



HAL
open science

**Contrôle des corrélats temporels et spectraux de la
quantité vocalique:de l'arabe syrien de l'Euphrate au
français de Savoie.**

Omran Allatif

► **To cite this version:**

Omran Allatif. Contrôle des corrélats temporels et spectraux de la quantité vocalique:de l'arabe syrien de l'Euphrate au français de Savoie.. Linguistique. Université Stendhal - Grenoble III, 2008. Français. NNT: . tel-00338010

HAL Id: tel-00338010

<https://theses.hal.science/tel-00338010>

Submitted on 10 Nov 2008

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.

UNIVERSITÉ GRENOBLE III - STENDHAL
U.F.R. DES SCIENCES DU LANGAGE

N° attribué par la bibliothèque
.....

THÈSE
Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ STENDHAL

Discipline: « Sciences du Langage »

Préparée à l'ex-Institut de la Communication Parlée (GIPSA-LAB)
dans le cadre de l'Ecole Doctorale:
« Langues, littératures et sciences humaines »

Présentée et soutenue publiquement

par

Omran ALLATIF

Le 3 octobre 2008

Titre:

**Contrôle des corrélats temporels et spectraux de la quantité vocalique:
de l'arabe syrien de l'Euphrate au français de Savoie.**

Sous la direction de:
Christian ABRY

JURY

M. Christian ABRY
M. Jean-François BONNOT
Mme Marie-Agnès CATHIARD
M. Mohamed EMBARKI

Professeur Directeur
Professeur Rapporteur (Président)
MCF Examineur
MCF HDR Rapporteur

Remerciements

En tout premier lieu, je voudrais remercier d'une manière toute particulière mon directeur de recherche, Monsieur Christian Abry, pour ses précieux conseils, son appui et sa patience dans les bons moments comme dans les périodes plus éprouvantes qui ont caractérisé cette longue entreprise. J'aimerais qu'il sache que j'aurai toujours beaucoup d'estime pour son travail et une grande admiration pour ses innombrables qualités de chercheur.

Je voudrais remercier Monsieur Jean-François Bonnot d'avoir accepté de faire partie du jury de ma thèse, avec toutes ses compétences en linguistique et phonétique qui me feront sans doute améliorer significativement la qualité des prolongements de ce travail de recherche.

Je remercie également Monsieur Mohamed Embarki d'avoir accepté de faire partie du jury de cette thèse, mais aussi pour la contribution scientifique de ses travaux dans le domaine de la quantité vocalique en arabe qui ont alimenté le débat autour de ce sujet.

Mes remerciements vont naturellement à Marie-Agnès Cathiard pour toute l'aide qu'elle a apportée à l'étude du français de la région de la Savoie, pour la recherche des locuteurs de la première génération et pour leur enregistrement, mais aussi pour la relecture d'une partie de cette thèse en sa qualité de psychologue. Je voudrais aussi la remercier pour la contribution humaine et morale qui était une grande source de courage pendant de nombreuses années.

Je dois témoigner de ma profonde gratitude à Jean-Luc Schwartz pour le temps qu'il m'a généreusement octroyé pour améliorer la qualité de ce travail de recherche et pour ses précieux conseils sur les seuils de la qualité vocalique. Je ne cache pas mon admiration pour ses recherches et pour ses contributions scientifiques dans le domaine de la parole qui m'ont inspiré certaines idées.

Un grand merci à Louis-Jean Boë d'avoir accepté de répondre à toutes mes questions durant ce travail, pour l'orientation vers les références qui intéressaient ce travail de recherche, pour les recherches qu'il a faites dans la base de données UPSID pour mon travail, et pour la méthodologie scientifique de recherche que j'ai acquise du fait d'être en contact avec lui. Merci Louis-Jean pour les programmes informatiques que tu as écrits sur le logiciel Matlab et merci pour l'initiation à l'utilisation de ce logiciel.

Je profite de ces lignes pour témoigner de ma gratitude à ceux qui, pendant de nombreuses années, ont partagé mon quotidien et je pense ici aux membres de l'ex-ICP, depuis 2007 Département Parole et Cognition de Gipsa-Lab, enseignants, chercheurs, techniciens, administratifs et étudiants, qui m'ont offert, chacun à sa façon, un soutien indéfectible. Je remercie ainsi Alain Arnal, pour l'initiation théorique et pratique à l'utilisation des outils d'enregistrement et pour la formation à la numérisation des bandes sonores. Je voudrais remercier aussi Anne Vilain pour le temps qu'elle m'a accordé pour répondre à mes questions durant ces années de thèse et pour sa relecture de certaines parties de ce travail de recherche. Je remercie également Hélène Loevenbruck pour l'initiation à l'utilisation du logiciel Praat. Mes remerciements vont aussi à Solange Rossato pour le temps qu'elle a passé pour ce travail et pour l'initiation au traitement statistique des données. Merci Nathalie Vallée pour nos discussions constructives en phonétique et autour de ce travail de recherche, nos échanges scientifiques m'ont permis d'avoir les idées plus claires. Je remercie également Elisabetta Carpitelli pour l'aide à l'établissement des références bibliographiques, mais aussi pour son attention et son humanité. Je tiens à remercier aussi Coriandre Vilain et Christophe Savariaux

pour toute l'aide technique et pour leurs interventions sur les rudes problèmes informatiques. Merci Isabelle Rousset pour ton aide. Je n'oublie pas les thésards de Gipsa-Lab DPC. Merci Nicolas Audibert d'avoir accepté de partager une boîte aux lettres avec moi sans rechigner, je pense aussi à Claire, Virginie, Aude, Fanny, Anne, Émilie, Hien, dont la présence et la solidarité étaient un soutien moral considérable.

Je dois témoigner de ma reconnaissance aux Syriens de Mayadin qui ont consacré un temps considérable pour l'enregistrement du corpus et pour l'identification des stimuli de l'expérience qui a consisté en la manipulation de la durée des voyelles.

Merci aussi à mes chers locuteurs du français de la région de la Savoie, Maurice, Denise, Germain, Sylvie, Charlène et Alane, qui ont bien voulu soit se déplacer soit m'accueillir chez eux pour les enregistrements.

A ma famille, sans qui je n'aurais pas pu faire ce long voyage, ni matériellement ni moralement.

Enfin à tous mes amis qui m'ont bien soutenu dans les moments difficiles.

Titre

Contrôle des corrélats temporels et spectraux de la quantité vocalique: de l'arabe syrien de l'Euphrate au français de Savoie.

Résumé

La quantité vocalique a deux corrélats principaux, temporel et spectral, seul le corrélat contrôlé jouant par principe un rôle phonologique pertinent. Nous avons étudié le contrôle de ces corrélats pour deux systèmes: l'arabe syrien de l'Euphrate et le français régional de la Combe de Savoie. Le corrélat temporel se traduit par la présence de deux catégories métriques: brève *vs.* longue. Tandis que le corrélat spectral se manifeste par la centralisation des brèves par rapport aux longues.

Afin de déterminer le corrélat principal de contrôle, nous avons pu expérimenter pour l'arabe de Mayadin sur deux mises en contexte ou perturbations naturelles cruciales. L'une, l'accélération du débit de parole réduit le plus directement la durée absolue des voyelles. L'autre, le focus interrogatif peut changer notablement leur structure spectrale (formants). Deux procédés donc aux antipodes en ce qui concerne leurs effets majeurs: réducteurs *vs.* magnificateurs (par rehaussement soit *enhancement*). Enfin, par manipulation digitale de la durée nous avons, pour une expérience perceptive, donné aux voyelles brèves les durées des longues et aux longues celles des brèves. Pour évaluer les résultats de ces expérimentations nous avons repris les débats de la littérature sur les seuils perceptifs des durées segmentales et des fréquences formantiques.

Nos résultats pour l'arabe syrien montrent que la différence temporelle entre brèves et longues est efficacement contrôlée. Quant à la centralisation, décrite jusqu'ici comme une simple réduction vocalique, elle s'est avérée un processus phonologique et non pas un processus articuloire qui dépend de la durée, soit un sous-produit (*by-product*) de la bréveté. Il ne s'agit pas en l'occurrence d'un geste qui manquerait sa cible. Les brèves sont ainsi en passe de devenir des cibles vocaliques à part entière, ce processus étant achevé pour le /i/ bref, qui reste [e], quel que soit le type de perturbation, même sous un contexte rehaussant comme en intonation interrogative. Le système de l'arabe de Mayadin ne se réduit cependant pas au triangle central des brèves (*dispersion suffisante*) qui auraient pu être dédoublées par la durée. Ce système garde des voyelles extrêmes pour les longues, avec des exemplaires [i] [a] [u], selon le principe universel de la *dispersion maximale*.

Le français régional de Savoie a conservé, mieux que le français standard, et comme beaucoup de dialectes arabes, des voyelles longues et brèves corrélées à des différences temporelles et aussi spectrales. Notre étude manipulant le débit de parole sur trois générations prouve que, chez les septuagénaires, le corrélat temporel est contrôlé pour les paires sur /A/ et sur /O/, le corrélat spectral l'étant par F1 et F2 sur /O/ et seulement par F2 sur /A/ (le contraste sur /E/ n'est pas robuste). Donc, comme en arabe, il ne s'agit pas non plus d'une simple réduction vocalique. Les résultats montrent aussi que chez la génération appartenant à la tranche d'âge 20-25 ans, seul le corrélat spectral reste contrôlé, pour la paire sur /O/, tout contraste étant perdu sur /E/ et /A/. À l'évidence, le changement intergénérationnel même en français de Savoie n'est pas conservateur des contrastes comme peut l'être l'arabe de la Syrie de l'Euphrate.

Du point de vue des universaux vocaliques aucun de ces changements ne porte atteinte au principe de la dispersion maximale. En conclusion, nous nous en assurons en reconsidérant le cas des systèmes vocaliques australiens, avocats d'une *dispersion suffisante*, puisque réputés défectifs sur les voyelles hautes [i] et [u]. Si la manipulation d'une intonation rehaussante ne fait pas réapparaître un [i] bref chez nos locuteurs de Mayadin, pas plus qu'un

débit plus lent ne restaurerait les contrastes qui disparaissent en français de Savoie, elle permet de retrouver un [i] chez des locutrices aborigènes. Ce qui nous a amené à considérer que la réversibilité d'un changement pouvait être sollicitée pour ne pas violer un principe universel comme la dispersion extrême des voyelles.

Mots clés

Contrôle quantité vocalique, seuils de perception (durée et formants), arabe de l'Euphrate (Syrie), français de Savoie (France), langues d'Australie (kayardild, dalabon et mayali), systèmes et réversibilité des changements vocaliques.

Title

Vocalic quantity and the control of its temporal and spectral correlates. From Syrian Arabic on the Euphrates to French in Savoy.

Abstract

Vocalic quantity corresponds to two main correlates: in the temporal and in the spectral domains. In principle, only controlled correlates play a relevant role in linguistic communication. We studied the control for these correlates in two different linguistic systems: dialectal Arabic of Mayadin on Syrian Euphrates, and regional French of the Combe de Savoie. While temporal correlates are mainly manifested by the presence of two metrical categories, short *vs.* long vowels, spectral correlates are reflected by the centrality of short with respect to long vowels.

In order to determine the principal controlled correlate, we tested for dialectal Arabic the effect of two contexts, or fundamental natural perturbations. The first one, speech rate acceleration, reduces most directly the absolute vowel duration. The second one, interrogative focus, is able to affect notably their spectral structure (formants). We used therefore two processes that are divergent concerning their major effects: reduction *vs.* enhancement. Finally, we assigned to short vowels, by digital manipulation of duration, for a perceptive experiment, the duration of long ones, and vice-versa. We revisited the literature on perceptive thresholds concerning segmental duration and formant frequency in order to evaluate our measurements.

Our results concerning Syrian Arabic show that the temporal difference between shorts and longs is efficiently controlled. As for centralization, described as a simple vocalic reduction, it turned out to be a phonological process and not an articulatory process that depends on duration: centralization is not a by-product of brevity. Hence, it is not a matter of an articulatory gesture undershooting its target. Shorts are becoming full-fledged vocalic targets. This process is now completed for short [i], which remains basically [e], whatever the type of perturbation is, even under the enhancing influence of interrogative intonation focus. However, this dialectal Arabic vocalic system is not reduced to the triangle of short vowel qualities (*sufficient dispersion theory*), which might be contrasted with the long ones by means of duration. According to the universal principle of the *maximum dispersion theory*, this system maintains extreme vowels for long ones, with true [i] [a] [u] exemplars.

Regional French from Savoy has maintained, better than standard French, and as well as many Arabic dialects, short and long vowel contrasts, which are correlated to temporal and

spectral distances. Our study, manipulating speech rate over three generations, demonstrated that speakers in their seventies do control temporal correlates for pairs in the /A/ and in the /O/ regions, the spectral correlates being controlled by F1 and F2 for the /O/ pair, and only by F2 for the /A/ pair (the contrast for the /E/ pair is not robust). So, like in Arabic, centralization of short vowels does not arise from a simple vocalic reduction, since spectral distances are also controlled. Results also show that for the generation in the age range of 20-25 years, only the spectral correlate remains controlled for the /O/ pair, all contrasts being lost for the /E/ pair and for the /A/ pair. Obviously, intergenerational changes in Savoy French are not as conservative of contrasts as for Arabic of Syrian Euphrates.

In conclusion, from a universal point of view, these vowel changes do not undermine the prediction of the *maximum dispersion* principle. To reinforce this statement we had to consider the special case of Aboriginal vowel systems of Northern Australia, which plead in favor of the *sufficient dispersion theory*, since they are classically presented as defective in high vowels, i.e. with no [i] and [u] exemplars. Enhancement of intonation does not induce short [e] to come back to [i] in Arabic speakers, not more than decelerating speech rate would be able to re-establish lost quantity contrasts in Savoy French. But intonational enhancement allows a recovering of [i] by Kayardild aboriginal women speakers. This finding leads us to believe that the reversibility of a change may be used in case of the violation (which does not occur in Arabic or French) of a universal principle such as the extremal vowel dispersion.

Keywords

Vocalic quantity control, perception thresholds (duration and formants), Euphrates Arabic (Syria), French of Savoy (France), Australian languages (Kayardild, Dalabon et Mayali), systems and reversibility of vowel changes.

Intitulé et adresse du laboratoire:

Gipsa-Lab, Département Parole et Cognition. UMR 5216 CNRS/INPG/UJF/Université Stendhal
(ex Institut de la Communication Parlée)
Université Stendhal, 1180 Av centrale,
BP 25-38040 Grenoble cedex 9
France
tél: +33 4 76 51 76 45
fax: +33 4 76 82 43 35

Introduction	7
I. Les seuils	15
I.1. Seuil de la qualité vocalique	16
I.1.1. Son pur	17
I.1.1.1. Rôle de la fréquence	17
I.1.1.2. Rôle de la durée	18
I.1.1.3. Rôle de l'intensité	24
I.1.2. Formant	30
I.1.3. Voyelle	44
I.1.3.1. Rôle des formants	45
I.1.3.2. Seuil relatif ou seuil absolu?	48
I.1.3.2.1. La non-uniformité de l'espace vocalique	48
I.1.3.2.2. Le seuil relatif	53
I.1.3.2.3. Le seuil absolu	56
I.1.3.3. Un seuil à valeur universelle?	64
I.1.4. Etude comparative entre plusieurs systèmes	68
I.1.4.1. Choix de la méthode de mesure	68
I.1.4.2. Comparaisons entre des distances dans plusieurs systèmes	72
I.1.4.2.1. L'anglais	73
I.1.4.2.2. Le français	74
I.1.4.2.3. L'arabe	75
I.1.4.3. Conclusion	77
I.2. Seuil différentiel de la perception de la durée vocalique.	78
I.2.1. Facteurs affectant le seuil de la durée	79
I.2.1.1. Facteurs extra-stimulus	79
I.2.1.1.1. le facteur âge	79
I.2.1.1.2. Sexe	84
I.2.1.2. Facteurs intra-stimulus	87
I.2.1.2.1. Fréquence, Intensité et durée initiale	87
I.2.2. Seuil pour les segments vocaliques.	93
I.2.2.1. Fréquence variable	93
I.2.2.2. Fréquence stable	97
I.2.3. Étude linguistique du seuil	100
I.2.4. Conclusion	108
II. La quantité vocalique	110
II.1. Aspects des langues à quantité vocalique	110
II.2. Différence temporelle et quantité vocalique	113
II.2.1. La durée est un trait secondaire	113
II.2.2. Spécification linguistique de la durée	117
II.3. Quantité vocalique en arabe	125
II.4. Espace vocalique de l'arabe	131
III. Expérience sur l'arabe dialectal de Mayadin (Syrie)	139
III.1. Introduction en repère proximal: quantité et qualité dans le dialecte arabe d'Amman (Jordanie).	139
III.2. Notre dialecte de Mayadin	157
III.3. Partie expérimentale	161
III.3.1. Caractérisation de la durée	161
III.3.2. Caractérisation de la qualité	162
III.3.3. Méthode	162
III.3.3.1. Stimuli	162

III.3.3.2. Sujets	165
III.3.3.3. Procédure et enregistrement	166
III.3.3.4. Dépouillement des données	167
III.4. Résultats	168
III.4.1. Durée	170
III.4.2. Les formants	173
III.4.2.1. Présentation numérique	173
III.4.2.2. Présentation graphique	175
III.5. Significativité des écarts	180
III.5.1. Significativité statistique	183
III.5.1.1. L'écart entre les durées	183
III.5.1.1.1. Schéma déclaratif	183
III.5.1.1.2. Schéma interrogatif	184
III.5.1.2. L'écart entre les valeurs des F1	188
III.5.1.2.1. Schéma déclaratif	188
III.5.1.2.2. Schéma interrogatif	191
III.5.1.3. L'écart entre les valeurs des F2	194
III.5.1.3.1. Schéma déclaratif	194
III.5.1.3.2. Schéma interrogatif	195
III.5.1.4. Résumé général des études statistiques	199
III.5.2. Significativité perceptive	201
III.5.2.1. Les écarts entre les durées (quantité)	201
III.5.2.2. Les écarts entre les formants (qualité)	203
III.6. Conclusions	209
IV. Le poids perceptif de la durée de la voyelle face à sa qualité acoustique dans la discrimination des voyelles longues et brèves	216
IV.1. Introduction	216
IV.2. manipulation instrumentale	227
IV.2.1. Introduction	227
IV.2.2. Méthode	228
IV.2.2.1. Stimuli	228
IV.2.2.2. Contrôle	229
IV.2.2.3. Sujets	229
IV.2.2.4. Procédure	230
IV.2.3. Résultats	231
IV.2.3.1. Signaux A (non-manipulés)	231
IV.2.3.2. Signaux B (manipulés)	232
IV.2.4. Discussion	233
IV.2.5. Conclusion	247
V. Durée et qualité vocaliques en français de Savoie. Étude acoustique intergénérationnelle	251
V.1. Introduction	251
V.2. Méthode	255
V.2.1. Stimuli	255
V.2.2. Sujets	256
V.2.3. Procédure et enregistrement	256
V.2.4. Dépouillement des données	258
V.3. Résultats	258
V.3.1. Durée	258
V.3.1.1. Première génération	258
V.3.1.2. Deuxième génération	271
V.3.1.3. Troisième génération	275
V.3.2. Qualité	279

V.3.2.1. Première génération	279
V.3.2.2. Deuxième génération	288
V.3.2.3. Troisième génération	291
V.3.3. Écarts perceptifs	294
V.3.3.1. Écart temporel	294
V.3.3.2. Écart spectral	297
V.4. Conclusions	302
VI. Conclusion	305
VI.1. Contrôle de la quantité vocalique	305
VI.1.1. Arabe dialectal	305
VI.1.2. Français de la région de la Savoie	314
VI.1.2.1. Première génération	314
VI.1.2.2. Deuxième génération	318
VI.1.2.3. Troisième génération	320
VI.1.3. Comparaison relative au comportement de la quantité vocalique en arabe dialectal de Mayadin et en français de la région de la Savoie	323
VI.2. Plasticité des systèmes vocaliques	329
VI.2.1. Langues des aborigènes d'Australie	329
VI.2.2. Vers un changement réversible vs. irréversible	336
Bibliographie	340
Annexes	359
Annexe I: exemple de segmentation des signaux.	360
Annexe II: Algorithme du calcul du deuxième formant effectif avec le logiciel Matlab.	361
Annexe III: déplacements des voyelles dans l'espace vocalique en fonction du pourcentage affecté à leurs formants.	363
Annexe IV: bande critique.	366
Annexe V: comparaison de deux méthodes de conversion des Hz vers l'échelle des Bark.	367
Annexe VI: résumé des résultats des études qui ont porté sur le seuil différentiel de la perception de la durée.	370
Annexe VII: Test des capacités des locuteurs du corpus de l'arabe à faire des minimales fondées sur la durée de la voyelle.	372
Annexe VIII: illustration graphique de la réduction/magnification de l'espace vocalique de l'arabe dialectal de Mayadin selon la condition expérimentale.	374
Annexe IX: manipulation non-formelle des durées des voyelles de l'arabe de Mayadin.	375
Annexe X: la durée de la syllabe en fonction de la durée de la voyelle en arabe de Mayadin (Syrie) pour les trois locuteurs.	381
Annexe XI: la durée de la consonne post-vocalique en fonction de la durée de la voyelle en arabe de Mayadin pour les trois locuteurs.	382
Annexe XII: la durée de la consonne pré-vocalique en fonction de la durée de la voyelle en arabe de Mayadin pour les trois locuteurs.	383
Annexe XIII: la fréquence fondamentale en fonction de la catégorie de la voyelle Brève/Longue en arabe de Mayadin pour les trois locuteurs	384
Annexe XIV: Intensité de la voyelle en fonction de sa catégorie Brève/Longue en arabe de Mayadin pour les 3 locuteurs.	385
Annexe XV: intensité pour chaque voyelles de l'étude de l'arabe de Mayadin, et effet du débit.	386
Annexe XVI: résultats concernant la partie qui a porté sur les mesures de la durée des voyelles en français de la Savoie pour chaque locuteur et chaque débit de parole à part.	389
Annexe XVII: voies de la transmission de l'information sur les identités des voyelles.	395
Annexe XVIII: Quelques exemples du rôle de la quantité vocalique dans la morphologie de l'arabe.	398

Introduction

Quand un train est lancé à toute allure sur une voie ferrée, il est possible que les feuilles d'un arbre tout proche de la voie ferrée s'agitent au passage du train, à cause de la turbulence aérodynamique engendrée par ce passage. Cette relation de cause à effet peut être *systématique* dans des conditions idéales. L'objectif du train n'est pas d'agiter les feuilles mais de se déplacer sur sa voie, parce que si un jour les feuilles ne se mettent pas en mouvement à son passage (alourdies par la pluie ou soumises à un vent très fort qui les maintient dans une position fixe), le train n'aura pas raté son objectif (sa cible). Les mouvements des feuilles sont *contingents* pour la cible, ils sont un sous-produit, car non contrôlé, du déplacement du train. Le mouvement de ce dernier est contrôlé, le train pouvant s'arrêter comme il peut se mouvoir. Un mouvement peut donc avoir des conséquences *systématiques* hors contrôle qui sont alors *contingentes*.

La parole est un ensemble de gestes ordonnés régis par une organisation temporelle (*timing*) d'une grande rigueur, vitesse d'élocution oblige. Le débit de la parole est en moyenne de 6 syllabes par seconde (6 syll./s). Le phénomène de sous-produits (*by-products*) non contrôlée existe en parole. L'épenthèse est un exemple, où nous observons la naissance d'unités phonétiques en raison de certaines contraintes biomécaniques. La naissance de la consonne épenthétique [p] dans le mot "Simpson" s'explique par des contraintes articulatoires, la co-articulation, qui nous oblige à anticiper les gestes de la consonne [s] alors que nous sommes en train de produire la consonne [m]. En fait, l'obligation d'aller vite (6 syll./s), fait que les gestes se chevauchent dans le temps, autrement dit une partie du [s] est anticipée, coarticulée, coproduite, dans le [m]. L'objectif de cet exemple est de montrer qu'il existe des sous-produits phonétiques qui se réalisent de manière non intentionnelle.

Nous avons deux objectifs dans cette thèse.

Notre premier objectif porte sur le contrôle de la quantité vocalique.

Nous cherchons à déterminer les corrélats acoustiques de la quantité vocalique en arabe dialectal de la ville de Mayadin dans la vallée de l'Euphrate en Syrie pour les trois paires de voyelles suivantes: /i/ vs. /i:/, /a/ vs. /a:/ et /u/ vs. /u:/, et nous les comparons avec le comportement de la quantité vocalique du français du comté de Savoie pour les quatre paires de voyelles suivantes: /a/ vs. /ɑ:/, /ɛ/ vs. /ɛ:/, /ɔ/ vs. /o:/ et /u/ vs. /u:/.

