores.

flavored juice, whereas a few others preferred the bitter one. Concerning the 3 replicates, it can be noticed that they are quite close on the map, which means that the consumers perceived more pronounced differences among the 5 variants of orange juice than among the 3 replicates.

Finally, Table 5 shows that there is no effect of the gender by sample interaction, but that there is a slight effect of the age by sample interaction. In order to examine this interaction, the individual liking scores were first standardized. Then, the distribution of these standardized liking scores was examined per age group for each of the 7 juices. The results showed that the interaction is mainly due to the fact that the oldest consumers tended to dislike the bitter juice less than the other age groups (Fig. 3) and that the youngest consumers liked the acidic juice more than the other consumers (Fig. 4). The first result may be explained by the hypothesis according to which there is an inborn rejection of bitterness (Coward, 1981) and a development of preferences over one's life span: rejection of bitterness can be overcome due to cognitive mechanisms, social influences and conformity paradigm (Lévy, 1998). Another hypothesis could be the decrease in sensitivity for bitterness with ageing (Weiffenbach, 1984) which may influence preferences and food choice (Rolls, 1993). The second result is in agreement with Liem and Mennella's findings (2003), who demonstrated that acidic taste preferences are heightened during childhood and that such preferences are related to children's food habits.

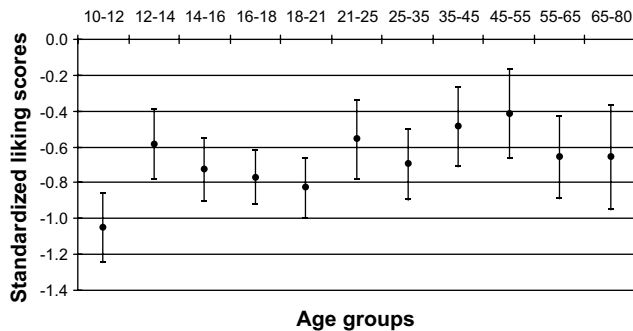


Fig. 3. Distribution of the standardized liking scores per age group for the bitter juice.

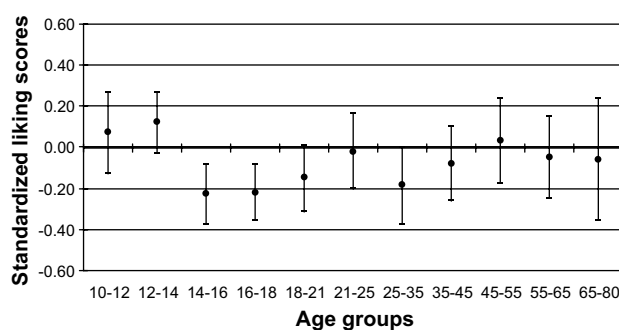


Fig. 4. Distribution of the standardized liking scores per age group for the acidic juice.

The analysis of variance performed on the liking scores for each of the other individual characteristics revealed few significant effects: the number of children in the household seems to have a positive effect on the overall mean liking scores. Also, the consumers whose BMI is higher than 22 (considering only the adults) gave higher scores than the others. Finally, the consumers who frequently consume orange juice (especially juice with no pulp) gave higher scores than the others. A few other characteristics were also significant, but it was difficult to find any satisfactory explanation.

3.2. Effect of the individual characteristics on the consistency of the liking scores

Regarding age and gender, Figs. 5–9 show the results of the analysis of variance on the 5 statistics, with means and confidence intervals at 5% by age and gender. For all statistics but RF, there was absolutely no interaction between age and gender effects. For RF, a small interaction ($p = 0.1517$) was found and interpreted. Concerning STD7 (Fig. 5), it appears that there is a significant effect of age and gender. Looking at the figure, it will be noticed that the overall dispersion decreases when age increases, with the exception of the

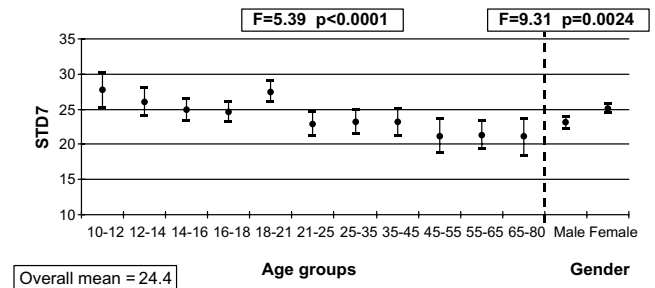


Fig. 5. Age and gender effects on STD7 (overall dispersion).

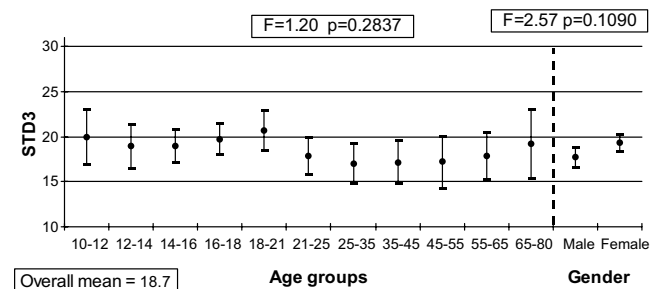


Fig. 6. Age and gender effects on STD3 (replication dispersion).

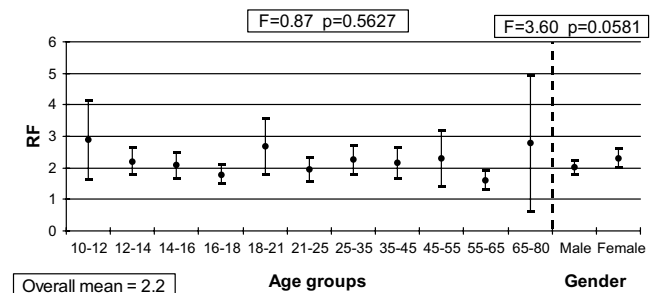


Fig. 7. Age and gender effects on RF (discrimination).

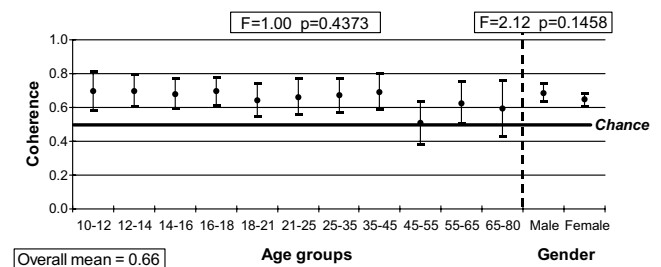


Fig. 8. Age and gender effects on COHER (coherence).

18–21 year old group. Thus, the youngest people seem to use a greater part of the scale than the oldest people. Again, the same pattern was found in the European study (Issanchou, 2003): the individual standard deviation

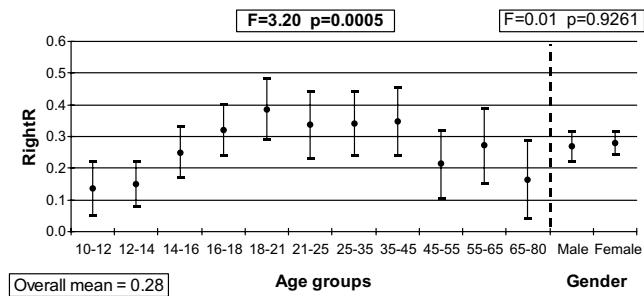


Fig. 9. Age and gender effects on RightR (right reason).

tions of the liking scores decrease with age, whatever the country and the type of product. Moreover, Fig. 5 shows that the overall dispersion is higher for women than for men.

Fig. 6 shows that STD3 tends to be higher for children. However, there is no significant effect of age nor of gender. What is also interesting in the comparison of Figs. 5 and 6 is that STD3 is quite close to STD7, although STD3 is always lower than STD7. This means that the consumers perceived higher differences among the 7 samples than among the 3 replicates, but to a rather small extent. Indeed, the standard deviation of the liking scores given for the 3 replicates represents about 20% of the range of the scale, which is rather large. However, this result cannot be used to draw conclusions about the consistency of the liking scores, as this depends on both overall and replication dispersions.

Consistency is more precisely described by RF. Fig. 7 shows that there is no age effect in average over genders. However, the confidence intervals are greater for the oldest group and for the youngest one. This is due to the smaller number of consumers in these groups and also to the higher standard deviation of RF for these groups. Furthermore, we see that there is a slight gender effect: women obtained a higher RF mean than men, which signifies that women had more consistent preferences than men in the first test. As said above, a slight gender by age interaction ($p = 0.1517$) was detected on RF. Actually, a one-way ANOVA of the age effect by gender was not significant in male ($p = 0.2967$) and significant in female ($p = 0.0966$). Based on contrast tests at $p = 0.05$, it is possible to demonstrate that the age effect in female was due to a higher discrimination in 65–80 (RF = 4.12) and in 10–12 (RF = 3.39). We have no clear explanation for that finding unless the fact that some individuals in these two groups were highly discriminative. Finally, the overall RF mean is 2.2, which means that the overall F-product mean is about 4.8: therefore, the variance of the liking scores from the 5 variants of orange juice is approximately 5 times as large as the variance of the liking scores from the 3 replicates. Although the replication dispersion was rather large

(STD3), the consistency of the liking scores appears to be quite good when taking the overall dispersion (STD7) into account. However, RF is strongly related to the choice of the products tested: when using less similar products, the RF mean would have been higher, and vice versa.

Concerning COHER (Fig. 8), it will be noticed that the means are all above the chance level which is 0.5. Only the 45–55 year old group obtained a coherence mean close to 0.5. However, there is no significant effect of age nor of gender. The total percentage of coherence is 66%, which means that 66% of the consumers chose the sample in the paired test which was consistent with the liking scores they gave in the first test. This value can be compared to those obtained by Schlich, Lespinasse, Navez, and Scandella (2003) and by Lévy and Köster (1999). In the experiment reported by Schlich et al., 361 consumers gave their overall liking scores for 10 varieties of tomatoes. Then, in the same session, they performed a paired preference test composed of the individual least and most preferred products. In that design, the percentage of coherence between the first test and the paired test was 85%. The difference between that coherence and the one obtained in our study is probably due to the fact that, contrarily to the orange juices, the visual aspect of the tomatoes could have helped the consumers being more coherent. In the study conducted by Lévy and Köster, two experiments were carried out: the first one with 3 soft drinks (Product A was a familiar product, whereas Products B and C were unfamiliar reformulations of Product A) and 70 consumers; the second one with 3 alcoholic beverages (Product U was a familiar beverage, Products S and T were unfamiliar ones) and 128 consumers. In each experiment, the subjects had to perform two tests during the same session: in the first test, hedonic scores were collected for the 3 products; in the second test, the consumers received the same products and had to choose the product they preferred (a pack of this product was promised to each consumer). In the first and second experiments respectively, 42% and 32% of the consumers made a final choice different from the one based on their first judgment, which corresponds to coherence values respectively equal to 58% and 68%. Several elements can explain the differences between these values of coherence and the one obtained in our study: first, the difficulty of the task performed by the subjects depends on the type of products, and on the level of differences between the samples. Second, in our study, all the juices were familiar products, contrary to the soft drinks and the alcoholic beverages used in the other study. Finally, in our study, the consumers had to choose one of two samples, whereas in the Lévy and Köster's study, they had to choose one of three samples. And yet, in Lévy and Köster's study as in our study, more than 30% of the subjects switched their preference during a session,

which is rather considerable: these results contribute to show that initial hedonic impression is a poor predictor of final liking (Köster, Couronne, Léon, Lévy, & Marcelino, 2003).

Finally, Fig. 9 illustrates the RightR means: here the age effect is dramatic. RightR increases with age up to 18 years old. Then there is a plateau until age 45. Afterward, RightR tends to decrease with age. Thus, the adolescents and the seniors had more difficulties in identifying the reason for their preference than the other consumers. Regarding the seniors, this result may be due to the loss of sensory abilities that occurs with ageing. Regarding the adolescents, it may be explained either by lower powers of concentration, compared to the other consumers or by an imperfect understanding of the proposed sensory attributes. In terms of gender, there is no significant effect. Besides age and gender effects, it is interesting to examine the RightR average for all the consumers: the percentage of consumers who gave a correct answer is 28%. Even if the chance level is 12.5%, this result is quite low and it confirms the hypothesis that most consumers have generally difficulty in describing what they perceive. Trying to overcome this limitation in consumer vocabulary by presenting them with a list of attributes assumes that all consumers are able to understand the meaning of each attribute and that all consumers interpret the terms in a same way. In order to determine what sort of confusion the consumers made when giving the reason for their preference, the distribution of the answers chosen was drawn up for each expected right answer. A correspondence analysis was then performed on these data (Fig. 10). It first appears that there was mainly a confusion of bitterness

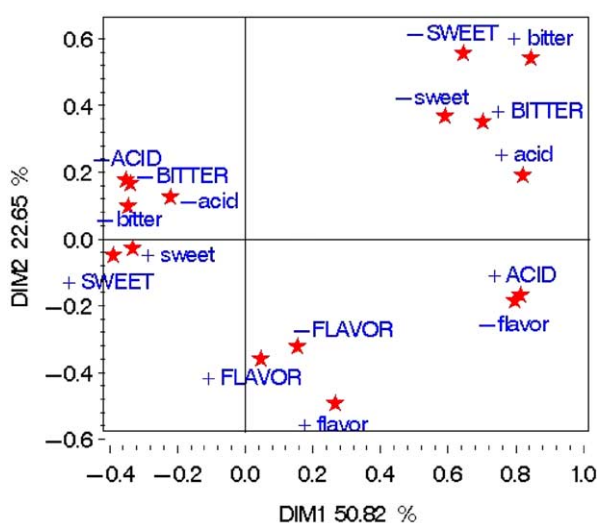


Fig. 10. Correspondence analysis of cross-frequency table between expected correct reasons and chosen reasons for preference in the contrast paired preference test (items in uppercase letters correspond to expected right reasons, whereas items in lowercase letters correspond to chosen reasons).

with acidity: when “more bitter” was the expected answer, some of the subjects answered “more acidic” and vice versa. This result is not surprising as even trained subjects commonly confuse the two terms. Also, for other consumers, it seems that the bitterness and the acidity masked the sweetness and vice versa. Again, this result is well-known for the numerous researchers who studied the interactions among taste compounds. Finally, the other confusions concerned the flavor: when the expected answer was “less flavored”, some consumers answered “more flavored”. These subjects who did not like that specific kind of orange flavor—which induced most of the preference segmentation in Fig. 2 refrained to characterize that juice as more flavored since the word “flavor” has certainly a positive hedonic value for everyone.

The results of the analysis of variance performed on the 5 statistics for each of the other individual characteristics revealed only two slight significant characteristics among the numerous ones noted on the consumer questionnaire: therefore, these observed effects may be due to chance and will not be further considered.

3.3. Further results concerning the two computed statistics, COHER and RIGHTR

Would COHER and RIGHTR depend on the magnitude of the difference between the basic and the other juice to which it was compared in the pair test (this juice will hereafter be called X)? In order to answer this question, a one-way analysis of variance was performed on the two statistics, according to the absolute value of the gap between the basic juice and X in the first test. No effect was found regarding RIGHTR, but it was found that COHER significantly increased when the gap increased (Fig. 11).

Are COHER and RIGHTR higher when X was either preferred or not preferred to the basic juice in the first test? In order to answer this question, a one-way analysis of variance was performed on RIGHTR, according to the sign of the gap between the basic juice

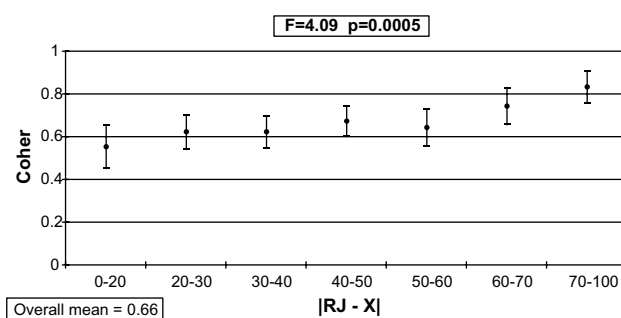


Fig. 11. Effect of the gap between the basic juice and X on COHER (coherence).

and X (sign of “X minus RJ”). The result showed that RIGHTR is significantly higher when the sign is negative, that is, when the basic juice was preferred to X in the first test (Fig. 12): it seems that the consumers had less difficulty to identify the reason for their preference when they had to answer “I prefer that sample because it is less sweet/bitter/acidic/flavored” than when they had to answer “I prefer that sample because it is more sweet/bitter/acidic/flavored”. Concerning COHER, a significant effect of the absolute value of the gap between the basic juice and X was found. As the basic juice obtained quite often higher hedonic scores than the modified juices, the mean gap between the basic juice and the negatively furthest juices was higher (49.9) than the mean gap between the basic juice and the positively furthest juices (36.9). Consequently, an analysis of covariance on COHER was performed, in order to test the effect of $\text{Sign}(X - \text{RJ})$ when taking the effect of the gap into account. The results showed that there is no effect of the subtraction sign on COHER.

Finally, do COHER and RIGHTR vary according to the sensory characteristic of X? As previously, an analysis of covariance on COHER was performed in order to test the effect of the sensory characteristic of X when taking the effect of the gap into account. It was found that this effect was significant ($F = 5.82$, $p = 0.0006$). Table 6 shows the least square means of COHER, that is, in a sense, the means corrected with the effect of the gap between the basic juice and X. It can be seen that coherence was higher when X was the bitter juice. An

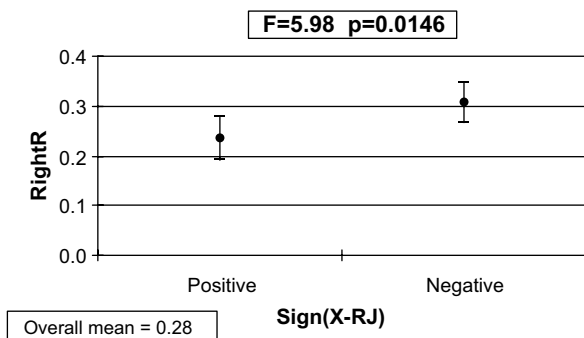


Fig. 12. Effect of $\text{Sign}(X - \text{RJ})$ on RIGHTR.

Table 6

Least square means of COHER according to the sensory characteristic of X

Sensory characteristic of X	LSMean of COHER	Group
Bitter	0.74	a
Flavored	0.64	b
Sweet	0.63	b
Acidic	0.57	b

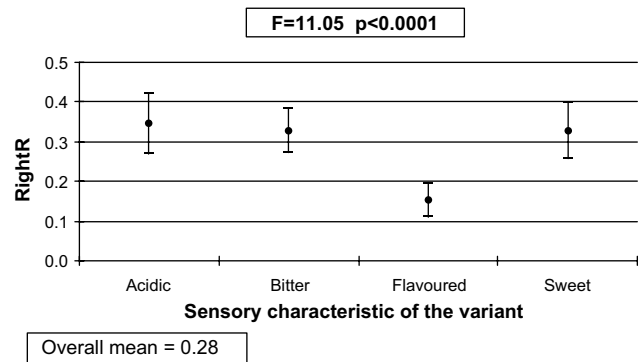


Fig. 13. Effect of the sensory characteristic of X on RIGHTR.

hypothesis could be that the bitter juice was more striking for the consumers who rejected it (or preferred it) in the first test than the other juices rejected or preferred by the other consumers. Concerning RIGHTR, a one-way analysis of variance was performed according to the sensory characteristic of X. Fig. 13 shows that RIGHTR was significantly lower when X was the flavored juice compared to the other tastes. This result was predictable, as consumers are more used to characterize food products using the common taste attributes in their daily life than using flavor attributes.

3.4. Influence of age and gender on the order effect in the first test

Table 7 shows the results of the analysis of variance. As in Table 5, the effects of age and gender on the overall liking scores for the 3 replicates are significant. These effects were previously explained in Section 3.1. Table 7 also shows that the order effect is highly significant. This means that the liking scores of the 3 replicates are significantly different, which underlines the order effect. Fig. 14 shows that the liking scores were higher for the third and to a lesser extent for the first replications than for the second one. The first position bias effect has been described by several authors (Earthy, MacFie, & Wakeling, 1995; MacFie, Earthy, & Deliza, 1996; Peyram & Pilgrim, 1957), but the last position bias effect observed in the study is less well known. This effect may be due to a feeling of weariness of the consumers at the end of the test and which express itself by an overvaluation of the last sample. In addition and not negligibly, this effect may also be due to a contrast effect towards the bitter juice—in average, the least preferred sample—presented next to last. Finally, Table 7 shows that there is no significant effect of the gender by order and the age by order interactions, which means that age and gender had no influence on the order effect and that our results about age and gender effects were hopefully not affected by the fixed order of presentation.

Table 7
Effects of age and gender on the order effect^a

Source	DF	Mean Sq	F-value	Prob
Gender	1	4065	4.66	0.0312
Age	10	6167	7.07	<0.0001
Gender * Age	10	1095	1.25	0.252
Consumer(Gender * Age)	895	873		
Order	2	6513	14.15	<0.0001
Gender * Order	2	152	0.33	0.7188
Age * Order	20	391	0.85	0.6522
Gender * Age * Order	20	191	0.41	0.9895
Residual	1790	460		

^a $R^2 = 0.5$. The effects in the first part of the table are tested against Consumer(Gender * Age), whereas those in the second part are tested against the usual Error term. A bold F indicates a significant effect ($p < 0.05$).

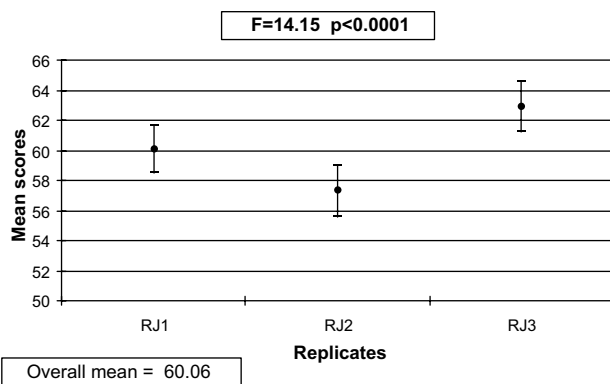


Fig. 14. Order effect on the mean liking scores of the 3 replicates.

4. Conclusion

To conclude on age and gender effects, the results of this experiment revealed that, regarding liking patterns, most of the consumers preferred the basic and the sweet juices over the other juices, that the oldest consumers disliked the bitter juice less than the other consumers and that the youngest consumers liked the acidic juice more than the other consumers. Regarding consistency, women were slightly more consistent over replicates within a session than men. Moreover, children and seniors had more difficulties than the other consumers in identifying the reason for their preference.

Besides age and gender effects, the contrast paired preference test notably suggested that coherence in liking evaluation was less difficult with bitterness compared to the other tastes. It also suggested that being able to name the reason for a preference was more difficult with flavor than with taste.

It could be now interesting to enlarge this study at other age groups: notably, the study should be reproduced with younger children, in order to confirm the tendencies observed in the liking patterns (noticeably, a higher liking for acidity) and in the results of the contrast paired preference test. In addition, it could be also interesting to reproduce that study in other countries,

since cultures may well influence the acceptance of bitterness for instance and the extent to which people are able to name the different basic tastes.

Acknowledgements

This research was conducted in the context of “Fête de la Science”, supported by CNRS. Thanks are also due to Eckes Granini Company for providing us with juices, and to Symrise Company for providing us with the flavoring. This study has also benefited from helpful discussions with Sylvie Issanchou. Finally, such a large experiment could not have been conducted without the help of numerous people at CESG who offered great help and who are deeply thanked.

References

- Cowart, B. J. (1981). Development of taste perception in humans: sensitivity and preference throughout the life span. *Psychological Bulletin*, 90(1), 43–73.
- Earthy, P. J., MacFie, H. J., & Wakeling, I. (1995). Effect of presentation orders on the internal preference mapping of 12 fragrances. *Appetite*, 24(3), 300.
- ESN (1996). A European sensory and consumer study. A case study on coffee. European Sensory Network, Chipping Campden, Gloucestershire.
- Gourillon, S., Issanchou, S., Schlich, P., Hansen, G., Kramer Vig, K., Nielsen, R., Hoyer, S., Seppelt, B., Simchen, U., Sheehan, E., & Delahunty, C. (2001). Impact of age and culture on food liking in three European countries. In The 4th Pangborn Sensory Science Symposium, Dijon, France, 22–26 July.
- Issanchou, S. (2003). Changing food preferences with ageing. Oral presentation in the workshop *How do age-related changes in sensory physiology influence food liking and intake, and what are the practical implications for the food industry?* The 5th Pangborn Sensory Science Symposium, Boston, 20–24 July.
- Köster, E. P. (1990). Les épreuves hédoniques. In SSHA & ISHA, *Evaluation Sensorielle: manuel méthodologique* (pp. 171–193). Lavoisier, Paris.
- Köster, E. P., Couronne, T., Léon, F., Lévy, C., & Marcelino, A. S. (2003). Repeatability in hedonic sensory measurement: a conceptual exploration. *Food Quality and Preference*, 14(2), 165–176.

- Lange, C., Martin, C., Chabanet, C., Combris, P., & Issanchou, S. (2002). Impact of the information provided to consumers on their willingness to pay for Champagne: comparison with hedonic scores. *Food Quality and Preference*, 13(7–8), 597–608.
- Lévy, C. (1998). Elaboration de methodologies adaptées à l'étude de la dynamique des preferences. Application aux boissons. Thèse de doctorat, Université de Bourgogne.
- Lévy, C., & Köster, E. P. (1999). The relevance of initial hedonic judgements in the prediction of subtle food choices. *Food Quality and Preference*, 10(3), 185–200.
- Liem, D. G., & Mennella, J. A. (2003). Heightened sour preferences during childhood. *Chemical Senses*, 28(2), 173–180.
- MacFie, H. J., Earchy, P., & Deliza, R. (1996). *Measurement of consumer preference*. The Netherlands: Freeman, M., Maarsen.
- Peyram, D. R., & Pilgrim, F. J. (1957). Hedonic scale method of measuring food preferences. *Food Technology*, 11(9), 9–14.
- Rolls, B. J. (1993). Appetite, hunger and satiety in the elderly. *Critical Reviews in Food Science and Nutrition*, 33(1), 39–44.
- Schlich, P., Lespinasse, N., Navez, B., & Scandella, D. (2003). Preference mapping of tomato varieties from the French market. A PrefMax consumer segmentation validated by a two-year study. Oral presentation. The 5th Pangborn Sensory Science Symposium, Boston, 20–24 July.
- Simchen, U., Gourillon-Cordelle, S., Issanchou, S., Schlich, P., Chabanet, C., Hoyer, S., Zunft, F., Jeppsen, L., Kramer Vig, K., Sheehan, E., & Delahunty, C. (2003). Contribution of culture, gender and age to preferences to different rice samples. In The 5th Pangborn Sensory Science Symposium, Boston, 20–24 July.
- Weiffenbach, J. M. (1984). Taste smell perception in aging. *Journal of Gerontology*, 3, 137–146.

Article 12

12. Rousset S., Deiss V., Juillard E., Schlich P. & Droit-Volet S. (2005). Emotions generated by meat and other food products in women. *British Journal of Nutrition*, **94**(4), 609-19.

Emotions generated by meat and other food products in women

S. Rousset¹*, V. Deiss², E. Juillard², P. Schlich³ and S. Droit-Volet⁴

¹*Institut National de la Recherche Agronomique, Laboratoire de Nutrition Humaine, BP 321, 58 rue Montalembert, 63009 Clermont-Ferrand Cedex 1, France*

²*Institut National de la Recherche Agronomique, Station de Recherches sur la Viande, Theix, 63122 Saint Genès Champanelle, France*

³*Centre Européen des Sciences du Goût, 15 rue Hugues Picardet, 21000 Dijon, France*

⁴*Laboratoire de Psychologie Sociale et Cognitive, UMR 6024 CNRS, 34 Avenue Carnot, 63000 Clermont-Ferrand, France*

(Received 2 November 2004 – Revised 3 June 2005 – Accepted 6 June 2005)

Eating behaviour depends partly on food preference, which is itself determined by different types of emotions. Among the emotions generated by food, disgust with red meat is common in women and can lead to reduced meat consumption. We tested the hypothesis that low meat intake is related to different negative emotions towards meat but does not affect the emotions expressed towards other food categories. Food intake of sixty women was followed throughout each day for 1 week and allowed us to assign women to two groups (low v. high meat-eating women). They were then invited to assess the intensity of twenty-six emotions described by words and induced by thirty food pictures. We determined the number of necessary dimensions to describe the space created by the twenty-six words. The results showed differences in emotions between the low and high meat-eating women. As expected, there were overall differences in the emotions generated by the thirty food pictures. Six clusters of emotions were necessary and sufficient to summarise the emotional space. These dimensions were described by 'disappointment', 'satisfaction', 'guilt', 'doubt', 'amused' and 'indifference'. As expected, the low meat-eating women felt more 'disappointment', 'indifference' and less 'satisfaction' towards meat than did the high meat-eating women. However, the low meat-eating women also stated other negative emotions such as 'doubt' towards some starchy foods. The only foods that they liked more than high meat-eating women were pears and French beans. In conclusion, low meat consumption was associated with specific negative emotions regarding meat and other foods.

Emotion: Dietary surveys: Meat consumption: Women

A previous nutrition study showed that about 25 % of French women are Fe deficient (Galan *et al.* 1998). This deficiency is due to an Fe intake lower than the minimum requirement. Several studies have shown that women ingest smaller quantities of Fe than men (Ortega *et al.* 1998). This coincides with the fact that meat consumption also tends to be higher in men (Ortega *et al.* 1998; Rousset *et al.* 2003). Indeed, meat and meat products are important sources of Fe, especially haeme Fe, which is easier to absorb than non-haeme Fe found in plants (Hulten *et al.* 1995; Shaw *et al.* 1995). However, Mooney & Walbourn (2001) showed that meat was the most commonly avoided food in female students. Thus, about half of the female students surveyed reported avoiding meat, and were four times more likely to avoid meat than vegetables. Even though women might have negative attitudes towards meat and red meat, in particular, and avoid consuming it, meat is well adapted for their needs because it is Fe-rich and they have high requirements in this micronutrient (16 mg/d).

The physical characteristics of food are important for determining the hedonic response (Kubberod *et al.* 2002). In the case of meat, tenderness is the main sensory criterion of acceptance within a species such as beef (Chambers & Bowers, 1993). Conversely, beef toughness is well known to raise objections.

However, when comparing meat of different species, appearance and colour (white and red) become important factors in discriminating consumer liking. Red meat colour, in particular, arouses strong emotion, depending on consumer gender. Men feel hedonic pleasure in seeing and eating red meat while women experience discomfort. The main reason reported by women for this is that the sight of blood is perceived as repulsive (Kenyon & Barker, 1998; Kubberod *et al.* 2002). Gregory (1997) and Guzman & Kjaernes (1998) have claimed that white meat such as chicken is less likely to be associated with blood or a living animal due to its white appearance and therefore is preferred to red meat.

Worsley & Skrzypiec (1998) and Lea & Worsley (2001, 2002) have shown not only the influence of pleasant feelings but also of beliefs on red meat consumption. Thus, the best predictor of the frequency of meat intake is the pleasant feelings experienced by young consumers. Conversely, negative beliefs towards meat decrease meat consumption. When red meat is thought to be fatty and unhealthy, consumers reduce or stop their meat consumption (Holm & Mohl, 2000; Lea & Worsley, 2001). Likewise, in the case of seafood, Olsen (1999) showed that attitudes towards eating this food, measured by endorsement of taste, explained the variation in the frequency of seafood consumption better than

health considerations or perceived convenience. Finally, certain attitudes towards meat as well as other objects are important determinants of behaviour. Psychological studies have shown that the consistency between attitude and behaviour is stronger when the subject has had direct experience with the object of specific attitudes (Regan & Fazio, 1977; Fazio & Zanna, 1981). Furthermore, the affective more than the cognitive component of the attitude is related to hedonic behaviour (Millar & Tesser, 1986, 1992). For example, meat-liking influences eating behaviour more than knowledge of meat-related nutritional and health characteristics.

The affective component of attitudes refers to emotions generated by a specific object or event. Emotions involve a set of interacting components and are commonly treated as a multifaceted phenomenon consisting of components such as behavioural reactions (action tendency), expressive reactions (facial expressions), physiological reactions (pulse rate, blood pressure) and subjective feelings. Moreover, each emotion is characterised by its valence and intensity. The valence is the hedonic quality of experience and deals with positive or negative evaluations of the emotion. Intensity is the magnitude (weak *v.* strong) of emotional perception. Because emotions are internal states that cannot be directly measured, many studies have used words that an individual uses to label his or her emotional states (Ekman *et al.* 1983; Russell, 1991; Baars, 1994; Scherer, 1998, 2005; Searle, 1998; Cabanac, 2002; Niedenthal *et al.* 2004).

The process consisting of labelling emotions is typical of man and exceeds physiological reaction. This process involves cognition and represents an active rather than a reactive role of the individual concerned (Leyens *et al.* 2000). In fact, feelings are the subjective experience of present or remembered emotional excitation, often conscious and communicated by words or emotional expressions. Feelings are exclusively studied by verbal labels (Scherer, 1998). Recently, Niedenthal *et al.* (2004) developed a large database on the French emotion lexicon in order to facilitate future research. Thus, 237 words denoting emotional states were translated into French from an initial list of Italian words used by Zammuner (1998). In the present study, valence and intensity for each emotion were assessed. Subjects determined that 'to like' has a positive valence while 'dislike' has a negative valence. Among words with a positive valence, 'enchantment' has a higher intensity than 'interest'. In a previous study (Juillard, 2003; V Deiss, E Juillard, S Droit-Volet & S Rousset, unpublished results), twenty-six emotional words were selected from the 237-word list to describe women's and men's emotions towards thirty different pictures of foods including vegetables, cheese, fruits, cereals, red and white meat, seafood, eggs and snacks (SU.VI.MAX, 1994). The results showed that there were differences in emotional responses to foods; high disgust and uneasiness towards offal such as kidney, tongue, black pudding, and great pleasure towards fish, pasta, chocolate and tarts. Understanding specific emotions towards nutrient-rich food in populations at risk of nutrient deficiency should help nutritionists and food companies to create and/or to propose specific liked food products, or to inform the population about the beneficial effects of consuming certain foods to avoid deficiency.

The aim of the present study was first to determine emotions associated with food pictures in young women, to categorise their emotional words, and to link their level of meat consumption to emotions produced by thirty different food pictures. Positive emotions towards meat products were expected to be related to

higher meat consumption. Conversely, it was assumed that negative emotions would be more closely associated with meat products in low meat-eating women. Moreover, we hypothesised that low meat-eating women might transfer their positive emotions to other food products.

Methods

Participants

In January 2003, a convenience sample was recruited in Clermont-Ferrand, France, using advertising in the press (the local newspaper, *La Montagne*, *Info Magazine*, FR3 TV) and displays in stores (supermarkets). The advertisement invited women between 20 and 40 years of age to participate in a survey dealing with eating habits. Suitable subjects for age and availability were selected among the women who had replied by telephone; 86 % of them actually agreed to complete the dietary survey and to assess food pictures. Sixty women with a mean age 29 (SD 6) years participated in the survey. They were not receiving medical treatment for any progressive illness.

Experimental design

Dietary survey. The consumption of foods at each meal was assessed by a questionnaire during seven consecutive days in winter 2003. The questionnaire, which was accompanied by instructions for completing it, comprised six headings, subdivided into several sub-headings: breakfast (beverages, bread, dairy products, jam or honey spread, sugar, other foods); a snack in the morning (beverage, sugar, biscuits or cakes, chocolate, other foods); lunch (aperitif, starter, sandwich, eggs, fish or sea food, meat, offal, ready-made meal, vegetable, pulses, starchy foods, potato, sauce, dairy products, dessert, bread, beverage, sugar, other foods); snack in the afternoon (same items as the morning snack); dinner (same as lunch); snack in the evening (same as the morning snack).

The quantities of each food were assessed by the subjects themselves from photographic standards, specifically designed, validated and used in a large French epidemiological study (Guinot *et al.* 2001). All participants were contacted after having received the questionnaire and before beginning the study to make sure that they did not encounter any difficulty in completing it. Fe intakes from meat were determined by using EPIBASE, version 1.1 H (1999). Participants were assigned to two groups on the basis of average meat consumption (1.6 mg meat Fe/d); low meat-eating and high meat-eating women.

Emotional assessment of food pictures. In our experiment, thirty pictures (A5 standard) were used, all taken from a specific book (SU.VI.MAX, 1994). Pictures consisted of vegetables (salad, French beans, potatoes, cauliflower, spinach), cheeses (Camembert, Roquefort), fruits (pears, bananas), starchy foods (bread, rice, pasta, pizza), sweets (fruit tarts, chocolate, pound cake, cream cake, i.e. cake having layers held together by a sweet filling and covered with frosting), red meat (roast beef, minced meat), white meat (turkey, rabbit, pork chop), fish (fillet, sea food), offal (tongue, kidneys, black pudding), eggs (omelette), processed pork (dry sausage) and pan bagnat (sandwich with tomato, egg, lettuce and tuna fish, commonly eaten in France).

The twenty-six key words for emotions towards foods were selected from the database of Niedenthal *et al.* (2004) that

includes 237 emotional French words applied to any event that had been evaluated for its emotional prototypicality on the basis of different indicators including valence and intensity of emotional state (Juillard, 2003; V Deiss, E Juillard, S Droit-Volet & S Rousset, unpublished results). In a pre-test, participants (different from the present study) were first asked to judge the relevance of each of the 237 emotional words to describe their emotions for food consumption, regardless of the food. Only words quoted by at least 50 % of the participants were selected. Seventy words met this criterion. Among those words, only twenty-six were statistically discriminant and not redundant (Juillard, 2003; V Deiss, E Juillard, S Droit-Volet & S Rousset, unpublished results). Among these words, there were thirteen words with a positive valence: 'to like'; 'content'; 'vigilant'; 'pleasure'; 'thrilled'; 'rejoicing'; 'satisfaction'; 'surprise'; 'serene'; 'amused'; 'delight'; 'pride'; 'interest'. There were thirteen other words with a negative valence: 'disgust'; 'hesitation'; 'indifference'; 'guilt'; 'uneasiness'; 'nostalgia'; 'impatience'; 'doubt'; 'frustration'; 'embarrassment'; 'disappointment'; 'lassitude'; 'regret'.

Each woman was presented with each of the thirty food pictures, in random order. Participants took part in two testing sessions at the laboratory, each between 11:00 hours and 14:00 hours or between 18:00 hours and 08:00 hours. Each woman assessed the intensity of twenty-six emotional words that were aroused by thirty food pictures at two different times, at a 1-week interval. Women were instructed to imagine actually having the food in front of them and being invited to consume it. This imagery task has been shown to successfully arouse emotions (Smith & Lazarus, 1993; Scherer, 1998). They then had to mark the intensity of each emotional word on a five-point scale ('0', I do not feel this emotion; '1', I feel it a little; '2', I feel it moderately; '3', I feel it a lot; '4', I feel it very strongly).

Statistical treatments

The present study generated 3600 observations (sixty participants; thirty pictures; two replicates). A multiple ANOVA (MANOVA) was first carried out on these observations to test the picture effect on all the participants. The MANOVA model was a two-way mixed model (picture and subject) in which the subject was declared as a random effect. A canonical variate analysis (CVA) showed the direction of emotional vectors, which contribute to the construct of the two main discriminant axes and the location of food, limited by their 90 % confidence ellipse on a plot. A likelihood ratio test determined the number of necessary and significant dimensions to show the major part of this discriminant emotional space. A clustering algorithm of variables was then used to classify the twenty-six emotional words into the number of clusters determined by the likelihood ratio test. For each cluster of words, a representative was chosen in order to maximise the correlation coefficient between itself and the average of the words from that cluster. The means of emotional variables within each cluster were then computed and identified by the variable which best represented the cluster as the 'mean emotional variables'. A principal component analysis (PCA) computed on the covariance matrix was carried out with the differences in the mean emotional variables between the low and high meat-eating women. This analysis showed the directions of emotional vectors which best discriminated the two types of par-

ticipants. Thus, the longer an arrow variable, the more discriminant the word was. PCA also showed the location of food products in this emotional space. A virtual product identified by a diamond symbol on the map with null coordinates for each mean emotional variable was superimposed on the PCA plot to show the location for which no difference was detected between low and high meat-eating women. Thus, the orthogonal projection of this point on the emotional vector discriminated the food into two groups; foods arousing more emotion in low or in high meat-eating women. In other words, when the orthogonal projection of a food picture on an arrow variable was located in the vector direction and well in front of the diamond projection, low meat-eating women felt more emotion towards that food. Finally, for each food product, a *t* test was performed to determine the effect of the group (low v. high meat-eating women) on the mean emotional variables.

Computation was done with the SAS system (SAS Institute Inc., 1998). The MANOVA and CVA were obtained using the GLM procedure with the MANOVA statement and the CANONICAL option. The variable clustering algorithm used was given by the VARCLUS procedure with the CENTROID option. The PCA was computed using the PRINCOMP procedure. The graphs from the CVA and PCA were plotted using specific SAS macros developed at the Institut National de la Recherche Agronomique (Dijon, France).

Results

Diet and social characteristics

Participants consumed meat with a mean frequency of six times per week. The total Fe intake varied between 6.2 and 19.3 mg/d and the mean was 10.8 mg/d. Since the recommended allowance is 16 mg Fe/d, 96 % of the participants had a lower intake than the French recommendations. Moreover, Fe intake from meat varied from 0 to 9.6 mg/d. Thus, thirty women had an intake lower than 1.6 mg/d (low meat-eating group) and thirty women had an intake higher than 1.6 mg/d (high meat-eating group). The meat consumption frequency among women in the high meat-eating group was double that of those in the low meat-eating group (7.2 v. 3.6; Table 1). However, Fe and energy intakes were similar in both groups ($P=0.25$; $P=0.84$). The energy intake was in accordance with the recommended allowances of 7530 kJ (1800 kcal)/d (Table 1). The low meat-eating women tended to have a higher energy intake from sweets ($P=0.06$) and vegetables ($P=0.13$) while the high meat-eating women consumed more energy from meat products ($P<0.0001$) and processed pork ($P=0.002$; Table 1).

The mean height of the subjects was 1.64 m. Their mean weight was 57.0 kg. The BMI was 21.1 kg/m². Of the young women, 8 % were overweight (BMI > 25 kg/m²) while 13 % were underweight (BMI < 18.5 kg/m²). No differences in height, weight and BMI distinguished the two groups of participants. Weight varied from 42 to 76 kg in low meat-eating women and from 40 to 70 kg in high meat-eating women. BMI ranged from 15.6 to 27.3 kg/m² in the low meat-eating group, and 17.0 to 26.2 kg/m² in the high meat-eating group.

The participant sample was composed of students (37 %), business employees (20 %), office workers (18 %), managers (11 %), artisans (4 %), unemployed (4 %) and homemakers (6 %). There was no difference in the participants' occupations between the two groups.

Table 1. Characteristics of women and their food consumption in low and high meat-eating groups
(Mean values and standard deviations)

	Low* (n 30)		High (n 30)		Statistical significance (P)
	Mean	SD	Mean	SD	
Age (years)	29.2	6.1	28.8	6.7	0.81
Height (m)	1.64	0.06	1.64	0.07	0.8
Weight (kg)	57.2	8.1	56.5	7.0	0.74
BMI (kg/m ²)	21.3	3.1	20.9	2.2	0.56
Meat consumption (frequency/week)	3.6	2.0	7.2	2.6	<0.0001
Total Fe intake (mg/d)	10.4	3.2	11.3	2.6	0.25
Meat Fe intake (mg/d)	0.9	0.5	3.0	1.5	<0.0001
Energy intake (kJ/d)	7481.4	1661.5	7385.6	1948.1	0.84
Meat products (kJ/d)	186.6	166.1	605.2	300.5	<0.0001
Processed pork (kJ/d)	144.3	162.3	354.2	324.5	0.002
Fish and seafood (kJ/d)	164.4	198.3	158.1	156.3	0.89
Fruit (kJ/d)	1080.3	741.5	892.8	573.0	0.27
Vegetables (kJ/d)	318.8	263.2	231.4	168.8	0.13
Sweets (kJ/d)	398.3	390.4	241.0	276.2	0.06
Pastries (kJ/d)	1425.9	1040.2	1205.9	941.5	0.39
Dairy products (kJ/d)	1192.4	642.7	1024.8	545.2	0.28
Eggs (kJ/d)	148.8	126.6	198.7	233.8	0.30
Ready-made meals (kJ/d)	752.4	557.2	648.4	496.3	0.45
Starchy foods (kJ/d)	1540.1	706.7	1664.9	723.7	0.5
Alcohol (kJ/d)	123.8	143.9	160.2	216.5	0.45

How food products were discriminated by emotions

MANOVA showed that there were overall significant differences in the emotions generated by the thirty food pictures ($P < 0.0001$). The emotional terms that best discriminated the food pictures were: 'disgust'; 'to like'; 'pleasure'; 'guilt'; 'uneasiness'; 'vigilant'; 'content'; 'doubt'; 'satisfaction'; 'delight'. Nevertheless, all the terms were significant ($P < 0.0001$; Table 2). The emotions felt towards these food products with the greatest intensity were 'to like'; 'pleasure'; 'content'; 'serene'; 'thrilled'; 'satisfaction'; 'delight'; 'rejoicing'; 'interest'. Thus, emotions towards food consumption were mainly positive. However, in the present study, the intensity of emotions was moderate since the mean score was never higher than '2' except for the emotional word 'to like'. However, as we can see in Table 3, differences in emotional intensity between food products were high, as shown by three examples: 'to like'; 'disgust'; 'guilt'.

These differences in the emotions generated by the thirty food pictures were shown on the main plot of the CVA (Fig. 1). The first two axes explained 64 % of the variance. The first discriminant axis shows a contrast between 'disgust', 'disappointment', 'doubt', 'uneasiness', 'vigilant', 'embarrassment' and 'serene', 'interest', 'to like', 'satisfaction', 'content', 'thrilled' and 'pleasure'. The second axis was mainly explained by 'guilt' and, to a lesser extent, by 'amused'. The confidence ellipses around picture means showed the variability of emotions reported by the participants. When the confidence ellipses did not overlap, emotions towards food pictures were different. The location of food on this mapping shows that participants were disgusted by offal such as kidneys, tongue and black pudding. Disgust was a strong negative emotion aroused first by kidneys (2.5) and second by tongue and black pudding (2.1 and 1.9; Table 3). Disgust for beef and other meats was low (1.3). Potato, salad, bread, French beans, rice, pasta and fish aroused the lowest level of intensity of disgust (0.1). Conversely, participants liked chocolate, pasta, tarts, salad, fish, bread, pizza and potatoes

(Fig. 1 and Table 3). Their mean scores of 'to like' were higher than 2.5. Obviously, the participants did not like offal very much. Thus, the scores for kidneys, tongue and black pudding were low (between 0.6 and 1.0; Table 3). The least liked foods after offal were red meats followed by white meats. Spinach,

Table 2. Rank of emotional words in decreasing order of picture discrimination
(Mean values and standard deviations)

	F value*	Mean	SD
Disgust	26.2	0.6	1.2
To like	19.3	2.1	1.2
Pleasure	18.8	2.0	1.3
Guilt	18.7	0.4	0.9
Uneasiness	16.9	0.5	1.0
Vigilant	16.2	0.7	1.1
Satisfaction	15.9	1.8	1.3
Content	15.6	1.8	1.3
Delight	15.4	1.7	1.3
Doubt	15.4	0.5	1.0
Thrilled	14.5	1.7	1.3
Rejoicing	14.5	1.6	1.3
Embarrassment	14.0	0.5	1.0
Serene	13.9	1.7	1.4
Interest	13.7	1.6	1.2
Disappointment	11.5	0.6	1.2
Impatience	8.8	0.9	1.2
Hesitation	8.4	0.6	1.1
Amused	7.3	0.5	0.9
Indifference	6.7	0.7	1.2
Astonishment	6.5	0.3	0.7
Regret	6.3	0.5	1.0
Pride	4.9	0.4	0.9
Frustration	3.9	0.4	0.9
Lassitude	3.8	0.4	0.8
Nostalgia	2.8	0.5	1.0

*All the F values of the Fisher test were significant ($P < 0.0001$).

Table 3. Three examples of emotion towards the thirty food pictures
(Mean values and standard deviations)

To like			Disgust			Guilt		
Food picture	Mean	SD	Food picture	Mean	SD	Food picture	Mean	SD
Chocolate	2.9	1.1	Kidney	2.5	1.7	Cream cake	1.2	1.5
Pasta	2.8	0.9	Tongue	2.1	1.8	Chocolate	1.2	1.4
Tart	2.8	1.0	Black pudding	1.9	1.7	Tart	1.0	1.4
Salad	2.7	0.9	Beef	1.3	1.6	Pizza	1.0	1.4
Fish	2.7	1.0	Seafood	1.0	1.4	Pound cake	0.9	1.3
Bread	2.7	0.9	Minced meat	1.0	1.5	Dry sausage	0.9	1.4
Pizza	2.6	1.0	Rabbit	0.9	1.4	Pan bagnat	0.7	1.1
Potato	2.6	0.9	Spinach	0.8	1.5	Camembert	0.6	1.1
French beans	2.5	0.8	Cream cake	0.8	1.2	Roquefort	0.5	1.0
Roquefort	2.4	1.2	Pork chop	0.7	1.2	Bread	0.5	0.9
Pear	2.4	1.1	Dry sausage	0.7	1.3	Pasta	0.4	0.8
Rice	2.4	0.8	Omelette	0.5	1.1	Rabbit	0.4	0.9
Banana	2.3	1.2	Camembert	0.5	1.2	Beef	0.3	0.8
Pound cake	2.3	1.1	Turkey	0.5	0.9	Pork chop	0.3	0.8
Camembert	2.2	1.2	Roquefort	0.5	1.2	Tongue	0.3	0.9
Pan bagnat	2.2	1.0	Banana	0.4	1.0	Kidney	0.3	0.9
Omelette	2.1	1.0	Cauliflower	0.3	0.8	Potato	0.3	0.7
Cauliflower	2.0	0.9	Pan bagnat	0.2	0.6	Banana	0.3	0.7
Seafood	2.0	1.4	Pear	0.2	0.8	Black pudding	0.2	0.8
Dry sausage	2.0	1.3	Quatre quart	0.2	0.6	Rice	0.2	0.5
Cream cake	2.0	1.5	Chocolate	0.2	0.7	Omelette	0.2	0.6
Turkey	1.9	1.0	Pizza	0.2	0.6	Minced meat	0.2	0.5
Rabbit	1.8	1.3	Tart	0.1	0.4	Turkey	0.1	0.4
Spinach	1.6	1.4	Fish	0.1	0.4	Seafood	0.1	0.3
Pork chop	1.6	1.1	Pasta	0.1	0.4	Spinach	0.0	0.3
Minced meat	1.6	1.1	Rice	0.1	0.4	Fish	0.0	0.2
Beef	1.5	1.3	French beans	0.1	0.3	Cauliflower	0.0	0.2
Black pudding	1.0	1.2	Bread	0.1	0.3	French beans	0.0	0.2
Tongue	1.0	1.2	Salad	0.1	0.3	Pear	0.0	0.2
Kidney	0.6	1.1	Potato	0.0	0.2	Salad	0.0	0.1

the least popular vegetable, was ranked near meat products. Other vegetables, both cheeses, both fruits, eggs, both cakes, rice, dry sausage and seafood obtained intermediate scores.

Guilt was the second most important category of negative emotion, independent of 'disgust'. However, its mean score was low; 0.4 (Table 2). First, cream cake and chocolate followed by fruit tart, pound cake, dry sausage and pizza evoked guilt (Table 3). Conversely, participants felt relatively little guilt about eating meat. Finally, salad, pears, French beans, cauliflower, fish and spinach caused no guilt. These food products were perceived with neutral or positive emotions such as 'to like'.

Discriminant power of the emotional words and dimensionality of the emotional space

The likelihood ratio test determined that a six-dimension emotional space was sufficient to differentiate between these pictures. The dimensions necessary to discriminate between the food pictures were described by 'disappointment', 'satisfaction', 'nostalgia', 'doubt', 'amused' and 'indifference'. Each of these words were best correlated with the average of the words from clusters 1, 2, 3, 4, 5 and 6 (Table 4). However, in the case of 'nostalgia', we preferred 'guilt' as representative of this dimension because this word was seven times more discriminant than 'nostalgia' (Table 2) and because the correlation coefficients between those words and the average of those two words were similar (0.76 and 0.80, respectively).

The negative emotions were scattered into four different clusters while positive emotions were found in only two clusters. Thus, negative emotional words were more variable than positive emotions because they expressed emotions as different as 'disappointment', 'guilt', 'doubt' and 'indifference'. In regards to negative emotions, the first cluster, 'disappointment', was composed of five words with 'disgust' being the most discriminant word (Table 2). The second negative cluster, 'guilt', contained 'nostalgia' and 'guilt'. The last word dealt with a strong negative emotion. The third one, 'doubt', contained four words, and 'uneasiness' was the strongest negative word among them. 'Indifference' and 'lassitude' composed the last cluster of negative emotions. Those last two words were less discriminant than most of the others. In regards to positive emotions, 'satisfaction' was the first representative of positive emotions and was expressed by 'to like', 'pleasure', 'content' and 'delight'. Among those words, 'to like' best discriminated food pictures (Table 2). The second positive representative was 'amused' that means 'to have fun' with a touch of 'astonishment' and 'pride'. However, this last cluster was less discriminating than that of 'satisfaction'.