Parmi les corrélats trouvés, nous déterminerons ceux qui sont contrôlés et ceux qui ne sont pas.

L'observation de la production de la parole dans des conditions normales (idéales) pourrait occulter des réalités linguistiques en attestant de la présence d'un geste qui s'actualise occasionnellement. En effet, certaines conséquences des gestes contrôlés, parce qu'elles sont *systématiques*, alors qu'elles sont *contingentes*, peuvent être décrites comme nécessaires à la communication et sont ainsi linguistiquement spécifiées. La soumission de la production de la parole à des conditions naturelles perturbatrices est un procédé qui permet de révéler les gestes obligatoires, parce que seuls les gestes à valeurs phonologiques, et par conséquent obligatoires, seront communs à des productions s'effectuant dans des conditions idéales et à des productions s'effectuant sous des conditions perturbatrices.

En opposant des voyelles longues à des voyelles brèves, les langues associent à la différence temporelle une différence qualitative conséquence d'un geste articulaire moins ample; chacune des voyelles brèves observe alors un écart temporel et une distance spectrale vis-à-vis de sa partenaire longue. Nos procédures, les perturbations rythmiques et prosodiques et les mesures du seuil différentiel de la perception de la durée et du seuil de la qualité vocalique (perceptible *vs.* catégoriel), ont de l'intérêt pour les systèmes vocaliques qui associent plusieurs facteurs pour la création d'un contraste distinctif et dont on cherche à déterminer les éléments contrôlés parmi les corrélats qui sont présents dans le contraste. Nous soulevons ici 2 questions relatives au contrôle des corrélats acoustiques de la quantité vocalique:

(1) la différence temporelle est-elle contrôlée? Autrement dit, en cherchant à créer des distances spectrales entre longues et brèves, l'appareil phonatoire produit-t-il de manière fortuite cette différence de durée? Il nous semble légitime de mettre en question cette différence temporelle, pourtant largement attestée par toutes les études qui ont décrit le système de voyelles de l'arabe standard et dialectal (Al-Ani, 1970; Ghazeli, 1979; Belkaid, 1984; Abou-Haidar, 1994; Newman, 2002; Barkat-Defradas, 2003; Allatif & Abry, 2004), parce que ces études-là n'ont pas testé la robustesse de l'écart temporel entre la brève et la longue en soumettant cette opposition à des conditions perturbatrices. Nous verrons qu'une objection de taille par rapport à cette différence de durée était soulevée dans des études élaborées par Embarki (2002, 2003, 2004) sur un dialecte de l'arabe marocain. L'étude de Jong & Zawaydeh (2002) sur l'arabe dialectal d'Amman (Jordanie) a perturbé cette opposition Brève/Longue, mais c'était uniquement pour les voyelles basses [a] et uniquement par la

prosodie (l'accent et le focus). Elle montre que l'écart temporel devient plus important (donc robuste) sous l'accent, elle montre aussi que le focus n'a pas d'effet sur cet écart, donc il se maintient, et il est par conséquent robuste. Cette tentative timide n'est pas suffisante pour prouver la présence de l'opposition temporelle sous la perturbation, parce qu'elle n'a pas eu recours à des perturbations réductrices, comme Minagawa *et al.* (2003) qui ont étudié les variations des durées des voyelles en japonais en fonction, entre autres conditions, du débit d'élocution, lent, rapide et très rapide (ce qui, nous le verrons, a été aussi fait pour l'arabe).

Nous nous posons la question sur la présence de cette différence temporelle d'autant plus que des études récentes avec une approche développementale (Embarki, 2002, 2003, 2004), ont montré qu'il n'y a pas de différence temporelle entre voyelles longues et voyelles brèves en âge préscolaire en arabe dialectal marocain de la ville de Ksar El Kébir, et que l'opposition temporelle demeure absente auprès des locuteurs adultes lorsqu'ils n'actualisent pas des formes phonétiques de l'arabe standard dans leur discours.

L'opposition au niveau qualitatif entre brèves et longues agit de telle sorte que les voyelles brèves se trouvent déplacées vers le centre de l'espace vocalique.

(2) ce déplacement (la réduction vocalique) est-t-il une sous-production du geste moins ample disposant de moins de temps pour se réaliser ou est-ce que c'est la production d'un geste rendu délibérément moins ample et qui prévoit l'emplacement des brèves dans le centre de l'espace? Autrement dit, est-ce que cette réduction vocalique est produite accidentellement ou est-ce qu'elle est produite de manière contrôlée? En ce lieu aussi nous sommes justifiés pour notre interrogation, puisque des perturbations naturelles "extrêmes", où l'exigence communicative n'était pas requise, ont montré que ces distances spectrales entre brèves et longues peuvent disparaître, en effet, Newman *et al.* (2002) ont décrit l'espace vocalique de l'arabe en termes de ses voyelles longues et brèves. Les auteurs ont procédé à l'analyse d'une récitation coranique (voix chantée), ils affirment que les distances spectrales entre brèves et longues ne sont pas significatives sous cette condition perturbatrice. de Jong & Zawaydeh (1999, 2002) ont trouvé aussi que les variations prosodiques, où, contrairement à Newman *et al.* (2002), l'exigence communicative était requise, peuvent avoir un effet significatif sur les qualités acoustiques des voyelles en arabe. Les auteurs ont examiné les F1 des voyelles basses [a] de l'arabe dialectal d'Amman (Jordanie), ils trouvent que lorsque la syllabe est frappée par l'accent, le F1 des voyelles longue et brève, devient significativement plus élevé, se déplaçant ainsi vers la périphérie de l'espace vocalique.

Ces deux 2 questions admettent des réponses partiellement incompatibles, puisque nous

ne pouvons pas nous attendre, pour chaque paire de voyelle (Brève vs. Longue) prise à part, à ce que les différences, temporelle et spectrale, soient toutes les 2 non contrôlées. Les réponses possibles sont 3: (i) soit la différence temporelle est contrôlée mais pas la différence spectrale, (ii) soit la différence spectrale est contrôlée mais pas la différence temporelle, (iii) soit les 2 différences sont contrôlées.

Pour répondre à ces questions, nous avons besoin de perturber les oppositions temporelles et qualitatives distançant perceptivement ces deux catégories de voyelles Brève/Longue.

Hypothèses

Si le geste qui place les voyelles brèves dans le centre de l'espace vocalique est contrôlé, nous devons alors nous attendre à ce que les oppositions qualitatives persistent sous des conditions expérimentales perturbatrices de la parole naturelle, puisque cette différence spectrale est un trait acoustique qui dispose d'une valeur phonologique, et par conséquent elle n'est pas autorisée à disparaître. Si ce déplacement est un sous-produit non contrôlé, il est légitime alors de s'attendre à ce que ces distances séparant les brèves des longues disparaissent partiellement ou totalement sous la perturbation. Le raisonnement est similaire en ce qui concerne les différences temporelles: si ces différences sont vraiment contrôlées, elles doivent se montrer robustes, c'est-à-dire qu'elles doivent se maintenir sous la perturbation pour assurer la survie de l'information linguistique.

Procédés

Perturbations naturelles:

Les systèmes vocaliques qui seront étudiés pour ce premier objectif sont celui de l'arabe dialectal de la ville de Mayadin dans l'est de la Syrie et celui du français de la région de Savoie.

En ce qui concerne l'arabe dialectal, deux types de perturbations, délibérément antagonistes, seront appliqués à ce système: (1) le débit rapide de la parole, qui, au niveau temporel, est connu pour son effet qui réduit la durée des voyelles et qui, au niveau spectral, est connu pour l'accentuation de la réduction vocalique (déplacement des voyelles vers le centre de l'espace vocalique). (2) l'intonation montante, marque de l'interrogation totale à l'oral: cette intonation frappera les voyelles qui font l'objet de notre étude. Ce schéma mélodique est réputé pour son effet qui s'apparente à l'effet de l'accent qui (cf. de Jong & Zawaydeh, 1999, 2002) magnifie la durée des voyelles et déplace les voyelles vers la

périphérie de l'espace vocalique. Le schéma mélodique interrogatif agit comme l'accent, il a cependant un effet plus ample que l'effet de l'accent. Ainsi, au niveau temporel, il pousse les voyelles à avoir des durées plus importantes par rapport à un schéma mélodique déclaratif, et au niveau spectral, il dilate l'espace vocalique (l'opposé de la réduction vocalique). Le croisement des deux débits de parole, normal et rapide, avec les deux schémas mélodiques, déclaratif et interrogatif, nous fournit 4 conditions expérimentales: 1) débit normal et schéma déclaratif; 2) débit rapide et schéma déclaratif; 3) débit normal et schéma interrogatif et 4) débit rapide et schéma interrogatif. Nous mesurons sous chaque condition l'écart temporel et la distance spectrale entre chaque voyelle brève et sa correspondante longue.

En ce qui concerne le français de la région de Savoie, un seul type de perturbation sera appliqué, c'est l'élocution rapide de la parole. En fait, notre étude de ce système s'inscrit dans 2 cadres: (i) comparatif, où nous avons pour but de savoir si d'autres systèmes opposant des voyelles longues et des voyelles brèves observent des réactions similaires à ceux que nous observerons avec le système de l'arabe dialectal de Mayadin lorsqu'ils sont soumis à la contrainte de la parole rapide. (ii) descriptif: dans ce cadre nous réalisons une étude acoustique intergénérationnelle (sur 3 générations) afin de confirmer, ou de réfuter, l'hypothèse selon laquelle le trait de la quantité vocalique dans le français de cette région, héritage du franco-provençal qui disposait du procédé de la quantité vocalique (Martinet, 1956 avec une description auditive, confirmée par Abry *et al.*, 1995 avec des mesures instrumentales), a réellement existé et a réellement disparu chez de la jeune génération.

Tests statistiques:

Pour les systèmes vocaliques étudiés, et toujours dans le but de retrouver les gestes contrôlés, nous effectuerons deux types de tests: statistique et perceptif (mesures des distances perceptives: temporelle et spectrale). Les tests statistiques ont pour but de montrer, ou non, la régularité de la production d'un geste articulatoire. Il faut que la production du geste soit *systématique* au préalable pour faire l'objet d'une étude, c'est ainsi, par exemple, que dans notre étude du français de la région de Savoie nous ne nous sommes pas intéressés à la production des voyelles de la troisième génération sous le débit rapide de la parole, parce qu'il s'est avéré, grâce aux tests statistiques, que les locuteurs de cette génération ne produisaient pas de contraste temporel entre voyelles dans des conditions normales. Lorsque la production se montre, par contre, régulière, elle sera examinée sous chacune des conditions expérimentales perturbatrices. De tels examens permettent d'éliminer la part du hasard dans la présence d'un geste, en établissant une relation de cause à effet significative entre la

production ou la sous-production acoustique et le geste articulatoire. Toute réalisation qui se montre aléatoire sous la perturbation sera alors considérée comme une production accidentelle, soit une sous-production.

Distances perceptives:

Nous consacrons une partie de notre travail à l'examen d'études qui ont porté sur deux types de seuils perceptifs: 1) le seuil différentiel de la perception de la durée vocalique; 2) le seuil absolu de "perception vs. catégoriel" de la qualité vocalique. Le premier seuil nous permet de savoir quelle doit être l'étendue d'une différence temporelle entre deux voyelles pour que cette différence puisse être exploitée linguistiquement, le deuxième seuil nous permet de savoir quelle doit être l'étendue d'une distance spectrale entre deux entités vocaliques pour que cette distance puisse être perçue (perception) ou qu'elle puisse fonder un contraste qualitatif entre les deux entités en question (catégoriel).

En fait, les mesures perceptives constituent une dimension secondaire qui étaye la méthode de perturbations et soutient sa cohérence. Nous nous attendons ainsi à ce que les distances qui ne soient pas robustes, donc non phonologiques, ne soient pas perceptibles non plus. En effet, les seuils nous indiquent en amont (en théorie) de combien doit être la différence pour fonctionner dans le système, les perturbations indiquent en aval (en pratique) que cela fonctionne ou non. Ainsi, nos jugements concernant le contrôle de la production des indices acoustiques seront doublement couverts.

Nous ne sommes pas certains que nous avons appliqué tous les procédés perturbateurs qui ont pour effet de révéler ce qui n'est pas sous contrôle. Une confusion peut naître si la sous-production résiste à la perturbation de la vitesse d'élocution de la parole et à l'effet prosodique du schéma mélodique interrogatif, dans la mesure où nous jugerons phonologique ce qui n'est pas sous contrôle. Les mesures perceptives constituent une démarche parallèle et un deuxième indice sur le contrôle des gestes articulatoires, parce que nous avons toute raison de penser que ce qui n'est pas perceptible, quoique systématique, ne peut pas être transmis via les voies de la communication auditive entre les sujets parlants, et ne peut être, par conséquent, que fortuit.

Perturbation par manipulation:

Nous procéderons aussi à une perturbation non naturelle, instrumentale, dans le but de retrouver les gestes contrôlés. En effet, nous procéderons à l'élimination de la différence temporelle entre la voyelle brève et la voyelle longue. En fait nous ferons plus que neutraliser la dimension temporelle dans le contraste, puisque nous ne nous contenterons pas de donner à la brève et la longue une même durée (une moyenne des deux durées), mais nous donnerons à la voyelle brève la durée de la longue et à la voyelle longue la durée de la brève. Donc, nous suggérerons aux sujets de prendre la brève pour une longue et la longue pour une brève, c'est le geste le plus extrême que nous puissions faire pour sonder l'influence de la différence temporelle. Notre hypothèse est que les termes des paires qui ne comptent que sur la dimension temporelle pour s'opposer devraient alors se confondre, alors que ceux des paires qui basent leurs oppositions sur d'autres corrélats acoustiques de la quantité vocalique devraient préserver leurs oppositions sous la manipulation. Cette manipulation vise en dernier lieu à savoir s'il existe d'autres corrélats acoustiques de la quantité vocalique en dehors des deux corrélats classiques (temporel et spectral) largement mentionnés dans la littérature sur la question. Les résultats de la manipulation nous permettront ainsi d'étayer ou de réfuter les résultats des mesures sous les conditions perturbatrices naturelles, ainsi les oppositions qui survivent à la disparition de la différence temporelle, en se maintenant, sont des oppositions qui disposent d'autres corrélats acoustiques, lesquels sont alors contrôlés puisqu'ils sont à valeurs phonologiques dans le système vocalique de l'arabe dialectal de Mayadin.

Notre deuxième objectif porte sur la plasticité des systèmes vocaliques.

Notre hypothèse est que les langues sont, le long de leur parcours évolutif (déjà non achevé), en quête de nouvelles stratégies qui réconcilient économie et efficacité; elles adoptent ainsi tout mécanisme qui semble mieux répondre à cette exigence. Nous pensons que la rencontre entre la langue et ce nouveau mécanisme naît sous la perturbation. Si cette quête-là n'existait pas nous n'aurions pas assisté à des évolutions comme celle du latin vers diverses langues romanes, où nous assistons à la disparition de certains procédés phonétiques autrefois utilisés. Par exemple, le latin avait un système vocalique faisant usage de la quantité vocalique. Le français, qui est l'un des aboutissements du latin vulgaire ayant transformé ces oppositions quantitatives en qualités, a retrouvé puis perdu ce procédé (ou est en train de le perdre). L'évolution de l'arabe standard vers les différents dialectes arabes est un autre exemple, où nous notons la tendance à la disparition de certaines coordinations articulatoires,

comme le coup de glotte (Hamza) en finale absolue, ou encore l'évolution de certaines diphtongues vers des qualités vocaliques stables. Les perturbations sont, pour nous, des contraintes qui poussent vers ces évolutions, puisqu'elles permettent aux langues d'explorer leurs potentiels organisationnels. Ainsi l'adaptation momentanée à une contrainte indique au système considéré qu'il est capable de recourir à de telles stratégies pour que l'information linguistique survive à la perturbation. Si cette adaptation momentanée se montre conforme au principe efficacité - économie, elle peut être alors adoptée définitivement comme un élément faisant partie intégrante du système en question. Nous résumons: l'évolution se fait par découvertes, les découvertes sont engendrées par des troubles, les bonnes découvertes seront adoptées comme des changements définitifs accomplis, pas de retour à la position d'avant troubles, alors que les mauvaises découvertes resteront ponctuelles et conditionnées par la présence des troubles.

En effet, les perturbations se posent comme des révélateurs non seulement du contrôle des gestes articulatoires en faisant disparaître les traits fortuits, mais aussi des révélateurs des propriétés organisationnelles des systèmes perturbés, les perturbations pouvant faire apparaître des traits distinctifs qui, dans des conditions normales de l'émission de la parole, ne se manifestent pas. Cette dernière réalité phonétique nous l'observons auprès de certaines langues des aborigènes du nord de l'Australie (le Kayardild, le Mayali et le Dalabon), et que nous discuterons à la lumière des résultats apportés par Fletcher & Butcher (2003), et à la lumière de la Théorie Dispersion - Focalisation (TDF) qui prévoit que les langues aient des exemplaires extrêmes [i a u], intercalant d'autres voyelles sur la périphérie [i e a o u], et que la variation des trois extrêmes [i a u] a pour effet, soit de remplir la périphérie par un nouveau prototype; soit de retrouver un extrême disparu.

Nous voudrions finalement démontrer que les systèmes vocaliques connaissent deux types d'adaptations (réorganisations), des adaptations locales réversibles qui s'actualisent sous la perturbation et qui disparaissent avec la disparition de la perturbation, et des adaptations locales irréversibles, qui étaient à un moment de l'évolution du système engendrées temporairement pour répondre à un besoin informationnel, et qui se maintiennent définitivement au-delà de la perturbation.

I. Les seuils

Nous cherchons à déterminer dans ce chapitre, par des recherches bibliographiques:

(1) le seuil au niveau spectral (qualitatif) pour les qualités vocaliques, et éventuellement les deux niveaux de seuil: (i) le seuil pour la perception d'un changement vocalique (Just Noticeable Difference, JND), et (ii) un seuil catégoriel de la qualité vocalique, c'est-à-dire la plus petite distance *robuste* et suffisante pour séparer 2 voyelles dans l'espace vocalique.

(2) le seuil différentiel de la durée des voyelles, c'est-à-dire le plus petit écart temporel *robuste* permettant à 2 voyelles de se différencier par les durées de leurs réalisations respectives.

Selon l'acception psychoacoustique du terme, le concept du seuil différentiel est défini théoriquement par la limite en dessous de laquelle l'oreille humaine ne parvient plus à discerner les différences acoustiques entre deux stimuli. La détermination des seuils est une question qui revêt une grande importance en raison de ses applications multiples, depuis les théories de la perception de la parole, la description des systèmes phonologiques, jusqu'à la synthèse, en passant par les différentes pathologies dont l'altération des seuils est le révélateur.

Ces capacités à la discrimination des variations des caractéristiques acoustiques du son, les seuils, seraient un héritage de nos ancêtres les primates. Ils seraient en changement, comme il est certain qu'ils ont changé au cours de l'évolution de l'espèce. En effet, Sinnott *et al.* (1976; 1987a; 1987b; 1991), qui ont élaboré des procédés expérimentaux dans le but d'établir une comparaison directe entre les capacités psychoacoustiques du singe et celles de l'homme, dans le cadre d'un long projet qui cherchait à caractériser les éventuels changements de l'audition de l'homme au cours de son évolution, affirment que l'homme a adapté l'héritage ancestral des primates, en ce qui concerne les capacités auditives, à ses besoins spécifiques actuels, telles la musique et la parole. Ces capacités-là sont représentées par les seuils différentiels de la perception de la durée, le seuil différentiel de la perception de la fréquence du son pur et celui de la perception fréquentielle des voyelles. Leurs résultats sont curieux, puisqu'ils montrent que, contrairement à l'homme, le seuil chez le singe augmente quand l'intensité augmente, alors que toutes les études menées sur des sujets humains normo-entendants et malentendant s'accordent sur le fait qu'une hausse de l'intensité du stimulus

contribue à une baisse du seuil (Nelson *et al.*, 1983; Wier *et al.*, 1977).

La capacité à la discrimination des voyelles chez le singe est très proche de celle de l'homme. Alors que pour le son pur le seuil chez le singe est 20 fois plus élevé que celui de l'homme, ce seuil chez le singe pour la voyelle, est de 2 à 3 fois plus élevé que celui de l'homme (Sinnott *et al.*, 1987b). Quant au seuil de la durée, il est de 2 à 3 fois celui des hommes. Ces résultats permettent d'avancer que, au cours de l'évolution, l'oreille humaine est devenue plus sensible aux variations acoustiques des caractéristiques phonétiques, au niveau de la durée et au niveau de la qualité. Les résultats des études de Sinnott *et al.* (1976, 1987a, 1987b, 1991) suggèrent que la discrimination des stimuli multidimensionnels complexes de la parole implique sans doute deux types de mécanismes: les premiers mécanismes ont été hérités de nos ancêtres les primates, et les deuxièmes sont devenus exclusivement spécifiques à l'homme.

1.1. Seuil de la qualité vocalique

Nous distinguons deux types de seuils, un seuil où la distance est juste perceptible (un seuil psychoacoustique), et un seuil où la distance est assez importante pour être robuste et supporter ainsi l'opposition qualitative de deux voyelles (distance phonétique), parce que le fait que la distance soit perceptible n'implique pas forcément le fait qu'elle soit suffisante (l'inverse n'est pas vrai).

Nous commençons par la recherche du seuil du premier type (psychoacoustique). Nous dénivelons par le bas. Ainsi nous cherchons, en premier lieu, la plus petite différence qualitative perceptible par l'oreille humaine pour un son pur, parce qu'un son pur a la structure acoustique la plus simple; plus la structure acoustique du son est complexe, et plus la variation de sa qualité doit être importante pour être perceptible, donc le son pur a le seuil, *a priori*, le plus fin. L'équation de Gagne & Zurek (1988) rend bien compte de cette vérité. Nous chercherons ensuite le seuil pour des structures acoustiquement plus complexes, en ce qui nous concerne, nous sommes intéressés par le seuil qualitatif du formant et de la voyelle (plusieurs formants évoluant simultanément).

Selon Moore (1971), plusieurs facteurs pourraient affecter l'amplitude du seuil, dont les plus importants sont la durée d'une manière intrinsèque, l'intensité, l'intervalle entre les stimuli (200 ms sont suffisantes pour espacer deux stimulus consécutifs et effacer l'impression auditive du stimulus précédent), et le mode de présentation. Moore affirme que l'entraînement

des sujets de l'expérience joue un rôle majeur; la performance atteint une limite au bout d'un certain nombre de jours d'entraînement, et elle peut se dégrader au bout d'un certain temps après avoir arrêté l'entraînement. Les facteurs susceptibles d'affecter le seuil peuvent donc être extra-stimulus.

Nous cherchons ici le seuil différentiel de la perception de la fréquence d'un son pur en fonction de plusieurs paramètres (fréquence, intensité et durée) du stimulus considéré. Ces interactions ont fait l'objet de nombreuses études (Moore, 1973/; Hall & Wood, 1984; Freyman & Nelson, 1986; Wier *et al.*, 1977). L'interaction des paramètres est évidemment plus simple à étudier quand nous considérons seulement 2 paramètres à la fois: fréquence et durée à intensité fixe, fréquence et intensité à durée fixe ou intensité et durée à fréquences fixes.

I.1.1. Son pur

I.1.1.1. Rôle de la fréquence

Moore (1973) a établi un rapport entre la fréquence d'un stimulus et la quantité de la variation de cette fréquence pour que cette variation soit perçue (*delta f*). Il a mené une étude exhaustive, laquelle résumait, en quelque sorte, les travaux de ses prédécesseurs. Contrairement aux études précédentes qui ne donnaient des résultats que sur une partie de l'échelle des fréquences, avec parfois une absence de l'information entre deux valeurs très éloignées, Moore a travaillé sur 8 valeurs, allant de 250 à 8000 Hz tout en fixant l'intensité à 60 dB, les deux stimuli étant séparés par un intervalle temporel de 200 ms, intervalle jugé suffisant (Moore, 1971) pour qu'il n'y ait pas d'interaction entre les 2 stimuli. Ses résultats ont montré que plus la fréquence était élevée et plus le seuil l'était aussi. Cependant, Moore trouve que la variation du seuil en fonction de la fréquence ne suit pas une fonction simple. Il ne parvient pas à établir une loi rendant compte de ce rapport: voir la figure 1, où nous avons choisi (le but étant de montrer la tendance et non pas donner des valeurs précises) de représenter le seuil de la fréquence en fonction de la fréquence pour trois durées différentes parmi les 7 durées étudiées par Moore (1973). En effet, nous pouvons résumer cette représentation graphique ainsi: quelle que soit la durée du stimulus, une hausse de fréquence entraîne une hausse de *delta f*.

I.1.1.2. Rôle de la durée

Nombreuses sont les études qui ont montré l'existence d'une relation entre *delta f* et la durée du stimulus (Koning 1957; Turnbull, 1944; Liang & Chistovich, 1961; Oettinger, 1959; Cardozo, 1962; Sekey, 1963; Kietz, 1963; pour une revue de questions, cf. Moore, 1973). Les résultats de ces études sont divergents à bien d'égards. Un point reste, cependant, commun à leurs résultats, c'est que pour une même fréquence, une baisse dans la durée entraîne une hausse en ce qui concerne *delta f*, autrement dit, le seuil est inversement proportionnel (jusqu'à une certaine limite), à la durée du stimulus. Nous avons opté de représenter le seuil en fonction de la durée sur la figure 2, où nous pouvons voir les 7 valeurs de la durée, de 6,25; 12,5; 25; 50; 100 et de 200 ms, combinées avec 8 valeurs de fréquences. Cette figure permet d'observer que plus la durée est importante et plus le seuil est faible et que cette tendance n'est pas toujours vraie, puisque nous remarquons qu'au-delà d'une certaine grandeur, la durée a tendance à perdre son effet. Effectivement, au-delà de 100 ms, la courbe qui représente *delta f* en fonction de la durée devient quasiment, mais pas tout à fait, plate. L'effet de la durée est d'autant plus important que la fréquence est importante, alors que dans les basses fréquences, l'effet de la durée est quasi inexistant.

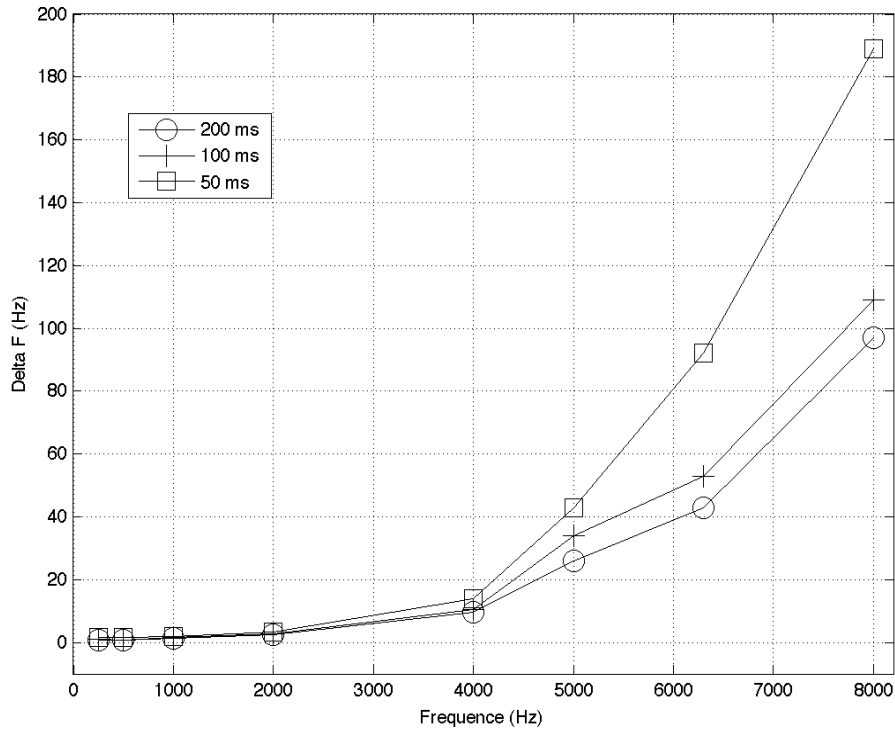


Figure 1. *Seuil de perception de la fréquence du son pur en fonction de la fréquence du stimulus. Test effectué pour trois durées, 50, 10, et 200 ms à 60 dB. D'après l'étude de Moore (1973).*

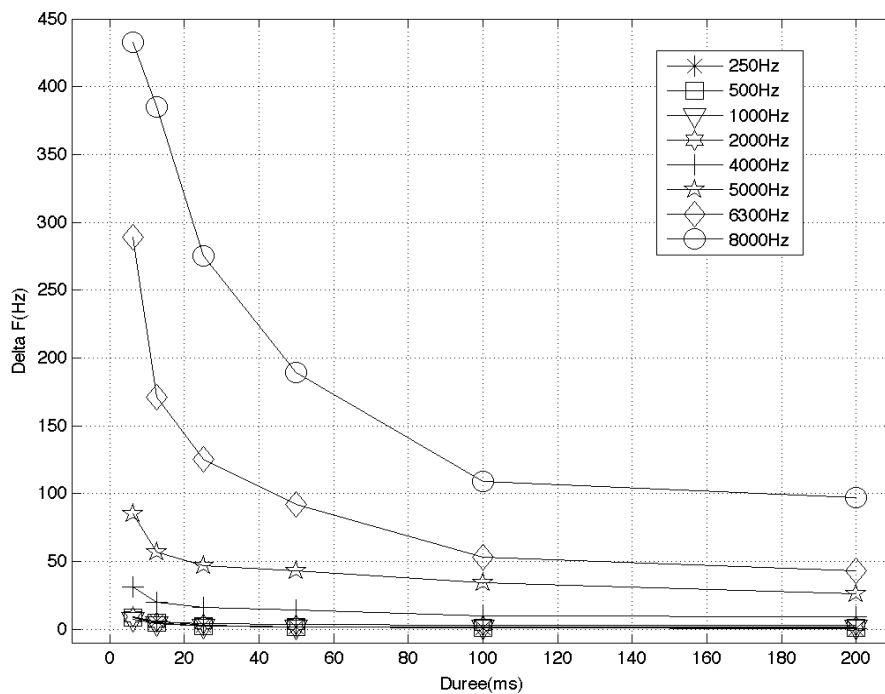


Figure 2. *Seuil de perception de la fréquence du son pur en fonction de la durée du stimulus. Test effectué pour 8 fréquences à 60 dB. D'après l'étude de Moore (1973).*

Nous retenons de cette étude que, globalement, la hausse de la durée baisse Δf alors que la hausse de la fréquence augmente Δf .

Cependant les travaux de Moore présentent un manque. Il n'a pas cherché à savoir ce qui se passerait pour des durées supérieures à 200 ms, ce qui nous donne l'impression que Δf baisse indéfiniment avec la hausse de la durée, alors que nous croyons à l'apparition d'une asymptote horizontale pour Δf en fonction de f .

Dans l'étude d'Hall & Wood (1984, figure 3), une étude dédiée plus à l'effet de la durée (5 valeurs) qu'à l'effet de la fréquence (2 valeurs), les stimuli étaient des sons purs de 2 fréquences, 500 et 1000 Hz. Les auteurs ont cherché à déterminer le seuil différentiel de la fréquence en fonction de la durée du stimulus, pour des malentendants cochléaires et pour des normo-entendants. Cette étude avait pour but d'apprécier la taille de la perte neurosensorielle chez les malentendants cochléaires en appréciant la taille de la perte du seuil différentiel perceptif entre les deux populations, normo-entendants et malentendants. Nous nous sommes intéressés aux 10 sujets normo-entendants de cette étude, dont la moyenne d'âge était de 38,8 ans. Les durées étudiées étaient de 5, 10, 20, 50 et 200 ms, l'intensité pour tous était de 90 dB¹. Deux stimuli consécutifs étaient séparés de 600 ms.

Une analyse de la variance avait montré que les effets de la fréquence et de la durée étaient significatifs; en effet, plus la fréquence était élevée et plus le seuil était élevé, le seuil augmente avec la baisse de la durée du stimulus.

Cette étude a montré aussi que la durée avait un effet drastique dans la résolution des variations de la fréquence pour des durées inférieures à 20 ms, en ce sens que la moindre hausse dans la durée se traduisait par une baisse très importante en ce qui concerne le seuil. Cet effet diminuait plus la durée était importante, vers 200 ms, la durée perdait son effet. Nous observons aussi que l'effet de la fréquence (rappel: + la f est élevée et + Δf est élevé) se maintenait pour des durées jusqu'à 200 ms. Cette étude, comme celle de Moore (1973), nous donne des informations sur le seuil pour des durées très petites, allant jusqu'à 5 ms. Ses résultats vont dans le même sens que les résultats obtenus par Moore et ils présentent le même défaut, s'arrêtant à 200 ms. Effectivement, nous ne savons pas si, pour une même fréquence, une hausse de la durée au-delà de 200 ms a ou non un effet sur le seuil en fonction de la durée.

¹ 90 dB SPL (Sound Pressure Level) ce niveau élevé était choisi pour que l'intensité des stimuli soit bien au-delà du niveau du seuil de perception de tous les sujets, normo-entendants comme malentendants.

C'est dans une étude postérieure à celle de Hall & Wood (1984), que nous trouvons des valeurs supérieures à 200 ms. En effet, Freyman & Nelson (1986, figure 4), ont effectué des mesures du seuil avec des sons purs dont la fréquence était de 1000 Hz. les durées des stimuli étaient 5, 10, 15, 25, 40, 60, 170 et 300 ms. Ils avaient pour objectif la comparaison entre les performances auditives des normo-entendants et celles des malentendants en ce qui concerne les seuils différentiels de la fréquence en fonction de la durée du stimulus.

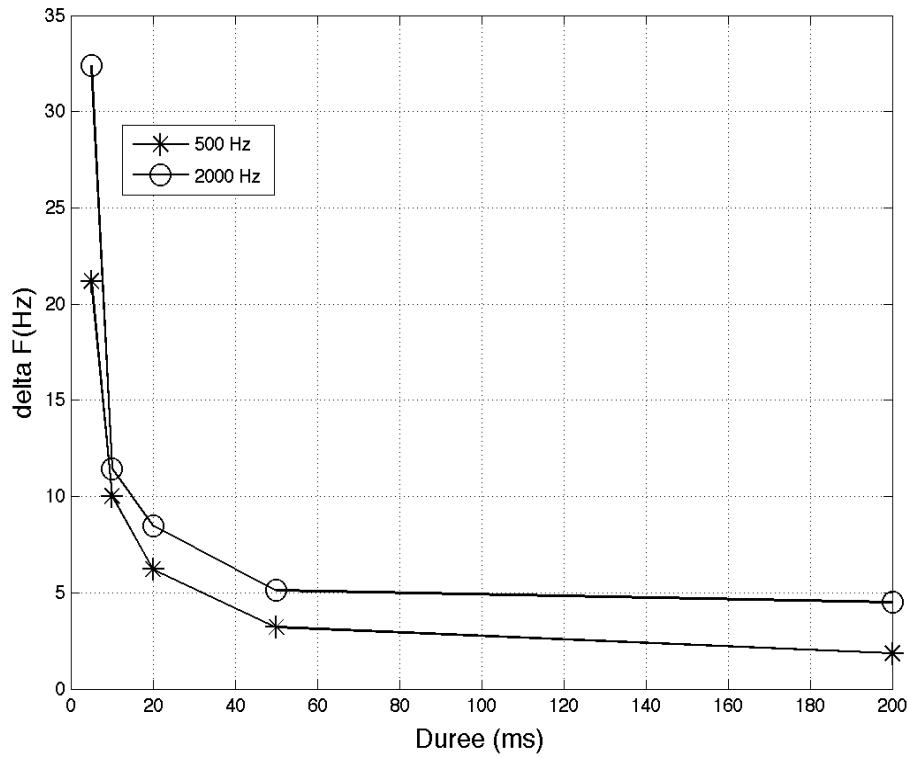


Figure 3. Seuil de perception de la fréquence du son pur en fonction de la durée du stimulus. D'après de Hall & Wood (1984).

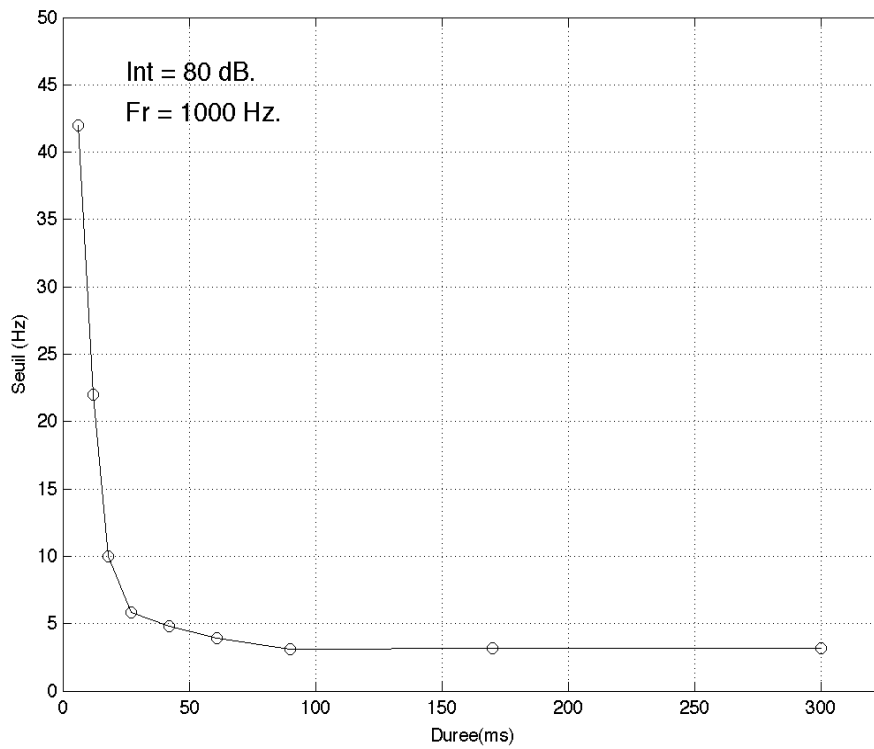


Figure 4. Seuil de perception de la fréquence du son pur en fonction de la durée du stimulus. D'après l'étude de Freyman & Nelson (1986).

L'expérience a porté sur 5 sujets normo-entendants et 7 sujets malentendants, les normo-entendants avaient tous des seuils d'audition inférieurs à 10 dB pour une plage de fréquence de 125 à 4000 Hz. Les signaux de l'expérience étaient des sons purs dont la fréquence était de 1000 Hz. Les résultats de cette étude, en ce qui concerne les normo-entendants, vont dans le même sens que les études précédentes, puisqu'elle met bien en évidence le fait que le seuil s'améliore avec l'augmentation de la durée, notamment pour les durées relativement courtes, et que cet effet de la durée se minimise et devient quasi inexistant pour les durées dont les valeurs dépassent les 200 ms.

Cette dernière étude prouve que le seuil atteint une valeur qui n'est pas sensible à la hausse de la durée. Effectivement, les auteurs de l'étude trouvent qu'il n'y a pas de différence entre 170 et 300 ms, la valeur du seuil étant la même pour ces deux durées-là, soit 3,2 Hz. Donc nous pouvons avancer qu'au-delà de 200 ms toute hausse de la durée n'a pas d'effet sur le seuil.

Cependant, une deuxième question se pose: est-ce que la fréquence a un effet quand la durée devient suffisamment élevée? Comme cette étude n'a porté que sur une seule fréquence, nous ne pouvons pas répondre à la deuxième question. Dans ce dessein, il nous faut une étude qui a porté sur des fréquences différentes pour des durées supérieures à 200 ms.

La réponse est réellement fournie par l'étude faite par Freyman & Nelson (1987). Les auteurs avancent que le processus de l'intégration temporelle du système auditif est quasiment achevé avec une durée de 200 ms. Dans ce but, ils examinent des durées supérieures à 200 ms pour comparer le seuil chez 7 sujets malentendants et 5 normo-entendants. Nous rendons compte uniquement des résultats des normo-entendants. Les stimuli étaient des sons purs de 50, 1000 ou 2000 Hz avec une durée de 5 ms ou de 300 ms. Le temps séparant deux stimuli consécutifs était de 550 ms. La fréquence standard était modifiée de 10%, avec une intensité de 80 dB (± 10 dB). La lecture de la figure 5 nous permet de constater que plus la durée est importante et moins l'effet de la fréquence sur le seuil est important (il n'y a pratiquement plus de différence de seuil à 300 ms entre les fréquences 500, 1000 et 2000 Hz).

Pour répondre à notre deuxième question: quand la durée devient assez élevée, la fréquence a tendance à perdre son effet sur le seuil. Nous précisons que c'est une tendance, car nous ne connaissons pas d'autres études qui ont porté sur des durées supérieures à 300 ms.

I.1.1.3. Rôle de l'intensité

Dans ce paragraphe, nous cherchons à savoir quels sont les rapports qu'entretiennent l'intensité et la fréquence d'un son en ce qui concerne la détermination du seuil de la fréquence.

Dans une étude de Wier *et al.* (1977), les auteurs de l'étude ont mesuré la discrimination de la variation de la fréquence pour 8 valeurs comprises entre 200 et 8000 Hz, chacune de ces valeurs était combinée avec 5 valeurs d'intensité comprises entre 5 et 80 dB. La durée des stimuli était de 500 ms suivie d'un silence de 500 ms. Les sujets de l'expérience étaient 4, ils avaient tous suivi 20 heures d'entraînement à la discrimination de la variation de la fréquence. Les sujets pouvaient être de sensibilités différentes pour la tâche, puisqu'ils ne pratiquaient pas tous le même métier. Cependant une analyse de la variance a montré qu'il n'y avait pas d'effet significatif des sujets, ce qui permettait de considérer la moyenne de tous les sujets plutôt que la moyenne de chacun d'eux.

L'analyse de la variance a montré un effet significatif de la fréquence (le seuil dépend de la fréquence du stimulus, comme vu dans le paragraphe précédent), un effet significatif de

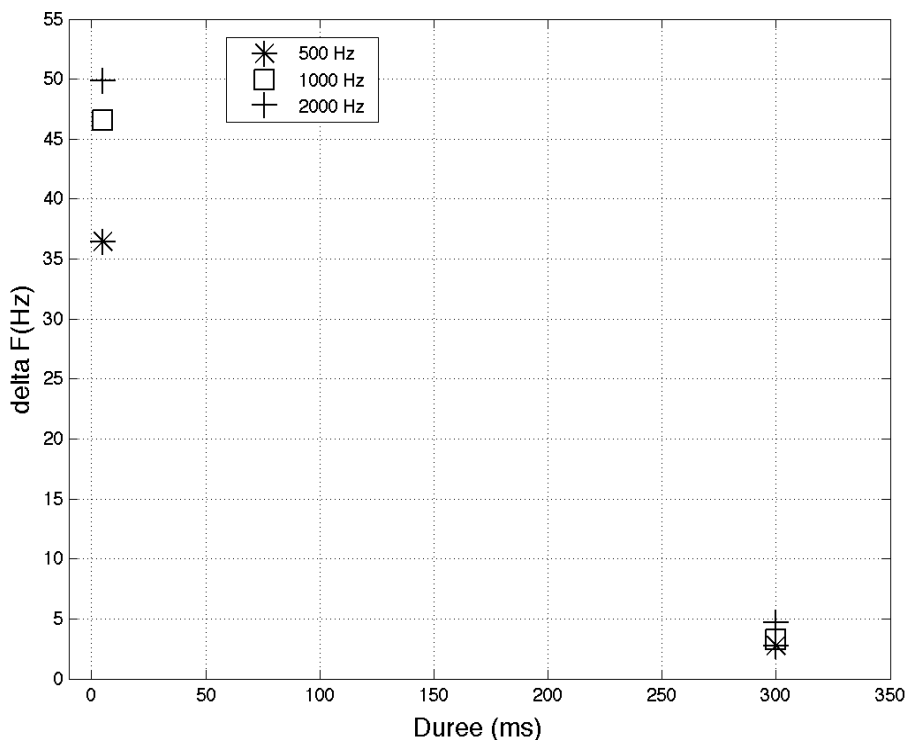


Figure 5. Seuil de perception de la fréquence du son pur en fonction de la durée du stimulus. D'après Freyman & Nelson (1987).

l'intensité (le seuil n'est pas le même selon l'intensité du stimulus), et une interaction significative entre la fréquence et l'intensité (2 stimuli de même fréquence et de 2 intensités différentes ne gardent pas la même différence entre les valeurs de leurs seuils respectifs en cas de changement de la fréquence).

Wier *et al.* proposent de représenter les résultats de leurs mesures non pas en valeurs absolues en Hz, mais, pour Δf , par son logarithme à base de 10, et pour la fréquence, par sa racine carrée (figure 6). Cela permettra, selon eux, d'exprimer mieux linéairement la variation de Δf en fonction de f , soit par des droites dont ils pourraient déterminer l'équation et, par conséquent, prédire le seuil à partir de deux informations: fréquence et intensité.

Les deux figures 6 et 7, où chaque point est la moyenne des 4 sujets de l'expérience, comparent les deux représentations possibles (1) f en fonction de Δf , et (2) la racine carrée de f en fonction de \log_{10} de Δf . Effectivement, la représentation de Δf en termes de son \log_{10} et f en termes de sa racine carrée aplatit les courbes et les rend quasiment parallèles (les droites de régression sont affichées en pointillé dans la figure 6 selon l'équation: $\log_{10} \text{Seuil} = a\sqrt{F} + b$).

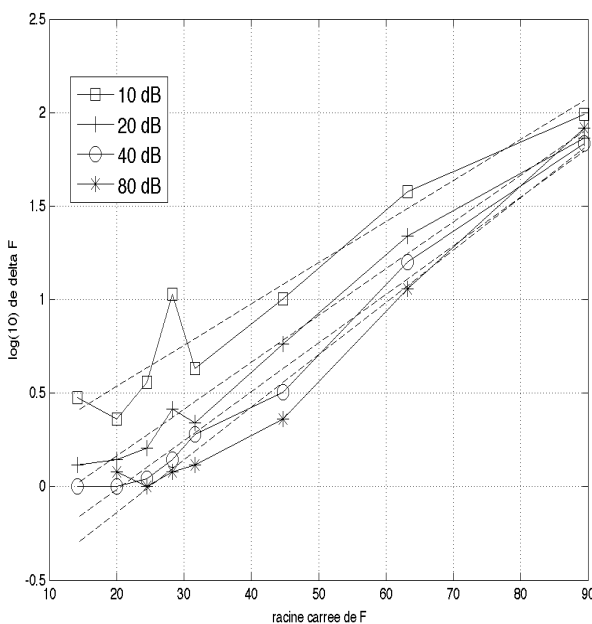


Figure 6. *Log₁₀ du seuil de perception de la fréquence du son pur en fonction de la racine carrée de fréquence du stimulus. Selon Wier et al. (1977).*

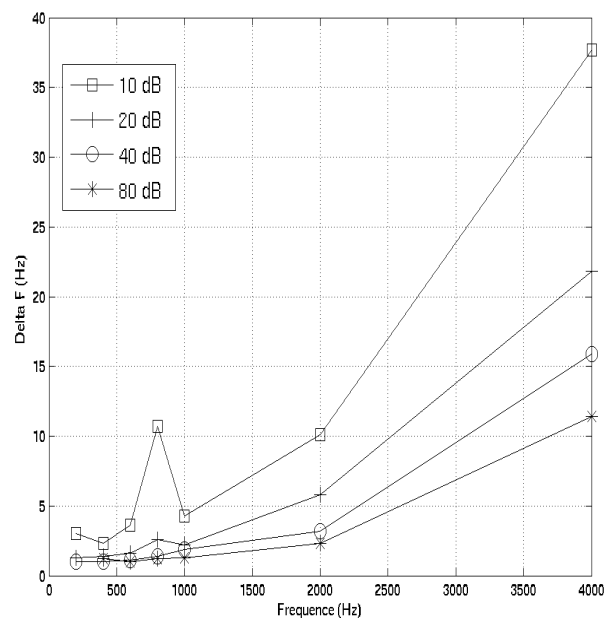


Figure 7. *Seuil de perception de la fréquence du son pur en fonction de la fréquence du stimulus. Selon Wier et al. (1977).*

Les travaux de Wier *et al.* sont absolument considérables, puisqu'ils montrent très clairement l'effet de l'intensité sur le seuil, et mettent bien en évidence le fait que le seuil est fonction de la fréquence et de l'intensité. En effet, pour une durée suffisamment élevée, le seuil ne dépendra que de la fréquence et de l'intensité, cette équation fournit le seuil si nous lui communiquons les valeurs de ces deux paramètres. Cependant, cette étude ne nous renseigne pas sur l'éventuel niveau d'intensité au-delà duquel une hausse ne produit pas d'effet sur le seuil. L'étude de Wier nous laisse penser que le seuil diminue indéfiniment plus l'intensité est élevée, notamment dans les hautes fréquences.