Difference in disgust and liking between low and high meat-eating women

Fig. 2 showed the most significant differences in scores of disgust and 'to like' towards the thirty food pictures given by low and high meat-eating women. Differences in disgust between low

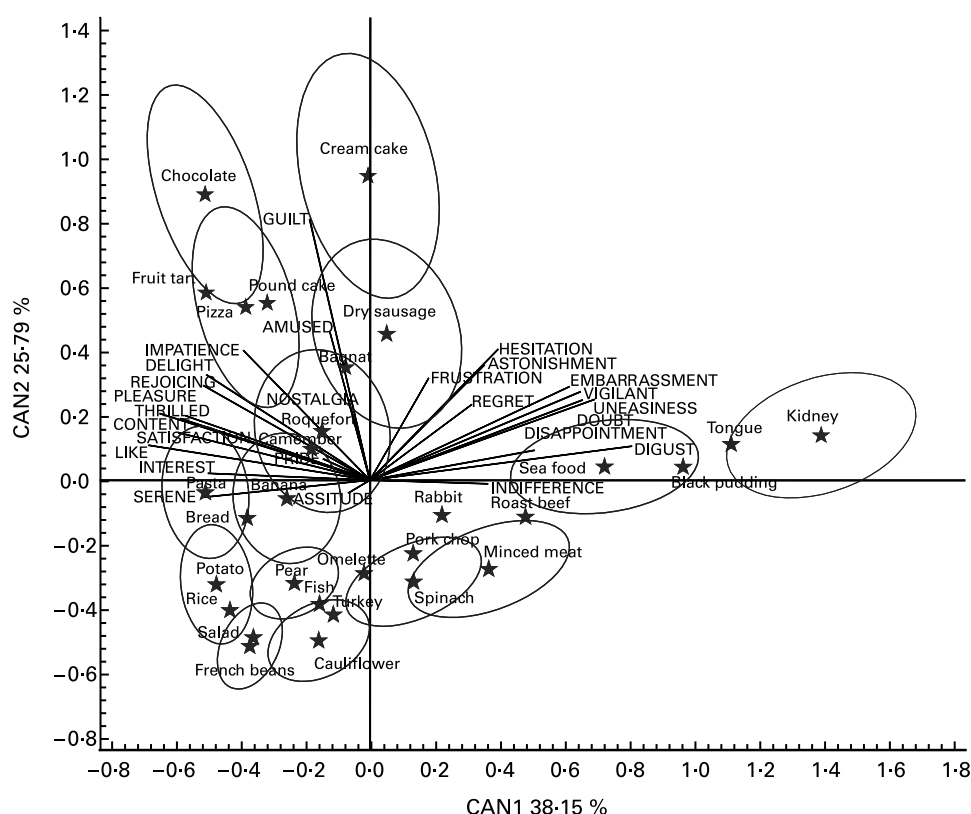


Fig. 1. Emotional discrimination of thirty food pictures by canonical (CAN) variate analysis. For legibility reasons, only fifteen out of thirty ellipses are shown. When two means (*) are very close together, their ellipses are partially superimposed.

Table 4. Cluster of emotional variables

Emotional cluster	Emotional words	Correlation with the average variable
1	Disgust	0.79
	Embarrassment	0.76
	Disappointment*	0.85
	Regret	0.76
2	Frustration	0.67
	To like	0.91
	Pleasure	0.92
	Content	0.92
	Satisfaction*	0.93
	Delight	0.87
	Thrilled	0.92
	Rejoicing	0.90
	Serene	0.63
	Interest	0.80
3	Impatience	0.71
	Guilt*	0.76
4	Nostalgia*	0.80
	Uneasiness	0.85
5	Vigilant	0.85
	Doubt*	0.86
	Hesitation	0.76
6	Amused*	0.83
	Astonishment	0.63
7	Pride	0.75
	Indifference*	0.88
	Lassitude	0.73

*Emotional word that best correlated with the average of the words from the cluster.

and high meat-eating women were evident for the offal, red and white meats, dry sausage, spinach and pasta that were more unpopular for the low meat-eating women. Only pears aroused less disgust in low as compared with high meat-eating women. As regards positive emotions, high meat-eating women liked kidneys and red meats more than the other women (Fig. 2), whereas low meat-eating women significantly preferred pears and French beans. Participants in both groups expressed a similar degree of liking for the other foods such as cheese, vegetables, fish, eggs, pastries or starchy foods. Differences in the other emotional words between both groups are described earlier.

Difference in mean emotional variables between low and high meat-eating women

The overall disagreements between both groups were shown on the PCA plot carried out with the mean differences of the six mean emotional variables between the low and high meat-eating women (Fig. 3). The first axis accounts for 81 % of the variance and shows the main source of disagreement between the two groups. On this axis, 'indifference' and 'disappointment' were opposed to 'satisfaction'. The second axis accounts for a lower percentage of variance (7 %) than the first axis and is mainly explained by 'guilt' and 'doubt'. Red meats, roast beef and minced meat located on the right-hand side aroused much more 'disappointment' in low than in high meat-eating women (Fig. 3 and Table 5). Conversely, pears, banana, fruit tart and French beans on the left-hand side aroused less 'disappointment' in low than in high meat-eating women. 'Indifference' was also

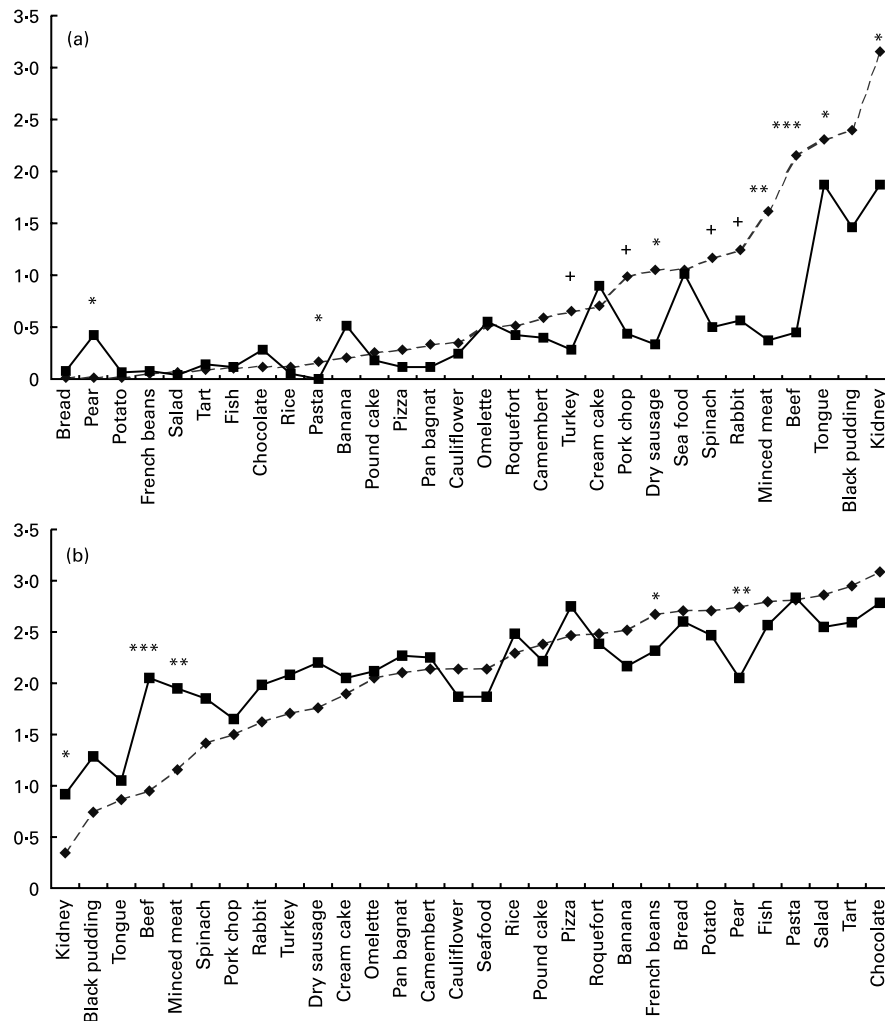


Fig. 2. Differences in disgust (a) and liking (b) between low (—◆—) and high (—■—) meat-eating women.

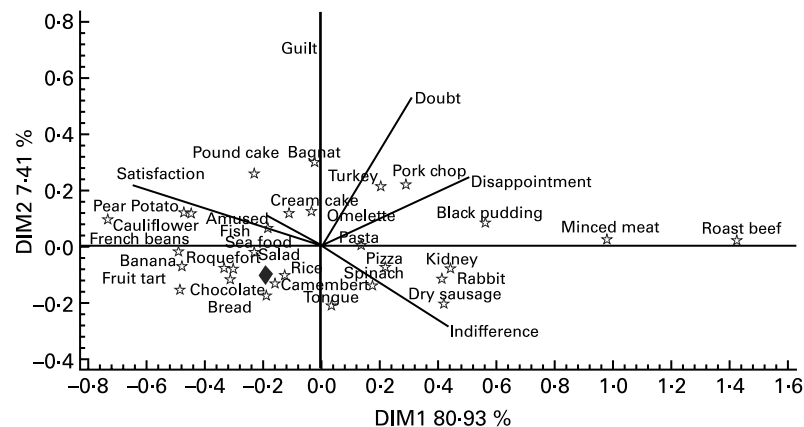


Fig. 3. Differences in mean (DIM) emotional variables between low and high meat-eating women. A virtual product identified by a symbol (◆) on the map with null coordinates for each mean emotional variable was superimposed on the principal component analysis plot to show the location for which no difference was detected between low and high meat-eating women. Thus, the orthogonal projection of this point on the emotional vector discriminated the food into two groups. In other words, when the orthogonal projection of a food picture on an arrow variable is located in the vector direction and well in front of the diamond projection, low meat-eating women felt more emotion towards that food.

Table 5. Significant differences in mean emotion variables felt by the low and high meat-eating women towards food products

	Satisfaction			Amused			Disappointment			Doubt			Indifference			Guilt		
	Low	High	P	Low	High	P	Low	High	P	Low	High	P	Low	High	P	Low	High	P
Roast beef	0.69	1.69	<0.001	0.23	0.44	NS	1.28	0.47	0.001	1.47	0.49	0.01	1.03	0.39	0.01	0.29	0.37	NS
Minced meat	0.83	1.56	0.00	0.23	0.29	NS	0.96	0.37	0.01	1.09	0.61	0.08	1.00	0.48	0.01	0.27	0.24	NS
Kidney	0.30	0.75	0.04	0.30	0.36	NS	1.75	1.42	NS	1.45	1.37	NS	1.03	0.74	NS	0.30	0.22	NS
Dry sausage	1.31	1.42	0.08	0.39	0.33	NS	0.77	0.56	NS	0.92	0.63	NS	0.72	0.47	NS	0.58	0.80	NS
Black pudding	0.59	1.03	0.09	0.34	0.45	NS	1.39	0.99	NS	1.20	0.89	NS	0.95	0.65	NS	0.36	0.22	NS
Tongue	0.66	0.82	NS	0.35	0.37	NS	1.25	1.33	NS	1.22	1.25	NS	0.92	0.58	0.08	0.37	0.31	NS
Pork chop	1.04	1.30	NS	0.32	0.22	NS	0.69	0.52	NS	0.94	0.48	0.04	1.03	0.81	NS	0.38	0.20	NS
Rabbit	1.22	1.57	NS	0.39	0.39	NS	0.81	0.59	NS	0.76	0.41	0.10	0.69	0.36	0.08	0.46	0.61	NS
Turkey	1.28	1.57	NS	0.24	0.18	NS	0.49	0.37	NS	0.70	0.25	0.01	0.81	0.81	NS	0.27	0.15	NS
Pear	2.11	1.70	0.07	0.54	0.38	NS	0.11	0.39	0.03	0.22	0.27	NS	0.34	0.56	NS	0.25	0.10	NS
Banana	2.00	1.85	NS	0.65	0.43	NS	0.17	0.42	0.06	0.26	0.20	NS	0.35	0.47	NS	0.27	0.35	NS
Fruit tart	2.39	2.23	NS	0.64	0.41	NS	0.23	0.46	0.06	0.30	0.36	NS	0.23	0.29	NS	0.84	0.93	NS
Pound cake	1.90	1.79	NS	0.58	0.34	NS	0.44	0.38	NS	0.56	0.37	NS	0.43	0.49	NS	0.90	0.68	NS
French beans	2.08	1.82	NS	0.43	0.37	NS	0.12	0.27	0.08	0.23	0.19	NS	0.32	0.48	NS	0.21	0.26	NS
Cauliflower	1.66	1.46	NS	0.45	0.31	NS	0.31	0.29	NS	0.37	0.23	NS	0.42	0.71	0.10	0.23	0.22	NS
Potato	2.10	1.95	NS	0.43	0.20	0.08	0.13	0.27	NS	0.23	0.11	NS	0.37	0.66	0.10	0.46	0.41	NS
Pizza	1.90	2.16	NS	0.50	0.41	NS	0.43	0.33	NS	0.70	0.31	0.03	0.61	0.43	NS	0.65	0.83	NS
Pasta	2.15	2.33	NS	0.39	0.48	NS	0.30	0.13	NS	0.30	0.09	0.10	0.52	0.44	NS	0.54	0.54	NS
Pan bagnat	1.76	1.76	NS	0.61	0.49	NS	0.46	0.28	NS	0.90	0.53	NS	0.45	0.49	NS	0.66	0.49	NS
Carnambert	1.73	1.79	NS	0.40	0.18	0.09	0.43	0.49	NS	0.51	0.37	NS	0.50	0.50	NS	0.46	0.63	NS

*P values are shown when $P \leq 0.10$ (t test).

strongly felt for red meats, tongue and rabbit, but weakly for cauliflower and potato in low meat-eating women (Fig. 3 and Table 5). Not only red meats (roast beef, minced meat and pork chop) and white meats (turkey and rabbit) but also pizza and pasta aroused more doubt in this last group. However, some food products aroused more positive emotions in low than in high meat-eating women. Thus, pears and French beans gave more satisfaction, and potatoes and Camembert amused the low more than the high meat-eating women. Red meats, kidney, dry sausage and black pudding mainly gave more satisfaction to the high meat-eating women (Table 5).

Discussion

The present study showed that, first, women generally felt more positive than negative emotions towards foods. Second, negative emotions were represented by more different words than positive ones since four and two dimensions sustained them, respectively. Third, the low meat-eating women generally felt more 'disappointment', 'indifference' and 'doubt' towards not only red meats but other food products as well such as white meats, dry sausage, tongue, pan bagnat, pasta, potatoes and pizza, suggesting a negative attitude towards meat products and some starchy foods.

Emotion has often been considered as a physiological reaction to a specific stimulus. However, emotion cannot be confined to the consequences of physiological reactions, as suggested by recent cognitive theories. As stated by Schachter & Singer (1962) and Cacioppo *et al.* (1993), physiological changes are not differentiated enough to account for variation in emotional states that can be labelled by verbal expressions. In the present study, analysis of emotional words associated with food products showed that not only 'disgust' or 'liking' are key words but more complex emotions such as 'uneasiness', 'guilt', 'doubt', 'embarrassment', 'disappointment', 'vigilant', 'satisfaction', 'delight', 'serene', 'interest', 'content' and 'lassitude', as well. These words, used to express subjective states, revealed that emotion linked to food intake was not just aroused by physiological reactions but also by cognitions. These cognitions are the result of the subjects' previous experience with objects or situations. In the present study, that previous experience probably played an important part since we asked participants to imagine eating the food shown on a picture.

Furthermore, emotion depends on the meaning given to an event or a stimulus by an evaluation process. The emotional evaluation is either positive (pleasant) or negative (unpleasant) and with a low or high intensity (Lazarus, 1991; Scherer, 1998). The results of the present study reveal the importance of these subjective states linked to the evaluation of food products. In general, the participants expressed positive emotions towards food except offal and, to a lesser extent, red meats. Positive emotions towards foods were associated with the pleasure of eating which was elicited by foods such as chocolate, pasta, tarts, salad, fish, bread, pizza and potatoes. The present study showed that although less strong, negative emotions towards food were more diversified than positive emotions. Among the twenty-six words related to emotions elicited by food, the participants used 'disappointment' (associated with disgust, embarrassment, regret and frustration), 'doubt' (associated with uneasiness, vigilant and hesitation), 'guilt' (associated with nostalgia) and 'indifference' (associated with lassitude) to characterise the negative features of food. In contrast, the positive words

were only 'satisfaction' and 'amused'. These negative concepts might help to explain food choice in terms of avoidance in women. Thus, 'disappointment' and 'guilt' differentiated well between food pictures, as well as between low and high meat-eating women, as we discuss later. As expected, offal and red meats were the most unpopular food products and aroused the most disappointment among the participants. Different causes of 'disappointment' for food should be advanced. As suggested earlier, not just appearance but other sensory components of food as well such as taste, texture or odour play an important part in preference. Kenyon & Barker (1998) and Kubberod *et al.* (2002) showed that the sight of blood or the red colour of meat products evokes images of animal death and induces disgust in women. Consequently, disgust should be related to cruelty refusal. It is also caused by a variety of political concerns and moral beliefs that constitute the basis of moral vegetarianism. This type of belief often induces strict avoidance of all meats (Rozin *et al.* 1997). Because meat is often considered as unhealthy, meat avoidance for health reasons should be the second ground for rejection (health vegetarianism). Moreover, Fessler *et al.* (2003) determined a third ground for meat avoidance that is induced by meat taste. Meat consumption was higher in meat-taste avoiders than in meat-health avoiders. The lowest intake was associated with the ethical and environmental meat avoiders. These three categories of meat avoiders did not differ from one another on overall disgust sensitivity. Unlike the results of the study by Fessler *et al.* (2003) there was an overall difference in disgust towards meat products between high and low meat-eating women in the present study. High meat-eating women were generally less disgusted by meats than low meat-eating women. This difference in results could be explained either by the difference in meat consumption frequency in both studies or by the test differences between the two studies. Meat consumption in the strongest meat-avoider group was 6.6 servings per week in the former study whereas it was 3.6 in the latter study. Meat consumption in the weakest meat-avoider group was 11.8 servings per week in the first study and 7.2 in the second study. Meat avoidance might not be strong enough (6.6 servings/d) in the participants in the Fessler *et al.* (2003) study to be related to real meat disgust. Another hypothesis is that the emotional test in the present study should be more directly linked to meat consumption than to the general disgust sensitivity test (Haidt *et al.* 1994).

The present study also showed that 'disappointment' and 'disgust' for red meats was higher than that for white meats, and higher in low meat than in high meat-eating women. Thus, the low meat-eating women liked meat less in general. Furthermore and surprisingly enough, negative emotions were not limited to meat products in the low meat-eating women. They were also 'disappointed' by and had 'doubts' about other food products such as pasta and pizza. Surprisingly, low meat-eating women liked only two more foods; pears and French beans. Thus, overall eating pleasure might have been lower in women who consumed few meat products, since negative emotions were strongly felt, not only for meat products but also for other food products. Contrary to our expectations, the low meat-eating women did not compensate for their 'disappointment' with meat products by a stronger liking for other food products except pears and French beans.

The 'doubt' for foods such as pasta or pizza could be explained by their reputation for making individuals gain weight. Knowledge of food nutrient content, even if it is false, can alter emotions towards food. Macht *et al.* (2003) observed that negative

emotions occurred more frequently when the energy content of the foods increased. The higher the energy content of the food, the more 'unhealthy' and 'dangerous' it was perceived. The tendency of individuals to view meat products as being high in energy and fat explains their rejection of these foods (Holm & Mohl, 2000; Lea & Worsley, 2001). Moreover, these overall negative attitudes towards foods might be caused by the multitude and extensiveness of health promotion programmes that highlight slimness. The consequences of these programmes are that food and eating have become a source of stress and substantial worry among women (Stein & Nemeroff, 1995). The pleasure of eating might be viewed as a 'vice' instead of being seen as one of life's harmless enjoyments. Moreover, women who were characterised by a high concern for controlling weight and dieting (restraint) showed higher emotional reactions to foods (Macht *et al.* 2003). In the case of pathological eating disorders, volunteers reported higher levels of disgust toward foodstuffs of animal origin, as well as towards the human body. Eating disorders are associated with increased disgust specific to food and the body (Troop *et al.* 2002).

Since meat is generally considered fatty and because it was consumed less by low meat-eating women, we initially assumed that their energy intake might be lower than that of other participants. However, in the present study, the energy intakes were similar in low and high meat-eating women. Since the energy intake was similar between the low and high meat-eating women in the present study, it is likely that the low meat-eating women compensated for their low meat consumption with a higher consumption of other foods. The low meat-eating women tended to have a higher energy intake from sweets and vegetables. In the present study, they showed a preference for some of these foods (fruit tart, chocolate, pound cake, pears, banana, potatoes, cauliflower, French beans). The reason for this consumption preference for fruit and vegetables might be found in the healthy image of those foods and in the association with weight loss (Oakes, 2003). Sweets are liked for the pleasure they bring, but they also generate guilt as well as conflicts between sensory appeal for good taste and expected implication for health or body image (Olsen, 1999). Thus, emotion towards foods is complex because some foods can be evaluated both positively and negatively at the same time. The negativity is less related to basic emotions such as pleasure or disgust than 'self-conscious' emotions such as guilt. The specificity of meat may be to arouse guilt for both its animal origin and its supposedly fatty content. However, in the present study, less guilt was felt towards meats than towards the other fatty food products; chocolate, cream cake, fruit tart, pizza and pound cake. Disgust, disappointment, doubt and indifference, more than guilt, could explain low meat consumption; the consequences of the consumption of fatty food on body weight could be the grounds for feelings of guilt.

In conclusion, the low meat-eating women, although belonging to a non-clinical population, generally felt more negative emotions towards meat products, and other food products as well, than their counterparts. This might be indicative of a problematic relationship with food and not only with meat products, although their energy intake was similar to that of high meat-eating women. The consequences of this general emotional state towards food are twofold. First, low meat-eating women more than high meat-eating women are at risk of Fe deficiency because of their lower consumption of haeme Fe-containing foods. Second, since women play a determinant part in food education,

there is a risk that they might pass their negative emotions towards foods on to their children. From the perspective of health education, it is crucial to continue research into the emotions elicited by food and their influence on eating behaviours.

Acknowledgements

We are grateful to Daniel Effron for his valuable advice, and to L. Debarge and C. Ferreira for their technical assistance.

References

- Baars BJ (1994) A global workspace theory of conscious experience. In *Consciousness in Philosophy and Cognitive Neuroscience*, pp. 149–171 [A Revonsuo and M. Kamppinen, editors]. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cabanac M (2002) What is emotion? *Behav Processes* **60**, 69–83.
- Cacioppo J, Klein D, Berntson G & Hatfield E (1993) The psychophysiology of emotion. In *Handbook of Emotion*, pp. 119–142 [M Lewis and JM Haviland, editors]. New York: Guilford Press.
- Chambers E IV & Bowers JR (1993) Consumer perception of sensory qualities in muscle foods. *J Food Sci* **46**, 1902–1906.
- Ekman P, Levenson RW & Friesen WV (1983) Autonomic nervous system activity distinguishes among emotions. *Science* **221**, 1208–1210.
- EPIBASE (1999) *Une Base de Données Biomédicales*, version 1-1H. Dijon, France: INRA Editions.
- Fazio RH & Zanna MP (1981) Direct experience and attitude-behavior consistency. In *Advances in Experimental Social Psychology*, pp. 162–202 [L Berkowitz, editor]. New York: Academic Press.
- Fessler DMT, Arguello AP, Mekdara JM & Macias R (2003) Disgust sensitivity and meat consumption: a test of an emotivist account of moral vegetarianism. *Appetite* **41**, 31–41.
- Galan P, Yoon HC, Preziosi P, *et al.* (1998) Determining factors in the iron status of adult women in the SU.VI.MAX study. *Eur J Clin Nutr* **52**, 383–388.
- Gregory NG (1997) Meat, meat eating and vegetarianism. A review of the facts. In *Proceedings of the 43rd International Congress of Meat Science and Technology*, pp. 68–85 [J Bass, editor]. Auckland, New Zealand: ICOMST.
- Guinot C, Latreille J, Malvy D, Preziosi P, Galan P, Hercberg S, Tenenhaus D (2001) Use of multiple correspondence analysis and cluster analysis to study dietary behaviour: food consumption questionnaire in the SU.VI.MAX cohort. *Eur J Epidemiol* **17**, 505–516.
- Guzman MA & Kjaernes U (1998) *Humans and Animals – a Qualitative Study*. SIFO Report no. 6. Lysaker, Norway: The National Institute for Consumer Research.
- Haidt J, McCauley CR & Rozin (1994) A scale to measure disgust sensitivity. *Pers Individ Dif* **16**, 701–713.
- Holm L & Mohl M (2000) The role of meat in everyday food culture: an analysis of an interview study in Copenhagen. *Appetite* **34**, 277–283.
- Hulten L, Gramatovski E, Gleerup A & Hallberg L (1995) Iron-absorption from the whole diet – relation to meal composition, iron requirements and iron stores. *Eur J Clin Nutr* **49**, 794–808.
- Juillard E (2003) *Lexique émotionnel et consommation alimentaire: le cas plus spécifique des produits carnés auprès de la population féminine (Emotional lexicon and food consumption: the specific case of meat products in women)*. MSc Thesis, University of Psychology, Clermont-Ferrand.
- Kenyon PM & Barker ME (1998) Attitudes towards meat-eating in vegetarian and non-vegetarian teenage girls in England – an ethnographic approach. *Appetite* **30**, 185–198.

- Kubberod E, Ueland O, Roodbotten M, Westad F & Risvik E (2002) Gender specific preferences and attitudes towards meat. *Food Qual Pref* **13**, 285–294.
- Lazarus RS (1991) *Emotion and Adaptation*. Oxford: Oxford University Press.
- Lea E & Worsley A (2001) Influences on meat consumption in Australia. *Appetite* **36**, 127–136.
- Lea E & Worsley A (2002) The cognitive contexts of beliefs about healthiness of meat. *Public Health Nutr* **5**, 37–45.
- Leyens JP, Paladino PM, Rodriguez-Torres R, Vaes J, Demoulin S, Rodriguez-Perez A & Gaunt R (2000) The emotional side of prejudice: the attribution of secondary emotions to ingroups and outgroups. *Pers Soc Psychol Rev* **4**, 186–197.
- Macht M, Gerer J & Ellgring H (2003) Emotions in overweight and normal-weight women immediately after eating foods differing in energy. *Physiol Behav* **80**, 367–374.
- Millar MG & Tesser A (1986) Effects of affective and cognitive focus on the attitude-behavior relation. *J Pers and Soc Psychol* **51**, 270–276.
- Millar MG & Tesser A (1992) The role of beliefs and feelings in guiding behavior: the mismatch model. In *The Construction of Social Judgements*, pp. 277–300 [LL Martin and A Tesser, editors]. Hillsdale, NJ: Erlbaum.
- Mooney KM & Walbourn L (2001) When college students reject food: not just a matter of taste. *Appetite* **36**, 41–50.
- Niedenthal P, Auxiette C, Nugier A, Dalle N, Bonin P & Fayol M (2004) A prototype analysis of the French category 'emotion'. *Emot Cogn* **18**, 289–312.
- Oakes ME (2003) Differences in judgments of food healthfulness by young and elderly women. *Food Qual Pref* **14**, 227–236.
- Olsen SO (1999) Strength and conflicting valence in measurement of food attitudes and preferences. *Food Qual Pref* **10**, 483–494.
- Ortega RM, Lopez-Sobaler AM, Requejo AM, Quintas ME, Gaspar MJ, Andres P & Navia B (1998) The influence of meat consumption on dietary data, iron status and serum lipid parameters in young women. *Int Vitam Nutr Res* **68**, 255–262.
- Regan DT & Fazio RH (1977) On the consistency between attitudes and behavior: look to the method of attitude formation. *J Exp Soc Psychol* **13**, 38–45.
- Rousset S, Patureau Mirand P, Brandolini M, Martin JF & Boirie Y (2003) Daily protein intakes and eating patterns in young and elderly French. *Br J Nutr* **90**, 1107–1115.
- Rozin P, Markwith M & Stoess C (1997) Moralization and becoming a vegetarian: the transformation of preferences into values and the recruitment of disgust. *Psychol Sci* **8**, 67–73.
- Russell JA (1991) Culture and the categorization of emotions. *Psychol Bull* **110**, 426–450.
- SAS Institute Inc. (1998) *SAS/STAT User's Guide*, release 8-01. Cary, NC: SAS Institute Inc.
- Schachter S & Singer JE (1962) Cognitive, social, and physiological determinants of emotional state. *Psychol Rev* **69**, 379–399.
- Scherer K (1998) Appraisal theories. In *Handbook of Cognition and Emotion*, pp. 637–663 [T Dalgleish and M Power, editors]. Chichester: Wiley.
- Scherer KR (2005) Emotion. In *Introduction to Social Psychology*, 3rd ed., pp. 1–46 [M Hewstone and W Stroebe, editors]. Oxford: Blackwell.
- Searle JR (1998) How to study consciousness scientifically. *Philos Trans R Soc Lond Biol Sci* **353**, 1935–1942.
- Shaw NS, Chin CJ & Pan WH (1995) A vegetarian diet rich in soybean products compromises iron status in young students. *J Nutr* **125**, 212–219.
- Smith CA & Lazarus RS (1993) Appraisal components, core relational themes, and the emotions. *Cogn Emot* **7**, 233–269.
- Stein RI & Nemeroff CJ (1995) Moral overtones of food: judgments of others based on what they eat. *Pers Soc Psychol* **21B**, 480–490.
- SU.VI.MAX (1994) Portions Alimentaires: Manuel Photos pour l'Estimation des Quantités (Food Portions: Photographic Manual for Estimating Quantity). [S Hercberg, M Deheeger and P Preziosi, editors]. Paris: Polytechnica.
- Troop NA, Treasure JL & Serpell L (2002) A further exploration of disgust in eating disorders. *Eur Eat Disord Rev* **10**, 218–226.
- Worsley A & Skrzypiec G (1998) Do attitudes predict red meat consumption among young people? *Ecol Food Nutr* **37**, 163–195.
- Zammuner VL (1998) Concepts of emotions: 'emotionness' and dimensional ratings of Italian emotion words. *Cogn Emot* **12**, 243–272.