D'autres études, postérieures à celle de Wier *et al.* étaient à la recherche d'un modèle unique qui permettra de prédire la valeur du seuil en fonction de la fréquence et de l'intensité du stimulus, ceci était pour éviter la multitude d'équations proposées par Wier *et al.* (1977), où il fallait reparamétrer a et b dans l'équation à chaque fois que le niveau d'intensité changeait. Nelson *et al.* (1983) se sont basés sur les travaux de Wier *et al.* (1977), en lui reprenant, surtout, son idée de la représentation du \log_{10} du seuil en fonction de la racine carrée de f . Ils ont cherché, contrairement à Wier *et al.* à trouver une seule équation pour toutes les fréquences et pour tous les niveaux d'intensité. Une équation qui minimise par ailleurs l'erreur entre le seuil prédit et le seuil mesuré. Comme Wier *et al.* ils précisent que le

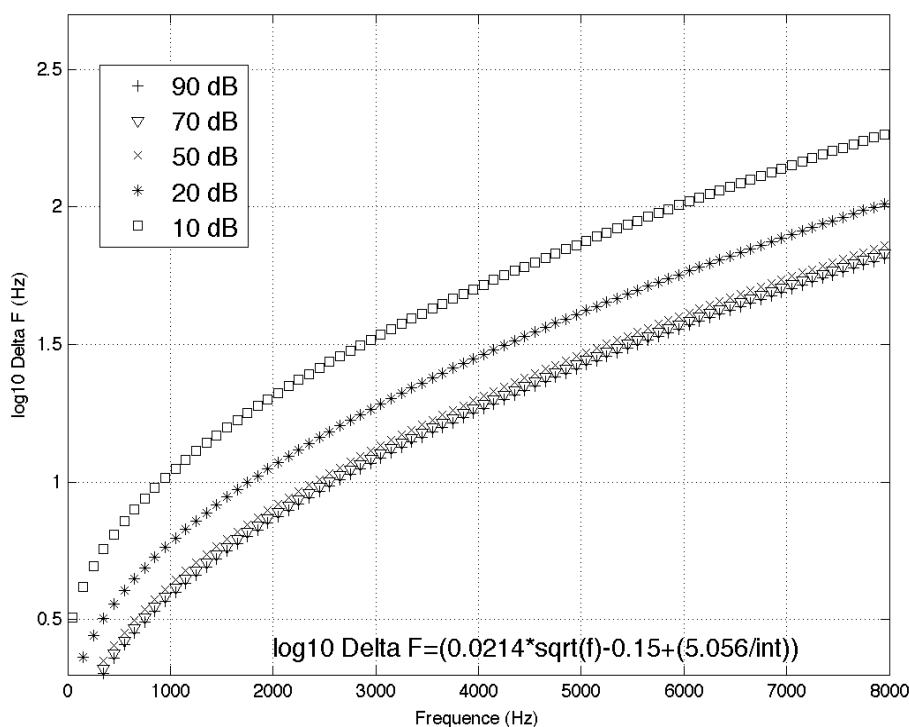


Figure 8. Seuil de perception de la fréquence du son pur en fonction de la fréquence et de l'intensité du stimulus. D'après Nelson *et al.* (1983).

seuil est fonction de deux variables

indépendantes: 1) il est proportionnel à la fréquence; 2) il est inversement proportionnel à l'intensité. Dans cette étude, la durée des signaux était fixée à 300 ms pour éviter l'intervention de la durée dans la détermination du seuil (comme vu au paragraphe précédent). Les chercheurs ont utilisé des fréquences de 300, 600, 1200, 2000, 4000 et 8000 Hz. Le changement était considéré à 71% (cf. Levitt, 1975) de bonnes réponses. L'intensité des stimuli était rangée entre 10 et 80 dB par pas de 10 dB.

En observant les résultats, figure 8, Nelson *et al.* ont remarqué l'apparition d'une limite pour la valeur du seuil, et ceci quand l'intensité s'approchait de 70 dB. Cette limite est le point que nous cherchions. Les auteurs de l'étude ont recalculé a et b qui sont des constantes qui furent déterminées à partir des données de trois études expérimentales². Une étude était faite pour savoir si un a' commun à toutes les données de toutes les études pourrait remplacer le a de chacune des études sans que ce a' ne soit significativement différent de a , c'était le cas, le a' pouvait alors représenter le a avec une perte inférieure à 2% de la variance calculée. Maintenant, il nous faut déterminer un b' qui pourrait rendre compte des différentes valeurs de b en fonction de l'intensité de toutes les différentes études. b' sera représenté par une équation qui rend compte du fait que le seuil diminue plus le niveau d'intensité augmente. $b' = m/Int + k$. m et k sont des constantes qui étaient déterminées à partir des données des trois études mentionnées en tableau 1. Ainsi, l'équation finale sera: $\log_{10} seuil = a' \sqrt{f} + k + (m / Int)$

f est la fréquence du stimulus et Int est son intensité, le tableau 1 donne les valeurs des trois constantes a' , m et k pour les trois études mentionnées ci-dessus.

	k	m	a'
(1) Nelson <i>et al.</i> (1983)	- 0.15	50,056	0.0214
(2) Wier <i>et al.</i> (1977)	- 0.24	4.295	0.0238
(3) Harris (1952)	- 0.43	3.298	0.0235

Tableau 1. Les valeurs de trois constantes permettant de prédire Δf à partir de la fréquence et de l'intensité du stimulus considéré.

Voici un exemple d'application de cette équation avec les données de ces trois études: Pour un stimulus d'une fréquence de 4000 Hz et d'une intensité de 50 dB, la hausse ou la baisse de la fréquence doit être d'une valeur de 20 Hz selon les données de l'étude (1), 22 Hz

² Nelson *et al.* (1983), Wier *et al.* (1977) et Harris (1952).

selon les données de l'étude (2) et 13 Hz selon les données de l'étude (3), pour qu'elle soit perçue par l'oreille humaine.

La figure 8 est affichée selon les données de l'étude (1), nous avons choisi de représenter 4 niveaux d'intensité selon le modèle théorique proposé dans Nelson *et al.* (1983). Nous remarquons que les courbes représentant le niveau 70 et le niveau 90 dB sont pratiquement confondues, ce qui signifie que ce modèle prend effectivement en compte le fait qu'une hausse de l'intensité au-delà de 70 dB ne produit aucun changement en ce qui concerne le seuil. Autrement dit, au-delà de 70 dB, le seuil ne dépend que de la fréquence du stimulus.

L'équation explique 99 % de la variance totale pour les données de l'étude de Nelson *et al.* (1983), 96 % de l'étude d'Harris (1952) et 93 % dans l'étude de Wier *et al.* (1977).

La variance totale prise en compte par l'équation, à travers tous les niveaux d'intensité et de fréquence, est de 98.9%. C'est-à-dire que l'équation explique 98.9% de la variance totale pour les données de l'étude de Nelson *et al.* (1983). Appliquée aux données d'autres études (Wier *et al.*, 1977), cette équation devient assez approximative. Toujours selon les auteurs de l'étude (1), cette équation donne une description exacte du seuil différentiel de la perception de la fréquence sur une large plage d'intensité (10-60 dB) et de fréquence (300-8000 Hz). Résultat: pour un son pur de 300 ms et de 70 dB, le seuil de sa variation ne dépendra que de la fréquence du stimulus.

Nous pouvons résumer ce chapitre en donnant les points communs de ces études examinées ci-dessus:

1) il n'y a pas de rapport simple entre le seuil de la fréquence et la durée du stimulus; pour des durées inférieures à 20 ms, la moindre hausse dans la durée se traduit par une définition meilleure du seuil (cf. figure 4).

Par ailleurs, au-delà de **200 ms**, une hausse de la durée ne produit aucun effet sur le seuil. Le seuil ne dépendra alors (pour une intensité suffisamment élevée) que de la fréquence.

2) le seuil est indéfiniment proportionnel à la fréquence. Il semble toutefois que la fréquence a tendance à perdre son effet sur le seuil, plus la durée est importante, pour converger ainsi vers une valeur unique, environ **4 Hz**, quelle que soit la fréquence (cf. figure 5, Freyman & Nelson, 1987, pour une durée de 300 ms.) Il est possible que des valeurs élevées de la fréquence (supérieures à 4000 Hz) ne tendent pas vers cette valeur de 4 Hz (cf. figure 2, Moore, 1973, pour une durée de 200 ms)

3) l'intensité influence le seuil jusqu'à **70 dB**. Au-delà de cette valeur, une hausse de l'intensité produit une baisse pratiquement négligeable en ce qui concerne le seuil différentiel de la perception de la fréquence.

Ces résultats, qui portent sur les sons purs, quoique pas définitifs et quelques peu différents d'une étude à l'autre, constituent avec beaucoup de vraisemblance la base de la performance de l'oreille humaine et nous servent de référence concernant le rôle de la durée, de la fréquence et de l'intensité dans la capacité à résoudre les variations de la fréquence (percevoir Δf). Ainsi nous nous attendons à ce que des structures acoustiques plus complexes aient des Δf supérieurs à ceux qui sont trouvés dans les études examinées ci-dessus.

I.1.2. Formant

Déterminer le changement fréquentiel minimal au niveau du formant entraînant une sensation perceptive différente: tel est notre objectif dans ce paragraphe. Nous examinons donc une structure acoustique plus complexe que celle du son pur, ceci dans le but d'en connaître le seuil comme nous avons fait pour le son pur, il s'agit des formants. Plusieurs études se sont intéressées à la question du seuil: "de combien faut-il augmenter ou abaisser la valeur d'un formant pour que ce changement (hausse ou baisse) soit perceptible?". Nous en citons quelques-unes dans le tableau 2 ci-dessous:

Fréquence (Hz)	ΔF (Hz)	$\Delta F/F$ (%)	
F1			
Flanagan (1955)	300 - 700	12 - 27	4.4
Mermelstein (1978)	350 - 600	33 - 50	9.6
Sinnott & Kreiter (1990)	270 - 390	5 - 12	2.4
Hawks (1994)	227 - 921	Non Rens.	1,8
F2			
Flanagan (1955)	1000 - 2000	20 - 90	3.4
Mermelstein (1978)	1780 - 2100	58 - 142	5.3
Sinnott & Kreiter (1990)	1990 - 2290	13 - 28	0,96
Hawks (1994)	764 - 2275	Non Rens.	1,97
Gagne & Zurek (1988)	300 - 800	7 - 20	2.8

Tableau 2. Seuil différentiel de la perception de la fréquence du formant selon plusieurs études.

Nous remarquons qu'aussi bien en valeur absolue qu'en valeur relative, les résultats sont assez différents. D'une étude à une autre, les stimuli, la méthode de présentation et le niveau d'entraînement des sujets ne sont pas les mêmes, ce qui expliquerait la divergence dans les résultats de ces études.

Déterminer la capacité de l'oreille humaine à résoudre les variations de la fréquence des formants est une question qui revêt beaucoup d'importance. Au départ, les recherches de James L. Flanagan, en 1955, étaient motivées par le fait que cette question avait une application fondamentale dans la synthèse de la parole, plus précisément des voyelles. La question était de savoir quelle était la marge de fluctuation de la valeur du formant, la réponse

permettrait d'éviter de se soucier, lors de la synthèse, de ce qui n'était pas perceptible par l'oreille humaine.

Les études qui avaient pour but de déterminer les limites différentielles de la variation des formants se sont poursuivies après celle de Flanagan pour diverses raisons: comparaison entre l'homme et le singe, comparaison entre normo-entendants et malentendants, perception de la parole, etc.

L'étude de Flanagan (1955) a inspiré beaucoup d'autres études, Mermelstein (1978), Nord & Sventelius (1979), Hawks (1994), Kewley-Port & Watson (1994), lesquelles se sont, plus ou moins, basées sur le modèle de Flanagan. C'est pour cette raison que nous accordons une importance particulière à cette étude pionnière et nous voudrions en décrire, brièvement, la méthode et les résultats.

Flanagan élaborait des tests pour déterminer les frontières perceptives pour les deux premiers formants F1 et F2, mais pas pour F3 ou F4 qu'il maintenait fixes. Quand il manipulait le F1, qui prenait les valeurs 300; 500 et 700 Hz, F2, F3 et F4 étaient fixes, et prenaient les valeurs suivantes: F2 = 150, F3 = 2500 et F4 = 3350 Hz. Quand il manipulait le F2 qui prenait les valeurs 1000; 1500 et 2000 Hz, F1, F3 et F4 sont fixes, et prenaient les valeurs suivantes: F1 = 50, F3 = 2500 et le F4 = 3350 Hz.

La variation utilisée pour la détermination des limites du F1 était entre -70 et +70 Hz par pas de 10 Hz; alors que la variation utilisée pour la détermination des limites du F2 était entre -175 et +175 Hz par pas de 25 Hz (figure 9). Le changement était considéré à 50% des réponses. La largeur de la bande du formant pour les 3 premiers, était respectivement: 130, 150 et 185 Hz.

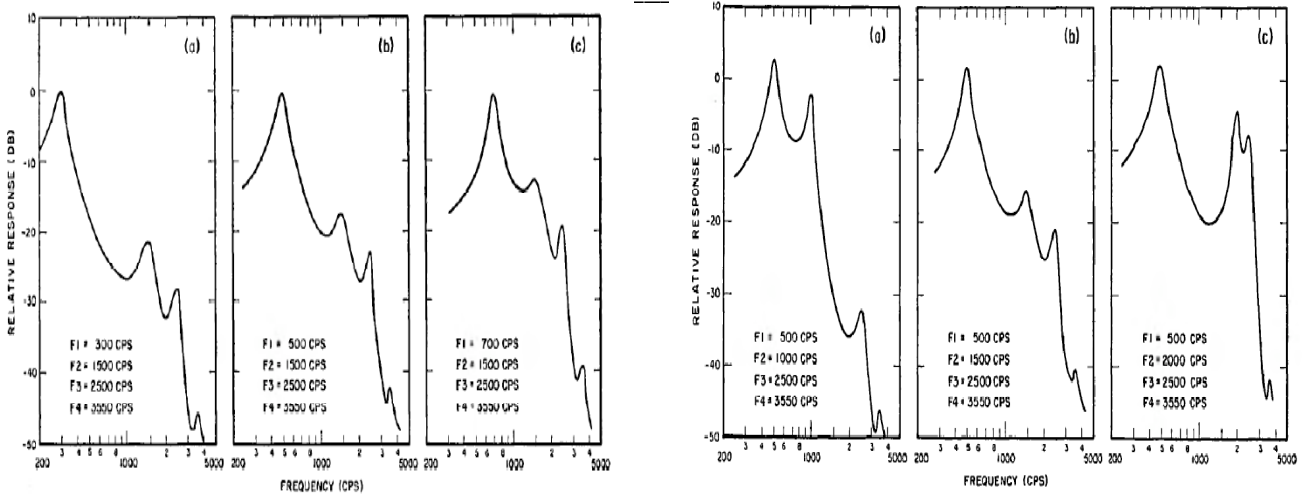


Figure 9. Spectre acoustique de la voyelle synthétique utilisée par Flanagan (1955) pour la détermination du seuil différentiel du premier formant (à gauche), [F1:300; 500 et 700 Hz, F2, F3 et F4 étaient fixes, F2 = 150, F3 = 2500 et F4 = 3350 Hz] et du deuxième formant (à droite), [F2: 1000; 1500 et 2000 Hz, F1, F3 et F4 étaient fixes, F1 = 50, F3 = 2500 et F4 = 3350 Hz].

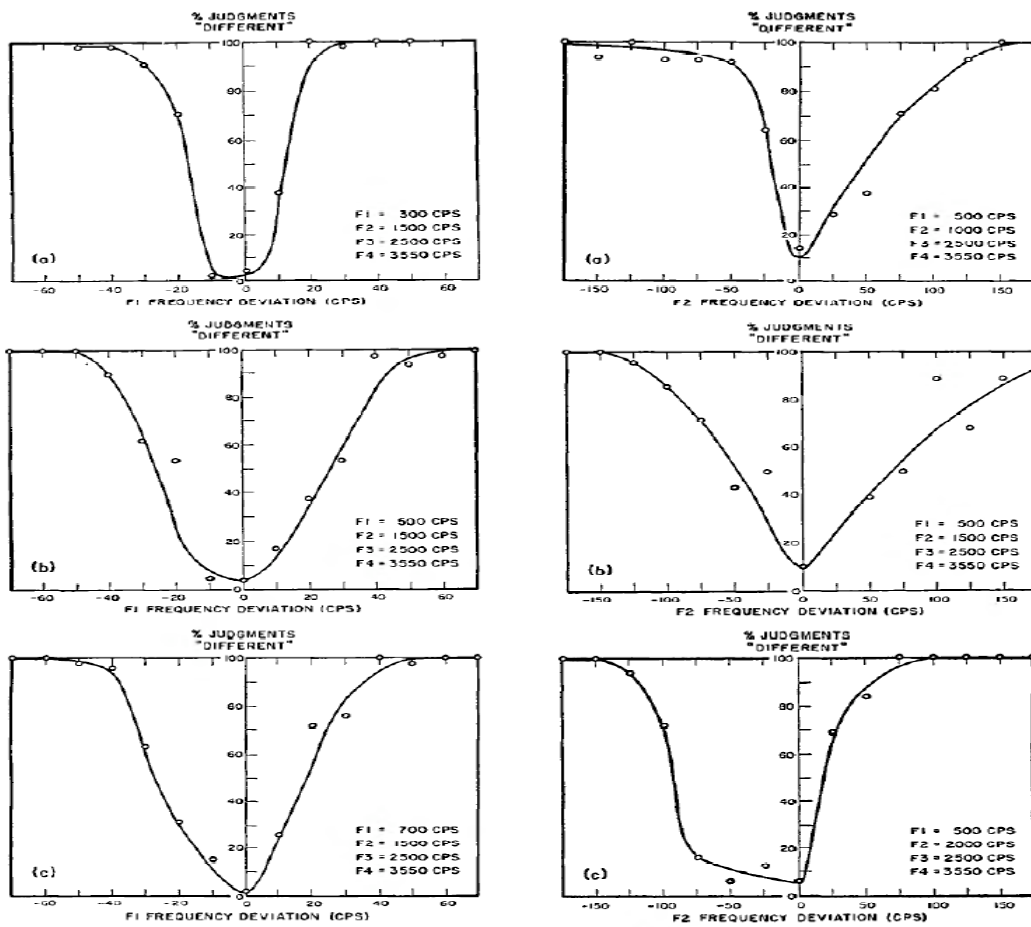


Figure 10. Taux de jugement "différent" en fonction du changement de la fréquence, pour F1 (colonne de gauche), et pour F2 (colonne de droite). Flanagan (1955).

La présentation était faite par paire (standard (A)/modifié (B)) dans un ordre aléatoire, le signal présenté avait une durée de 750 ms, le temps séparant 2 stimuli consécutifs d'une paire était de 500 ms, le temps séparant 2 paires consécutives était de 5 secondes. Le test a consisté en 20 paires aléatoires (A/B) et 6 paires identiques (A/A). Les sujets étaient 4 hommes adultes. La question était: "Est-ce que les deux termes de la paire sont les mêmes ou est-ce qu'ils sont différents? "

Les résultats montrent que la détection du changement dans la qualité (le jugement "*différent*") avait lieu à partir des 20 paires (A/B). Les ogives (figure 10), montrent le taux de jugement "*différent*" en fonction du changement de F1 (idem pour F2). Flanagan considérait que le changement avait lieu à 50% de jugement "*différent*". Flanagan présentait ses résultats sous formes d'ogive, où il représentait la variation de la fréquence des formants sur l'axe des abscisses et les réponses des sujets sur l'axe des ordonnées. Ce modèle était suivi par d'autres chercheurs, nous le retrouvons dans d'autres études, comme celle de Nord & Sventelius (1979).

Selon Flanagan, si l'on considérait le jugement au niveau individuel, la variabilité lors du jugement des seuils semblerait une variabilité très faible, d'où l'intérêt du regroupement des données.

Certaines ogives étaient asymétriques par rapport à l'axe des ordonnées. Flanagan notait que le seuil n'était pas symétrique et que l'asymétrie n'était pas systématiquement du même côté, tantôt la valeur de la hausse était supérieure à celle de la baisse et tantôt le contraire. Flanagan attribuait cette asymétrie au fait que deux formants se trouvaient proches l'un de l'autre. Cette interprétation ne faisait pas l'unanimité des chercheurs (cf. Kewley-Port & Watson, 1994).

Les résultats numériques de la détection du changement sont présentés dans le tableau 3.

F1 (Hz)	seuil négatif (Hz)	seuil positif (Hz)
300	17	12
500	25	27
700	27	19
La différence moyenne (tous seuils confondus) est de 4,4%		
F2 (Hz)	seuil négatif (Hz)	seuil positif (Hz)
1000	20	50
1500	45	75
2000	90	20
La différence moyenne (tous seuils confondus) est de 3,4%		

Tableau 3. Les valeurs où la perception de la "différence" se présente à 50%. La différence moyenne (tous seuils confondus) est de 3,4%.

En résumé: Flanagan trouve que le seuil est de 4,4 % pour F1 et de 3,4 % pour F2. Le seuil, tous formants confondus, est de 3,9 %.

Nord & Sventelius (1979) se sont basés sur le modèle d'expérimentation proposé dans l'étude de Flanagan (1955), dans le sens où leur étude a porté sur F1 et F2 séparément. Ils ont trouvé des valeurs similaires à celles données par Flanagan: la valeur moyenne du seuil était de 3,3 %, tous formants confondus, alors qu'il était de 3,9 % chez Flanagan. Bien que les résultats d'autres études, comme celle de Nord & Sventelius (1979), aillent dans le même sens que les résultats de Flanagan (1955), les résultats de son étude ont été discutés par la suite. En effet, dans une étude de Mermelstein (1978), l'auteur était en accord avec Flanagan sur le sens de la variation du seuil, le rapport entre le seuil et la fréquence du formant pour F1 et F2, et sur le fait qu'il y avait de l'asymétrie entre hausse et baisse. Cependant, Mermelstein (1978) ne trouvait pas les mêmes valeurs pour le seuil. Pour F1, le seuil était de 50 Hz à 350 Hz, alors que Flanagan trouvait qu'il était de 15 Hz à 300 Hz. À 600 Hz, les résultats des deux auteurs se rapprochent et sont comparables. Mermelstein trouvait aussi que le contexte phonétique environnant la voyelle avait un effet sur le seuil, un aspect ignoré par Flanagan, puisque Flanagan n'a travaillé que sur des voyelles isolées. Les résultats de Mermelstein pour la voyelle [I], montraient qu'il n'y avait pas de différence de seuil pour F1 entre voyelle isolée (le seuil était de 50 Hz) et voyelle en contexte [gVg] ou [bVb] (le seuil était de 49 Hz). Par contre, pour F2 le seuil pour la voyelle isolée était de 142 Hz, alors qu'il était de 186 Hz pour la voyelle insérée dans un contexte phonétique, soit une hausse de 31%. Quant à l'asymétrie mentionnée par Flanagan, Mermelstein lui accordait peu de considération.

D'autres études affichaient leur désaccord avec les résultats de Flanagan en soulignant le fait qu'il effectuait ses mesures sur des formants isolés, un seul formant à la fois. En effet, ceux qui n'étaient pas en accord avec ce procédé expérimental avançaient l'argument que dans la parole naturelle il n'en était pas ainsi, puisque la variation pour tous les formants se faisait simultanément, Hawks (1994). Ce dernier remarquait l'absence totale d'études du F3. Mais il n'était sûrement pas au courant d'une étude antérieure réalisée par Sinnott *et al.* (1991), dont le but fut de mesurer la différence entre les seuils différentiels des continua vocaliques chez l'homme et chez le singe, et comparer cette différence à celle qu'ils ont trouvée entre les seuils différentiels du son pur. Sinnott *et al.* faisaient varier plus d'un formant en même temps, et faisaient intervenir le F3 pour l'un des continua (cf. tableau 4).

Sinnott *et al.* (1991) ont travaillé sur 3 continua vocaliques synthétiques de l'anglais, [ɪ - i], [æ - ε] et [ɑ - Λ]; ces voyelles avait une durée de 300 ms, une intensité de 80 dB avec une F0 qui commençait à 120 Hz et décroissait linéairement jusqu'à la fin de la voyelle pour atteindre 80 Hz.

Pour la transition [ɪ - i], ils faisaient varier les 3 premiers formants simultanément, par pas de 20, 50 et 75 Hz respectivement, ce qui donnait un continuum de 7 stimuli.

Alors que pour les transitions [æ - ε] et [ɑ - Λ], seuls les deux premiers formants étaient changés par pas de 20 Hz chacun, et émis simultanément, ce qui donnait 7 stimuli pour [æ - ε] et six stimuli pour [ɑ - Λ]. Les signaux étaient émis au rythme de 1 par seconde (300 ms durée stimulus + 700 ms d'intervalle entre deux signaux). Les seuils étaient choisis à 50 % des réponses. Pour tous ces continua, des expériences supplémentaires qui consistaient à varier les valeurs fréquentielles des formants en des pas plus petits (10, 5 et 2 Hz) ont été effectuées aussi, mais cela n'a pas donné des résultats meilleurs.

	[ɪ - i]	[æ - ε]	[ɑ - Λ]
F1	390 => 270	660 => 540	730 => 630
F2	1990 => 2290	1720 => 1840	1090 => 1190
F3	2550 => 3000	2410 => 2410	2240 => 2240
$\Delta F1$	5 à 11	8 à 16	8 à 14
$\Delta F2$	13 à 28	8 à 16	8 à 14
$\Delta F3$	19 à 41	stable	stable
$^3\Delta F$	20,8	12,1	10,9

Tableau 4. Seuil du formant (toutes valeurs en Hz). Résumé des résultats pour l'homme.

Cette étude se rapproche beaucoup de la condition naturelle de l'émission de la parole, où tous les formants évoluent simultanément. Cependant, nous observons l'absence du contexte phonétique dans cette étude, lequel a pourtant un effet significatif sur le seuil (Mermelstein, 1978), le F3 n'est pris en compte que pour le seul continuum vocalique [ɪ - i].

En réalité, le fait de considérer tous les formants à la fois, même si c'est la condition naturelle, ne nous permet de comparer que des seuils pour un même continuum vocalique, cela nous empêche de nous rendre compte de l'importance de chaque changement pour chaque formant à part, d'où l'intérêt pour nous de nous appuyer, tout de même, sur les études qui ont étudié les formants séparément et qui nous renseignent sur le seuil pour chaque formant et pour chaque contexte.

D'autres recherches sur le seuil ont fait intervenir d'autres facteurs dont on ignorait, jusqu'alors, l'effet exact. Nous pouvons citer à ce sujet l'étude menée par Gagne & Zurek (1988), laquelle cherchait à déterminer, comme ses précédentes, la capacité de l'oreille pour 4 sujets à détecter le plus petit changement perceptible (le seuil) de la fréquence d'un formant, et les facteurs dont cette capacité dépendait. Dans ce dessein, ils considéraient, d'un côté, la nature de la source, et de l'autre côté, la largeur de la bande du formant. Le signal de source avait deux natures différentes, un signal harmonique ou un bruit blanc aléatoire; Le signal harmonique avait 2 types de F0: : 1) un F0 fixe, soit de 100 Hz, soit de 250 Hz; 2) ou bien un F0 qui variait sur une étendue (max - min) de 50 Hz. Ce dernier F0 augmentait de 50 Hz sur les premiers 100 ms du signal, puis descendait sur les 300 ms qui suivaient. Les deux valeurs initiales de la F0 (100 vs. 250 Hz) étaient combinées avec les 2 types de F0 (fixe et variable), ce qui créait 4 conditions expérimentales. Les valeurs des formants mesurés étaient 300; 475; 650 et 800 Hz, des valeurs qui couvraient toute, mais seulement, la région de F1. Ces mêmes

³ Tous formants confondus.

valeurs ont été testées avec un bruit blanc comme source, en plus d'un formant à 2000 Hz. Les tendances restaient les mêmes malgré le changement de la nature de la source, et toutes peuvent être bien résumées par la formule (1).

Gagne & Zurek (1988) introduisaient dans leur calcul un nouveau facteur. En effet, pour chaque formant ils changeaient la valeur du facteur de qualité⁴, Q, en changeant la largeur de la bande du formant. Q prenait alors les valeurs suivantes: 1; 2,8; 10 et 36 unités SI. Parce que, selon eux, le facteur de qualité intervient dans la détermination de la nature du changement, par exemple, lorsque la source de l'excitation est un bruit blanc (le spectre est continu) le changement de la fréquence de résonance est perçu différemment selon la largeur de la bande (un changement de fréquence en cas de bande étroite et un changement d'intensité ou de timbre dans le cas d'une bande large), d'où l'intérêt de le considérer dans la détermination du seuil.

Les durées des stimuli et des intervalles séparant les stimuli étaient de 400 ms.

Les résultats de l'étude ont montré que la nature de la source avait très peu d'impact sur la valeur du seuil. Une analyse de la variance pour la source harmonique montrait que la valeur initiale de la F0 (100 vs. 250 Hz) n'influçait pas la valeur du seuil; d'autre part, la détection du seuil entre une F0 qui variait et une F0 qui était fixe était significativement différente, mais ceci était vrai dans une seule condition (F0 = 100 Hz et F0 est fixe). Pour les autres conditions, l'effet de la variation de la F0 était très légèrement, mais systématiquement, différent de l'effet de la stabilisation de la F0. Ces analyses ont montré aussi qu'il y avait un effet pour la fréquence du formant et un effet de Q. Le plus petit changement susceptible d'être perçu (seuil) a une valeur proportionnelle à la fréquence et inversement proportionnelle à Q (proportionnelle à la largeur de la bande).

Les résultats avec une source constituée d'un bruit blanc sont mieux ordonnés, ils vont dans le même sens que les résultats obtenus avec le signal harmonique. Les tendances

⁴ Le facteur de qualité est défini par: la fréquence du formant divisé par la largeur de la bande. $Q = f_{\text{Formant}}/BW$.

$$\begin{aligned} \Delta F &= 0.079 \times F_{\text{formant}} / \sqrt{Q} \\ &= 0.079 \frac{\sqrt{F_{\text{formant}}} \times \sqrt{F_{\text{formant}}}}{\sqrt{F_{\text{formant}}} \div \sqrt{BW}} \\ &= 0.079 \frac{\sqrt{F_{\text{formant}}} \times \sqrt{F_{\text{formant}}} \times \sqrt{BW}}{\sqrt{F_{\text{formant}}}} \\ &= 0.079 \times \sqrt{F_{\text{formant}}} \times \sqrt{BW} \\ &\text{d'où} \\ \Delta F &= 0.079 \times \sqrt{F_{\text{formant}}} \times \sqrt{BW} \end{aligned}$$

observées avec le signal harmonique (effet de la valeur initiale et de la variation de la F0) sont mieux marquées avec un bruit blanc.

Selon cette étude, Δf est proportionnel à la fréquence, cela est en total accord avec les précédentes études que nous avons examinées plus haut (Moore, 1973; Wier *et al.*, 1977, Nelson *et al.*, 1983; Freyman, 1987). Δf sera, selon Gagne & Zurek (1988), inversement proportionnel au facteur de qualité Q. Pour toutes natures de source, Gagne & Zurek (1988) résument leurs résultats concernant Δf par l'équation suivante: où F est la fréquence du formant considéré et Q est le facteur de qualité.

$$\Delta F = 0.079 \times F_{\text{formant}} / \sqrt{Q}$$

Formule (1) Prédiction du seuil différentiel de la perception de la fréquence d'un formant en fonction de la fréquence et la largeur de la bande du formant considéré.

Nous avons réécrit ci-dessous la formule (1) de manière à faire apparaître la largeur de la bande du formant.

$$\Delta F = 0.079 \times \sqrt{F_{\text{formant}} \times BW}$$

Avec BW la largeur de la bande, cette réécriture montre que Δf est proportionnel à la largeur de la bande du formant et proportionnel à la fréquence. La représentation graphique (figure 11), est faite à partir de l'équation proposée par Gagne & Zurek (1988). Bien que leur étude n'ait porté que sur un seul formant, nous avons choisi de représenter une plage qui couvre les 3 premiers formants, et ceci pour pouvoir comparer graphiquement par la suite leurs résultats avec des résultats d'autres études qui ont porté sur une plage de fréquence plus large.

En résumé de cette étude, nous pouvons considérer les nouveautés suivantes:

- La nature de la source, harmonique *vs.* aléatoire, a très peu d'impact sur la valeur du seuil.
- La valeur de la F0 (100 *vs.* 250), n'influence pas le seuil.
- La variation de F0 (F0 fixe *vs.* F0 variable), a un effet significatif sur le seuil.
- Le seuil est proportionnel à la fréquence du formant.
- Le seuil est proportionnel à la largeur de la bande du formant.

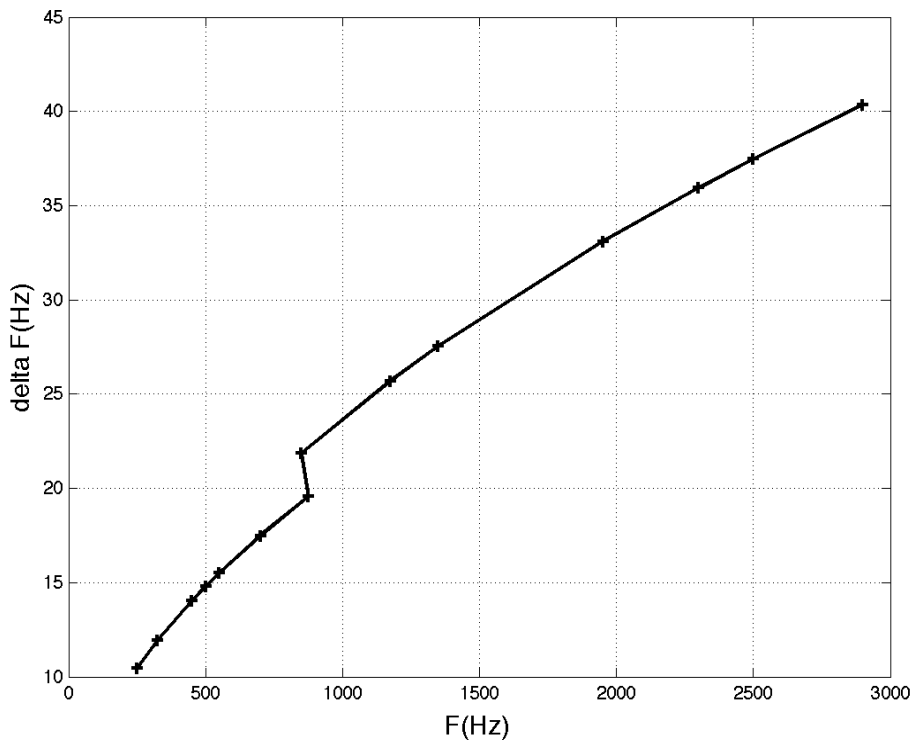


Figure 11. Seuil différentiel de la fréquence du formant en fonction de la fréquence et de la largeur de la bande du formant. D'après Gagne & Zurek (1988)

Parmi les travaux qui ont été réalisés dans ce domaine, nous nous sommes particulièrement intéressés à l'étude plus exhaustive que ses précédentes, effectuée par Kewley-Port & Watson (1994). Cette étude a porté sur le seuil perceptif, relatif et absolu, par hausse et par baisse pour le premier formant d'un côté et du deuxième formant de l'autre.

L'étude a porté sur 10 voyelles monophthongues de l'anglais [i, ɪ, e, ɛ, æ, ɑ, ʌ, o, ʊ, u], lesquelles ont été synthétisées d'après la voix d'une femme adulte avec une F0 de 200 Hz et une durée de 160 ms. L'expérience a consisté en la hausse des valeurs de F1 et de F2, le seuil de différentiel était fixé à 71% de réponses, idem pour la baisse des valeurs de F1. Au total ils ont effectué 40 mesures (10 voyelles * 2 formants * 2 conditions). Comme chez Flanagan, ou encore chez Mermelstein, Kewley-Port *et al.* ont observé à leur tour une asymétrie dans la détermination du seuil, entre hausse et baisse, laquelle se manifestait pour plusieurs mesures par un seuil anormalement élevé.

Flanagan pensait que cette asymétrie était due au fait que deux formants d'une voyelle se rapprochaient. Kewley-Port & Watson (1994) ont effectué, dans une étude secondaire, une comparaison du seuil pour des voyelles qui avaient des F1 et F2 relativement proches, telles

/o, ɑ, u/ ou /i/ où F2 et F3 sont relativement proches, avec le seuil pour des voyelles qui avaient des formants relativement éloignés. Ils ont trouvé que le seuil était de 19 Hz pour le première type de voyelles, alors qu'il était de 20,5 Hz pour le deuxième. Cela mettait bien en évidence le fait que ce n'était pas la proximité des formants qui était la cause de l'asymétrie.

Mermelstein n'était pas en accord non plus avec ce que pensait Flanagan en ce qui concerne l'asymétrie, il pensait que cette asymétrie se produisait lorsque la F0 n'était pas alignée avec l'harmonique centre du formant. Les données de Kewley-Port & Watson (1994) sont en complète contradiction avec ce qu'avancait Mermelstein, parce que cette asymétrie est, justement, fortement prononcée dans le cas de l'alignement. En effet, Kewley-Port & Watson (1994) ont attribué la cause à l'alignement de la fréquence fondamentale avec l'harmonique centre du formant; alignement veut dire que l'harmonique centre du formant est un multiple entier de la F0.

Donc l'asymétrie entre la hausse et la baisse se produisait lorsque le centre du formant était une harmonique alignée avec la fréquence fondamentale (un multiple entier de 20, dans l'expérience de Kewley-Port & Watson (1994)). Cet alignement produit, par ailleurs, des seuils anormalement élevés, une fois que les seuils de ces formants alignés sont éliminés, il n'y a pratiquement plus d'asymétrie, et les valeurs des seuils obtenues par hausse comme par baisse peuvent être considérées comme provenant de la même condition expérimentale.

En effet, quand les valeurs des seuils pour les 5 formants alignés avec la F0 étaient éliminées, c'était le cas du F1 de la voyelle /ɛ/ (600 Hz), F1 de la voyelle /æ/ (1000 Hz), F2 de la voyelle /o/ (800 Hz), F2 de la voyelle /ʌ/ (1400 Hz) et le F2 de la voyelle /ɛ/ (2200 Hz), la moyenne de la valeur absolue de la différence entre seuil-hausse et seuil-baisse pour les 15 formants qui restaient, était de 3 Hz. Cette moyenne était beaucoup plus importante pour les 5 formants alignés avec la F0, elle était de 14 Hz.

Une autre expérience supplémentaire a montré que la F0 avait une grande influence sur la valeur du seuil. En effet, contrairement à Gagne & Zurek (1988), Kewley-Port & Watson (1994), ont trouvé que la valeur de la F0 influençait le seuil, la moyenne du seuil avec une F0 = 160 Hz était de 19 Hz, alors qu'elle était de 24 Hz pour des F0 = 200 ou F0 = 240 Hz. Encore une fois, contrairement à Gagne & Zurek (1988), Kewley-Port *et al.* ont trouvé qu'il n'y avait pas de différence significative entre le seuil obtenu avec une F0 stable et *vs.* une F0 qui varie. La variation de la F0 lors de l'émission de la voyelle a l'avantage d'éviter l'alignement et par conséquent l'asymétrie, mais elle n'influe pas sur la valeur du seuil.

Une expérience parallèle était menée afin de déterminer si le seuil était réellement lié à la F0. Les résultats ont montré qu'il y avait une interaction significative ($p < 0,05$) entre la variation de la F0 et la variation du seuil. Une F0 plus élevée a entraîné une hausse au niveau du seuil, et il y a eu, par ailleurs, une interaction significative ($p < 0,02$) entre la F0 et les valeurs des formants. Le seuil est mieux décrit en termes des valeurs absolues qu'en termes des valeurs relatives, du moment où $100*(\Delta f/f)$ ne suit pas une fonction évidente.

La représentation graphique du seuil en fonction de la fréquence du formant (figure 12), montre que le seuil pour F1 est très similaire à travers les sujets, le seuil avait une valeur quasi constante, d'où l'intérêt de le représenter en termes de sa valeur absolue.

Alors que pour F2 le seuil n'était pas constant, mais relatif à la fréquence. Leurs observations expérimentales les ont amenés à la conclusion suivante: L'expression du *delta f* en fonction de la fréquence peut être représenté par une fonction affine en 3 morceaux (*piecewise*), 3 coefficients directeurs pour 3 intervalles sur les abscisses:

1) de 200 à 550 Hz; $y = 0*x + 14.5$. coefficient directeur nul: Fonction constante à 14.5.

2) de 550 à 800 Hz; $y = 0.0034*x + 12.63$.

3) de 800 à 3000 Hz; $y = 0,010*m + 7.35$.

Le seuil sur la plage des fréquences qui couvrent le premier formant est constant $\approx 14,5$ Hz. Sur l'intervalle qui couvre les valeurs du F2, au-delà de 800 Hz, le seuil est fonction de la fréquence et vaut $\approx 1,5 \%$.

En termes de $100*\Delta f/f$ le seuil différentiel sur la plage des fréquences qui couvrent le premier formant n'est pas constant, Δf pour F1 étant insensible à la hausse de f . Le seuil est proportionnel à la fréquence pour l'intervalle qui couvre les valeurs de F2, au-delà de 800 Hz, ce qui fait qu'il est constant en termes de $100*\Delta f/f$.

Pour résumer:

Kewley-Port & Watson (1994) considèrent que le seuil est mieux exprimé en termes de valeur absolue pour les fréquences inférieures à 800 Hz (la région de F1), parce qu'il est constant, il est de 14.5 Hz; au-delà de cette valeur (la région de F2), le seuil est mieux exprimé en valeur relative, $100*(\Delta f/f)$, parce qu'il est constant, en pourcentage, soit 1,5%.

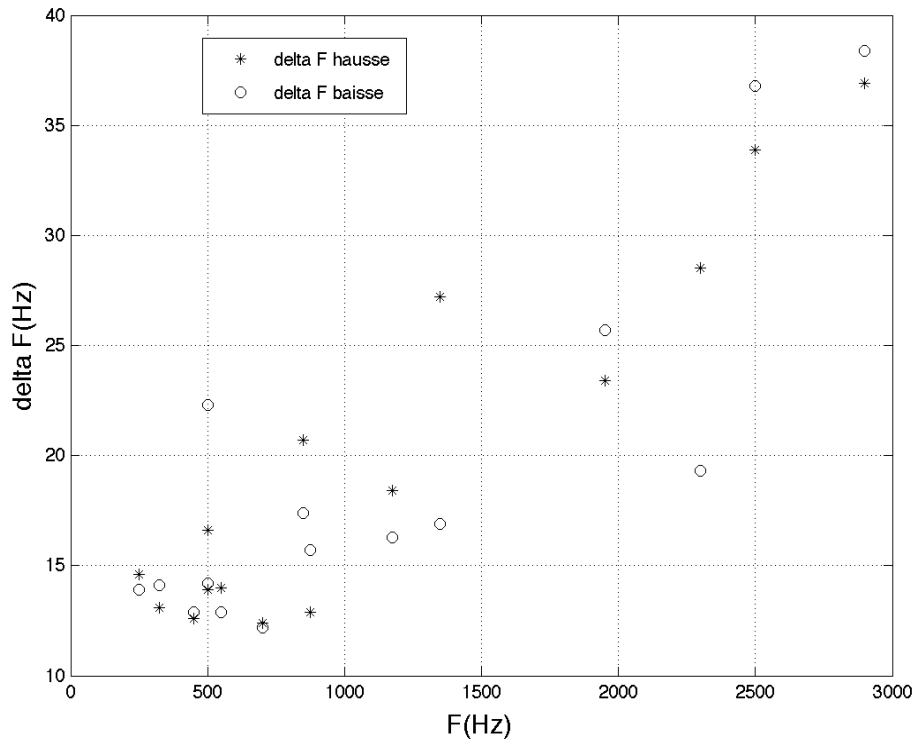


Figure 12. Seuil absolu de perception de la fréquence du formant en fonction de la fréquence. Kewley-Port & Watson (1994).

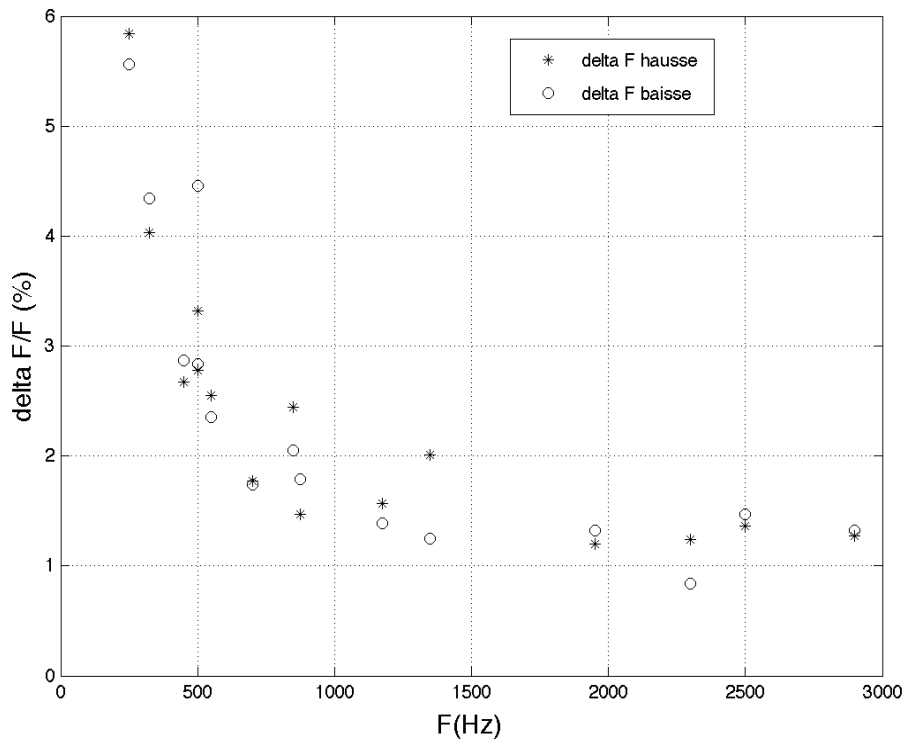


Figure 13. Seuil différentiel relatif de perception de la fréquence du formant en fonction de la fréquence. Kewley-Port & Watson (1994).

Nous voudrions comparer graphiquement ici (figure 14), les résultats de Kewley-Port & Watson (1994) présentés : (i) soit d'après les mesures expérimentales; (ii) soit selon le modèle théorique proposé par Gagne & Zurek (1988). Ces deux représentations seront comparées avec les résultats concernant le son pur, donnés par Moore (1973).

La prédiction théorique fournie par l'équation de Gagne & Zurek (1988) donne des résultats qui sont très proches des résultats des mesures effectuées par Kewley-Port & Watson (1994), cette similarité est mieux prononcée pour les fréquences en dessous de 800 Hz, où la valeur moyenne du seuil est de 14.1 Hz dans la première étude, et de 14.5 Hz dans la deuxième étude.