Article 13

13. Schlich P. Pineau N., Brajon D., Cordelle S. (2006). Le projet SensoBase: construire une base de profils sensoriels pour documenter les performances des panels d'analyse sensorielle. In *9èmes Journées Agro-industrie et Méthodes Statistiques*, Montpellier. 101-106.

9^{èmes} Journées Européennes

AGRO-INDUSTRIE
et MÉTHODES
STATISTIQUES

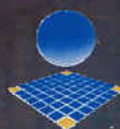
**AGRO-INDUSTRIE et
MÉTHODES STATISTIQUES**

Actes des 9^{èmes} journées

Montpellier

25 - 27 janvier 2006

Organisées par la
Faculté de Pharmacie de l'Université Montpellier 1
sous l'égide de la
Société Française de Statistique



Le projet Sensobase : construire une base de profils sensoriels pour documenter les performances des panels d'analyse sensorielle

P. Schlich, N. Pineau, D. Brajon & S. Cordelle

Centre Européen des Sciences du Goût. 15 rue Hugues Picardet. 21000 Dijon
E-mail : schlich@cesg.cnrs.fr - Web : <https://liris.cesg.cnrs.fr/>

Résumé

Cette présentation présente le projet Sensobase qui a déjà été réuni et organisé en une base de données de 307 jeux de données de profils sensoriels provenant de 25 laboratoires de 7 pays différents. Cette base de données a pour but de documenter les applications de l'analyse sensorielle, d'évaluer les performances du jury et de permettre de nombreuses comparaisons des techniques de sensométrie. Pour chaque jeu de données chargé sur <https://liris.cesg.cnrs.fr>, le contributeur reçoit une analyse gratuite des performances des sujets des panels et des différences entre produits. Cette analyse présente la dernière version du système CAP, une table FLASH et une carte CVA des produits. Cette présentation a délibérément pour but d'attirer de nouveaux contributeurs à la Sensobase.

Mots-clés :

profil sensoriel, performances des panels, base de données, fouille de données

Abstract

The paper introduces the **Sensobase** project which has already gathered and arranged in a database 307 datasets of sensory profiles from 25 laboratories in 7 countries. This database aims to document the practices of sensory analysis, to benchmark the panellist performances and to allow for extensive comparisons of sensometrics techniques. For each dataset uploaded at <https://liris.cesg.cnrs.fr>, the contributor receives a free analysis of the panellist performances and of the product differences. This analysis features the last version of the CAP system, a FLASH table and a CVA map of the product. The paper deliberately aims to attract new contributors to the **Sensobase**

Keywords :

sensory profiling, panel performances, database, datamining

1 Introduction

Beaucoup de méthodes statistiques ont été proposées pour le contrôle des performances des sujets des panels d'analyse sensorielle (Schlich, 1994 ; Dijksterhuis, 1995 ; Couronne, 1997 ; Naes, 1998 ; Brockhoff, 1998 ; Piggott & Hunter, 1999 ; Qannari *et al.*, 1999 ; Martens *et al.*, 2000 ; Qannari & Meyners, 2001 ; ...), mais aucune étude ne se prononce sur le niveau réel que peuvent atteindre ces performances. Seul le projet **ProfiSens** a essayé d'établir des critères de comparaison pour des panels de pays différents (McEwan *et al.*, 2002). Le panel d'analyse sensoriel descriptif (dit profil sensoriel) est un outil sophistiqué et précieux dont finalement, nous ne connaissons pas la précision véritable. Il est vrai que la diversité des situations en jeu dans l'exercice du profil sensoriel, générée aussi bien par le type de descripteur évalué (apparence, odeur, texture, saveur, arômes ...) que par le type de produit décrit, fait que l'établissement de normes de précision à atteindre est délicat.

Le projet **Sensobase**, développé dans le laboratoire des auteurs au fil de ces 2 dernières années, vise à combler ces lacunes de connaissance du fonctionnement de l'outil de mesure qu'est le panel d'analyse sensoriel. Notre

approche consiste à échantillonner l'ensemble des situations de profils sensoriels afin de pouvoir établir des distributions objectives de valeurs de performances rencontrées dans la réalité de l'analyse sensorielle.

Comme son nom l'indique, la **Sensobase** est donc une base de données de profils sensoriels suffisamment documentés pour nous permettre d'agréger des sous-ensembles de cette base dans des approches de type fouille de données desquelles nous espérons faire émerger une connaissance cachée.

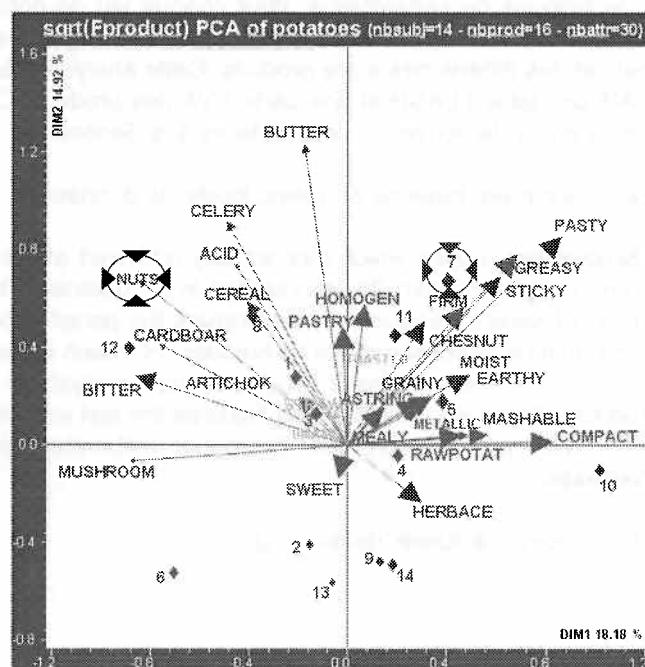
2 Construction et taille de la Sensobase

Afin d'« attirer » des jeux de données vers notre base, nous offrons une analyse statistique sophistiquée - et bien entendu gratuite - à tout possesseur de jeux de données de profil sensoriel qui accepte de laisser ses données dans notre base. Cette analyse est en fait une batterie d'analyses dédiées aussi bien à l'étude des performances des sujets et des panels qu'à la mise en évidence des différences entre les produits de l'étude. L'ensemble représente l'aboutissement d'une dizaine d'années d'étude de ce thème par le premier auteur de cette communication. Nous pensons que cette batterie d'analyse est suffisamment originale pour attirer véritablement des jeux de données d'origines très diverses. Cet échange « données contre analyses » est réalisé de manière semi-automatique par l'intermédiaire d'un site internet - <https://liris.cesg.cnrs.fr> - que le lecteur est vivement encouragé à consulter.

En fin d'année 2005, la **Sensobase** contient 307 jeux de données provenant de 25 laboratoires ou entreprises différentes situés dans 7 pays différents.

3 Analyse standard offerte pour chaque jeu de données

Cette batterie d'analyse est constituée de 6 éléments. Les quatre premiers relèvent du contrôle des performances des sujets alors que les deux derniers traitent de la mise en évidence des différences sensorielles entre les produits au niveau du panel. Ces 6 éléments sont donnés dans le Tableau 1.



ACP non normée du tableau sujet x descripteur contenant les racines carrées des F produit des ANOVA individuelles. Couleurs et tailles des symboles sujets et descripteurs renvoient au niveau des performances.

FIG. 1 – Carte CAP.

- **Table CAP** (Tableau 1). Un diagnostic pour chaque sujet et chaque descripteur établissant si le sujet perçoit des différences entre les produits, si celles-ci sont en accord avec celles trouvées par le reste du panel et si ce sujet est significativement moins répétable sur ce descripteur que les autres sujets.
- **Carte CAP** (Figure 1). Une analyse en composantes principales des racines carrées des rapports de Fisher de l'effet produit dans les analyses de la variance individuelles, résumée par une représentation biplot où les

ATTRIBUTE	FPROD	FINTER	MEAN	RMSE	7	5	11	3	1	8	12	4	10	14	6	9	2	13
FIRM	17.17	1.61	37.6	18.8	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MASHABLE	13.91	0.88	40.2	25.5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MOIST	11.65	1.20	46.9	21.1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
GRAINY	11.57	1.10	26.9	20.6	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MEALY	11.44	0.83	42.3	25.2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SWEET	9.49	1.10	42.4	21.9	+	+	+	+	+	+	+	+	+	+	+	+	+	+
COMPACT	8.57	1.77	48.6	18.6	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HOMOGEN	7.46	1.17	59.0	22.4	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CHESNUT	6.02	1.10	41.1	24.3	+	+	+	+	+	+	+	+	+	+	+	+	+	+
EARTHY	5.78	1.07	29.8	21.7	+	+	+	+	+	+	+	+	+	+	+	+	+	+
GREASY	5.28	1.47	15.3	15.7	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PASTRY	4.42	1.61	19.2	20.1	+	+	+	+	+	+	+	+	+	+	+	+	+	+
PASTY	3.67	1.32	34.7	22.8	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HERBACE	3.23	1.33	30.5	21.2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
STICKY	2.95	1.30	27.6	22.5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BITTER	2.87	1.28	26.9	20.2	+	+	+	+	+	+	+	+	+	+	+	+	+	+
RAWPOTAT	2.81	1.09	22.7	21.9	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ARTICHOK	2.62	0.95	39.7	24.7	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ASTRING	2.58	1.00	29.3	21.9	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CEREAL	2.32	1.05	26.3	22.5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CELERY	1.86	1.39	16.9	15.9	+	+	+	+	+	+	+	+	+	+	+	+	+	+
METALLIC	1.77	1.13	14.9	17.7	+	+	+	+	+	+	+	+	+	+	+	+	+	+
BUTTER	1.73	1.52	16.8	17.3	+	+	+	+	+	+	+	+	+	+	+	+	+	+
NUTS	1.63	1.05	16.3	16.5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
STRAW	1.62	1.05	29.0	21.5	+	+	+	+	+	+	+	+	+	+	+	+	+	+
CARDBOAR	1.45	1.30	9.07	13.8	+	+	+	+	+	+	+	+	+	+	+	+	+	+
TOASTEDB	1.43	1.06	8.71	15.8	+	+	+	+	+	+	+	+	+	+	+	+	+	+
MUSHROOM	1.35	1.29	12.0	14.9	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ACID	1.34	1.39	14.7	15.8	+	+	+	+	+	+	+	+	+	+	+	+	+	+
UMAMI	0.99	1.08	14.8	17.4	+	+	+	+	+	+	+	+	+	+	+	+	+	+
rankf					5.57	6.13	6.50	6.67	6.73	7.03	7.10	7.37	7.60	7.90	8.23	8.87	9.20	10.10

Diagnostics CAP : + signifie discriminant en accord avec le groupe, - signifie discriminant en désaccord avec le groupe, = signifie non discriminant et ! signifie moins répétable que la moyenne du panel. FPROD et FINTER sont les F produit et interaction produit par sujet de l'ANOVA globale à deux facteurs en modèle mixte. RMSE est la racine carrée du carré moyen des erreurs de ce modèle. MEAN est la moyenne des scores du descripteur (ATTRIBUTE).

TAB. 1 – Table CAP.

sujets sont les observations et les descripteurs les variables. Cette carte est interactive dans la mesure où un click sur chaque sujet ou chaque descripteur provoque l'apparition du positionnement des performances de ce sujet ou de ce descripteur dans la distribution de l'ensemble des performances.

- Tableau de performances multidimensionnelles basées sur des **MANOVA** individuelles (Tableau 2). Ces indices permettent de comparer les sujets sur la base de l'intensité et de la complexité des différences perçues entre les produits sur l'ensemble des descripteurs simultanément. (Schlich *et al.*, 2004)
- Tableau de performances multidimensionnelles basées sur 3 approches du **coefficient RV** (Tableau 3). Ces indices permettent de comparer les sujets sur la base des configurations multidimensionnelles des produits, des structures de corrélation entre descripteurs et des moyennes des covariances entre descripteurs homologues. (Schlich, 1996 ; Ledauphin *et al.*, 2006)
- **Table Flash** (tableau 4). Présentation compacte de l'ensemble des analyses de la variance au niveau du panel et des comparaisons des moyennes de chaque produit à la moyenne de l'ensemble des produits. Descripteurs triés selon une classification VARCLUS de SAS®. (Schlich, 1995)
- Carte sensorielle par **analyse en variables canoniques** (Figure 2). Carte des produits et des descripteurs avec superposition des ellipses de confiance de la moyenne des produits (Schlich, 2004).

Ces sorties sont organisées en page html parmi lesquelles l'utilisateur navigue à partir d'une page d'accueil qui lui donne aussi accès à une notice d'interprétation. Ces analyses sont modulables à l'aide d'un ensemble de paramètres dont l'utilisateur peut modifier les valeurs par défaut lors de la commande d'une seconde analyse d'un même jeu de données qui peut, de plus, ne porter que sur une sélection des données de départ. Tout utilisateur régulièrement inscrit dans le projet **Sensobase** peut consulter la liste de ses jeux de données présents dans la **Sensobase** afin de commander en ligne des analyses supplémentaires. Il reçoit par messagerie électronique les pages html de résultats sous un délai moyen de 1 à 2 jours.

PANELLIST	NATTSEL	F	NCAN
12	15	924.61	2
14	11	20.32	4
2	10	11.51	4
3	11	10.84	3
10	7	8.25	4
7	11	8.14	4
1	9	7.57	4
6	7	6.28	3
13	5	4.92	3
5	5	4.71	4
9	3	4.69	2
8	7	4.24	4
4	6	3.50	3
11	6	3.34	4

TAB. 2 – MANOVA individuelles

NATTSEL est le nombre de descripteurs sélectionnés par la procédure STEPDISC de SAS® au seuil $p=0.15$. F est le test de Hotelling-Lawley de l'effet produit. NCAN est le nombre de variables canoniques significatives (rapport de vraisemblance, $p=0.10$)

PANELLIST	NRV	NALPHA	NRVD
1	2.68	3.45	3.59
10	2.92	3.73	5.96
11	1.94	3.18	5.41
12	2.57	3.15	5.49
13	1.79	2.81	6.32
14	3.32	4.02	4.00
2	3.75	4.34	6.35
3	3.23	3.74	7.23
4	3.17	3.74	6.16
5	3.20	3.88	7.55
6	0.97	2.45	4.29
7	3.22	4.24	6.50
8	2.36	2.92	5.42
9	2.06	3.21	5.03
MEAN	2.66	3.49	5.66

TAB. 3 – Accord multidimensionnel de chaque sujet avec le panel

Coefficients RV normalisés entre configurations de produits (NRV) et entre matrices de corrélation des descripteurs (NRVD). NALPHA est un coefficient normalisé de la moyenne des covariances par descripteur. Au dessus de 2, il y a accord ; entre 1 et 2, il y a juste une tendance à l'accord ; en dessous de 1, il y a désaccord.

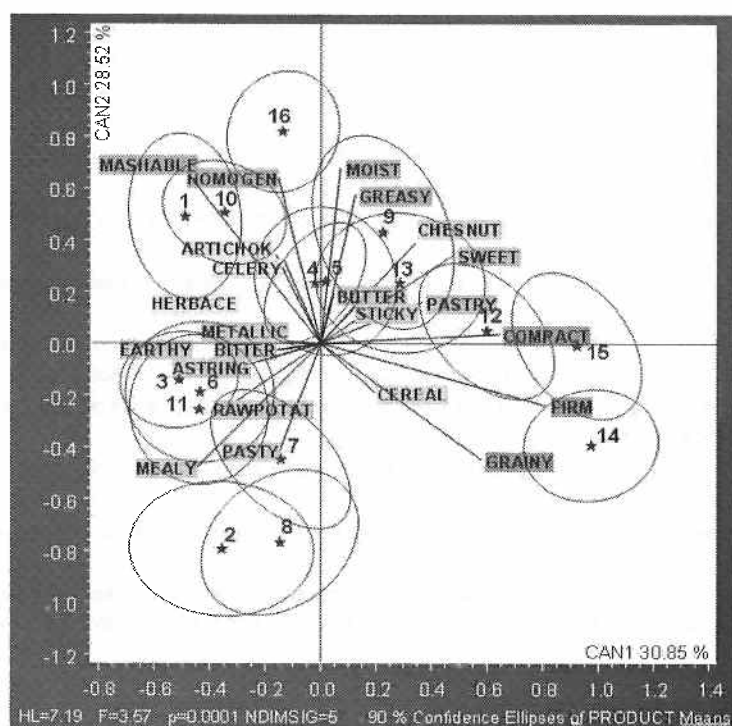
ATTRIBUTE	F	PROB	GMEAN	16	3	10	1	6	2	8	5	11	4	13	9	7	12	15	14
FIRM	17.2	0.000	37	19-	21-	21-	22-	22-	24-							49+	63+	69+	77+
COMPACT	8.58	0.000	49		29-		40-	27-	39-	36-			59+				60+	73+	73+
GRAINY	11.6	0.000	27	6-	14-	12-	15-	18-		52+							39+	51+	46+
PASTRY	4.43	0.000	19				10-			10-						4-	28+	44+	
SWEET	9.49	0.000	42	54+	30-	56+	27-	32-		29-	33-			54+	56+	22-	51+	69+	
BUTTER	1.74	0.047	17								25+						25+		9-
CHESNUT	6.02	0.000	41						29-	21-		30-		53+	58+		63+	58+	
CEREAL	2.33	0.004	26				17-												44+
ARTICHOKE	2.63	0.001	40		52+				23-										28-
CELERY	1.86	0.029	17		29+										24+				
HERBACE	3.23	0.000	31						20-		49+					42+		20-	13-
EARTHY	5.78	0.000	30				40+	42+		41+	43+		21-			47+	19-	14-	
ASTRING	2.58	0.001	29													44+	18-		
BITTER	2.88	0.000	27						18-	36+				17-		46+			
RAWPOTAT	2.81	0.001	23				32+			38+								10-	
METALLIC	1.78	0.04	15				23+	26+											
PASTY	3.68	0.000	35				22-	47+	56+								23-	24-	
MEALY	11.4	0.000	42	29-	60+			54+	78+	55+	31-	54+		32-			26-	27-	31-
STICKY	2.95	0.000	28	37+	16-						17-				47+				
MASHABLE	13.9	0.000	40	69+		58+	66+		29-	25-	51+					26-	27-	16-	11-
HOMOGEN	7.41	0.000	59	78+		72+	70+		42-	31-	69+		70+		69+			45-	
GREASY	5.27	0.000	15	26+	7-			8-	2-	7-	26+	7-	25+		25+	8-			
MOIST	11.7	0.000	47	68+		63+	69+	38-	17-	29-	58+	35-				33-	56+		34-

Moyennes des produits significativement ($p=0.05$) différentes de la moyenne générale (GMEAN). Classification des descripteurs selon la procédure VARCLUS de SAS (option CENTROID) ; les différents groupes sont surlignés avec des couleurs différentes. F et risque associé (PROB) de l'effet produit dans l'ANOVA à deux facteurs en modèle mixte.

TAB. 4 – Table FLASH.

4 Les premiers enseignements apportés par la Sensobase

Une première fouille de la **Sensobase**, basée uniquement sur 171 jeux de données et présentée récemment au **Pangborn Sensory Science Symposium** (Pineau *et al.*, 2005), nous a déjà permis de suggérer que le pouvoir discriminant des sujets des panels d'analyse sensorielle ne dépend pas du sexe du sujet, atteint son meilleur niveau



Les couleurs des descripteurs renvoient à la classification de la table FLASH. Les ellipses de confiance des moyennes des produits sont basées sur la loi bi-normale en utilisant les projections individuelles (une seule par sujet) pour estimer la matrice de covariance. Les descripteurs sont représentés à l'aide de leur coefficient de corrélation total avec les axes.

FIG. 2 – Analyse en variables canoniques.

vers l'âge de 30-35 ans et décroît ensuite très lentement mais continuellement. D'autre part, une ventilation des performances selon le type de descripteurs sensoriels et le type de produits suggère que les meilleures répétabilités sont obtenues pour les descripteurs visuels et les moins bonnes pour les descripteurs de texture (Tableau 5). Le trop petit nombre de laboratoires (22) ayant fournis des données nous interdit pour le moment d'accorder une confiance absolue aux comparaisons de répétabilité par type de produits.

	Taste	Aroma	Visual	Odour	Texture	Moyenne	N	Ecart-type
Meat/Fish	0,96	0,99	1,52	1,93	1,34	1.26 A	808	0,54
ReadyMM	1,16	1,27	1,37	1,19	1,73	1.34 B	684	0,49
Bakery	1,25	1,25	1,42	1,44	1,86	1.44 C	375	0,54
Fruit/Veg	1,49	1,44	1,3	1,39	1,63	1.48 C	1180	0,57
Beverage	1,48	1,55	1,13	1,52	1,35	1.49 C	805	0,68
Dairy Pdt	1,54	1,63	1,26	1,27	1,87	1.62 D	454	0,71
Moyenne	1.34 A	1.38 AB	1.40 AB	1.42 B	1.63 C	1,43	.	.
N	1084	1071	593	681	877	.	4306	.
Ecart-Type	0,6	0,59	0,49	0,6	0,62	.	.	0,6

Les moyennes suivies d'une même lettre ne sont pas significativement ($p=0.05$) différentes.

TAB. 5 – Ecart-type moyen entre répétitions basés sur les 171 premiers jeux de données de la Sensobase.

La présentation orale actualisera ces résultats sur la base des 307 jeux de données désormais disponibles dans la Sensobase. Nous espérons que cette communication convaincra les producteurs de profils sensoriels d'alimenter le flux de données vers la Sensobase.

Références

- Brockhoff, P. M. (1998). Assessor modelling. *Food Quality and Preference*, 9(3), 87-9.
- Couronne, T. (1997). A study of assessors' performance using graphical methods. *Food Quality and Preference*, 8(5/6), 359-365.
- Dijksterhuis, G. (1995). Multivariate data analysis in sensory and consumer science : an overview of developments. *Trends in Food Science & Technology*, 6(6), 206-211.

- Ledauphin, S., Hanafi M. & Qannari, E.M. 2006. Assessment of the agreement among the subjects in fixed vocabulary profiling. *Food Quality and Preference*, 17(3/4), 277-280
- Martens, M., Bredie, W. L. P., & Martens, H. (2000). Sensory profiling data studied by partial least squares regression. *Food Quality and Preference*, 11(1/2), 147-149.
- McEwan, J. A., Hunter, E. A., van Gemert, L. J., & Lea, P. (2002). Proficiency testing for sensory profile panels: measuring panel performance. *Food Quality and Preference*, 13(3), 181-190.
- Naes, T. (1998). Detecting individual differences among assessors and differences among replicates in sensory profiling. *Food Quality and Preference*, 9(3), 107-10.
- Piggott, J. R., & Hunter, E. A. (1999). Evaluation of assessor performance in sensory analysis. *Italian Journal of Food Science*, 11(4), 289-304.
- Pineau, N., Brajon, D., Cordelle, S., Schlich, P. (2005). Mining a database of sensory descriptive analyses to understand panellist performances. *6th Pangborn Sensory Science Symposium*, Harrogate, U.K., 7-11 August 2005
- Qannari, E. M., MacFie, H. J. H., & Courcoux, P. (1999). Performance indices and isotropic scaling factors in sensory profiling. *Food Quality and Preference*, 10(1), 17-21.
- Qannari, E. M., & Meyners, M. (2001). Identifying assessor differences in weighting the underlying sensory dimensions. *Journal of Sensory Studies*, 16(5), 505-515.
- Schlich, P. (1994). Grapes: a method and a SAS® program for graphical representations of assessor performances. *Journal of Sensory Studies*, 9(2), 157-169.
- Schlich, P. (2004). L'analyse en variables canoniques des données de profils sensoriels. Applications pour le contrôle des performances des sujets et pour la comparaison des produits. *8èmes Journées Agro-Industries et Méthodes Statistiques*, Rennes, 10-12 mars 2004. p33-43
- Schlich P. (1996). Defining and validating assessor compromises about product distances and attributes correlations. In *Multivariate Analysis of Data in Sensory Science*. T. Naes & E. Risvik (Eds.). Elsevier Science B.V., 259-306

Article 14

14. Pineau N. Chabanet C., Schlich P. (2007). Modeling the evolution of a sensory panel: a mixed-model and control chart approach. *Journal of Sensory Studies*, **22**, 212-241.

MODELING THE EVOLUTION OF THE PERFORMANCE OF A SENSORY PANEL: A MIXED-MODEL AND CONTROL CHART APPROACH

NICOLAS PINEAU^{1,3}, CLAIRE CHABANET² and PASCAL SCHLICH¹

¹*Laboratoire d'Interface Recherche-Industrie et de Sensométrie
Centre Européen des Sciences du Goût
15 rue Hugues Picardet, 21000 Dijon, France*

²*FLAveur, Vision, Comportement du consommateur
Institut National de la Recherche Agronomique-Dijon
Dijon, France*

Accepted for Publication July 28, 2006

ABSTRACT

A sensory panel is often used to profile the same type of product with the same set of attributes for many years. We are interested in characterizing the evolution of the performance of such a panel (and its panelists) over time. This article presents a methodology based on a mixed-model approach that takes into account the evolution of both panel and panelist in the same model. At the panel level, linear and quadratic evolutions of the performance are tested. At the panelist level, the method allows detection of whether some panelists perform better than others, and whether this difference remains the same or evolves over time. This mixed-model approach is followed by a graphical representation using a control chart method to identify occasional outliers. Data used to illustrate this methodology are eight sensory profiling data sets collected on ready-made frozen meals between 1997 and 2001 (every 6 months). The performance index chosen as an example in this study is the individual repeatability measured by standard deviation over replicates.

INTRODUCTION

Descriptive analysis provides quantitative information on the sensory properties of a set of products. To perform such an analysis, the first step consists of the training of a panel of assessors. During this period, a set of

³ Corresponding author. TEL: +33-3-80-68-16-38; FAX: +33-3-80-68-16-32; EMAIL: pineau@cesg.cnrs.fr

sensory attributes is established to describe the products. Panelists also work on the identification of their sensory perceptions and learn how to score the attributes on the scale (ISO 8586-1 1993). Although training was shown to have no significant impact on panel performance in a few studies (Moskowitz 1996), most of the studies that deal with the importance of training observed a significant increase of the performance compared to that of untrained panels. Training was found to increase discrimination (Wolters and Allchurch 1994) and consensus within a panel (Roberts and Vickers 1994; Chollet and Valentin 2001). In addition to these effects, training was observed to increase reliability (Labbe *et al.* 2004). Because the quality of measurements depends on the training period, the duration of this first step is often long. This is also one of the reasons why companies that use descriptive analysis often conduct several experiments over time with the same panel, even though a few assessors may leave the panel and be replaced by new panelists. Once panelists are considered sufficiently trained, the measurement phase of the descriptive analysis begins.

Like any other measurement instrument, panel performance must be controlled. In the literature, many statistical methodologies are dedicated to the control of assessor performances for sensory data collected in one experiment. Several studies are dedicated to one aspect of panel performance, such as repeatability (REP) (Rossi 2001) or agreement between panelists (Qannari *et al.* 1999). Other methodologies are based on several indexes or analyses of the performance (Næs and Solheim 1991; Schlich 1994; Couronne 1997). Nevertheless, the evolution of panel performance over time is seldom reported in the literature. A longitudinal study was investigated to determine whether aging, experience and exposure effect have an influence on panelist performance (Bitnes *et al.* 2007). However, it dealt with a basic taste identification task rather than descriptive analysis.

Investigations on the longitudinal study of the performance are of high importance for panel leaders as they make it possible to assess whether the panel performance increases or decreases over time. The evolution of each panelist's performance can also be assessed and compared to the panel.

The aims of this study were to present a methodology to characterize the evolution of a performance index over time at panel and panelist levels, to identify the best and worst panelists over a long period and to detect occasional but important variations in performance level. This article presents a statistical methodology based on a mixed-model approach to characterize the evolution of a REP index over time, which is the standard deviation (SD) over replicates. Mixed-model analysis was chosen because it makes it possible to model both panel and panelist evolution simultaneously. Moreover, a mixed model accommodates missing values. This univariate analysis is followed by graphical displays based on control chart techniques

TABLE 1.
PANELIST PARTICIPATION TO THE DATA SETS

Panelists	Data sets over time (year)								nbD
	1	1.5	2	2.5	3	3.5	4	4.5	
101	×	×	×		×	×	×	×	7
102	×	×	×	×					4
103	×	×	×	×	×	×	×	×	8
104	×		×	×	×	×	×	×	7
105		×	×	×	×	×	×	×	7
106	×	×	×		×				4
107		×		×		×	×	×	5
108	×	×	×	×	×	×	×	×	8
109	×	×	×		×	×	×	×	7
111	×	×	×	×	×	×		×	7
112	×		×	×	×	×	×		6
113	×		×	×	×	×	×	×	7
115	×	×		×	×				4
Nbpan	11	10	11	10	11	10	9	9	

Nbpan, number of panelists at each point in time; nbD, number of data set participation of each panellist.

to identify outliers. This methodology is exemplified using a succession of eight data sets (involving 13 panelists) over a period of 4 years.

MATERIALS AND METHODS

Longitudinal Data

The data used to illustrate this methodology are eight data sets from eight sensory descriptive analyses with common characteristics, performed every 6 months between 1997 and 2001. In each experiment, four ready-made frozen meals were tasted twice and scored over 30 attributes by 9 to 11 panelists among a set of 13 panelists (Table 1). From one experiment to another, the attributes were always the same (except for three attributes that were not assessed in the last three experiments), whereas the set of samples included the same brands or were different production batches from the same process. Attributes were scored on 0 to 10 continuous scales. Among the 30 attributes, there were seven appearance attributes (coded from ap_1 to ap_7), seven aroma attributes (from ar_1 to ar_7), seven odor attributes (from od_1 to od_7), three taste attributes (from ta_1 to ta_3), two trigeminal attributes (tr_1 and tr_2) and four texture attributes (from tx_1 to tx_4). During each 6-month

period between two consecutive measurements, panelists performed sensory analyses on other products and participated to sensory trainings.

REP Index

To follow the evolution of REP over time, the selected index was the root mean square of error (RMSE) obtained from a one-way analysis of variance (ANOVA) of the product effect, performed for each panelist, attribute and data set. It corresponds to the estimation of the SD over replicates. The higher this index is, the lower is the REP. One advantage of RMSE is that it has the same unit as the scoring scale. For example, an RMSE equal to 1.2 on a 0 to 10 continuous scale means that a replicate score deviates from the product mean from 12% of the total scale length. Moreover, because REP is the measurement of the variability within products, RMSE is not related to the variation between product samples. Therefore, variations in RMSE values do not depend on product variability from one point in time to another. This would not have been the case for a discrimination index or an agreement index, which is related to the ability to describe the product space, and hence, is product dependent. In the following, the RMSE scores are changed to their negatives and are called REP to display higher values for better REP scores. For each attribute, data are repeated measurements of this REP index over time.

Mixed Model

Mixed-model methodology was chosen to analyze this type of repeated measurement data because it makes it possible to take simultaneously into account the evolution of the panel as a whole and the evolution of the panelists. Moreover, mixed models accommodate missing values. For each attribute, the general model can be written as

$$y_{jt} = \mu_t + \text{panelist}(j, t) + \varepsilon_{jt} \quad (1)$$

where y_{jt} is the REP score of the panelist j at time t , μ_t is the time fixed effect, $\text{panelist}(j, t)$ is the panelist random effect and ε_{jt} is the error term.

Random variables from a given model are independent and normally distributed with specific variance. The fixed part models the evolution of the whole panel, whereas the random part models the panelists' behavior. The aim of the selection of the fixed part was to determine whether panel performance increases, decreases, reaches a minimum or a maximum or remains at a constant level. This is assessed by specifying a linear time effect, a quadratic time effect or no time effect in the fixed part of the mixed model. The selection of the random part makes it possible to determine whether the performance

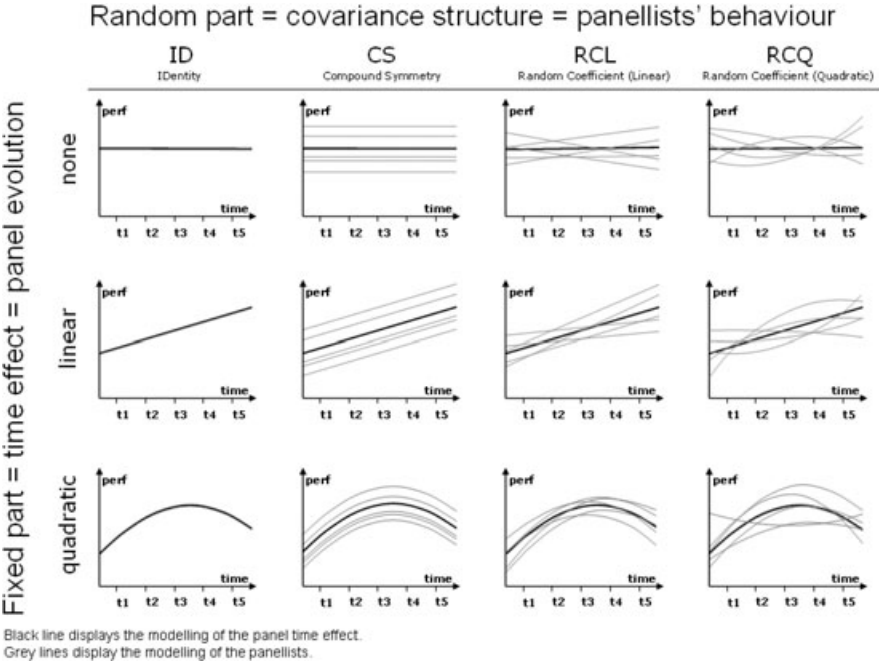


FIG. 1. GRAPHICAL DISPLAYS OF THE MIXED MODELS ACCORDING TO THE CHOICE OF THE FIXED AND RANDOM PARTS

level of some panelists increases or decreases faster than others, or increases faster at the beginning and then decreases progressively. This is assessed by specifying the covariance structure of the random part of the mixed model, as explained in “Choice of the Covariance Structure.” According to the selected covariance structure, the panelists’ behavior is assumed to be very close to panel evolution or not. Figure 1 shows the variety of the mixed models tested in this study according to the selected fixed and random parts, and displays a graphical representation of each case. The black curve is the overall panel evolution, and each gray curve represents the evolution of one panelist. For example, if the fixed time effect is quadratic (third row) and the covariance structure of the random effect is the compound symmetry (CS) (second column), it implies that panel performance reaches a maximum (or a minimum) and the evolution of each panelist has the same shape as the panel, but a different performance level at the beginning. This means that each panelist is supposed to have an intrinsic potential level, which is stable over experiments.

To select the model that fits the best to the data, the first step consists of the choice of covariance structure of the random part of the mixed model (Wolfinger 1996; Pinheiro and Bates 2000). The second step is to determine the fixed part.

Choice of the Covariance Structure. The choice of the covariance structure is linked to the supposed panelists' behavior compared to the panel evolution. For the selection of the random part covariance structure, the fixed part must be the same for every covariance structure tested, and must be as complete as possible. Thus, the time effect is chosen as a categorical factor. Among the large collection of covariance structures for repeated measures (Wolfinger 1996), four nested structures are tested: identity (ID), CS, linear random coefficient (RCL) and quadratic random coefficient (RCQ). ID structure is associated to no correlation over time. The choice of this structure implies that there is no significant difference between panelist performances over time. In other words, each panelist has no specific behavior and is supposed to evolve like the panel. Therefore, there is no panelist effect in the mixed model:

$$\begin{aligned} y_{jt} &= \mu_t + \varepsilon_{jt} \\ \varepsilon_{jt} &\sim N(0, \sigma_e^2) \end{aligned} \quad (2)$$

The CS structure implies a constant correlation between two measures, whatever the time interval between the two measures. If this structure fits the data, it means that the shape of the performance evolution of each panelist is the same as the panel, but the performance level at the beginning is different for each panelist. In other words, the curve of the evolution of each panelist is the same, but the intercept is different. It means that each panelist is supposed to have an intrinsic potential level at the beginning and to keep it over experiments. Graphical displays related to this structure are presented in the second column of Fig. 1. The CS structure corresponds to a random panelist effect. The expression of the mixed model becomes

$$\begin{aligned} y_{jt} &= \mu_t + \alpha_j + \varepsilon_{jt} \\ \alpha_j &\sim N(0, \sigma_{\text{int}}^2) \\ \varepsilon_{jt} &\sim N(0, \sigma_e^2) \end{aligned} \quad (3)$$

where α_j is the panelist random effect, measuring the intrinsic value of panelist j compared to the panel.

The RCL structure is characterized by a random intercept (as for CS) and a random slope. In this case, the speed of the evolution of the performance can

be different among the panelists. The graphs corresponding to this structure in Fig. 1 (third column) show that each panelist has his or her own intercept and slope, whatever the fixed time effect. The corresponding mixed model is

$$\begin{aligned}
 y_{jt} &= \mu_t + \alpha_j + \beta_j \cdot \text{time} + \varepsilon_{jt} \\
 \alpha_j &\sim N(0, \sigma_{\text{int}}^2) \\
 \beta_j &\sim N(0, \sigma_{\text{slope}}^2) \\
 \varepsilon_{jt} &\sim N(0, \sigma_e^2)
 \end{aligned} \tag{4}$$

where α_j is the panelist random effect, and β_j is the random slope parameter describing the speed of the evolution of the individual performance.

The RCQ structure adds a random panelist curvature effect. Individual performance levels can increase (resp. decrease) faster than the panel at the beginning and then decrease (resp. increase) at the end compared to the panel evolution. This structure is exemplified in the fourth column of Fig. 1. The corresponding mixed model is

$$\begin{aligned}
 y_{jt} &= \mu_t + \alpha_j + \beta_j \cdot \text{time} + \gamma_j \cdot \text{time}^2 + \varepsilon_{jt} \\
 \alpha_j &\sim N(0, \sigma_{\text{int}}^2) \\
 \beta_j &\sim N(0, \sigma_{\text{slope}}^2) \\
 \gamma_j &\sim N(0, \sigma_{\text{curv}}^2) \\
 \varepsilon_{jt} &\sim N(0, \sigma_e^2)
 \end{aligned} \tag{5}$$

where α_j is the panelist random effect, β_j is the random slope parameter and γ_j is the random curvature parameter.

The covariance structures being hierarchically organized, each model is tested against a model of higher order to determine the best covariance structure according to the data; the underlying principle being that a parsimonious model should be retained unless a more complex model (with more parameters) leads to a significant improvement in the fit. From this perspective, likelihood ratio tests, which make it possible to compare nested models, are performed (Pinheiro and Bates 2000). For each structure, the likelihood is computed by REstricted Maximum Likelihood (REML). This method provides a correction for the bias found in the usual maximum likelihood estimates (for further details, see Pinheiro and Bates 2000). Under the null hypothesis of no difference, likelihood ratio logarithm for two nested models with the same fixed part follows a chi-square distribution, which number of

degrees of freedom is the difference between the numbers of parameters of the models. When the test is significant ($P < 0.05$), the more complex model fits significantly better.

Another approach to compare the models is to compute an index such as the Akaike Information Criterion (AIC) (Pinheiro and Bates 2000):

$$\text{AIC} = -2\log(\text{likelihood}) + 2k \quad (6)$$

where k is the number of random parameters of the model.

This index, also based on the likelihood, is particularly appropriate in the sense that it trades off between the goodness of fit and the complexity of the covariance structure of the parameter estimates. Both methods are used in this study to compare the models selected by both approaches for each attribute.

Choice of the Fixed Part. The aim of the selection of the fixed part was to determine whether the evolution of the whole panel over time is linear, quadratic or if the performance remains at a constant level. The mixed model can be written as

$$\begin{aligned} y_{jt} &= \mu + \alpha \cdot \text{time} + \beta \cdot \text{time}^2 + \text{panelist}(j, t) + \varepsilon_{jt} \\ \varepsilon_{jt} &\sim N(0, \sigma_e^2) \end{aligned} \quad (7)$$

where $\text{panelist}(j, t)$ is the general expression of the random panelist effect, specified in “Choice of the Covariance Structure.” The parameter μ is the constant term (or the intercept) of the time effect. The parameters α and β correspond, respectively, to the slope and the curvature effects.

The significance of the linear and quadratic effects related to the parameters α and β is tested using conditional Fisher’s test (Pinheiro and Bates 2000) with type I mean squares. If the parameter α is significantly different from zero, panel performance increases or decreases over time according to the sign of the estimate. If the parameter β is significant, panel performance reaches a maximum (negative parameter estimate) or a minimum (positive parameter estimate). In Fig. 1, the evolution of the panel corresponds to the black curves. The graphs in the first row represent situations for which the time effect is not significant ($\alpha = \beta = 0$); the second row displays graphs for which time effect is linear ($\alpha \neq 0$ and $\beta = 0$) and the third row shows quadratic evolutions of the performance of the panel ($\beta \neq 0$).

Individual Parameters

Individual solutions for random effects (which correspond to the estimation of the α_j , β_j and γ_j parameters involved in models 3, 4 and 5) are calculated

in order to model the evolution of the performance of each panelist. Individual parameters make it possible to highlight panelists whose evolution is significantly different from that of the whole panel. These individual solutions for random effects and their associated standard errors are estimated by best linear unbiased predictors. The word “best” refers to minimum mean squared error (Pinheiro and Bates 2000; SAS 2003). According to the covariance structure selected in the mixed model, the number of parameters estimated for each attribute is different. There is no parameter for ID structure because there is no random effect (except the residual variance, which is estimated in every mixed model). Intercept parameters are estimated for the CS, RCL and RCQ covariance structures, the slope parameters are estimated for the RCL or RCQ structure and the curvature parameters are only estimated for the RCQ structure. Each parameter is tested using a Student’s t -test. When the estimation of a parameter is significant, the evolution of the given panelist is different from that of the panel. For example, considering that CS structure has been selected, panelists for whom the intercept parameter estimation is significantly higher than 0 are panelists for whom the performance level over time is higher than that of the panel. These tests make it possible to monitor individual performances over time.

Control Charts

The control chart is a graphical tool used to monitor the mean value μ of a quality characteristic and its variability σ over time (Montgomery 2001). This variability is used to determine control limits (CL) for this characteristic. The general expression for control chart limits is

$$CL = \mu \pm 3\sigma \quad (8)$$

The value of “3” is widely used for control charts; it represents a risk of 3 over 1,000 to wrongly reject an observation for a bilateral test if the variate is normally distributed. This very conservative value is used for industrial quality control processes to minimize the false-alarm risk. However, as the aim of the study was to identify points in time for which panel or panelist performance is different from the overall evolution and not only to detect highly drifting values, selected value for CL is lowered to 1.96 to obtain a confidence interval of 95%.

Generally, μ and σ are unknown parameters estimated from an ANOVA. For this study, because a mixed model was determined previously on the data, control chart parameters are estimated from the mixed-model parameters. Parameter estimation depends on the information monitored on the control chart. When the aim is to monitor performance variability at panel level,

represented data are the mean of the individual REP values at each point in time. The estimation of σ is $\hat{\sigma}_e/\sqrt{n}$, the residual root mean square of the mixed model divided by the square root of the number of panelists, n . The estimation of μ , called $\hat{\mu}_t$, is derived from the estimation of the parameters related to time effect in the mixed model. If the linear or quadratic fixed time effect is significant, then $\hat{\mu}_t$ is not constant over time. Therefore, the CL for panel REP is

$$CL_{\text{panel}} = \hat{\mu}_t \pm 1.96\hat{\sigma}_e/\sqrt{n} \quad (9)$$

where $\hat{\mu}_t = \hat{\mu}$ if neither the linear nor quadratic time effect is significant in the mixed model, $\hat{\mu}_t = \hat{\mu} + \alpha \cdot \text{time}$ if the linear time effect is significant and $\hat{\mu}_t = \hat{\mu} + \alpha \cdot \text{time} + \beta \cdot \text{time}^2$ if the quadratic time effect is significant.

The mean REP values outside CL are points in time for which the panel REP level does not conform to the modeled REP of the evolution of the panel.

When the aim is to monitor REP at the panelist level over time, represented values are individual REP values at each point in time. The estimation of μ , is still $\hat{\mu}_t$, but the estimation of σ depends on the covariance structure selected in the mixed model. Nevertheless, whatever the type of covariance structure, this variability $\hat{\sigma}_t$ is not divided by $\sqrt{n}$ because we are interested in a confidence interval around individual samples and not around mean samples. Therefore, CL for panelist REP is

$$CL_{\text{panelist}} = \hat{\mu}_t \pm 1.96\hat{\sigma}_t \quad (10)$$

where $\hat{\mu}_t$ is the same as Eq. (9), $\hat{\sigma}_t = \hat{\sigma}_e$ if the correlation structure is ID, $\hat{\sigma}_t = \sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_{\text{int}}^2}$ if the correlation structure is CS, $\hat{\sigma}_t = \sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_{\text{int}}^2 + \text{time}^2 \cdot \hat{\sigma}_{\text{slope}}^2}$ if the correlation structure is RCL and $\hat{\sigma}_t = \sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_{\text{int}}^2 + \text{time}^2 \cdot \hat{\sigma}_{\text{slope}}^2 + \text{time}^4 \cdot \hat{\sigma}_{\text{curv}}^2}$ if the correlation structure is RCQ.

Values outside CL point out panelists for whom the REP level is significantly different from panel REP at a given time.

Although the methodology presented in this article is unidimensional, a multidimensional insight is presented in a contingency table in which the rows correspond to the panelists and the columns to the points in time. Reported values are the numbers of attributes for which the panelist performance is significantly lower than the panel (example in the Table 2). This contingency table highlights the overall performance of each panelist with respect to all the attributes as well as points in time for which the number of detections of a bad performance is high.

All the analyses were conducted using SAS 9.1.3 software. Specific macros were developed to automatically select the best model and to display

TABLE 2.
SUMMARY OF THE OUTLIERS OVER TIME

Time	102	106	104	111	108	103	115	112	101	105	113	109	107	Sum	Ratio†
1	6	2	2	7	2						1			20	1.82
1.5	2	2		1	4	2	2							13	1.30
2	5	1	1	3		2			1					13	1.18
2.5	4		5	1	5									15	1.50
3		6	2	1	1									10	0.91
3.5			3	2			1			1				7	0.70
4			4		1									5	0.56
4.5			1											1	0.11
Sum	17	11	18	15	13	4	2	1	1	1	1	0	0		
Ratio*	4.25	2.75	2.57	2.14	1.63	0.5	0.5	0.17	0.14	0.14	0.14	0	0		

* Ratio = sum/number of experiments performed by the assessor.
† Ratio = sum/number of assessors who participated in the experiment.
Each cell contains the number of attributes for which panelist repeatability level was significantly lower than panel level. Gray cells indicate that a panelist does not participate to an experiment.

the corresponding control charts. The part of these macros dealing with the settings of the mixed models is available in Appendix 1.

RESULTS

Mixed Model: Choice of the Correlation Structure

For each attribute, mixed models related to each of the four covariance structures (ID, CS, RCL, RCQ) were performed and the choice of the covariance structure was made according to nested model tests (Table 3). In this table, each row corresponds to an attribute. It gives, for each hypothesis test between two nested models, the number of degrees of freedom for the chi-square statistics (df), the likelihood ratio logarithm (log lik) and the associated probability (prob). When prob < 0.05, the more complex model is considered to fit better. The best covariance structure, according to the nested model tests, is given in the last column (the full strategy used to select the best structure is explained in Appendix 2). For most of the attributes, CS is the covariance structure that fits the best to the data. The implication of this finding is that the evolution of the performance of each panelist is the same, but their initial level is different and this difference between panelists remains over experiments, whatever the overall evolution of the panel. Nevertheless, this general result can be modified depending on the type of attribute. CS was the most often selected structure for aroma, odor and taste attributes, but for the other types of

TABLE 3.
NESTED MODEL TESTS FOR COVARIANCE STRUCTURE

Descr	ID versus			CS versus						RCL versus			Choice						
	CS			RCL			RCQ			RCQ									
	df	log lik	prob	df	log lik	prob	df	log lik	prob	df	log lik	prob							
ap_1	1	4.06	0.0439	2	4.06	0.1313	3	4.06	0.255	1	<0.01	0.9753	2	<0.01	0.9995	1	<0.01	0.9999	CS
ap_2	1	0.33	0.5656	2	0.73	0.6941	3	0.73	0.8661	1	0.40	0.527	2	0.40	0.8187	1	<0.01	0.9999	ID
ap_3	1	3.98	0.0461	2	5.97	0.0505	3	6.42	0.0928	1	1.99	0.1583	2	2.44	0.2952	1	0.45	0.9297	CS
ap_4	1	<0.01	0.9785	2	0.03	0.9785	3	0.03	0.9785	1	0.03	0.9785	2	0.03	0.9785	1	<0.01	0.9785	ID
ap_5	1	10.68	0.0011	2	10.68	0.0047	3	10.68	0.0135	1	0.89	0.3454	2	0.89	0.6408	1	<0.01	0.9999	CS
ap_6	1	0.30	0.5838	2	1.75	0.4168	3	1.88	0.5976	1	1.45	0.2285	2	1.58	0.4538	1	0.12	0.9893	ID
ap_7	1	4.62	0.0316	2	8.44	0.0146	3	8.67	0.0340	1	3.82	0.0506	2	4.05	0.1319	1	0.23	0.9726	CS
ar_1	1	19.15	< 0.001	2	24.43	< 0.001	3	24.43	< 0.001	1	5.28	0.0215	2	5.28	0.0713	1	<0.01	0.9999	RCL
ar_2	1	6.72	0.0095	2	10.26	0.0059	3	10.26	0.0164	1	3.55	0.0595	2	3.55	0.1694	1	<0.01	0.9999	CS
ar_3	1	5.74	0.0165	2	5.74	0.0566	3	5.74	0.1249	1	<0.01	0.9999	2	<0.01	0.9999	1	<0.01	0.9999	CS
ar_4	1	5.60	0.0179	2	5.6	0.0608	3	5.6	0.1327	1	<0.01	0.9999	2	<0.01	0.9999	1	<0.01	0.9999	CS
ar_5	1	8.29	0.0039	2	9.42	0.0090	3	9.42	0.0241	1	1.14	0.2856	2	1.14	0.5655	1	<0.01	0.9999	CS
ar_6	1	4.75	0.0292	2	4.75	0.0930	3	4.75	0.1910	1	<0.01	0.9999	2	<0.01	0.9999	1	<0.01	0.9999	CS
ar_7	1	9.86	0.0016	2	9.87	0.0071	3	15.73	0.0012	1	0.01	0.9203	2	5.90	0.0523	1	5.90	0.0151	CS
od_1	1	11.04	0.0008	1	11.04	0.0008	3	11.04	0.0115	1	<0.01	0.9999	2	<0.01	0.9999	1	<0.01	0.9999	CS
od_2	1	0.61	0.4347	2	1.11	0.5740	3	1.11	0.7746	1	0.5	0.4795	2	0.50	0.7788	1	<0.01	0.9999	ID
od_3	1	23.30	< 0.001	2	23.59	< 0.001	3	23.63	< 0.001	1	0.29	0.5902	2	0.32	0.8521	1	0.03	0.9986	CS
od_4	1	21.09	< 0.001	2	24.42	< 0.001	3	25.18	< 0.001	1	3.33	0.068	2	4.10	0.1287	1	0.76	0.859	CS
od_5	1	8.21	0.0041	2	9.02	0.0109	3	9.02	0.0290	1	0.81	0.3681	2	0.81	0.6669	1	<0.01	0.9999	CS
od_6	1	10.20	0.0014	2	10.20	0.006	3	10.31	0.0161	1	<0.01	0.9999	2	0.10	0.9512	1	0.10	0.9918	CS
od_7	1	6.90	0.0086	2	10.40	0.0055	3	16.10	0.0011	1	3.50	0.0614	2	9.20	0.0101	1	5.70	0.0170	RCQ
ta_1	1	5.90	0.0151	2	6.38	0.0411	3	6.42	0.0928	1	0.49	0.4839	2	0.52	0.771	1	0.04	0.9978	CS
ta_2	1	8.04	0.0045	2	8.06	0.0177	3	11.09	0.0112	1	0.02	0.8875	2	3.05	0.2176	1	3.03	0.3870	CS
ta_3	1	8.52	0.0035	2	8.52	0.0141	3	8.52	0.0364	1	<0.01	0.9999	2	<0.01	0.9999	1	<0.01	0.9999	CS
tr_1	1	16.23	< 0.001	2	16.23	0.0002	3	25.07	< 0.001	1	<0.01	0.9999	2	8.84	0.012	1	8.84	0.0314	RCQ
tr_2	1	11.91	0.0005	2	11.94	0.0025	3	11.94	0.0075	1	0.04	0.8414	2	0.04	0.9801	1	<0.01	0.9999	CS
tx_1	1	26.39	< 0.001	2	26.39	< 0.001	3	39.95	< 0.001	1	<0.01	0.9999	2	13.56	0.0011	1	13.56	0.0035	RCQ
tx_2	1	45.51	< 0.001	2	45.51	< 0.001	3	45.51	< 0.001	1	<0.01	0.9999	2	<0.01	0.9999	1	<0.01	0.9999	CS
tx_3	1	7.83	0.0051	2	12.90	0.0015	3	12.9	0.0048	1	5.07	0.0243	2	5.07	0.0792	1	<0.01	0.9999	RCL
tx_4	1	16.28	< 0.001	2	17.87	0.0001	3	17.87	0.0004	1	1.59	0.2073	2	1.59	0.4515	1	<0.01	0.9999	CS

ID, identity; CS, compound symmetry; RCL, linear random coefficient; RCQ, quadratic random coefficient; df, number of degrees of freedom for the chi-square statistics; log lik, likelihood ratio logarithm; prob, associated probability.

attributes, other structures can be selected as well. For appearance attributes, the ID structure was selected for three attributes (ap_2, ap_4 and ap_6) out of seven. In addition, the probability values associated with nested model tests between the ID and CS structures for attributes ap_1 and ap_3 are very close to the significance limit of 0.05 ($\text{prob}_{\text{ID-CS}} = 0.044$ for ap_1 and $\text{prob}_{\text{ID-CS}} = 0.046$ for ap_3), so the ID structure may be well adapted to two additional appearance attributes. In this case, considering ID as a potential structure for appearance attributes, this means that the individual REP levels are not different among panelists over time for this type of attribute; thus, the REP level of the panelists is more homogeneous for appearance attribute. Concerning the trigeminal and the texture attributes, the CS structure was not selected as frequently as for the other types of attributes: half of the selected structures are RCL or RCQ. This could be an argument to consider that panelist performance evolution (compared to panel evolution) is more complex for trigeminal and texture attributes, because each panelist is able to have his or her own slope and/or curvature parameter compared to the panel.