Cela montre, en dépit des divergences sur beaucoup de points concernant notamment le rôle de la F0 dans la détermination du seuil, que ces deux dernières études affichent une certaine cohérence quant à la valeur du seuil en fonction de la fréquence du formant. Le seuil

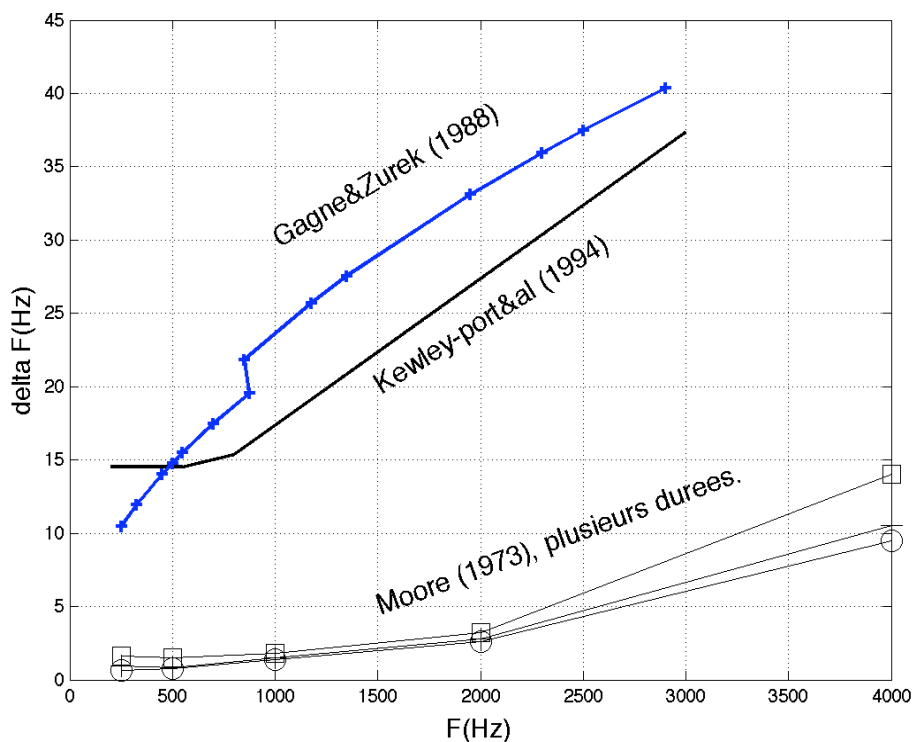


Figure 14. Comparaison des seuils de perception de la fréquence, (1) une étude psychoacoustique expérimentale qui a porté sur le son pur, Moore (1973), (2) un modèle théorique pour les formants, Gagne & Zurek (1988) et selon une étude linguistique expérimentale pour les formants, Kewley-Port et al. (1994).

du son pur est considéré comme le seuil le plus bas dans la performance du système auditif, ce qui est prédit par le modèle théorique de Gange & Zurek (1988), où le seuil tend vers 0 quand BW tend vers 0. Cette réalité peut nous servir de référence, c'est-à-dire que nous ne devrions pas trouver, pour des signaux plus complexes que le son pur, des seuils dont les valeurs sont inférieures ou égales aux valeurs des seuils obtenues pour des sons purs; d'où l'intérêt de comparer les résultats de Kewley-Port & Watson (1994) avec les données de Moore (1973).

Une simple comparaison, donnée dans la figure 14, montre que l'allure de la courbe représentative est la même pour le son pur et pour les formants, les deux courbes sont construites par des fonctions affines par morceaux (*piecewise*). Cependant le seuil pour les formants est de 8 à 10 fois plus élevé que le seuil d'un son pur. De fait cette différence est moins importante que ce que nous aurions pu penser, que ce soit au niveau quantitatif (seulement 8 fois), ou en ce qui concerne l'allure des courbes.

Ce que Kewley-Port & Watson (1994) et Gagne & Zurek (1988) nous fournissent dans leurs études sont les premiers résultats concrets qui intéressent directement notre problématique concernant l'opposition entre voyelles longues et brèves, puisqu'ils permettent de savoir à partir de quel changement dans le F1 et dans le F2 une discrimination auditive devient possible. Mais sans toutefois élaborer un calcul au niveau de la voyelle, un calcul qui intègre le seuil pour F1 et pour F2.

Ces résultats restent cependant des premiers pas encore trop élémentaires pour apporter une réponse à notre questionnement, ces études s'étant limitées à la tâche de la discrimination de formants isolés sans contexte. Par ailleurs, les conditions de l'émission et de l'identification des stimuli dans ces études étaient des conditions idéales (sans perturbation). Réellement élémentaires, parce que dans la parole la tâche ne se limite pas à la simple distinction entre 2 réalisations acoustiques, mais il faut que ce contraste soit d'une étendue suffisante pour qu'il puisse se maintenir sous des conditions d'émission que connaît la parole naturelle. D'où l'intérêt des études qui ont travaillé sur ces seuils des formants en tant que composantes intégrées de la voyelle, dans de telles conditions.

I.1.3. Voyelle

Quand nous nous intéressons à un système vocalique donné, nous le décrivons traditionnellement en termes des valeurs des formants de ses voyelles. Nous nous posons la question sur l'existence d'autres facteurs, à part les formants, qui seraient responsables de

l'identification de la voyelle, parce que le seuil doit être déterminé à la lumière des indices pertinents pour la perception de la voyelle.

Quels sont les facteurs qui influencent le seuil?

I.1.3.1. Rôle des formants

Nous voudrions examiner ici une étude capitale dans la littérature de la perception des voyelles, et qui montre la place fondamentale qu'occupent les formants dans l'identification des voyelles.

En effet, des évaluations ont été effectuées par Klatt (1982c) sur la distance au sein d'une paire de voyelles statiques (monophongues) synthétiques où les stimuli contenaient des changements au niveau des formants mais aussi au niveau de certains indices acoustiques jugés peu pertinents (pente spectrale, amplitudes relatives des formants, le filtre passe-haut ou passe-bas). En effet, ces indices peuvent être utilisés pour évaluer la distance métrique spectrale par des méthodes techniques que nous ne discuterons pas ici, la question qui nous intéresse étant principalement de savoir quel est l'indice qui possède le poids perceptif majeur.

L'expérience de Klatt (1982c) a consisté en la synthèse d'une série de 66 voyelles acoustiquement très proches de timbre [æ], en ajoutant à la fois des harmoniques à des fréquences précises, avec des amplitudes et des relations de phase différentes. Les stimuli ont été présentés en ordre aléatoire et les sujets avaient pour tâche d'estimer la distance perceptive entre chacun des stimuli et un stimulus référence (une voyelle). La moyenne de la distance subjective était comparée aux petits changements sur la fréquence des formants, la largeur de leur bande, la pente spectrale et le filtre de la bande passante.

Concernant la distance phonétique perçue dans le voisinage du [æ], les jugements des distances en termes de signaux psychophysiques sont différents des jugements perceptifs subjectifs (phonétiques). Parmi toutes les manipulations psychophysiques effectuées sur les signaux, seuls les changements sur les formants avaient un effet radical sur la perception. Bien que la largeur de filtre et la pente spectrale changent, selon Klatt (1982c), substantiellement le spectre, ces changements sont cependant ignorés à un niveau perceptif.

Les mêmes manipulations ont été effectuées pour la voyelle [a]. Concernant la distance phonétique perçue dans le voisinage du [a], les résultats étaient très semblables à ceux de la voyelle [æ], les changements des valeurs formantiques influençant plus que tout autre changement la perception de la qualité de la voyelle, pente spectrale, filtre, amplitude, etc. Il

semble, par ailleurs, que le changement du F3 a peu d'importance pour les voyelles arrières.

Dans une 3^e expérience Klatt a généré un [æ] sans voisement, dont il a augmenté la largeur de la bande pour F1, pour simuler la perte causée par une ouverture de la glotte, phénomène qui touche les voyelles basses. Les résultats sont superficiellement les mêmes que pour un [æ] voisé, dans la mesure où c'était la modification des formants qui était suivi d'un basculement qualitativement drastique. Klatt nuance par ailleurs que les changements de F1 étaient moins importants, et qu'il n'y avait pas de doute que cela est dû en partie à l'élargissement de la bande de F1 ce qui le rendait moins intense. Les sujets percevaient cette voyelle en tant que consonne fricative [ʃ], pour d'autres ce n'était pas de l'anglais. Toujours dans le but de déterminer les indices acoustiques pertinents pour la perception des voyelles, Klatt a cherché à déterminer les raisons de la perception de la fricative. Le même stimulus était présenté à d'autres locuteurs qui le jugeaient plutôt voyelle ou plutôt fricative, Klatt a trouvé que le stimulus était perçu comme une fricative quand il y avait une concentration de l'énergie dans les hautes fréquences ou quand il y avait deux formants proches situés dans les hautes fréquences, alors que le stimulus qui était jugé en tant que voyelle avait moins d'énergie dans les hautes fréquences que la voyelle de référence. La pente spectrale a plus d'importance pour les non voisés que pour les voisés, Klatt pense que, peut-être, c'est à cause de la proximité des frontières entre la plutôt voyelle et la plutôt consonne.

Les résultats des trois expériences montrent que certains changements ne sont pas perceptibles. Dans les trois expériences c'est le changement des formants qui était le plus important au niveau de la perception. L'intensité du formant, elle aussi, jouait un rôle, puisque Klatt trouvait que lorsque un pic d'un formant n'était pas aussi intense que le pic d'un autre formant dans la bande critique du spectre, son changement avait moins d'effet sur la perception.

Les résultats de cette étude mettent bien en évidence le rôle majeur que les formants jouent dans la reconnaissance du timbre de la voyelle, et le rôle très secondaire, voire négligeable que jouent les autres indices. Ces résultats nous permettent, lors de notre recherche du seuil, de ne considérer qu'un seul facteur, les formants. Nous nous intéresserons alors aux distances entre voyelles dans l'espace acoustique défini en termes de F1 et de F2.

Dans une étude postérieure à celle de Klatt (1982c), Jacewicz (2005) s'est mis à tester la supposition classique selon laquelle les valeurs fréquentielles des formants sont le principal facteur dans la détermination de l'identité de la voyelle, alors que les autres facteurs, telles la largeur de la bande du formant, la pente spectrale ou l'amplitude des formants sont considérés

comme secondaires (Klatt, 1982). Le travail de Jacewicz (2005) cherchait à apporter plus de précision sur le rôle que jouaient les amplitudes des formants dans la perception des différences à l'intérieur d'une seule qualité vocalique et entre des qualités. Dans ce dessein, elle a examiné dans deux expériences la sensibilité des locuteurs à la perception de la variation de l'amplitude au sein d'une même qualité et entre deux qualités, ceci était testé pour les deux voyelles synthétiques [i] et [ɪ]. La première expérience testait le jugement de "Naturel vs. Pas naturel", la deuxième expérience testait l'effet de l'amplitude sur l'identité de la voyelle. L'amplitude relative de F2 et F4 ont été modifiées progressivement le long du continuum acoustique, ces deux voyelles ont été créées conformément aux valeurs trouvées dans la parole naturelle.

Dans la première expérience, les deux voyelles synthétisées [i] et [ɪ] étaient insérés dans des mots isolés de type CVC. Ci-dessous sont données les valeurs en dB des amplitudes des 4 premiers formants pour les 2 voyelles:

Voyelle	A1	A2	A3	A4
[i]	64.0	41.6	40.5	41.2
[ɪ]	66.4	52.3	43.4	36.3

Seules A2 et A4 pour les deux voyelles étaient manipulées, une amplitude à la fois. Les amplitudes changeaient en pas de 2 dB.

Les valeurs des formants (Hz) étaient synthétisées comme ci-dessous:

Voyelle	F1	F2	F3	F4
[i]	310	2030	2970	3400
[ɪ]	400	1780	2578	3400

Les résultats de la première expérience montrent que, pour les deux voyelles, les locuteurs étaient sensibles aux changements de l'amplitude A2. Ils étaient insensibles aux variations de A4 de [i] mais sensibles aux variations de A4 de [ɪ]. Ces résultats, soutenus par des tests statistiques (ANOVA), montrant que, à part pour le F4 de [i], les locuteurs sont sensibles aux variations de l'amplitude des formants.

La deuxième expérience testait l'influence qu'exerçait la variation de l'amplitude sur l'identification (identité) de la voyelle. Tout est resté inchangé par rapport à la première expérience, seule différence était la question posée aux sujets de l'expérience: est-ce [i] ou [ɪ],

Comme lors de l'expérience 1, la variation de A4 n'avait pas d'effet sur l'identification du [i], mais avait de l'effet quant au jugement relatif à [ɪ], lequel tend à être perçu comme un

[i] quand A4 est élevée et comme [ɪ] quand elle est basse. les ANOVA montrent que pour [i] l'effet de A2 est significatif mais pas celui de A4. Pour [ɪ], il y a eu un effet significatif, pour les deux A2 et A4.

Selon Jacewicz, ces résultats montrent que la variation des patrons des amplitudes des formants affecte le jugement des locuteurs quant à la question naturelle ou non, et quant à la question de l'identité aussi. L'auteur pense que cela signifie que l'information acoustique présente dans l'amplitude du formant joue un rôle dans l'identité de la voyelle. Pour nous, le rôle que joue les amplitudes, quoique de telles variations ont été contrôlées sous des conditions idéales de laboratoire, a échoué à être systématique, et pour la question du Naturel et pour la question de l'identité. Finalement, l'information véhiculée par les amplitudes des formants, comme le précisait Klatt (1982c), demeure secondaire.

I.1.3.2. Seuil relatif ou seuil absolu?

Nous pensons que le seuil ne doit pas être une valeur relative, mais une valeur absolue, parce que nous croyons que l'espace vocalique n'est pas un espace uniforme. Donc l'idée de Flanagan, ou plus récemment, Kewley-Port *et al.* (pour F2), d'exprimer le seuil en % de la valeur du formant ne peut pas être valide pour toutes les voyelles, abstraction faite du timbre de la voyelle. Nous proposons deux raisons:

1) l'espace acoustique des voyelles n'est pas uniforme, les voyelles n'ayant pas la même prégnance perceptive selon leur positionnement dans le triangle vocalique défini en termes de F1 et de F2.

2) Une modification en pourcentage des valeurs initiales n'a pas le même impact perceptif selon la voyelle.

I.1.3.2.1. La non-uniformité de l'espace vocalique

Nous discutons ici des études qui ont rapporté des résultats qui montrent que l'espace acoustique perceptif des voyelles n'est pas un espace homogène, ce qui soutient notre idée, selon laquelle le seuil doit avoir une valeur absolue et non pas relative.

Les études qui révèlent l'asymétrie des relations perceptives dans différents domaines, comme la couleur, l'orientation des lignes, les nombres et les figures géométriques (cf. Tversky & Gati, 1978), suggèrent fortement que les asymétries perceptives pourraient fournir une clé importante pour comprendre la nature et la structure de l'espace perceptif. Nous

pensons que les asymétries perceptives sont un indice fort de la non-uniformité de l'espace acoustique des voyelles.

Dans une revue de questions sur la catégorisation, Medin & Barsalou (1987) rapportent que l'asymétrie se manifeste quand la frontière catégorielle entre deux stimuli change en fonction de l'ordre de la présentation. C'est exactement ce que Polka & Bohn (2003) ont remarqué. Ils soulignaient que le fait qu'un changement dans la qualité de la voyelle se discerne plus facilement quand le passage d'une qualité à l'autre se produit dans un sens plutôt que dans l'autre sens, est un signe de l'asymétrie dans la perception des voyelles.

Polka & Bohn (2003) ont essayé dans leur étude de trouver des explications à ce phénomène qui s'observait déjà chez le nouveau-né. Ces derniers pensent que les relations asymétriques sont en contradiction avec la supposition classique selon laquelle l'espace perceptif dispose d'une symétrie régie par le calcul euclidien (Schwartz *et al.*, 1997b). Ils pensent que si l'espace perceptif est supposé euclidien, et que cette supposition continue à être vraie, c'est tout simplement parce qu'il n'y a pas eu d'étude perceptive qui montre le contraire.

Ce questionnement sur les asymétries perceptives venait des résultats d'une étude faite par Polka & Werker (1994) où les deux chercheurs examinaient l'effet de l'âge sur la perception des voyelles chez le nouveau-né. Dans cette étude-là, ils testaient la discrimination du contraste [u - y] et celui du [ʊ - ʏ] en allemand par des enfants anglophones, sur deux groupes d'âge de 6-8 et 10-12 mois. Le protocole utilisait le test du "changement/pas de changement". La présentation était contrebalancée, c'est-à-dire que la moitié de chaque groupe était testée avec l'un des deux sens de changement. Pour le changement du [u] versus [y], la moitié a assisté à une variation qui partait de [u] vers [y] et l'autre moitié a assisté à un changement qui partait de [y] vers [u]. Les résultats ont mis en évidence un effet robuste de l'ordre de la présentation, en ce sens que les deux groupes avaient des performances différentes selon le sens du changement, et pour les deux continuum, [y] à [u] et pour [ʏ] à [ʊ].

Kuhl (1991) rapporte que les asymétries dues à l'ordre de présentation sont en harmonie avec la notion de l'effet de l'aimant perceptif. Selon Kuhl, la voyelle étrangère qui ressemble le plus à la voyelle native semble agir comme un aimant perceptif pour les autres voyelles étrangères du même système. Une telle assertion était émise suite aux résultats de l'identification des voyelles de l'allemand par des anglais. L'expérience de Kuhl a consisté en la présentation des voyelles [u] et [y] de l'allemand à des locuteurs anglophones, ces deux

voyelles ont été perçues, respectivement comme une bonne et mauvaise réalisation de la voyelle [u] de l'anglais. Ce fait montre que chacune des voyelles de la paire est perçue comme bonne versus mauvaise réalisation appartenant cependant à (ou fusionnant dans) une même qualité vocalique de l'anglais, alors que le [y] aurait pu être identifié comme une voyelle étrangère au système de l'anglais plutôt qu'une mauvaise réalisation du [u]. De même les voyelles [ʊ] - [ʏ] de l'allemand sont perçues respectivement comme bonne et mauvaise voyelle [ʊ] de l'anglais.

Dans une étude ultérieure à celle de Kuhl, Polka & Bohn (1996), avec un test lié directement à l'hypothèse qui stipule que les asymétries de la perception des voyelles chez le nouveau-né sont dues au fait qu'il y a des voyelles qui sont relativement plus familières que certaines d'autres, ces voyelles ont un effet d'aimant perceptif. Dans cette étude, ils ont examiné la discriminabilité de l'opposition des voyelles [ɛ - æ] de l'anglais (une opposition qui n'existe pas en allemand), et la discriminabilité de l'opposition des voyelles [u - y] de l'anglais (une opposition qui n'existe pas en allemand) par 2 groupes d'enfants allemands et anglais de 6 à 8 mois et de 10 à 12 mois, les nouveaux-nés anglais écoutaient les voyelles de l'allemand et inversement. Le but était d'observer les asymétries perceptives dans la discrimination des contrastes auprès de non natifs (parce qu'ils sont trop petits pour les considérer en tant que des locuteurs qui ont acquis le système phonologique de leur milieu) qui écoutent des contrastes qui n'existent pas dans leur système phonologique.

Pour [u - y] de l'allemand, les nouveaux-nés anglais et allemands dans les deux groupes d'âge montraient la même asymétrie perceptive, avec un changement de [y] à [u] plus discriminable que de [u] à [y]. Pour le contraste en anglais de [æ] - [ɛ], la discrimination était nettement plus facile lorsque le changement se faisait de [ɛ] à [æ] que dans le cas contraire, et ceci était indépendant de la langue et de l'âge des enfants.

Contrairement à ce qui était prévu en ce qui concerne l'effet d'aimant perceptif, le [æ] de l'anglais (homologue en allemand) agit comme une voyelle de référence pour les enfants allemands et anglais à la fois. Plutôt que soutenir l'hypothèse des voyelles familières, les asymétries découvertes dans cette comparaison directe entre enfants anglais *vs.* enfants allemands révéleraient une tendance linguistique universelle.

Les 3 contrastes pour lesquels Polka & Bohn (1996) ont trouvé une asymétrie, [u - y], [ɛ - æ] et [ʊ - ʏ] n'étaient pas les seuls qui présentaient des statuts: "référence"/"satellite". Les auteurs rapportent que plusieurs explications ont tenté de donner une réponse à l'existence de cette asymétrie, mais que ce n'était ni le statut phonologique de la voyelle, ni sa fréquence dans les langues du monde, ni son intensité, ni sa durée ni l'ordre de changement qui pouvaient expliquer cet effet d'aimant perceptif.

Le fait est que la discrimination des voyelles est plus facile pour les nouveaux-nés quand elles sont présentées comme un changement de la moins périphérique vers la plus périphérique. Cette donnée met en évidence la réalité d'une tendance qui pourrait être universelle; les voyelles qui sont relativement plus périphériques serviraient comme des voyelles de référence quand il s'agit d'un contraste entre deux voyelles qui n'ont pas le même degré de périphéricité.

Polka & Bohn (2003) ont testé l'hypothèse de la périphéricité par une autre expérience. Le premier but de cette étude était d'examiner les rôles de la cible spectrale, de la dynamique spectrale et de la durée dans la discrimination de beaucoup de voyelles natives présentées dans le contexte [dVt] auprès d'enfants allemands. Ils se sont basés sur la discrimination de 3 contrastes en allemand [i - e], [e - ɪ] et [o - ʊ], via la procédure d'habituation à tourner la tête. Chaque enfant, pour pouvoir participer à l'expérience, devait pouvoir arriver à un certain niveau de réponses correctes au préalable. Il répondait à 20 stimuli, 10 de chaque sens de changement. Le critère était: à combien d'enfants faut-il présenter le contraste pour en trouver 10 qui réagissent au changement? Plus le nombre d'enfants testés est grand et moins le contraste est marqué.

Exemple: si pour un contraste A il faut tester 90 enfants pour en retenir 10 seulement, et que pour un contraste B il faut tester 25 pour en retenir 10, c'est que le contraste B est plus marqué que le contraste A.

Partant de l'hypothèse de la périphéricité, les auteurs s'attendaient à ce que le [i] soit la référence dans l'opposition [i - e], que [e] serait la référence dans [e - ɪ] et [o] dans [o - ʊ].

Alors que les différences se sont produites comme prévu, les analyses statistiques ont montré un effet significatif de la direction du changement pour le couple [e - ɪ], mais le chi² ne donnait pas d'effet significatif pour les deux autres couples. Enfin, plus d'enfants ont été testés pour [i - e] (25 pour en retenir 10) que pour [e - ɪ] (18 pour en retenir 10). Plus d'enfants ont été testés pour [o - ʊ] (21 pour en retenir 10) que pour [ʊ - o] (19 pour en retenir

10). Les auteurs notent que les différences, au niveau des formants, pour les contrastes [o - ʊ] et [i - e], sont plus petites comparées avec l'autre contraste qui a montré une nette asymétrie.

Les auteurs ont eu recours à la littérature de l'asymétrie pour puiser quelques exemples qui vont dans le sens de leurs résultats. En effet, ils rapportent que l'expérience effectuée par Swoboda *et al.* (1978), sur des voyelles synthétiques de 60 ms, le long du continuum [i - ɪ] avec des enfants anglais de 8 semaines, a montré que le changement du stimulus en direction de la voyelle périphérique [i] est mieux discriminé que le changement vers la voyelle moins périphérique [ɪ]. Dejjardin & Trainor (1998) donnent des résultats similaires concernant l'asymétrie dans la discrimination des voyelles synthétiques [i - ɪ]. Polka & Bohn affirment que cette asymétrie était observée pour des voyelles de la parole naturelle. Ils rapportent, en outre, que Best *et al.* (1997) trouvent de l'asymétrie pour le couple de voyelles naturelles [y - ʉ] en Norvégien.

Le travail de Repp *et al.* (1979), qui a porté sur la série [i - I - ɛ], a montré que les asymétries des voyelles, selon leurs directions, apparaissent chez des adultes qui percevaient leur langue maternelle, mais pas chez des enfants. Il semble que les nouveaux-nés sont prédisposés à réagir différemment aux voyelles qui occupent différentes positions dans l'espace articulatoire/acoustique (défini en termes de F1 et de F2) et que le sens de l'asymétrie (vers la périphérie), montre une tendance universelle.

Les résultats des précédentes observations avancent plusieurs pistes pour permettre de comprendre entièrement cette tendance dans la perception des voyelles chez les nouveaux-nés. D'abord, ces asymétries sont-elles des tendances inhérentes à notre mécanisme auditif? Polka et Bohn pensent qu'il est possible que les positions les plus évidentes soient les positions extrêmes de l'espace vocalique. Si cela est vrai, les effets relatifs à la direction peuvent être expliqués en partie.

Les auteurs pensent que l'asymétrie dans la perception des voyelles pourrait refléter une tendance du système de perception des sons du langage chez l'homme. Il est en effet possible que cette tendance soit un résultat de l'exposition à la parole, en l'occurrence aux voyelles, parce que les voyelles extrêmes se présentent avec peu, voire pas, de chevauchement. Leurs ellipses présentent moins de chevauchement par rapport à celles des voyelles centrales: elles sont relativement moins confondues. Les extrêmes délimitent l'espace vocalique, et comme elles sont perceptivement plus évidentes et relativement plus stables, toutes ces qualités acoustiques les rendent des voyelles particulièrement prégnantes.

L'asymétrie dans la perception des voyelles est une tendance linguistique générale qui apparaît tôt après la naissance en réponse aux aspects globaux de la parole, la perception des voyelles est structurée ainsi de manière innée, ce que Polka & Bohn soutiennent. Ils exposent un ensemble de raisons que nous ne discuterons pas ici. Polka & Bohn rapportent que les asymétries dans la perception des voyelles de l'anglais, par exemple: [y - u]; [Y - Ū]; [ɛ- æ]; [i - y] ou [I - i] ne sont pas expliquées par le fait que ces voyelles sont de la langue native (la voyelle référence, parfois, n'existe pas dans l'environnement linguistique du nouveau-né), ni par le fait que la voyelle soit universellement favorisée, ni par des effets de l'intensité ou de la durée.