Table 4 presents the AIC values for each attribute and each covariance structure. The lower this index is, the closer is the model fit. The choice of the covariance structure with the AIC index compared to the choice with nested model tests is different for four attributes only (ap_7, ar_2, ar_7, od_4). For these attributes, the covariance structure selected with AIC is more complex (RCL or RCQ rather than CS). However, it is worth mentioning that differences in covariance structure selection would have disappeared if the probability value for nested model tests was 0.07 instead of 0.05. Thus, AIC criterion leads to the choice of slightly more complex models in few cases, but conclusions are very close for both selection methods. In this article, the first method (nested model tests) is retained to keep more parsimonious models as often as possible.

Mixed Model: Choice of the Fixed Time Effect

For each attribute, the best covariance structure according to Table 3 was selected. Then, the fixed part of the model (Eq. 7) is selected. The intercept effect is significant for every attribute of each model, so the mean value of the REP index was always significantly different from 0. Time effect (linear or quadratic) is significant only for five attributes so the REP level of the panel usually does not evolve over time; panel performance is very likely to stay at the same level for most of the attributes. Table 5 presents the results of the parameter estimates for the attributes for which the time effect was significant ($\text{prob} < 0.05$). Four attributes had a linear time effect (ap_2, ap_5, od_2, od_6). The estimates of the slopes for both appearance attributes were negative so the REP level was decreasing over experiments. On the contrary, the slope

TABLE 4.
AKAIKE INFORMATION CRITERION (AIC) INDEX FOR
COVARIANCE STRUCTURES

Attribute	ID*	CS*	RCL*	RCQ*
ap_1	154.1	152.0	154.0	156.0
ap_2	130.2	132.2	134.2	136.2
ap_3	133.7	131.7	133.7	133.2
ap_4	117.6	119.6	119.5	121.6
ap_5	148.7	140.1	142.1	142.1
ap_6	188.7	190.4	190.9	188.8
ap_7	166.8	164.2	162.3	164.1
ar_1	185.5	168.4	165.1	167.1
ar_2	179.8	175.1	173.5	175.5
ar_3	188.8	185.1	187.1	189.1
ar_4	107.9	104.3	106.3	108.3
ar_5	116.1	109.8	110.7	112.7
ar_6	172.6	169.8	169.8	169.8
ar_7	180.6	172.7	174.7	170.9
od_1	152.5	143.4	145.4	147.4
od_2	163.1	165.1	167.1	169.1
od_3	98.3	77.0	78.7	80.7
od_4	184.6	165.5	164.2	165.4
od_5	157.3	151.1	152.3	152.3
od_6	164.1	155.9	155.9	157.8
od_7	207.9	203.1	201.6	197.9
ta_1	127.8	123.9	125.4	127.4
ta_2	153.9	147.9	149.9	148.8
ta_3	177.0	170.5	172.5	174.5
tr_1	153.5	139.3	141.3	132.5
tr_2	173.1	163.2	165.2	167.2
tx_1	191.0	166.7	169.7	155.1
tx_2	229.7	186.2	188.2	190.2
tx_3	173.3	167.5	164.4	166.4
tx_4	191.9	177.6	178.0	180.0
nb sel	4	17	5	4

* Best covariance structure is bold for each attribute; smaller AIC is better.

nb sel, number of selection for the corresponding covariance structure.

estimates for both odor attributes were positive; panel performance was increasing significantly. Only one attribute (ap_4) had a significant quadratic effect and the corresponding parameter estimate was positive, meaning that the REP had a minimum. However, the parameter estimate of the linear effect was almost significant (prob = 0.0707) and it was positive, indicating the general

TABLE 5.
PARAMETER ESTIMATES FOR FIXED TIME EFFECT (ONLY FOR THE ATTRIBUTES FOR WHICH LINEAR OR QUADRATIC EFFECT IS SIGNIFICANT)

Descr	Parameter*	Estimate†	SE	df	t Value	prob
ap_2	Intercept	-1.1751	0.08414	78	-13.97	<0.0001
	Slope (time)	-0.1206	0.04966	78	-2.43	0.0174
	Curvature (time ²)	-0.00270	0.04890	78	-0.06	0.9561
ap_4	Intercept	-1.4287	0.08064	78	-17.72	<0.0001
	Slope (time)	0.08722	0.04760	78	1.83	0.0707
	Curvature (time ²)	0.1130	0.04686	78	2.41	0.0183
ap_5	Intercept	-1.1611	0.1219	12	-9.52	<0.0001
	Slope (time)	0.1052	0.05156	66	2.04	0.0453
	Curvature (time ²)	0.02277	0.04926	66	0.46	0.6454
od_2	Intercept	-0.7860	0.1088	78	-7.22	<0.0001
	Slope (time)	0.1703	0.06425	78	2.65	0.0097
	Curvature (time ²)	-0.03292	0.06326	78	-0.52	0.6042
od_6	Intercept	-1.3022	0.1342	12	-9.70	<0.0001
	Slope (time)	0.1126	0.05490	66	2.05	0.0442
	Curvature (time ²)	-0.00999	0.05237	66	-0.19	0.8493

* Estimated parameter, corresponding fixed effect is in brackets.
† Estimate of the parameter related to the fixed effect.
SE, standard error related to the estimate; df, number of degrees of freedom; t Value, Student's statistic value; prob, probability associated to Student's statistic value.

increasing tendency of the REP level of the panel. It is worth mentioning here that a linear effect should never be dropped from a quadratic model, even though it is not significant, to maintain the hierarchical nature of the model. Thus, the REP level was more likely to remain constant at the beginning and then to increase progressively. The graphical display associated to the evolution of the attribute ap_4 will be presented and commented later in the control chart part (Fig. 2A).

Mixed Model: Individual Parameters

According to the mixed model selected for each attribute, corresponding individual parameters are estimated. In this study, the results are presented for only three attributes (ar_4, tx_3 and od_7), for which covariance structures are, respectively, CS, RCL and RCQ. Table 6 presents the parameters estimated for each panelist. For the attribute ar_4, the intercept parameter is significantly different from zero for the panelist 102 only. As the estimation is negative (-0.47), the REP level of this panelist is globally lower than that of the panel over the experiments. For the attribute tx_3 (RCL covariance structure), three panelists have a significant slope parameter. For panelists 102 and 104, the slopes are positive, so they improve their REP levels over time compared to the

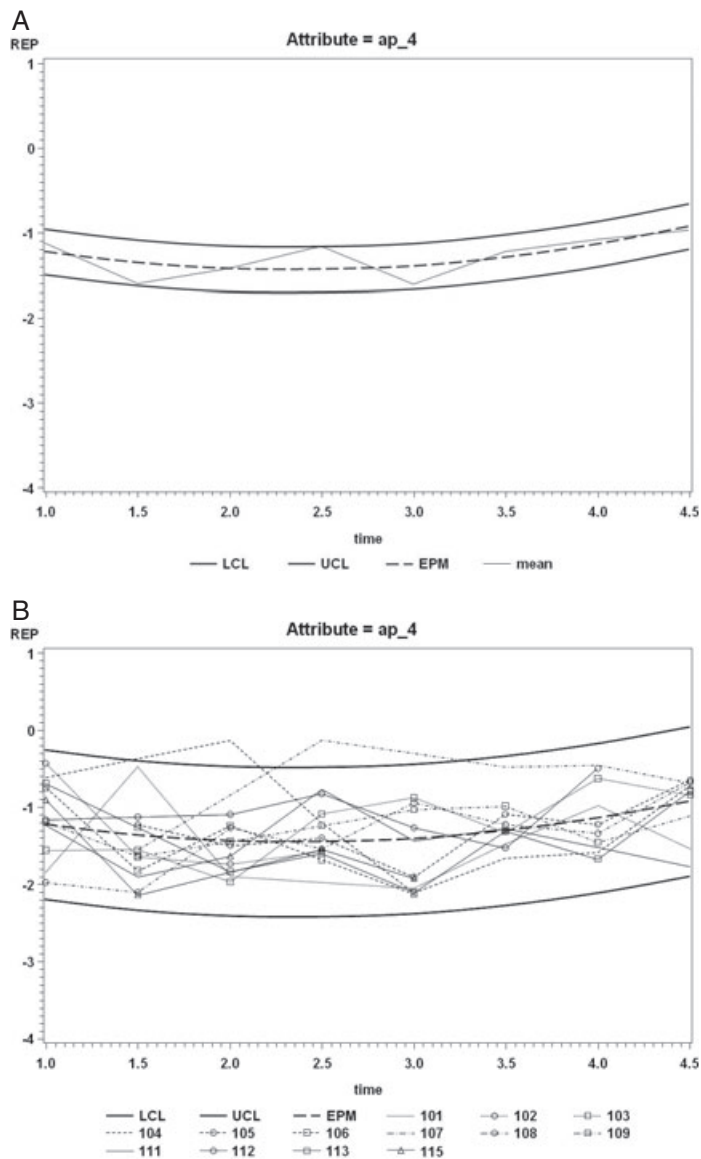


FIG. 2. CONTROL CHART OF THE EVOLUTION OF THE (A) PANEL AND (B) PANELISTS FOR ATTRIBUTE AP_4

(A) Significant quadratic time effect; no data outside the control limits. (B) Covariance structure = identity; significant quadratic time effect; 104 and 107 outside control limits once (upper side).

TABLE 6.
INDIVIDUAL PARAMETER ESTIMATES

Panelist	ar_4		tx_3		od_7	
	CS structure		RCL structure		RCQ structure	
101	Intercept	−0.131	Intercept	0.049	Intercept	0.003
			Slope	0.067	Slope	−0.03
					Curvature	−0.067
102	Intercept	−0.470*	Intercept	−0.337	Intercept	−0.001
			Slope	0.320**	Slope	0.043
					Curvature	−0.475*
103	Intercept	0.352	Intercept	0.129	Intercept	−0.021
			Slope	−0.17	Slope	0.087
					Curvature	−0.169
104	Intercept	−0.243	Intercept	−0.226	Intercept	0.003
			Slope	0.269**	Slope	−0.098
					Curvature	0.073
105	Intercept	0.257	Intercept	0.288	Intercept	−0.02
			Slope	0.23	Slope	−0.016
					Curvature	−0.344**
106	Intercept	−0.233	Intercept	−0.267	Intercept	0.001
			Slope	−0.098	Slope	0.039
					Curvature	−0.374**
107	Intercept	−0.121	Intercept	−0.081	Intercept	0.012
			Slope	0.054	Slope	0.035
					Curvature	0.222
108	Intercept	0.046	Intercept	0.114	Intercept	−0.016
			Slope	−0.055	Slope	−0.087
					Curvature	−0.02
109	Intercept	0.359	Intercept	0.441*	Intercept	0.035
			Slope	−0.052	Slope	0.022
					Curvature	0.407*
111	Intercept	−0.336	Intercept	−0.057	Intercept	−0.029
			Slope	−0.261**	Slope	0.009
					Curvature	0.017
112	Intercept	0.214	Intercept	0.317	Intercept	0.015
			Slope	−0.04	Slope	0.031
					Curvature	0.142
113	Intercept	0.147	Intercept	−0.354	Intercept	0.036
			Slope	0.024	Slope	−0.009
					Curvature	0.316**
115	Intercept	0.161	Intercept	−0.015	Intercept	−0.015
			Slope	−0.051	Slope	−0.026
					Curvature	0.273

* $P < 0.05$.
** $P < 0.10$.
Significant estimates are bold.

evolution of the panel. In contrast, the slope of panelist 111 is negative; the REP level of this panelist deteriorates over time compared to the evolution of the panel. Concerning the attribute *od_7* (RCQ covariance structure), five panelists have a significant curvature parameter, but the slope parameter is not significant. The REP of the panelists 102, 105 and 106, for whom curvature estimates are negative, increases faster than the panel REP at the beginning and then decreases faster. It is the opposite phenomenon for the panelists 109 and 113 for whom curvature estimates are positive. Such findings indicate that the behavior of these panelists is substantially different from the average evolution of the panel, so the average evolution does not well reflect each panelist's behavior.

Control Charts

According to the mixed model selected for each attribute, corresponding control chart parameters are calculated. In this study, control charts for only 4 out of 30 attributes are displayed in Figs. 2–5 corresponding to attributes *ar_4*, *tx_3*, *od_6* and *ap_4*. Upper CL (UCL) and lower CL (LCL) are displayed in black and bold. The panel mean estimated from the mixed model is called Estimated Panel Mean (EPM). It is also displayed in black and bold, but with a dotted line.

The control chart for attribute *ar_4* is an example for which the mixed-model covariance structure was CS and the time effect was not significant. This is the mixed model selected for most of the attributes. Because the time effect is not significant, CL values are constant over time. Figure 3A is the control chart for the panel mean level. Panel performance is never outside the CL, so panel REP is homogeneous over time. Figure 3B is the control chart for panelist REP. It evidences that, although global panel performance was not outside the CL at $t = 2$, three panelists (102, 104 and 111) had a significantly lower REP level than other panelists at this time.

Figure 4A,B displays the results for attribute *tx_3* for which the covariance structure was RCL and the time effect was not significant. Contrary to the first example, the REP level of the whole panel is significantly better at time $t = 2.5$ and $t = 4$, but none of the panelists is significantly better than the others at these times. Nevertheless, the REP level increases for most of the panelists at $t = 2.5$ in Fig. 4B. Moreover, two panelists (102 and 104) are pointed out as having lower REP than the panel for $t = 1$. Because variability calculated for CL depends on time if the RCL or RCQ covariance structure is selected, as presented in Eq. (10), CL values change over time.

Figure 5A,B displays the results for attribute *od_6* for which the covariance structure was CS and the linear time effect was significant. Therefore, the estimation of the REP mean value evolves over time and the CL as well. Linear evolution over time corresponds to a significant increase in the REP level over

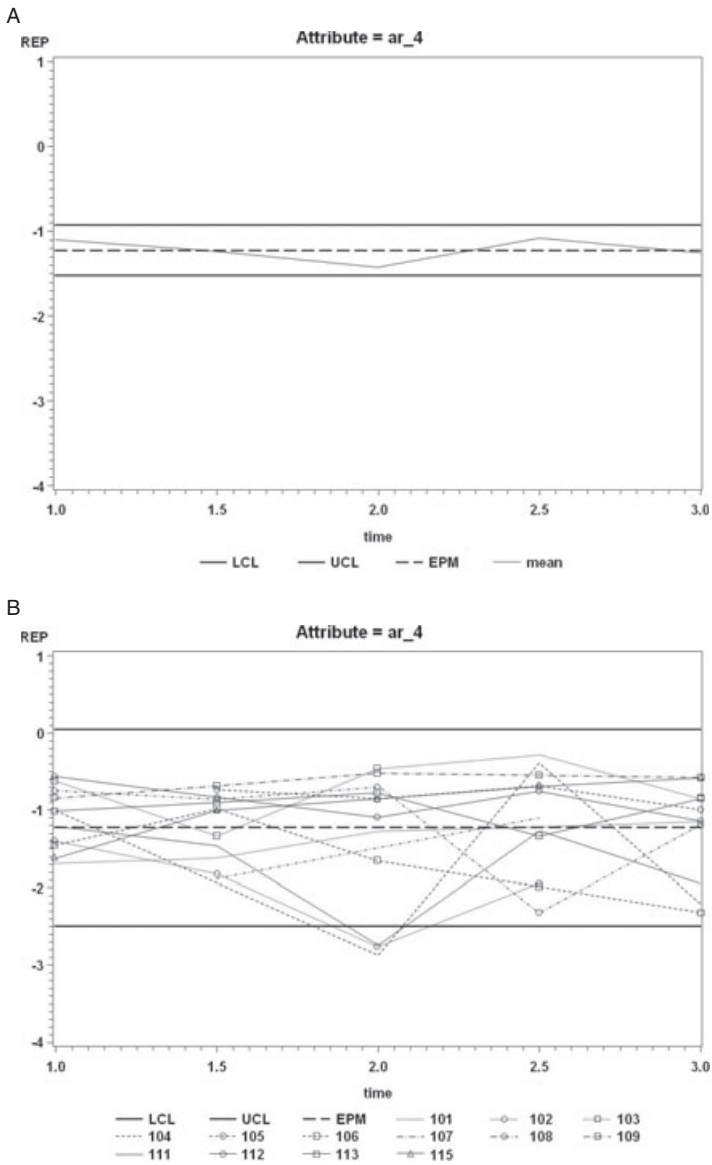


FIG. 3. CONTROL CHART OF THE EVOLUTION OF THE (A) PANEL AND (B) PANELISTS FOR ATTRIBUTE AR_4

(A) Time effect n.s.; no data outside the control limits. (B) Covariance structure = compound symmetry; Time effect n.s.; 102, 104 and 111 outside the control limits at time $t = 2$.

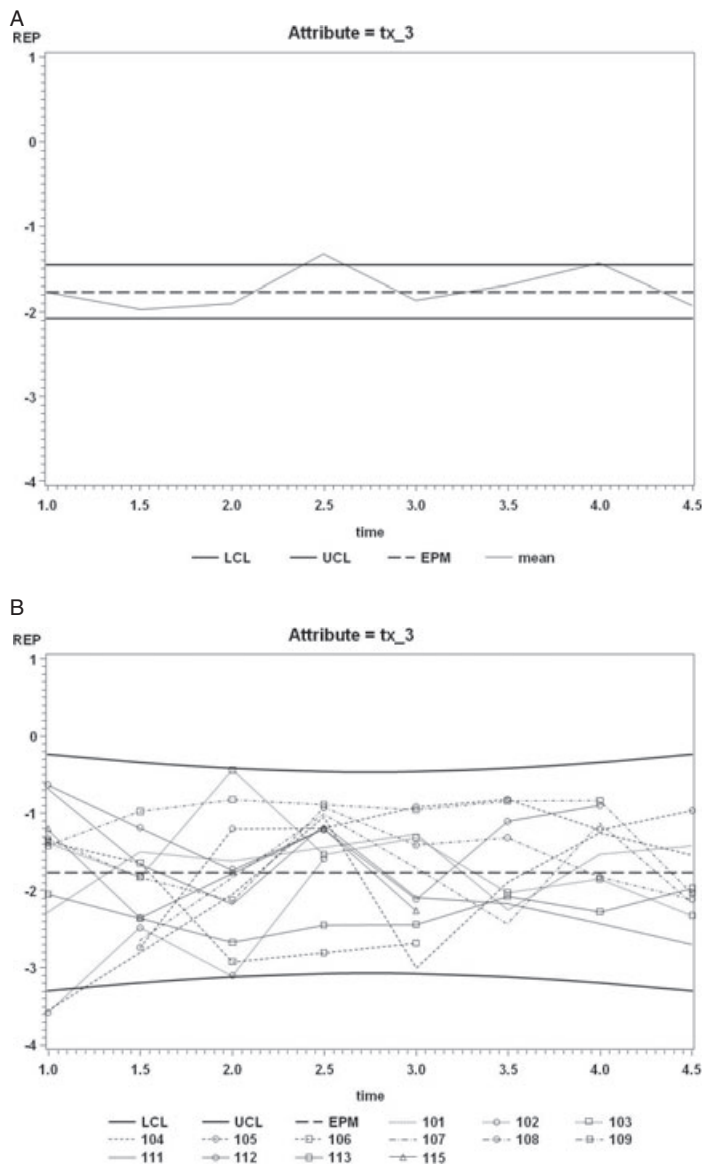


FIG. 4. CONTROL CHART OF THE EVOLUTION OF THE (A) PANEL AND (B) PANELISTS FOR ATTRIBUTE TX_3

(A) Time effect n.s.; panel performance outside the control limits at $t = 2.5$ and $t = 4$.
(B) Covariance structure = linear random coefficient; time effect n.s.; 102 and 104 outside the control limits at $t = 1$.

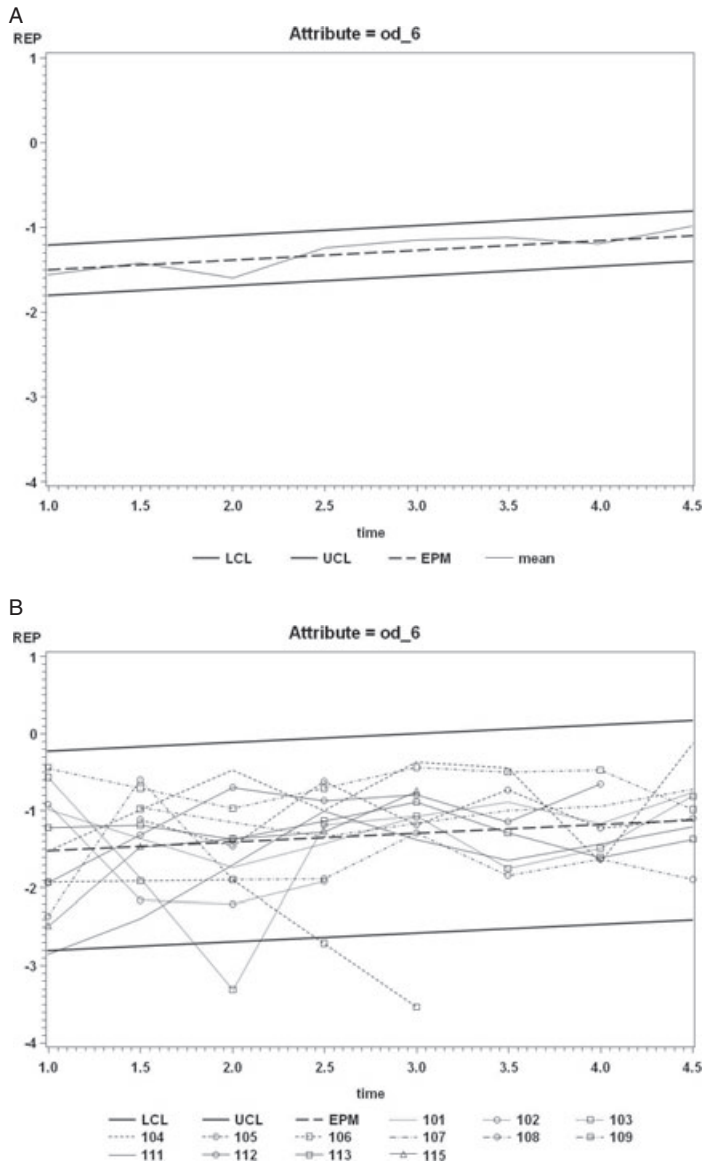


FIG. 5. CONTROL CHART OF THE EVOLUTION OF THE (A) PANEL AND (B) PANELISTS FOR ATTRIBUTE OD_6

(A) Significant linear time effect; no data outside the control limits. (B) Covariance structure = compound symmetry; significant linear time effect; 103, 106, 111 outside the control limits once.

time (slope estimation = 0.113, $\alpha = 0.042$). Three of the panelists are declared to be outside CL at different points in time (panelists 103, 106 and 111) in Fig. 5B.

Figure 2A,B displays the results for attribute ap_4 for which the covariance structure was ID and the quadratic time effect was significant. Like for Fig. 5A,B, the estimation of the REP mean value evolves over time but with a quadratic aspect. The panel REP level tends to slightly decrease at the beginning before slightly increasing from $t = 3$ to $t = 4.5$. Two panelists are declared to have better REP than the others (panelist 104 for $t = 2$ and panelist 107 for $t = 2.5$).

To make a synthesis of the alerts generated by the control charts across the attributes, Table 2 shows panelist detections when the performance level is significantly lower than the panel level. Some of the panelists are often detected to be outliers (panelists 104 and 102), whereas others have their REP level above the lower limit for all the attributes and all points in time (panelists 109 and 107). Furthermore, the number of detections decreases over time, meaning that the panelists' REP levels seem to be more homogeneous over time.

DISCUSSION

The methodology presented in this article makes it possible to take into account both panel and panelists' variations in the same model by determining, respectively, the fixed part and the random part. For each attribute, the selected mixed model gives information on the overall evolution of the panel and the behavior of the panelists over experiments. Control charts at panel level identify points in time for which panel performance is significantly different from the selected mixed model and control charts at panelist level identify, for each point in time, the panelists whom performance level is significantly different than the panel.

Concerning the fixed part of the mixed model, results show that the time effect is not significant for the majority of the attributes on these data. Thus, the panel REP seems to remain at a constant level over time. This finding can be related to the fact that the data did not contain the first sessions assessed by the panelists because the protocol was different during the first 2 years and no replications were performed, hence, individual REP cannot be calculated. Therefore, we can suppose that most of the improvement of the panel may have happened during the early sessions. The constant REP level achieved in the data used in this article give nonetheless information on the long-term evolution of the panel, showing that the performance does not decrease over years whereas weariness could be expected.

Furthermore, the choice of the covariance structure and the time effect can be discussed. In this article, the best model is chosen for each attribute, according to nested model tests and Fisher's test. This kind of selection gives the possibility to present the variety of conclusion one can obtain by using this mixed-model and control chart approach. However, if one model is chosen for each attribute, it is assumed that panel and panelist evolution can be completely different from one attribute to another. In sensory analysis, the selection of the same model for every attribute, or at least by type of attribute, might be more practical.

The methodology presented in this article is exemplified for a REP index, but other indices can be used. Nevertheless, one must pay attention to the "side factors" that could influence the evolution of the index under consideration. For example, a discrimination index does not only depend on panel performance, but also on product variability over time. Moreover, the selected index must satisfy several assumptions. In particular, the residuals of the selected mixed model must follow the normal distribution. In our case, the Kolmogorov–Smirnov goodness-of-fit test showed that this assumption was fulfilled by only 16 attributes out of 30. Nevertheless, non-normality of the residuals resulted from the presence of outliers for most of the attributes. Once these values are discarded, normality of the residuals was assessed for 25 out of 30 attributes. Another possibility would be to transform the data to normality before the analysis while still reporting the results and presenting the charts in the original units of measure. In the case study, the transformation by the square root of the RMSE would reduce the number of attributes for which non-normality is evidenced. However, interpretation of the covariance structure and time effect would be more difficult because they would not be directly related to the REP index. Thus, we did not feel confident to make covariance structure choice on the basis of the transformed data. Therefore, the RMSE index without transformation was kept in order to facilitate the interpretation.

CONCLUSION

The major aim of this study was to highlight the value of a mixed-model approach for the analysis of the evolution of REP over time. This methodology takes into account both panel and panelist effects simultaneously. Mixed-model analysis is followed by graphical displays based on control chart methodology that uses parameters of the mixed model. This makes it possible to highlight occasional outliers. Concerning the choice of the random part of the mixed model, the CS structure is the most representative covariance structure across the attributes of the data sets analyzed in this study, meaning that each

panelist seems to have his or her own intrinsic level. Nevertheless, these differences among panelists tended to disappear for appearance attributes (structure ID), whereas the covariance structure seems to be more complex for texture and trigeminal attributes (RCL or RCQ). Concerning the fixed part, the REP level remains constant over time for the majority of the attributes. However, this conclusion is obtained for a panel that is used to performing sensory analysis. It could be interesting to use such an analysis on REP indices computed during training sessions to assess the evolution of the performance level.

ACKNOWLEDGMENTS

We thank Technological Center Aériol, which provided the data that served to illustrate our approach, and El Mostafa Qannari for his comments.

REFERENCES

- BITNES, J., MARTENS, H., UELAND, Ø. and MARTENS, M. 2007. Longitudinal study of taste identification of sensory panellists: Effect of ageing, experience and exposure. *Food Qual. Prefer.* 18, 230–241.
- CHOLLET, S. and VALENTIN, D. 2001. Impact of training on beer flavor perception and description: Are trained and untrained subjects really different? *J. Sensory Studies* 16(6), 601–618.
- COURONNE, T. 1997. A study of assessors' performance using graphical methods. *Food Qual. Prefer.* 8(5/6), 359–365.
- ISO 8586-1. 1993. *Sensory Analysis – General Guidance for the Selection, Training and Monitoring of Assessors – Part 1: Selected Assessors*, AFNOR, Paris, France.
- LABBE, D., RYTZ, A. and HUGI, A. 2004. Training is a critical step to obtain reliable product profiles in a real food industry context. *Food Qual. Prefer.* 15(4), 341–348.
- MONTGOMERY, D.C. 2001. *Introduction to Statistical Quality Control*, 4th Ed., Wiley, Hoboken, NJ.
- MOSKOWITZ, H.R. 1996. Experts versus consumers: A comparison. *J. Sensory Studies* 11(1), 19–37.
- NÆS, T. and SOLHEIM, R. 1991. Detection and interpretation of variation within and between assessors in sensory profiling. *J. Sensory Studies* 6(3), 159–177.

- PINHEIRO, J.C. and BATES, D.M. 2000. *Mixed-effects Models in S and S-PLUS*, Springer-Verlag, New York, NY.
- QANNARI, E.M., MACFIE, H.J. and COURCOUX, P. 1999. Performance indices and isotropic scaling factors in sensory profiling. *Food Qual. Prefer.* 10(1), 17–21.
- ROBERTS, A.K. and VICKERS, Z.M. 1994. A comparison of trained and untrained judges' evaluation of sensory attributes intensities and liking of Cheddar cheeses. *J. Sensory Studies* 9(1), 1–20.
- ROSSI, F. 2001. Assessing sensory panelist performance using repeatability and reproducibility measures. *Food Qual. Prefer.* 12(5–7), 467–479.
- SAS. 2003. *SAS/STAT User's Guide (Version 9.1.3), MIXED Procedure*, SAS Institute, Cary, NC.
- SCHLICH, P. 1994. Grapes: A method and a SAS program for graphical representations of assessor performances. *J. Sensory Studies* 9(2), 157–169.
- WOLFINGER, R.D. 1996. Heterogeneous variance–covariance structure for repeated measures. *JABES* 1(2), 205–230.
- WOLTERS, C.J. and ALLCHURCH, E.M. 1994. Effect of training procedure on the performance of descriptive panels. *Food Qual. Prefer.* 5(3), 203–214.

APPENDIX 1

Using SAS software, the MIXED procedure is available to perform ANOVA with mixed effects. The most complex model, for which fixed time effect is quadratic and covariance structure is RCQ, is carried out using this program:

```
ODS output solutionF=SolF
              covparms=CovP
              tests1=Tst1
              solutionR=SolR;

proc MIXED data=example;
  CLASS panellist;
  MODEL REP= time0 time0*time0 / htype=1 solution;
  RANDOM int time0 time0*time0 / subject=panellist
  solution;
run;
```

“ODS output” is not mandatory but this statement creates the SAS data sets named SolF, CovP, Tst1 and SolR (in this example) that contain the output results.

Panelist is specified in the statement CLASS (classification variable).

The fixed part of the model is determined in the MODEL statement. REP is the variable that contains the repeatability values for each panelist at each point in time. “time0 time0*time0” corresponds to the quadratic time effect. Time variable is called time0 as a reminder that time values must be centered (mean = 0) to obtain correct estimations of the random parameters (proc mixed user’s guide SAS 2003). In this example, time values ranged from 1 to 4.5 (years) and the average time was 2.25, so the time values were turned into values from -1.75 to 1.75. htype=1 lead to the use of Type I mean squares for the estimation of the fixed effects (results are stored in the Tst1 data set). The option “solution” enables to get the estimates of the fixed parameters (stored in the SolF data set).

The random part of the model is determined in the RANDOM statement. The RCQ structure is specified by “int time0 time0*time0 / subject=panellist.” The covariance parameter estimates are stored in the data set CovP. The option “solution” provides the estimates of the individual parameters (stored in the SolR data set).

Results from the output window for the attribute od_7 are displayed below:

----- attribute = od_7 -----

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	214.24460768	
1	4	199.73678559	.
2	3	199.36443773	.
3	1	199.30991603	0.00005311
4	1	199.30839397	0.00000008
5	1	199.30839179	0.00000000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Intercept	panellist	0.007872
time0	panellist	0.01578
time0*time0	panellist	0.1017
Residual		0.4793

Fit Statistics

-2 Res Log Likelihood	199.3
AIC (smaller is better)	207.3
AICC (smaller is better)	207.9
BIC (smaller is better)	209.6

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	-1.2963	0.1184	12	-10.95	<.0001
time0	-0.01446	0.08300	12	-0.17	0.8646
time0*time0	-0.04270	0.1125	12	-0.38	0.7109

Solution for Random Effects

Effect	panellist	Estimate	Std Err Pred	DF	Value	Pr > t
Intercept	101	0.002614	0.08662	42	0.03	0.9761
time0	101	-0.02971	0.1104	42	-0.27	0.7893
time0*time0	101	-0.06673	0.1566	42	-0.43	0.6721
Intercept	102	-0.00123	0.08721	42	-0.01	0.9888
time0	102	0.04344	0.1226	42	0.35	0.7249
time0*time0	102	-0.4755	0.1947	42	-2.44	0.0189
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Intercept	115	-0.01541	0.08699	42	-0.18	0.8602
time0	115	-0.02569	0.1229	42	-0.21	0.8354
time0*time0	115	0.2726	0.1951	42	1.40	0.1696

Type 1 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
time0	1	12	0.02	0.8886
time0*time0	1	12	0.14	0.7109

The table **Iteration History** presents the iterations for the calculation of the likelihood estimated by REML (-2 Res Log Like). At the end of the iterations, -2 Res Log Like = 199.3. This value is reported in the table **Fit Statistics** in which AIC and two additional information criteria (AICC and BIC) are calculated. According to Eq. (6), one can check that $AIC = -2 \text{ Res log likelihood} + 2 \cdot k = 199.3 + 2 \cdot 4 = 207.3$, k being the number of random parameters of the model (intercept, time0, time0*time0 and the residual variance). The result of the AIC for the attribute od_7 is not the same as the one presented in Table 4 because the fixed part taken in this example is a quadratic time effect whereas the time effect was chosen as a

categorical factor for the results displayed in Table 4. The corresponding result was $-2 \text{ Res log likelihood} = 189.9$, so $\text{AIC} = 189.9 + 2 \cdot 4 = 197.9$.

The table **Covariance Parameter Estimates** gives the estimates of the variance–covariance parameters associated with the selected random effect. Since RCQ was chosen in this example, the variances of three parameters related to the panelist effect (intercept, time0 and time0*time0) and the variance of the residual are estimated.

The table **Solution for Fixed Effects** provides the estimate and the associated Student's statistics for each fixed effect parameter. In this example, neither the slope parameter (estimate = -0.01446 , $P = 0.8646$) nor the curvature parameter (estimate = -0.04270 , $P = 0.7109$) is significantly different from zero.

Individual parameter estimates are displayed in the table **Solution for Random Effects**. Curvature parameter for the panelist 102 is significantly lower than that of the panel (estimate = -0.4755 , $P = 0.0189$). The parameter estimates of every panelist are also displayed in the last column of Table 6.

The last table presents the tests of the fixed effects calculated with Type I mean squares. Linear and quadratic time effects are not significant.

To carry out an analysis with a linear time effect or no time effect (fixed part), model statement in the SAS program should be, respectively, replaced by:

```
model REP= time0 / htype=1 solution; (linear time effect)
model REP= / solution; (no time effect; data set Tst1 is not available)
```

RCL and CS covariance structures are carried out by replacing the random statement, respectively, by:

```
random int time0 / subject=panellist solution; (RCL structure)
random int / subject=panellist solution; (CS structure)
```

To carry out the covariance structure ID, random statement must be dropped out (data set SolR is not available in this case).

It is worth mentioning that when mixed models are carried out to select the best correlation structure (Choice of the Covariance Structure), time effect must be chosen as a categorical factor so time0 is declared in the statement CLASS. The SAS program for the MIXED procedure becomes:

```
proc MIXED data=example;
  CLASS time0 panellist;
  MODEL REP= time0 / htype=1 solution;
```



```

RANDOM int t0 t0*t0 / subject=panellist solution;
run;

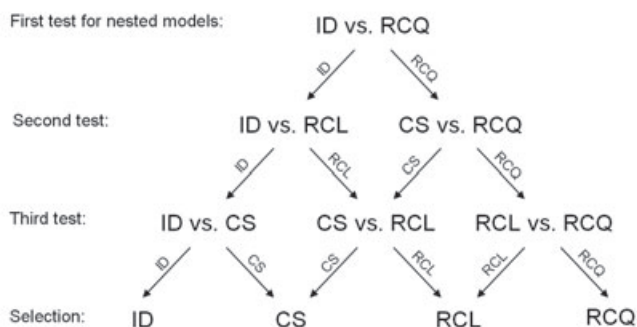
```

As the time variable must be defined as a categorical factor for the MODEL statement, another variable that is exactly the same as time0 (and called t0) must be created in the input data set and used in the RANDOM statement.

The comparison between two covariance structures (for models with the same fixed part) is performed by computing the corresponding likelihood ratio logarithm. Under null hypothesis of no difference, this ratio follows a chi-square distribution, which number of degrees of freedom is the difference between the numbers of parameters of the models. For example, to compare RCL and RCQ covariance structures for the attribute `od_7`, both models are performed with a fixed time effect chosen as a categorical factor. $-2 \text{ Res Log Likelihood} = 189.9$ for RCQ and $-2 \text{ Res Log Likelihood} = 195.6$ for RCL. The logarithm of the ratio of the likelihoods, $R \log \text{lik}$, is equal to the difference between the likelihood logarithms: $R \log \text{lik} = 195.6 - 189.9 = 5.7$. The number of degree of freedom is equal to $4 - 3 = 1$. The probability corresponding to the chi-square statistics for a value of 5.7 with one degree of freedom is $P = 0.017$ and thus the RCQ model is selected. These values are reported in the Table 3, at the end of the row corresponding to the attribute `od_7`.

APPENDIX 2

The strategy for the selection of the best covariance structure according to the nested model tests is represented in the following decision tree:



At each step, the model that is not rejected is tested versus another model. In this strategy, ID and RCQ structures are less likely to be selected

because they must succeed in three nested model tests whereas CS and RCL can be selected even though they fit better to the data for only one test. Though this strategy is unbalanced, inappropriate selections occur rarely with real data because the selected structure is most often better than any other one.

Article 15

15. Pineau N., Cordelle S., Mathonière C., Etiévant P., Issanchou I., Imbert A., Rogeaux M., Schlich P. (2006). Temporal dominance of sensations: a new technique to record several sensory attributes simultaneously over time. *Food Quality and Preference*. Under review.

Temporal Dominance of Sensations

a new technique to simultaneously record several sensory attributes over time

N. Pineau^a, P. Schlich^a, S. Cordelle^a, C. Mathonnière^a, S. Issanchou^b, A. Imbert^c, M. Rogeaux^d,
P. Etiévant^b

^a: *Laboratoire d'Interface Recherche-Industrie et de Sensométrie, Centre Européen des Sciences du Goût, 15 rue Hugues Picardet, 21000 Dijon, France*

^b: *UMR FLAVIC, Institut National de Recherche Agronomique, 17 rue Sully, BP 86510, 21000 Dijon cedex, France*

^c: *Fromageries Bel, Vendôme, France*

^d: *Danone Vitapole, Palaiseau, France*

Abstract

Temporality is essential in the sensory perception of food and beverages. Time-intensity (TI) methodology is useful for studying the temporal aspects of the perception of a given sensory attribute in a product. However, when TI is used for several attributes, it quickly becomes very time-consuming because one run is needed by attribute. This paper proposes a new method of sensory evaluation, called Temporal Dominance of Sensations (TDS), that makes possible to collect simultaneously temporal data for up to ten attributes on complex food products. The method consists in presenting the attributes simultaneously on a screen and asking the subject to select and score the sensory attribute that is perceived as dominant over the others. During the tasting of one sample, every time the subject thinks another attribute becomes dominant, he or she selects and scores this attribute, until perception ends. Based on two studies on dairy products, this paper explains how to describe the temporality of the sensations in a product thanks to the TDS curves. In the first study, scores obtained by the TDS method provide information which recovers the information obtained from another panel having performed regular descriptive analysis of the same products. In addition, TDS data give supplementary information on the sensory temporality of the products. In the second study, TI and TDS were

performed on the same products. When a succession of sensations was observed in TI, the same sequence was exemplified in the TDS curves. Moreover, in some products, TI did not find a succession of attributes, whereas TDS did.

Keywords: temporal dominance of sensations, time-intensity, sensory attributes, dairy products

1. Introduction

The evolution of sensations in sensory analysis of food products is increasingly studied. To obtain such a temporal information, the most often used method is time-intensity (TI) which consists in recording the evolution of the intensity of a given sensory attribute over time (Lee III & Pangborn, 1986). This method has been used on a large variety of products such as model solutions (Bonnans & Noble, 1995), beer (King & Duineveld, 1999), wine (Ishikawa & Noble, 1995) and with solid food, where texture properties were also evaluated, like meat (Duizer, Gullett & Findlay, 1993), or cheese (Wendin, Janestad & Hall, 2003). In all these studies, TI methodology is performed only on a small number of attributes or with a limited number of products since only one attribute is evaluated at a time. Nevertheless, studies on a large number of products and with several attributes have been undertaken (Chaya, Perez-Hugalde, Judez, Wee & Guinard, 2004) but they are time-consuming because of the requirement of a large run number. A method called dual-attribute was developed by Duizer, Bloom & Findlay, 1996) to score simultaneously two attributes by moving a cursor on an X,Y map in which each axe refers to one of the two attributes. In addition to a reduction of the time needed to perform the analysis, Zimoch & Findlay, 1998) show, in an application on meat, a reduction of the “halo-dumping” effect (Clark & Lawless, 1994) with this method. However, this method could not be extended to more than two attributes. Another possibility to reduce halo-dumping effect is to use discontinuous TI (Clark & Lawless, 1994). In this method, subjects are asked to give an intensity score to an attribute at several moments chosen by the experimenter during the evaluation. The attributes can be different for each point in time during one tasting. (Pionnier et al., 2004) used this method on cheese products to register temporal data concerning up to five attributes on the

same run. Nevertheless, the number of runs needed to obtain sufficient temporal information is as important as with classical TI.

In order to reduce the duration of the experiment and to avoid the halo-dumping effect, this paper presents a new method, called TDS for Temporal Dominance of Sensations. It consists in presenting to the panellist the complete list of attributes on a computer screen. Thereafter, the panellist is asked to assess which of the attributes is perceived as dominant and to give a score to this dominant sensation. In the course of the evaluation, when the panellist considers that the dominant attribute has changed, he or she is asked to score the new dominant attribute, and so on, until the perception ends. This new methodology provides curves of the evolution of each attribute over time that represent the temporality of the sensations associated with the product. In order to check the validity of the scores given by TDS methodology, the results of a first study based on 12 dairy products were compared with conventional profiling data performed on the same products. A second study, using five different dairy products, compares the results given by TDS with those provided by TI.

2. First experiment: Conventional profiling versus TDS

2.1 Materials and methods

2.1.1 Panelist selection and training

In order to compare conventional profiling and TDS procedures, two panels (called CP group and TDS group) were trained on one of each methodology. Selection tests were similar for both groups. This includes salivary flow measurement, triangular test on dairy products, term generation on dairy products, taste identification, descriptive test of two dairy products, ranking test, flavour recognition and chewing efficiency. At the end of the tests, two groups of 16 panellists were selected and trained. In the first twelve training sessions, panellists learned to recognize tastes and textures and drew up a list of 10 attributes to describe dairy products. This list had to be the same for both groups in the study. Three other sessions were needed to establish a consensual definition for each attribute. Thereafter, three sessions were devoted to familiarizing the panellists with the use of the computer system for CP group and 6 sessions for TDS group.

2.1.2 Experiment

The samples were presented monadically and assessors were asked to rinse their mouth with water and to eat bread between each sample. The presentation orders of the products were balanced using mutually orthogonal Latin squares. The tests were run in red light conditions to avoid colour identification of the products.

The two groups tasted twelve dairy products (AF, AFKA, AFKG, AFKI, EF, EFKA, EFKG, EFKI, KA, KG, KGKA, KGKI). The attributes selected by the panel were: astringent, diacetyl, sweet, elastic, firm, fluid, melting, lactic, pasty and salty. Total duration of the tasting of each sample was 90 s. Respectively 6 and 3 replicates of each product were made by TDS group and CP group. TDS and conventional profiling data were registered using Fizz software (Biosystem®, France).

2.1.3 Conventional profiling

The attributes were shown on a computer screen and assessors were asked to score intensity on an unstructured scale. Subjects could taste each product several times if necessary. Since the aim of this study was to assess differences between products at panel level, data were summed up into a product by attribute matrix where each entry is the average over the panellists.

2.1.4 Temporal dominance of sensations

- **general procedure**

The TDS computerized system shows the panellist the entire list of attributes on a PC screen. The subject is then asked to click on the start line scale as soon as the product is into his or her mouth, to consider which of the attributes is perceived as dominant, to score its intensity. Thereafter, the subject starts again by choosing which attribute is dominant and so on, until the perception ends. During each tasting, the computer records for each event (click), the time elapsed since the start button was clicked, the name of the attribute chosen and the intensity scored. Each attribute could be selected several times in a given test. Conversely, an attribute may not be selected at all. During the training session, a dominant attribute was defined as the

attribute one is just thinking about at a given time, or the most important attribute to characterize the product at a given time and not necessarily the one with the highest intensity. For example, in a beer, the strongest sensation can be the bitterness all over the tasting. Nevertheless, one can perceive other slight sensations, such as caramel, that are less intense but correspond to the best attribute to describe the product at a given time.

- ***TDS data analysis based on three parameters: Time, Duration and Score***

Each time a panellist clicks on an attribute scale, the computer records the intensity score (Score parameter) and the corresponding time (Time parameter). Moreover, a Duration parameter was defined as the time elapsed until the next event (the following click on a scale). In this way, we consider that an attribute is dominant in a product as long as no other attribute is chosen. This definition of the dominance will be also used for the construction of the TDS curves in the next section.

If an attribute is chosen several times in the same run by a panellist, there are several values for Time, Duration and Score parameters for this attribute. In this case, the Time parameter is set to the first time of citation, the durations are summed up and the scores are averaged with corresponding duration as weights in order to give more importance to the "persistent" sensations, as illustrated in figure 1.

These new values are then averaged for each of the three parameters over the panel in order to obtain a single value by product and attribute. If an assessor does not select an attribute for a given product, the corresponding parameter values are considered as missing values. In this case, averaging was only performed with the available values. TDS data were now summed up into three (product by attribute) matrices; each matrix being associated with a parameter (T, D, S). Non normalized Principal Component Analysis (PCA) was performed on each of these matrices.

- ***TDS curves***

TDS curves computation is explained in figure 2. Let us focus on each sensory attribute at a time. For each point of time, the proportion of runs (subject x replication) for which the attribute was assessed as dominant was computed. These proportions, smoothed using the TRANSREG procedure of SAS®, are plotted against time and called TDS curves. Transformation used in the TRANSREG procedure is noniterative smoothing spline used with parameter smooth=30. Thereafter, TDS curves from a given product are displayed on the same graph. It can be noticed that intensity scores given by subjects during the session are not taken into account in this representation. These curves rely solely on the selection of an attribute as dominant or not.

In order to get more insight into the TDS graphical display, two other curves are drawn (figure 3). The first one, called “Chance level”, is the dominance rate that an attribute can obtain by chance. Its value, P_0 , is equal to $1/P$, P being the number of attributes.

The second one, called “Significance”, is the limit of significance, for this proportion value, calculated using the confidence interval of a binomial proportion based on a normal approximation (1).

$$P_t = P_0 + 1.645 \sqrt{\frac{P_0(1 - P_0)}{n_t}} \quad (1)$$

P_t : lowest significant proportion value ($\alpha=0.05$) at point in time t for a TDS curve

n_t : number of subject*replication having a dominant attribute at time t

It is worth mentioning that for each replication, one subject has a dominant attribute from the first time of his or her citation of an attribute to the end of the perception (when he clicks on a stop button). As a matter of fact, n_t value is lower at the beginning of the tasting because all the subjects have not scored an attribute yet. It is also the case at the end because perception ends earlier for some of the panellists. Therefore P_t value is higher at the beginning and at the end. Furthermore, when a confidence interval of a proportion, based on a normal approximation, is calculated, it is recommended that the number of observations multiplied by the proportion value under chance is not lower than five. In this paper, $P_0=0.10$ so minimum number of

observations should be 50. It is one of the reasons why each of the 16 panellists performed four replications of each product.

- **TDS differences curves**

The TDS graph displays the P attribute curves on the same plot. In order to compare two products, one can display the superimposition of the curves of the differences of the dominance rates. These differences are plotted only when significantly different from zero, highlighting the differences between the products over time. Limit of significance for each curve of differences over time is obtained using the usual test to compare two binomial proportions (2):

$$P_{t\text{diff}} = 1.96 \sqrt{\frac{P_{\text{moyt}}(1-P_{\text{moyt}})}{n_1 + n_2}} \quad \text{where } P_{\text{moyt}} = \frac{P_{1t} + P_{2t}}{2} \quad (2)$$

With:

$P_{t\text{diff}}$: least significant difference ($\alpha=0.05$) for proportion difference at time t

P_{1t} : proportion for product 1 at time t

P_{2t} : proportion for product 2 at time t

n_1 : number of subject by replication having already cited a first attribute at time t and having not yet stopped to rate for product 1

n_2 : the same as n_1 for product 2

2.1.5 Comparison between TDS and conventional profiling

Since conventional profiling is not a temporal method, this comparison is made on the basis of the TDS Score parameter only. RV and normalized RV coefficients (Schlich, 1996), were computed between both product by attribute matrices where each value is the mean over the panel and the replications, as defined in the parametric approach part (see section 2.1.4). Non normalized PCA were also performed on the same matrices.

Moreover, a PCA of each of the two additional TDS parameter (Time and Duration) was performed to observe what additional information was provided by these parameters.

2.2. Results and discussion

2.2.1 TDS curves

Figure 4 shows the TDS graphs for two products of the first study among twelve. TDS curves are on the left side whereas TDS curves of the difference to the means of every product are on the right side. For example, the curves of the product KGKI show that the first dominant attribute is pasty from 0 to 18 s (selected by 29% of the subjects at 12 s) followed by melting from 18 to 40 s (selected by 28% of the subjects at 23 s). Beneath these two curves, the sensation of saltiness seems to be less dominant but lasts longer (between 15 s and 45 s). It is easy to follow the dominant attribute over time on each graph, and to characterize the temporality of the product. Moreover, the chance_level and the significance curves improve the interpretation of this graph. They show that only firm, pasty, melting and salty are significantly dominant at least once. Lactic (at about 25 s) and astringent (at about 50 s) are slightly dominant but not significantly. The other attributes were never considered as dominant in this product.

On the difference curves, it can be seen that the AF product is sweeter and more melting than the mean of the products at the beginning. Then, it becomes more diacetyl and finally, slightly more astringent than the other products. AF is also less firm and elastic at the beginning. It is important to note that when the curve of a given attribute is under zero, this reflects the fact that this attribute was less frequently selected as dominant in this product than in the other products. Therefore, the sensation corresponding to the attribute may still exist (and be experienced) in the product, but it is not as dominant as in the other products.

2.2.2 Comparison between TDS and conventional profiling parameters

Figure 5 shows the conventional profiling PCA (figure 5a) and the Score parameter PCA (figure 5b). In both cases, the first axis is positively correlated with firmness, pasty and elastic attributes and negatively correlated with fluid sensations. The second axis is also correlated with the same attributes for both methods: salty and sweet. Product locations on the maps are nearly the

same, in particular for those close to the borders: EF, AF and KA which have high scores for attributes fluid, sweet and firm respectively. Considering the close similarity between these two graphical displays, we can conclude that the information given by the Score parameter during a TDS session is nearly the same as that obtained by conventional profiling. RV and normalized RV coefficients, which measure the similarity between two configurations, were computed between the Score parameter matrix and the conventional profiling matrix and confirm the visual conclusion ($RV=0.93$, $NRV=7.12$). Thus, although the task is more difficult in TDS, the intensity scores recorded in TDS are in agreement with those from conventional profiling.

Moreover, PCA performed on the Time and Duration parameters give further information on the temporality of the products that is quite different, as shown by the normalized RV coefficient between Score and Duration ($NRV_{S-D}=1.85$) and between Score and Time ($NRV_{S-T}=1.62$). On the time parameter PCA (figure 5c), all the attributes (except Fluid) are positively correlated with the first axis, which can thus be interpreted as the global temporality of the products. Products on the left (EF, AF, EFKG, KG) have small Time values because sensory sensations appear at the very beginning for them whereas new sensations can be detected later in the products on the right side of the map (AFKI or AFKA). The second axis reveals the opposition between fluid and melting attributes, showing that two opposite texture attributes cannot be perceived at the same time in a product. On the duration parameter map (figure 5d), other information is given for certain products. For example, AFKG can be characterized by the duration of its lactic taste and its elasticity. EF and AF have a longer duration of the fluid and astringent sensations than the other products. These two maps are useful in comparing products in a temporal way and in characterizing dynamics of the sensations in the mouth. Moreover, they show differences or similarities between products that cannot be seen on the Score map. For example, scores given to EF and AF are very different as these products appear far removed from each other on the Score PCA map, but they are close from a temporal point of view because the sensations produced by these products have similar durations for each attribute in mouth (EF and AF are close on the Duration PCA map). On the contrary, whereas AFKI and AFKA are close on the

Score map, Duration map reveals that salty, astringent and fluid sensations last longer for AFKI than for AFKA.

3 Second experiment: TDS versus TI

3.1 Materials and methods

3.1.1 Panellist selection and training

In order to compare TDS and TI acquisition methods, two panels (called TDS group and TI group) were trained simultaneously on one of each methodology. Selection tests were the same as in the first experiment. At the end of the tests, two groups of 16 panellists were selected and trained. In the first twelve training sessions, panellists learned to recognize taste and texture and drew up a list of ten attributes to describe dairy products. As in the first experiment, this list had to be the same for both groups in the study. Three other sessions were needed to establish a consensual definition for each attribute. Thereafter, six sessions were devoted to familiarizing the panellists with the use of the computer system for TI and TDS groups.