Le but principal de cette étude exhaustive était de mettre en évidence les asymétries perceptives qui soulignent une tendance dans la perception des voyelles, une tendance qui est observée dans la discrimination chez les enfants et qui apparaît chez l'adulte dans certains contextes. Les données convergent vers une conclusion globale que les périphériques ont des propriétés uniques dans la perception de la parole, et que finalement l'espace vocalique n'est pas uniforme.

C'est ce dernier point qui affirme, abstraction faite des raisons qui pourraient en être à l'origine, la non-uniformité de l'espace vocalique qui est l'objectif final de notre examen des études qui ont porté sur l'asymétrie. Cette donnée nous permet de dire que: puisqu'il y a des voyelles plus prégnantes que d'autres, un même pourcentage n'aura pas la même valeur selon qu'il est calculé à partir d'une voyelle peu ou très prégnante. De même que si l'on prélève un même volume de deux matières qui n'ont pas la masse volumique, ce même volume n'aura pas le même poids, parce que son poids dépendra de la densité de la matière où il était prélevé.

1.1.3.2.2. Le seuil relatif

Nous abordons maintenant le deuxième point qui nous permettra d'avancer que le seuil devrait avoir une valeur absolue plutôt qu'une valeur relative.

Nous avons modifié les valeurs des 2 premiers formants des 3 voyelles cardinales de l'espace vocalique: [i, a, u] en 4 pas, et nous avons observé leurs déplacements en fonction du pourcentage affecté. Nous avons appliqué les pourcentages ci-dessous:

1) F1*5%; F2*10%. Ce qui se lit: F1 augmente sa valeur de 5%, F2 augmente sa valeur de 10%. Les trois lignes ci-dessous se lisent de la même manière:

- 2) $F1*10\%$; $F2*10\%$.
- 3) $F1*20\%$; $F2*20\%$.
- 4) $F1*30\%$; $F2*30\%$.

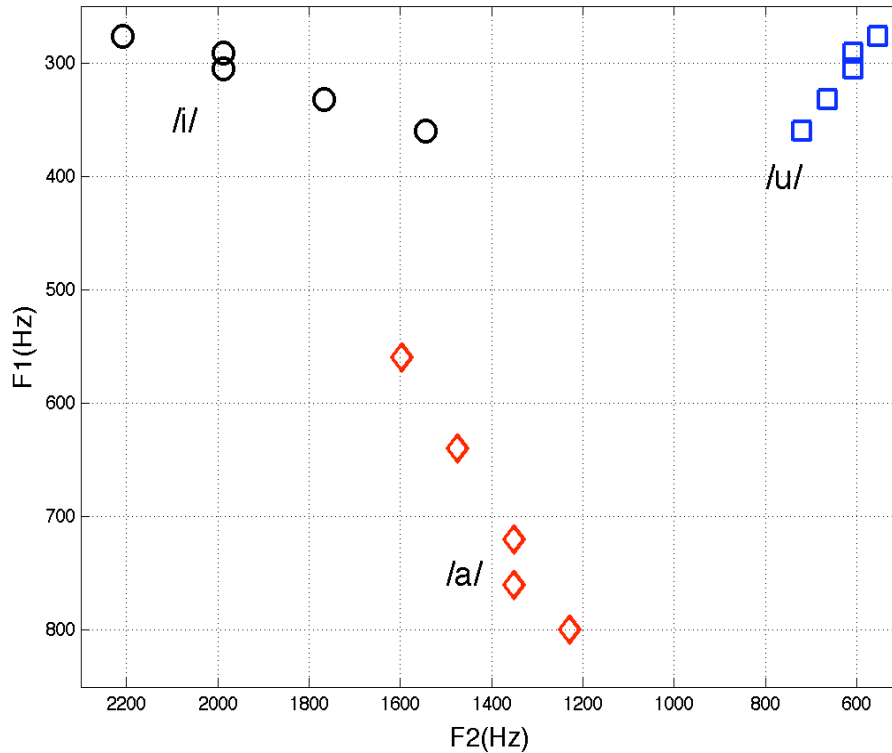


Figure 15. Changement des valeurs fréquentielles, en termes de F1 et F2 (en Hz), des voyelles [i, a et u] en cinq étapes.

Au total, cinq valeurs pour chaque voyelle (4 changements + la valeur de référence).

Cas de la voyelle [i]. Nous prenons comme valeurs de référence des valeurs prototypiques basées sur l'inventaire de la base de données UPSID pour les voyelles des langues du monde (idem pour les autres voyelles, cf. Schwartz *et al.*, 1997b). Les valeurs sont: $F1 = 277$ Hz, $F2 = 2208$ Hz, $F3 = 3079$ Hz et $F4 = 3560$ Hz. La modification se faisait de telle sorte que les voyelles se déplacent vers le centre de l'espace (dans le sens de l'*undershoot*). Pour le premier pas: $F1$ baisse de 5%, $F2$ augmente de 10%, $F3$ et $F4$ restent inchangés. Pour chacune des étapes et pour toutes les valeurs de référence après modification, voir annexe n° III.

La figure 15 montre le déplacement des voyelles dans l'espace vocalique. Ce déplacement est le résultat du changement qui a affecté les deux premiers formants seulement, puisque nous représentons l'espace F1-F2.

Nous remarquons que les voyelles ne se déplacent pas les unes autant que les autres dans les mêmes proportions de changement. La représentation graphique nous montre que le pourcentage n'est pas la stratégie à adopter pour déterminer la distance nécessaire, parce qu'il n'a pas le même effet selon la voyelle.

Nous avons voulu savoir si l'intégration des autres formants (F3 et F4) pouvait rendre uniformes les déplacements des voyelles dans l'espace acoustique en fonction du pourcentage affecté, pour cela nous avons eu recours à la représentation en termes de F1 et de F'2 (le deuxième formant effectif⁵) qui intègre les trois derniers formants. Le F'2 étant calculé plus traditionnellement en Bark, nous avons converti les mêmes valeurs des voyelles de la figure 15 à l'échelle des Bark. Les pourcentages qui étaient affectés étaient les suivant:

- 1) F1*5%; F2*10%; F3*10% et F4 inchangé,
- 2) F1*5%; F2*10%; F3*10% et F4*10%,
- 3) F1*10%; F2*10%; F3*10% et F4*10%,
- 4) F1*20%; F2*20%; F3*20%; F4*20%,
- 5) F1*30%; F2*30%; F3*30% et F4*30%.

Comme nous avons procédé pour les valeurs en Hz, nous avons procédé avec nos valeurs en Bark. Pour la voyelle [i] par exemple, nous prenons les valeurs prototypiques de référence, et nous modifions les formants de telle sorte qu'ils se déplacent vers le centre de l'espace. La figure 16 rend compte de ces changements par rapport à des valeurs de référence qui constituent les points extrêmes de l'espace. Ici, le déplacement des voyelles dans l'espace vocalique est le résultat du changement qui a affecté les quatre premiers formants, puisque F'2 intègre les trois derniers.

⁵ F'2: Deuxième formant effectif, cf. Schwartz et al. (1997b). Voir annexe II pour l'algorithme qui calcule le F'2 à partir des 4 premiers formants.

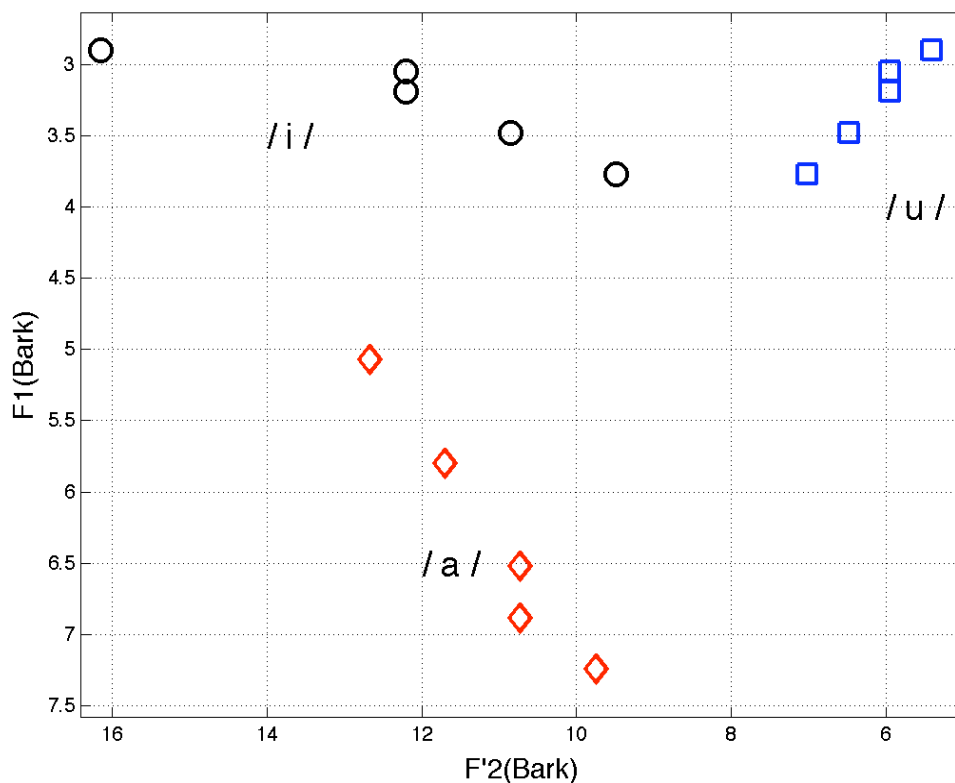


Figure 16. Changement des valeurs fréquentielles, en termes de F1 et F'2 (en Bark), des voyelles /i/, /a/ et /u/ en cinq étapes.

Conclusion: l'intégration de F3 et de F4 n'a pas pu rendre le déplacement plus uniforme, nous remarquons toujours que le [a] se déplace plus que toute autre voyelle sur l'axe de F1 et que le [i] se déplace le plus sur l'axe de F'2. Cela plaide en faveur de notre thèse selon laquelle le seuil de la distance ne doit pas être donné en termes de valeur relative à une valeur de référence, parce qu'elle ne peut être appliquée à toutes les voyelles. La valeur relative n'est vraie que pour une valeur de référence et une seule; c'est-à-dire que 10% appliqué sur les valeurs initiales du [a] ne produit pas la même distance quand il est appliqué sur les valeurs initiales du [i].

1.1.3.2.3. Le seuil absolu

Les études précédentes nous ont bien donné des informations sur les seuils pour des formants isolés (Kewley-Port & Watson, 1994 et Gagne & Zurek, 1988), sur les seuils pour les formants insérés dans des voyelles (Flanagan, 1955), sur les seuils pour des formants insérés dans des voyelles qui sont à leur tour insérées dans des syllabes (Mermelstein, 1978). La syllabe est le plus long contexte étudié jusqu'à présent. Mais par ailleurs, aucune étude ne nous montre comment procéder pour mesurer les distances entre les voyelles dans l'espace

acoustique, comment intégrer F1, F2 et éventuellement F3 et F4 en une seule composante, c'est-à-dire qu'aucune étude jusqu'à maintenant ne nous permet finalement de comparer directement des distances, entre voyelles, mesurées dans un système phonologique de voyelles avec des distances entre des voyelles mesurées dans un autre système. Autrement dit aucune étude ne nous fournit pour l'instant de réponse à la question:

Quelle est la distance minimale entre deux voyelles dans l'espace vocalique qui permet leur opposition qualitative robuste?

Dans ce sens, nous discutons amplement l'étude de Kewley-Port & Zheng (1999), parce qu'elle apporte des réponses d'une importance majeure quant à notre questionnement concernant le seuil.

Les valeurs des voyelles synthétisées qui ont fait l'objet de l'expérience sont montrées dans le tableau 5. La figure 17 donne la représentation graphique qui nous permet de mieux visualiser les emplacements de ces voyelles.

Voyelle	F0(baisse linéaire)	Valeurs standards				Variation	
		F1	F2	F3	F4	F1	F2
[i]	220 devient 180	450	2300	3000	4000	455-522	2315-2513
[ɛ]	220 devient 180	600	2200	3000	4000	605-671	2210-2341
[ʌ]	220 devient 180	700	1400	2600	4000	705-771	1410-1542
[æ]	220 devient 180	1000	1950	3000	4000	1007-1107	1960-2091
LB ⁶		70	90	170	200		

Tableau 5. Ce tableau montre plage de la variation de la F0, les valeurs des 4 premiers formants des 4 voyelles de l'étude, plage de la variation de F1 et de F2 où le seuil est susceptible de se produire.

Dans le souci de simuler au mieux la parole naturelle et d'obtenir des résultats plus fiables que les résultats obtenus dans des conditions idéales, Kewley-Port & Zheng (1999) ont manipulé trois conditions dans cette étude:

1) l'entraînement des sujets: les auteurs de l'étude ont constaté dans une expérience secondaire que l'augmentation du temps de l'entraînement de 7 heures à 15,5 heures a significativement amélioré les performances des locuteurs, et ceci en baissant la valeur du seuil et en réduisant la variance en ce qui concerne la performance des sujets. Selon le protocole expérimental, les sujets choisis ont suivi un entraînement, mais ceux qui ont

⁶ La largeur de la bande passante en unité SI.

participé à l'expérience avaient une performance moyenne, parce que ceux qui avaient une très bonne performance et ceux qui avaient une très mauvaise performance étaient exclus.

2) le contexte phonétique: le corpus de cette étude était constitué de 4 voyelles synthétisées d'après la voix d'une femme adulte, lesquelles étaient présentées (1) en isolation, le formant qui est l'objet de la mesure est inséré dans une voyelle, mais la voyelle est isolée de tout contexte (2) en syllabe [bVd], les 4 voyelles choisies produisent alors les quatre mots suivants: "bid", "bed", "bad" et "bud" (3) en une phrase de 3 mots "say [bVd] again" ou (4) en une phrase de 9 mots "Tell Sam to say [bVd] again on the tape". Les recherches sur l'identité de la voyelle en anglais ont bien montré que la durée de la voyelle et les mouvements de ses formants étaient deux indices fondamentaux, mais le premier indice reste l'indice spectral (Nearey, 1989).

L'analyse de la Variance (Anova) pour 2 facteurs (contexte phonétique et formant) a montré qu'il y avait un effet significatif du contexte phonétique, de la valeur du formant et de l'interaction des deux. Comme prévu et comme l'ont bien montré les études précédentes (discutées plus

haut), la valeur du seuil augmentait avec l'augmentation de la valeur du formant. L'effet du contexte phonétique n'était pas simple, parce qu'il n'était pas le même selon où l'endroit où le stimulus se trouvait sur l'échelle des fréquences. En effet, il était tout simplement inexistant pour F1, c'est-à-dire que le F1 n'était pas affecté par la taille du contexte phonétique. Pour F2, il n'y avait pas de différence de seuil entre la phrase de 3 mots et la phrase de 9 mots; alors que cette différence de seuil était très importante pour F2 entre celui de la syllabe d'un côté et celui de la phrase (de 3 ou de 9 mots) de l'autre, cette différence augmentant avec la fréquence. L'Anova réalisée pour les trois facteurs (contexte phonétique, voyelle et formant), montrait que seul l'effet de la voyelle n'était pas significatif, c'est-à-dire que les voyelles ne

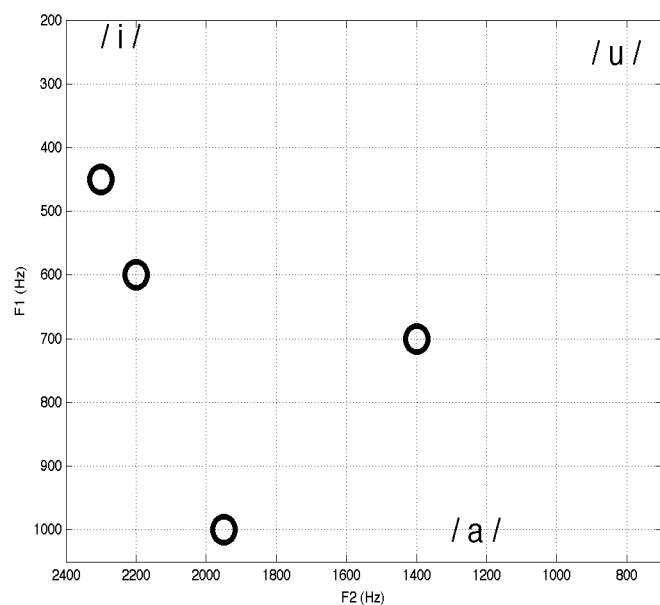


Figure 17. Les 4 voyelles de l'expérience de Kewley-Port & Zheng (1999), chaque cercle représente une voyelle.

réagissent pas de manière différente les unes par rapport aux autres selon le contexte et selon les valeurs de leurs formants

3) la charge cognitive: il y avait 3 niveaux de difficulté selon le nombre de choix dans la phrase de 9 mots: 2 choix alternatifs forcés, 4 choix alternatifs forcés et 16 choix alternatifs forcés. Les sujets entraînés et ceux qui sont novices ont trouvé l'identification avec 16 choix très difficiles. Un test était effectué pour déterminer s'il y avait une différence entre les 4 niveaux de difficulté: 1, 2, 8 ou 16 interprétations possibles pour la phrase de 9 mots. L'analyse de la variance montre que l'identification des 9 unités de la phrase n'était pas affectée par le niveau de complexité de la phrase. Cette expérience met en évidence le fait que la discrimination du changement dans la qualité d'une voyelle insérée dans une phrase porteuse n'est pas affectée par la compréhension du sens de la phrase, parce que quel que soit le nombre d'interprétations possibles d'une phrase la valeur du seuil reste la même. Ce qui change la valeur du seuil, c'est la longueur de la phrase et non pas sa compréhension.

Le seuil était fixé à 71% de réponses correctes (selon Levitt, 1971). La longueur de la syllabe était de 250 ms, quand elle était seule, et de 220 quand elle était insérée dans une phrase de 3 ou de 9 mots. Dans cette expérience (1), l'identification n'a porté que sur les formants, sans contexte acoustique adjacent, c'est-à-dire que la charge cognitive qui accompagnait l'identification du changement était minimale.

Les résultats montrent que l'effet du contexte, l'effet du formant et celui de leur interaction, sont tous significatifs. Comme prévu, le seuil augmente en augmentant la fréquence du formant. Un test statistique (Tukey HSD) confirme que la valeur moyenne du seuil (tous formants et toutes voyelles confondus) quand le contexte est représenté par la voyelle (contexte-voyelle) est **48,9 Hz**, il est significativement inférieur à celui du contexte-syllabe où il est de **68,4 Hz**, à celui de contexte-phrase 3 mots, où il est de **75,8 Hz** et à celui de contexte-phrase 9 mots où il est de **73,1 Hz**. La moyenne du contexte-syllabe est à son tour inférieure à celles du contexte-phrase 3 mots et à la moyenne du contexte-phrase 9 mots. Les moyennes des contexte-phrase 3 et 9 mots ne sont pas différentes entre elles. Cela peut être résumé comme ceci:

"seuil contexte-voyelle" < " seuil contexte-syllabe" < " seuil contexte-phrase 3 mots" = " seuil contexte-phrase 9 mots".

L'un des principaux objectifs de cette expérience était de déterminer l'effet de la taille du contexte phonétique sur la détection des seuils des formants. Le seuil pour les voyelles

isolées est plus fin (valeur basse) que pour tout autre contexte phonétique. L'effet du contexte est plus fort en passant de la voyelle à la syllabe (36% de réduction dans la résolution de la fréquence du formant). Puis en passant de la syllabe à la phrase de 3 ou de 9 mots (8% de réduction).

Or pour déterminer le seuil sur la totalité de la voyelle (F1 et de F2 à la fois), il faut éliminer la différence entre les valeurs des seuils entre les 2 formants, une différence due au fait que le F2 a une fréquence plus élevée que celle du F1 et par conséquent un seuil plus élevé. Les auteurs de l'étude ont testé plusieurs méthodes pour pouvoir considérer le seuil de F1 et celui de F2 comme provenant de la mesure d'une même composante acoustique. Quand nous représentons le seuil en fonction de la fréquence, nous observons une courbe qui monte avec la fréquence (voir figure 18). Sur le plan de la représentation graphique, la méthode recherchée doit permettre d'aplatir la courbe du seuil en fonction de la fréquence, c'est ainsi qu'il sera possible de traiter les seuils de F1 et de F2 à titre égal.

Parmi les choix qui se présentaient, Kewley-Port & Zheng (1999) avançaient que la conversion de la fréquence de l'échelle des Hz à l'échelle des Bark, était une méthode particulièrement efficace. L'échelle des Bark rend compte de la faible importance perceptive des composantes hautes fréquences qui sont plus facilement masquées. Pour des exemples sur la procédure du calcul du seuil en Bark, voir tableaux "*Exemples du calcul du seuil en Bark*" disponibles en annexe V.

En fait, il est établi que l'oreille humaine regroupe des excitations sonores ayant des fréquences voisines dans des bandes fréquentielles, appelées des "bandes critiques", où la largeur est déterminée par la plage de fréquences qu'une fréquence de 60 dB peut masquer. Le champ auditif (20 à 20000 Hz) est divisé traditionnellement en 24 bandes critiques (annexe n° IV), où la largeur de chaque bande est de 1 Bark, alors que la largeur en Hz dépend du numéro de la bande critique; par exemple, la largeur de la sixième bande est 120 Hz, la largeur de la vingtième est 1100 Hz, alors que chacune des largeurs vaut 1 Bark, elles sont par conséquent égales entre elles. Effectivement, quelle que soit la valeur de la fréquence marquante de départ, la largeur de la bande sera toujours, approximativement, égale à 1 Bark.

En convertissant les valeurs des résultats de l'étude de Kewley-Port & Watson (1994) de l'échelle des Hz (figure 18), à l'échelle des Bark (figure 19), Kewley-Port & Zheng (1999) montrent que les valeurs élevées du seuil du F2 ont eu des valeurs proches de celles du seuil

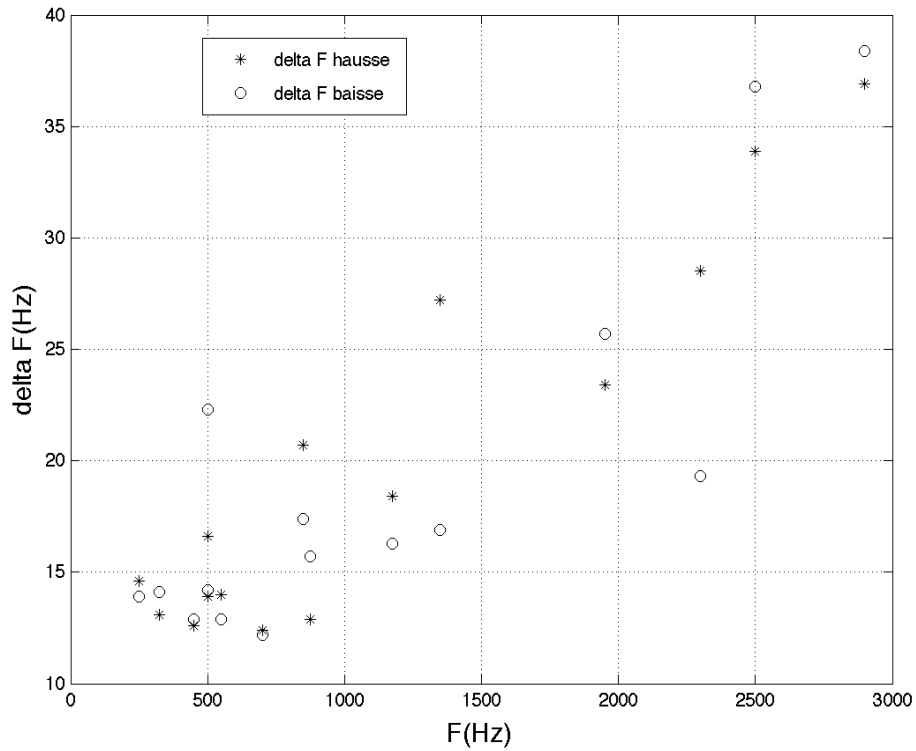


Figure 18. Seuil différentiel de F1 et de F2 pour 10 voyelles de l'anglais, Kewley-Port & Watson (1994). Les valeurs sont Hz. Les formants sont isolés (sans contexte).