3.1.2 Experiment

The protocol was the same as for the first experiment. Two groups tasted five dairy products (FC, FFFK, FFSY, FFV and FF61). The attributes selected by the panel were: acid, diacetyl, sticky, melting, fruity, smoky, fatty, nutty, pasty, salty and sweet. The total duration of one TDS evaluation was 120 s. One group used the TI methodology with three replicates for each product and each attribute, whereas the other group used the TDS methodology with four replicates for each product.

3.1.3 Data acquisition (TI and TDS)

For each evaluation of an attribute with TI methodology, when the panellist put the product into his or her mouth, he or she clicked on the left extremity of a horizontal unstructured scale (corresponding to no sensation). Then, the panellist was asked to move the cursor along the scale as the sensation evolved in mouth until the end of the perception. Data were recorded at a

frequency of 2 Hz. For each sample and each attribute, average TI curves were drawn by averaging the data at each point of time across the 16 subjects and the three replicates. No specific averaging method was used. TDS methodology is presented in section 2.1.4. The TI and TDS data were recorded using Fizz software (Biosystem®, France).

3.1.4 Comparison between TDS and TI

We are interested in the comparison of the temporal curves given by both methods. After a visual comparison, several parameters were considered for both methods to summarize these curves:

- VMAX: maximum value on Y axis (which corresponds to a mean intensity score in TI and to a dominance rate in TDS)
- TMAX: time to reach VMAX from the beginning of the tasting
- DMAX: duration period when the value on Y axis is higher than $0.9 \times \text{VMAX}$.

Pearson correlation coefficients were calculated for each product and each parameter over the attributes. These coefficients evaluate whether a parameter evolves in the same way with both methods.

3.2 TI versus TDS curves

Figure 6 shows the temporal curves obtained with the TDS method on the left and with the TI method on the right. The curves of the 10 attributes are plotted on the same graph for each sample and each method. In general, the more important attributes are the same for both methods.

Pearson correlation coefficients (corrP) were then calculated from the parameters of these curves between the methods (Tab.1). On average over the products, the correlation coefficients are significant ($p < 0.05$) for VMAX so the highest and lowest curves are globally and significantly the same for both methods. This means that maximums of intensity in TI are comparable to maximum dominance rates in TDS although units are different. The more intense an attribute is perceived with TI, the more dominant with TDS. For example, for the product FC

($\text{corrP}_{\text{VMAX}}=0.84$), pasty and diacetyl are most often chosen as dominant with TDS and also have the highest intensity scores with TI. On the contrary, nutty and smoky are the lowest curves for both methods. Middle curves fruity and sticky are close. It is the same for fatty and salty. For product FFFK, correlation between TI intensities and TDS dominances is also easy to read on the graphs. Nevertheless, high intensity in TI is not always correlated with high dominance in TDS. For product FFV, the perceived intensity of saltiness is high in TI whereas it is not considered as a dominant attribute by the TDS group. For the product FFSY, $\text{corrP}_{\text{VMAX}}$ is not significant. It could be the consequence of the proximity of all the VMAX values for both methods for this product.

The $\text{CorrP}_{\text{TMAX}}$ value is less significant on average but a tendency is still present. $\text{CorrP}_{\text{TMAX}}$ for FFV is significant at $p=0.05$, at $p=0.15$ for FC and FFFK and it is not significant for the two other products (FFSY and FF61). Thus, ranking of the attributes according to TMAX values is not always alike for TI and TDS. For example, on TDS graph for product FFSY, ranking of the most dominant attributes is melting, sticky, fatty, salty, acid and smoky, whereas on the TI graph, melting does not reach TMAX first and smoky is not the last attribute but smoky. Moreover, it is less obvious to visually determine the sequence of the attributes with TI than with TDS. One reason why $\text{CorrP}_{\text{TMAX}}$ are less significant than $\text{CorrP}_{\text{VMAX}}$ might involve data acquisition. While TI method lets the panellist describe one attribute over time, the TDS method asks him or her to make a choice among all the attributes. So the temporality of the perceptions is exhausted with TDS because the subject is obliged to separate his or her sensations over time. In this way, while TMAX values are close in TI (because individual sensory sensations seem to evolve in the same manner), TDS focuses on the dominant attribute and is able to separate and highlight the most important attributes.

$\text{CorrP}_{\text{DMAX}}$ values are never significant between the methods. This could be the result of the choice made in the construction of the TDS curves. Indeed, an attribute is considered dominant until another attribute is chosen, so the curve of an attribute rises the more often it is chosen as dominant. On the other hand, the decreasing phase is only due to the increasing designation of other attributes. It does not depend on the concerned attribute itself. Duration notions are

therefore different between the methods so it is not surprising that there is no correlation between the methods for this parameter.

Conclusion

This paper presents a new method called TDS for Temporal Dominance of Sensations, which makes it possible to obtain simultaneous temporal information for up to ten attributes. Above this limit, panellists said that it is difficult to keep in mind all the attributes simultaneously during an evaluation. Moreover, time to familiarizing with TDS procedure was not as long as expected. Three sessions would have been sufficient for this task, instead of six. Thanks to a few replications of each product (four at least) and a usual number of panellists in sensory analysis (twelve at least), it is possible to draw the curves of the dominance of each attribute over time on the same graph for each product. These curves provide a characterization of the temporality of the sensations in food products. The differences between individual product curves and the average curve of the products highlight the specificities of each product.

In the first case study, we showed that TDS procedure provides useful information for products characterization which, in particular, recover the information obtained by means of the conventional sensory profiling scores. Therefore, although the TDS task is difficult, panellists are able to give marks that have the same significance as those given in conventional profiling. Additionally, Time and Duration parameters provide useful information on the temporality of the sensations.

Comparison between TDS curves and TI curves recorded for the same products and the same set of attributes shows that general aspects of the curves are similar for some of the products but really different for others. VMAX values are significantly correlated between methods for the whole set of products, but TMAX correlation is only a tendency and DURATION values are not correlated. These differences may stem from variability inherent in panels or in training sessions. Nevertheless, the lack of correlation may also be due to the differences between methods themselves. An assessor is not focusing at exactly the same thing when he has to follow the intensity of only one attribute for about a hundred seconds and when he has to

constantly select the dominant attribute of a product at various points of time. With the second method, the panellist is obliged to make a choice to select the dominant attribute and to separate sensations over time. In fact, he has to characterize the main temporal aspects of the products, simultaneously integrating all the sensations he or she perceives. Indeed, TDS is a true multi-attribute temporal method, whereas TI only focuses on one attribute at a time.

It is worth mentioning that the results presented in this paper are always analyzed at the panel level and not at individual level because we are interested in product characteristics. At the panellist level, performance comparison is difficult because each panellist does not give the same number of data in the same run. Some generally click about 5 or 6 times in an evaluation, whereas others generally click 15 to 20 times in an evaluation. Nevertheless, these differences are not a problem for the comparison of the products at the group level, because each subject retains his or her way of using the TDS methodology over the evaluations.

Based on these first studies, our objectives are now to improve this method in order to get more homogeneous data across panellists. A recent study on manipulated Italian red wines highlights the differences between TI and TDS, showing that TDS is better adapted to multi-attribute temporal data acquisition. At the same time, we want to formulate other statistical treatments of the TDS data in order to take into account the scores given by the subjects in the elaboration of the TDS curves, which is not the case for the moment.

Acknowledgement

We thank Mostafa Qannari for his general involvement and advice in this paper.

References

- Bonnans, S. R., & Noble, A. C. (1995). Interaction of salivary flow with temporal perception of sweetness, sourness and fruitiness. *Physiology & Behavior*, 57(3), 569-574.
- Chaya, C., Perez-Hugalde, C., Judez, L., Wee, C. S., & Guinard, J.-X. (2004). Use of the STATIS method to analyze time-intensity profiling data. *Food Quality and Preference*, 15(1), 3-12.
- Clark, C. C., & Lawless, H. T. (1994). Limiting response alternatives in time-intensity scaling : an examination of the halo-dumping effect. *Chemical Senses*, 19(6), 583-594.
- Duizer, L. M., Bloom, K., & Findlay, C. J. (1996). Dual-attribute time-intensity measurement of sweetness and peppermint perception of chewing-gum. *Journal of Food Science*, 61(3), 636-638.
- Duizer, L. M., Gullett, E. A., & Findlay, C. J. (1993). Time-intensity methodology for beef tenderness perception. *Journal of Food Science*, 58(5), 943-947.
- Ishikawa, T., & Noble, A. C. (1995). Temporal perception of astringency and sweetness in red wine. *Food Quality and Preference*, 6(1), 27-34.
- King, B. M., & Duineveld, C. A. A. (1999). Changes in bitterness as beer ages naturally. *Food Quality and Preference*, 10(4/5), 315-324.
- Lee III, W. E., & Pangborn, R. M. (1986). Time-intensity: The temporal aspects of sensory perception. *Food Technology*, 40(11), 71-78 & 82.
- Pionnier, E., Nicklaus, S., Chabanet, C., Mioche, L., Taylor, A. J., Le Quéré, J. L., & Salles, C. (2004). Flavor perception of a model cheese: relationships with oral and physico-chemical parameters. *Food Quality and Preference*, 15(7-8), 843-852.
- Schlich, P. (1996). Defining and validating assessor compromises about product distances and attribute correlations. In T. Naes, & E. Risvik (Eds), *Multivariate Analysis of Data in Sensory Science*, (pp. 259-306). Amsterdam: Elsevier Science.
- Wendin, K., Janestad, H., & Hall, G. (2003). Modelling and analysis of dynamic sensory data. *Food Quality and Preference*, 14(8), 663-671.
- Zimoch, J., & Findlay, C. J. (1998). Effective discrimination of meat tenderness using Dual Attribute Time Intensity. *Journal of Food Science*, 63(6), 940-944.

Figures and table

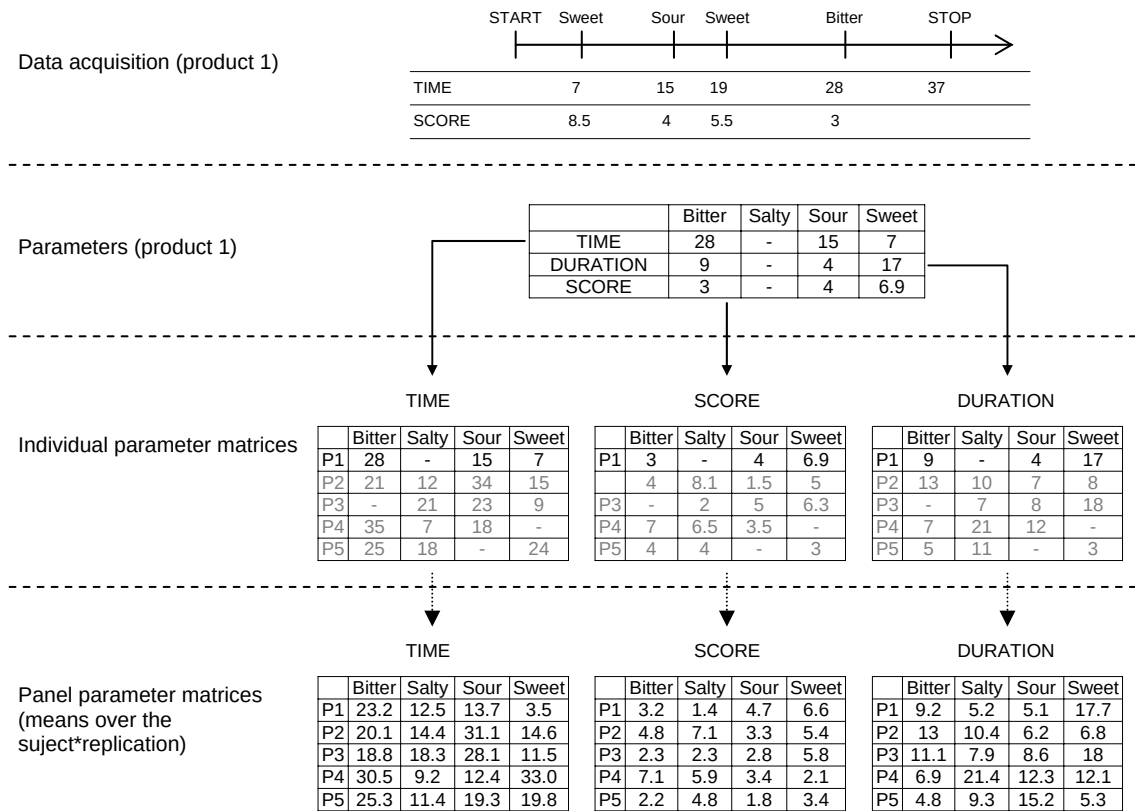


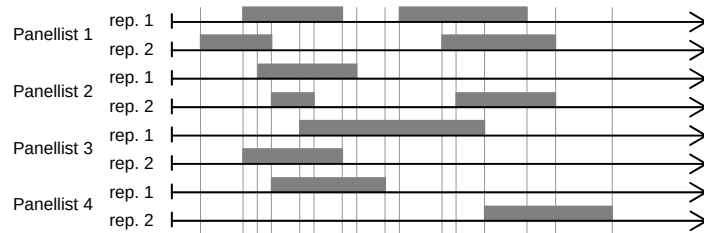
Figure 1: TDS parameters computation

Data for one evaluation by one panellist

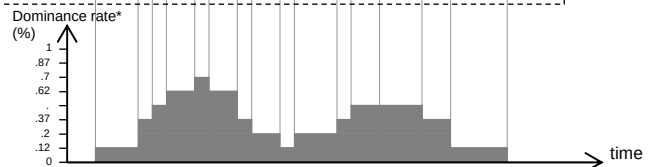


■ : Periods where sweet sensation is dominant

Periods where sweet sensation is dominant for one product at panel level (4 panellists x 2 replications in this example)



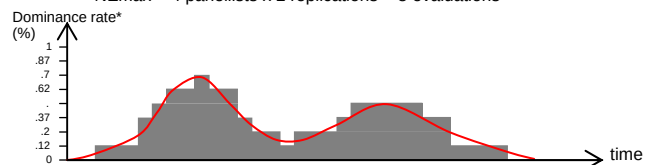
number of evaluations, NE, where the attribute Sweet is dominant in the course of the evaluation



Computation of the dominance rates for the panel

* Dominance rate = NE / NEmax
NEmax = 4 panellists x 2 replications = 8 evaluations

TDS curve: smoothing of the dominance rates for Sweet (TRANSREG, SAS ®)



Superimposition of the different attributes

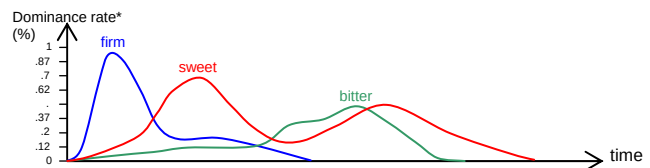


Figure 2 : methodology to compute TDS curves

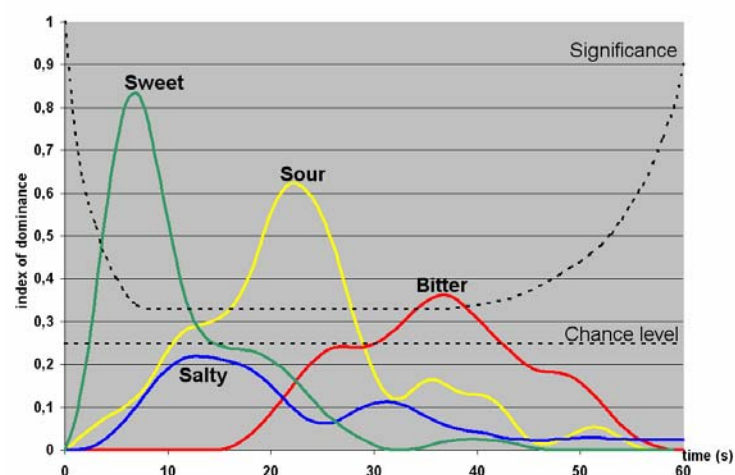


Figure 3: Example for significance and chance level curves on TDS graphical representation for four attributes evaluated for one product by a sensory panel.

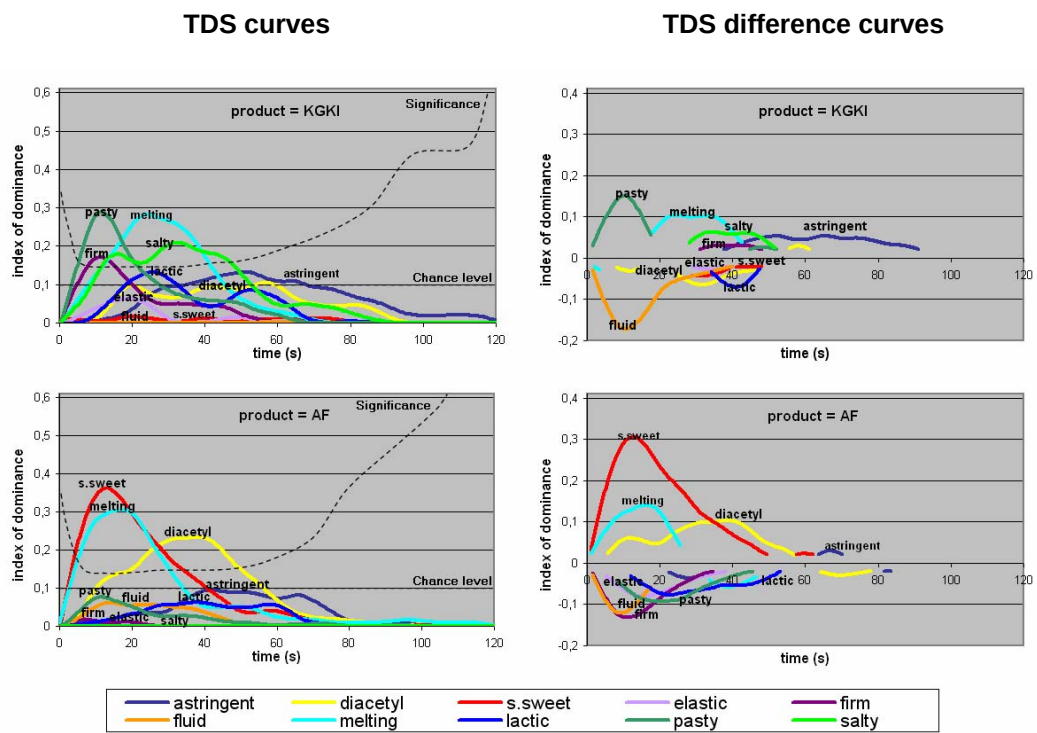


Figure 4: TDS curves and TDS difference curves for 2 dairy products.

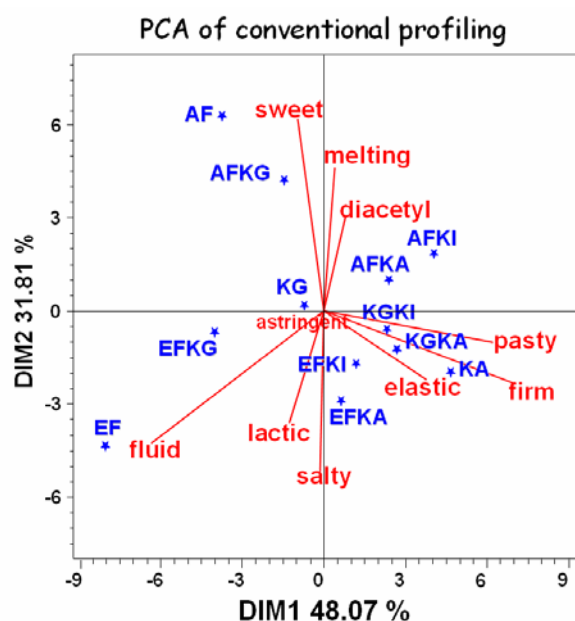


Figure 5a: Non normalized PCA of the conventional profiling data

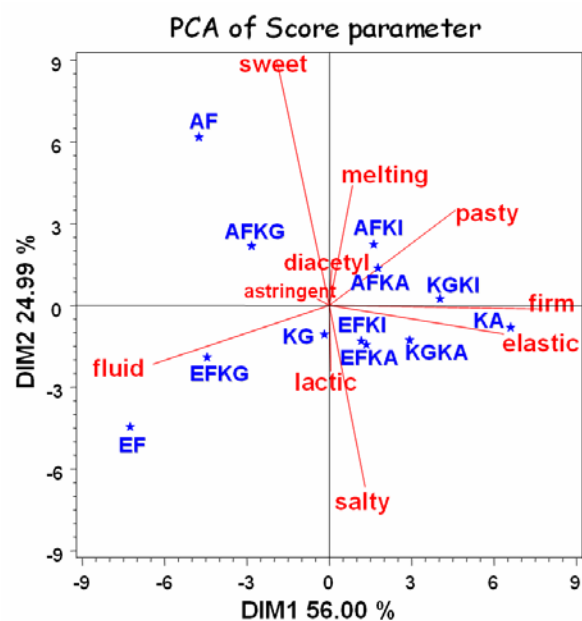


Figure 5b: Non normalized PCA of the TDS Score matrix

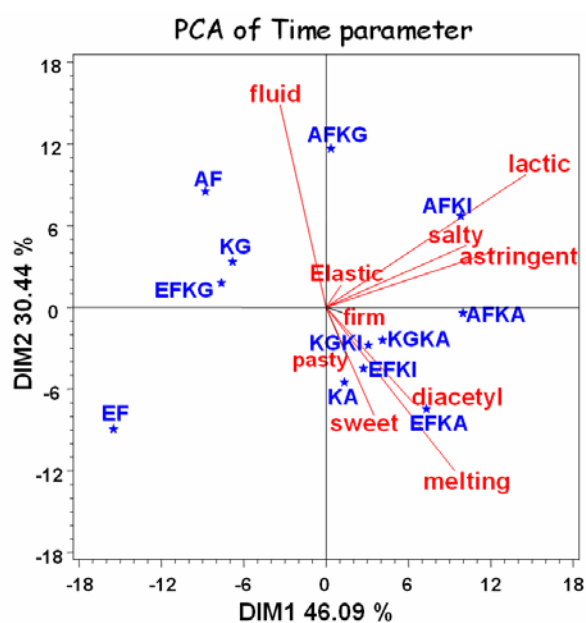


Figure 5c: Non normalized PCA of the TDS Time matrix

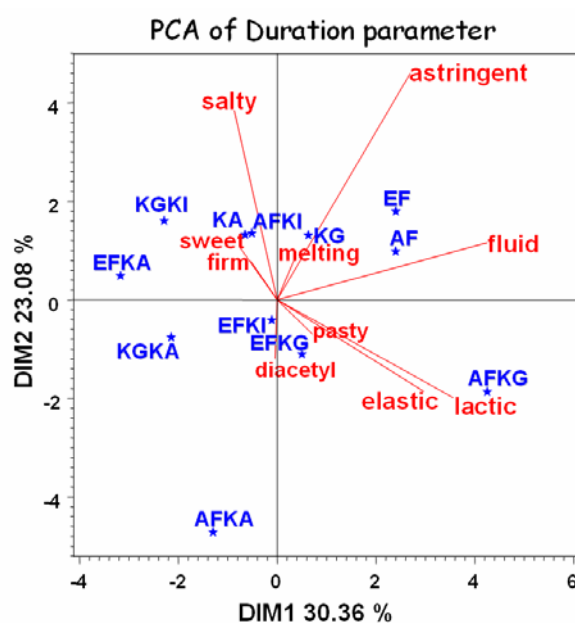


Figure 5d: Non normalized PCA of the TDS Duration matrix

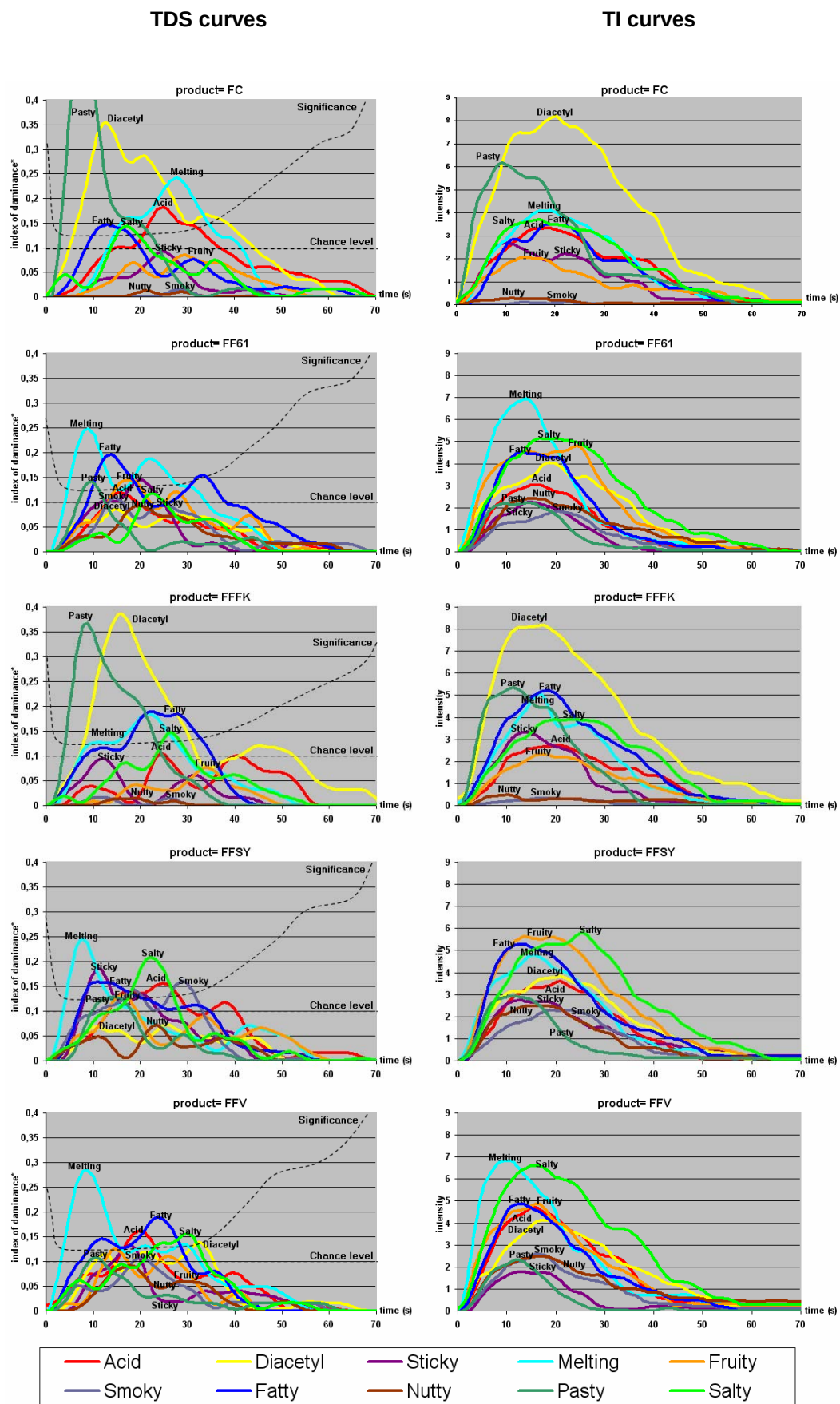


Figure 6: comparison between TI and TDS curves on 5 dairy products

Table 1: Pearson correlation coefficients between TI and TDS methodology

Product	Pearson correlation coefficients		
	Duration	lmax	Tmax
FC	0,34	0,84**	0,51(*)
FFFK	-0,39	0,91***	0,61(*)
FFV	0,39	0,72*	0,66*
FFSY	-0,16	0,35	0,48
FF61	0,45	0,57(*)	0,32
Mean	0,13	0,68**	0,47(*)

*** : p<0.001 ** : p<0.01 * : p<0.05 (*) : p<0.15