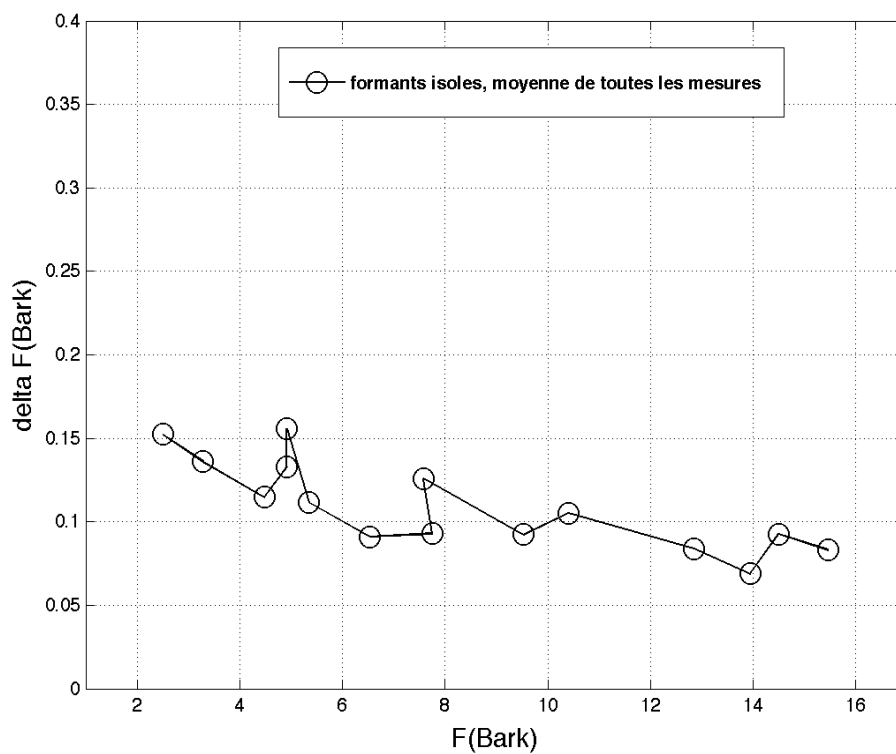


Figure 19. Conversion en Bark des valeurs du seuil différentiel de F1 et de F2 pour 10 voyelles de l'anglais de l'étude de Kewley-Port & Watson (1994). Les formants sont isolés (sans contexte), valeurs obtenues lors de l'étude de Kewley-Port & Watson (1994).

du F1. Le procédé était le suivant: à la valeur fréquentielle du formant en Hz "F", était ajoutée la valeur du Δf , le résultat de la somme était converti en Bark (selon l'équation (6) de Traunmüller, 1990)⁷, la valeur de "F" en Bark était ensuite soustraite à ce résultat. Ce deuxième résultat est le Δf en Bark, lequel a presque la même valeur quelle que soit la valeur de la fréquence du formant.

La disparition, relative, des différences entre les seuils des formants permet d'obtenir une moyenne qui sera relative à la totalité de la voyelle. Ce procédé permet réellement de considérer la voyelle comme une seule entité et de comparer non seulement l'effet de la taille du contexte phonétique sur la perception du seuil, mais de comparer aussi des distances entre voyelles dans un espace vocalique, ou encore comparer des distances entre deux espaces vocaliques. Cette découverte constitue parfaitement une réponse à notre question.

La valeur moyenne du seuil est de **0,28 Bark**, toutes conditions expérimentales de la parole naturelle confondues: (i) avec ou sans reconnaissance du sens; (ii) avec changement de la taille du contexte phonétique autour du formant à discriminer; (iii) en variant le niveau de l'ambiguïté de la phrase, de faible à élevé, cette valeur est de **0,28 Bark**. C'est une moyenne⁸ de F1 et de F2 qui correspond à la distance minimale **prédite** qui devrait séparer deux voyelles de l'espace vocalique pour que chacune de ces deux voyelle puisse constituer une qualité indépendante.

Une investigation, qui cherchait à déterminer cette même distance par l'observation, a consisté en la création d'une base de donnée à partir de 16 locuteurs de trois études différentes qui mesuraient les valeurs formantiques des voyelles de l'anglais des Etats-Unis, Bradlow (1993) pour la région de New York, Hillenbrand *et al.* (1995) pour la région de Michigan et une troisième (non publié) faite par Kewley-Port et son équipe dans le laboratoire où les auteurs de l'étude effectuaient leur travail de recherche, dans l'Indiana. Cette étude-là a établi la mesure des distances entre les voyelles de l'espace vocalique. Elle a montré que la distance, **observée**, la plus petite qui existe était de **0,56 Bark**, laquelle est égale à 2 fois la distance prédite. La distance observée, bien que largement au-dessus de la distance prédite, reste plausible, parce que la distance prédite était une moyenne où les auteurs intégraient des valeurs théoriques pour le seuil (i.e. seuil pour voyelles isolées, seuil pour des syllabes, sujets entraînés à la discrimination perceptive, etc.). Nous remarquons aussi que cette valeur prédite

⁷ $Z = ((26.81 * f) / (1960 + f)) - 0.53$. (Z est en Bark et f est en Hz).

⁸ Toutes valeur en Bark; $\text{Seuil} = ((|F1V2 - F1V1|) + (|F2V2 - F2V1|)) / 2$.

du seuil se rapproche sensiblement de la valeur observée (tableau 6), quand le seuil est établi dans des conditions plus authentiques pour les situations de communication; en effet, le seuil est de 0,434 Bark pour des sujets normo-entendants, non entraînés, avec des contextes phonétiques de tailles importantes.

Contexte phonétique	Seuil (Bark)
Formant	0,109
Voyelle	0,242
Syllabe	0,330
Phrase (3 mots et plus)	0,357
Phrase (non entraînés)	0,434

Tableau 6. Il montre la valeur moyenne du seuil en fonction de la taille du contexte phonétique.

La représentation graphique de ces résultats est illustrée dans les figures 20 et 21, où nous présentons, comme nous avons fait avec les données de Kewley-Port & Watson (1994) pour les formants isolés, les valeurs en Hz dans une figure (figure 20), et les mêmes valeurs en Bark dans une autre figure (figure 21).

Nous remarquons que la conversion en Bark permet de traiter les valeurs du seuil comme provenant, dans certaines limites, du même formant, c'est-à-dire que nous pouvons faire abstraction de l'origine de la valeur. La valeur la plus vraisemblable à adopter nous semble être la valeur obtenue par observation, et cette valeur observée dans la parole naturelle

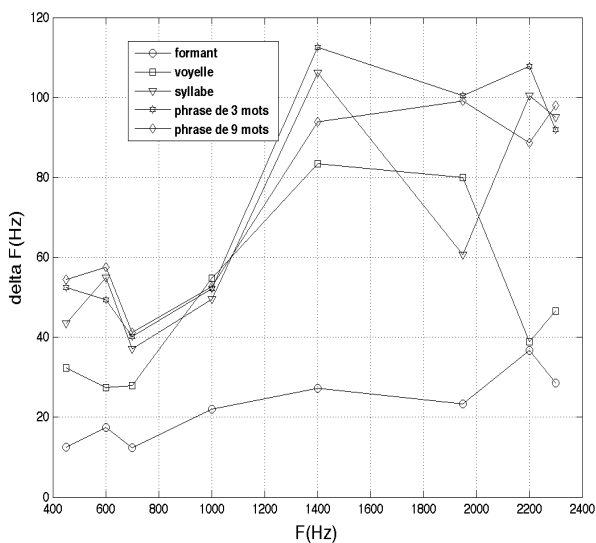


Figure 20. Seuil différentiel de F1 et de F2 pour 4 voyelles de l'anglais, pour cinq tailles de contexte, les valeurs sont en Hz. Kewley-Port & Zheng (1999).

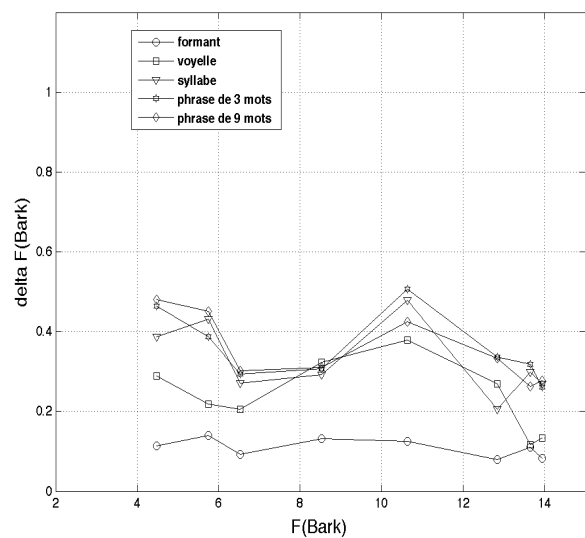


Figure 21. Seuil différentiel de F1 et de F2 pour 4 voyelles de l'anglais, pour cinq tailles de contexte, les valeurs sont en Bark. Kewley-Port & Zheng (1999).

est d'autant vraisemblable qu'elle se rapproche le plus de la valeur obtenue pour le contexte le plus long avec des sujets non entraînés. C'est la valeur de **0,56 Bark**

I.1.3.3. Un seuil à valeur universelle?

Suite aux résultats obtenus par l'étude de Kewley-Port & Zheng (1999) pour le système des voyelles de l'anglais américain, il y a une question qui s'impose. Nous voulons savoir si ces valeurs-là peuvent ou non être transposées sur d'autres systèmes, si nous pouvons les appliquer sur le système de l'arabe par exemple, sachant qu'il n'a pas autant de voyelles? Autrement dit, ces résultats sont-ils universels? Pour cela, nous avons besoin de savoir si le nombre de voyelles dans un système conditionne ou non leurs production et/ou leur perception. Nous pouvons trouver une réponse partielle dans l'étude menée par Meunier *et al.* (2003), où les auteurs cherchaient à savoir si le nombre de voyelles dans un système phonologique donné pouvait avoir de l'influence sur la capacité de production ou de perception des voyelles.

L'étude de Meunier *et al.* (2003) était effectuée sur trois langues, anglais, français et espagnol.

Meunier *et al.* rapportent que Manuel & Krakow (1984) ont pu montrer que la taille de l'ellipse (points de dispersion sur le triangle vocalique) d'une voyelle est inversement proportionnelle au nombre de voyelles d'un système dont la voyelle en question fait partie.

À la lumière des études précédentes, Meunier *et al.* (2003) traitent des questions relatives au processus cognitif de l'intégration des sons de la parole en fonction de l'organisation du système phonologique. Les langues ne possédant pas les mêmes inventaires de voyelles, la question est de savoir si les locuteurs d'une langue, dont le système vocalique possède peu de voyelles, perçoivent des catégories phonétiques plus larges que celles qui sont perçues par les locuteurs d'une langue dont le système vocalique possède beaucoup de voyelles. Y aura-t-il un rapport entre l'espace acoustique associé à la production d'une voyelle et l'espace perceptif qui représente sa qualité vocalique? Ils posent une autre question relative à la production des voyelles, laquelle nous importe particulièrement, puisqu'elle cherche à savoir si la production des voyelles est influencée ou non par la densité du système. C'est cette question que nous allons suivre en premier dans cette étude.

Leur recherche a consisté en la mesure acoustique des systèmes vocaliques de trois langues, l'anglais, le français et l'espagnol. Il y avait 3 locuteurs par langue, 5 répétitions par locuteur et 2 contextes pour chaque voyelle, soit $5 \times 3 \times 2 = 30$ réalisations par voyelle et par

langue. L'analyse a porté sur F1 et F2. Un tiers de la durée seulement était considéré pour les mesures acoustiques; ils justifient cela par le fait que les locuteurs anglophones avaient tendance à produire des diphtongues.

Les auteurs partent à la vérification de 3 hypothèses:

(1) Si la taille de l'ellipse de dispersion de la voyelle est fonction de la densité du système, alors l'ellipse d'une voyelle dans un système qui possède peu de voyelles aura une taille plus importante que la taille de l'ellipse de la même voyelle lorsque cette voyelle se trouve dans un système qui possède, relativement, beaucoup de voyelles.

(2) L'espace de la voyelle est le même quelle que soit la densité du système; ainsi dans des systèmes peu denses, il y aura un espace non exploité entre les voyelles du système.

(3) La taille de l'ellipse peut être la même d'un système à un autre, alors que c'est l'emplacement de l'ellipse qui change en fonction du nombre des voyelles dans le système en question, en ce sens que les systèmes qui ont peu de voyelles ont leurs voyelles dans une zone centrale du triangle.

Par exemple la taille de l'ellipse d'un [i] français sera la même que celle d'un [i] espagnol, alors que ce sont les emplacements des ellipses qui sont différents.

Les résultats des mesures acoustiques ont montré qu'il y avait des caractéristiques qui étaient communes à toutes les langues; à savoir, les voyelles avant et les voyelles ouvertes présentaient des ellipses dont les tailles étaient relativement plus importantes, par rapport à celles des voyelles arrière et les voyelles fermées.

Concernant la première hypothèse: il y avait peu de chevauchement de qualités vocaliques (peu d'intersection d'ellipses de dispersion) pour l'espagnol, un peu plus pour le français, et beaucoup plus pour l'anglais à tel point que l'ellipse de certaines voyelles sont incluses dans les ellipses d'autres voyelles (le [I] est à l'intérieur de [e]). Cependant, les tailles des ellipses des voyelles sont similaires d'une langue à une autre. Cela va à l'encontre de l'hypothèse (1) qui propose que l'ellipse de la voyelle est d'autant plus grande que le système a moins de voyelles.

La première hypothèse, qui supposait que les locuteurs adaptaient leurs productions en fonction du nombre des voyelles qu'ils avaient dans leur système pour occuper la totalité de l'espace, est infirmée. En effet, les chercheurs trouvaient que la taille de l'ellipse de chaque voyelle restait quasiment la même d'un système à l'autre, la taille de l'ellipse était insensible à la densité du système.

Ces données ont tendance à aller dans le sens de la deuxième hypothèse.

Par rapport à l'hypothèse 3, les triangles de l'espagnol et de l'anglais présentaient beaucoup de similitude, en termes de leurs F1 et F2, alors que celui du français semblait être plus postérieur et plus fermé. Cependant, les triangles des trois langues avaient quasiment la même aire. Cette dernière donnée réfute l'hypothèse (3), puisque l'espagnol, avec un système qui a relativement peu de voyelles, n'a pas centralisé ses voyelles. Finalement, le nombre de voyelles dans un système n'influence pas la taille de l'espace de dispersion.

Les auteurs de l'étude observaient un décalage de triangles vocaliques sur l'ensemble de l'espace vocalique, mais ils n'observaient pas plus ou moins de centralisation en fonction de la densité du système étudié, ce qui réfutait la troisième hypothèse.

En ce qui concerne la production, les auteurs concluent donc que la densité d'un système ne joue pas de rôle sur la disposition des voyelles, confirmant ainsi la deuxième hypothèse.

Ces résultats, à propos de l'hypothèse (1) et à propos de l'hypothèse (3) ne sont pas pourtant en accord avec les résultats de Barkat-Defradas *et al.* (2003). Dans cette dernière étude, qui avait pour but de savoir si l'organisation des voyelles dans un espace dépendait de leur nombre, Barkat-Defradas *et al.* comparaient deux espaces vocaliques de l'arabe, l'arabe jordanien (AJ) et l'arabe marocain (AM). Ils trouvent, en effet, que les ellipses de l'AM en ce qui concerne les voyelles [i; a; u] sont plus grandes que leurs correspondantes en AJ. Or l'AJ dispose de 8 voyelles, alors que l'AM en a 5. Ils soutiennent ainsi l'idée que moins une langue a des voyelles et plus il y a de la variation pour chaque réalisation vocalique (l'ellipse est plus grande), ce qui va dans le sens de l'hypothèse (1), laquelle était réfutée par Meunier *et al.* (2003).

Toujours dans l'étude de Barkat-Defradas *et al.* (2003), les résultats montrent aussi que l'espace vocalique de l'arabe marocain est plus centralisé que celui de l'arabe jordanien, et cela était vrai aussi bien pour les longues que pour les brèves. Après avoir étudié la variabilité des voyelles dans les 2 dialectes arabes, les auteurs soutiennent que la répartition des voyelles dans l'espace vocalique est tributaire de leur nombre; ce qui est en accord avec l'hypothèse (3) de Meunier *et al.* (2003) et qui était infirmée par ces derniers.

Meunier *et al.* (2003) rapportent que les locuteurs espagnols ont identifié les voyelles de leur langue sans ambiguïté. Leur perception du français correspondait à la réalité acoustique, à savoir, le français était perçu comme plus haut et plus postérieur, à l'exception du [a]. Les

voyelles centrales du français, [y] et [ø] qui sont absentes du système de l'espagnol, ont été souvent assimilées à des voyelles postérieures. Alors que le [œ] était assimilé à un [a].

Les voyelles de l'anglais qui correspondaient à celles de l'espagnol [a; e; i; o; et u] ont été correctement identifiées. Cela est en conformité avec les résultats de leurs mesures acoustiques, puisque les espaces occupés par les deux langues sont similaires. En ce qui concerne les voyelles de l'anglais absentes de l'espagnol, leur assimilation pose une petite difficulté, à part le [ɛ] qui était catégorisé soit en tant [a] soit en tant [e].

Les voyelles [ʊ] et [u] ont été assimilées aux voyelles antérieures. Les locuteurs anglophones proposaient un grand nombre de voyelles différentes pour chaque stimulus, ce qui reflétait la difficulté qu'ils avaient avec la tâche.

Les sujets espagnols étaient capables de catégoriser des stimuli qui n'appartenaient pas à leur système, ce qui signifie que l'espace de la perception est plus grand que l'espace de la production.

Meunier *et al.* (2003) ont observé une grande cohérence entre les données acoustiques et les résultats qui perceptifs. Les français et les espagnols, selon eux, produisaient des qualités vocaliques distinctes, et n'avaient pas de difficultés à catégoriser des stimuli qui ne faisaient pas partie de leur système vocalique. D'autre part, les sujets anglais montraient un chevauchement acoustique entre les qualités, et leurs résultats expérimentaux montraient une grande difficulté dans la tâche de catégorisation.

Les auteurs concluent de leur étude que la densité d'un système ne joue pas de rôle sur la production, mais peut-être sur la perception, pour les systèmes dont les voyelles ne se différencient qu'en termes de leurs formants. Cela voudra dire, dans la limite des langues étudiées, que pour les systèmes qui ont peu d'unités vocaliques, leurs unités sont séparées par des distances plus importantes par rapport à un système qui a beaucoup d'unités vocaliques, et par conséquent, il est fort probable que les locuteurs natifs d'un système à peu d'unités n'aient pas la même sensibilité que des locuteurs natifs d'un système à beaucoup d'unités.

En réalité nous ne pouvons pas parler d'une vraie distance stable pour séparer deux qualités vocaliques, mais plutôt, nous parlerons d'un intervalle stable, parce que la sensibilité aux variations qualitatives des voyelles n'est pas seulement fonction du système, mais aussi qu'elle peut, au sein d'un même système, changer d'une personne à une autre selon l'entraînement et l'expérience linguistique. En effet, l'étude menée par Vallée & Kandel

(2003), qui cherchait à tester la robustesse de la perception de l'ouverture (corrélée à F1) et celle de la perception de la position avant-arrière de la langue (corrélée à F2), a montré que les personnes qui s'intéressaient à la phonétique étaient significativement plus compétentes à discerner des distances perceptives plus fines que les personnes naïves en termes de phonétique.

Les résultats de ces études nous conduisent à conclure que le seuil trouvé dans un système donné est propre au système où il a été obtenu, puisque la perception des voyelles est influencée par leur nombre. Par conséquent, les distances obtenues dans le système de l'anglais américain ne sont valides que pour l'anglais américain, et dans les meilleurs cas, valides pour les systèmes qui ont autant de voyelles. Cependant, même si les résultats sur l'anglais ne sont pas transposables sur d'autres systèmes, ils nous donnent un ordre d'idée très réaliste sur la distance robuste qui sépare deux voyelles et qui est de l'ordre de 0,56 Bark pour un système relativement dense avec 10 qualités. Ainsi cette distance pourrait constituer la borne inférieure de l'intervalle, et nous croyons que la valeur de la borne supérieure peut être obtenue en répliquant les travaux de Kewley-Port et son équipe sur le seuil de la qualité vocalique sur un système très peu dense, tel le système de l'arabe.

I.1.4. Etude comparative entre plusieurs systèmes

Nous comparons dans ce paragraphe deux méthodes de mesures de distances dans l'espace vocalique, celle proposée par Schwartz *et al.* (1997b) et celle proposée par Kewley-Port & Zheng (1999). Nous avons commencé par la comparaison des 2 méthodes de conversion en Bark pour nous assurer que l'éventuelle différence dans les résultats ne provient pas de la formule de conversion. Schwartz *et al.* (1997b) utilisent la formule proposée par Schroeder *et al.* (1979): $1\text{Bark} = 7 * \text{asinh}(f \text{ en Hz}/650)$; alors que Kewley-Port & Zheng (1999) utilisent la formule proposée par Traunmüller (1990): $1\text{Bark} = ((26,81 * f \text{ (en Hz)})/(1960 + f \text{ (en Hz)})) - 0,53$. La figure de comparaison des 2 formules est disponible en annexe n° V. Nous choisirons l'une des deux méthodes pour effectuer les mesures des distances de plusieurs systèmes à densités différentes, notre but étant d'observer si la distance de 0,56 Bark est vérifiée comme borne inférieure du seuil. Nous essayerons aussi de déduire une borne supérieure pour l'intervalle du seuil.

I.1.4.1. Choix de la méthode de mesure

Sur les 31 prototypes vocaliques (figure 22), établis d'après l'inventaire de voyelles d'UPSID (Schwartz *et al.*, 1997b), nous avons procédé à la mesure de toutes les distances

possibles⁹. Il y en a 465, dans cet espace vocalique théorique selon les deux méthodes ci-dessus. Schwartz *et al.* (1997b) proposent une distance euclidienne pondérée comme suit: $dViVj = [(F1i-F1j)^2 + 0,3^2 (F'2i- F'2j)^2]^{1/2}$, où $dViVj$ est la distance entre deux voyelles. Cette distance, à la différence de celle proposée par Kewley-Port & Zheng (1999), intègre les quatre premiers formants, dont les trois derniers sont intégrés dans le deuxième formant effectif ($F'2$), lequel est basé, à son tour, sur le seuil de 3,5 Bark proposé par Chistovich & Lublinskaya (1979). Nous donnons quelques résultats à titre illustratif. La plus petite distance parmi les distances possibles est celle qui existe entre [œ] et [ɜ] et vaut 0,1650 Bark, alors que la distance maximale se trouve entre le [a] et le [i] et vaut 4,7549 Bark suivi par la distance entre [a] et [u] 4,5415 et puis [a] et [y] 4,4579. Ce calcul montre qu'il y a 18 distances parmi les 465 possibles qui sont inférieures à 0,56 Bark. Il nous permet d'exclure l'existence de certaines qualités en la présence de certaines autres, notamment celles qui ont une distance très petite, de l'ordre de 0,2 ou 0,3 Bark.

La méthode de mesure utilisée par Kewley-Port & Zheng (1999) consiste tout simplement à calculer une moyenne arithmétique qui base la distance uniquement sur F1 et F2. Cette méthode est résumée comme suit: $dViVj = [(|F1V2 - F1V1| + |F2V2 - F2V1|)/2]$. Une comparaison graphique entre les résultats des deux méthodes est disponible dans la figure 23. La méthode de Kewley-Port & Zheng (1999) semble produire des résultats très similaires, et par conséquent comparables, aux résultats produits par la méthode de Schwartz *et al.* (1997b) pour des distances inférieures à 1 Bark (là se trouve le seuil). Cette dernière méthode est sûrement plus précise parce qu'elle prend en compte les valeurs de F3 et de F4 lors de l'établissement de la distance, mais elle nécessite plus de calculs au préalable, notamment le calcul du deuxième formant effectif $F'2$. En fait, Kewely-Port & Zheng parviennent à estimer la valeur de la distance nécessaire pour séparer perceptivement deux qualités vocaliques avec leur méthode (nous l'avons examinée en I.1.3.2.3.), une démarche que Schwartz *et al.* n'ont pas accomplie, en effet, nous ne savons pas quelle doit être la valeur du seuil si nous utilisons la méthode proposée par Schwartz *et al.* (1997b). Ainsi la méthode de Kewely-Port & Zheng s'impose non pas pour sa précision mais pour les distances que ces chercheurs ont mesurées avec cette méthode. Cela nous permet d'évaluer les distances que

⁹ Il s'agit du nombre des combinaisons possibles de 2 éléments parmi 31. $C_{31}^2 = \frac{31!}{(31-2)! \times 2!} = 465$

nous observerons dans les espaces vocaliques lorsque nous étudierons l'arabe de Mayadin en III.4.2.2. et le français de la région de la Savoie en V.3.3.2.

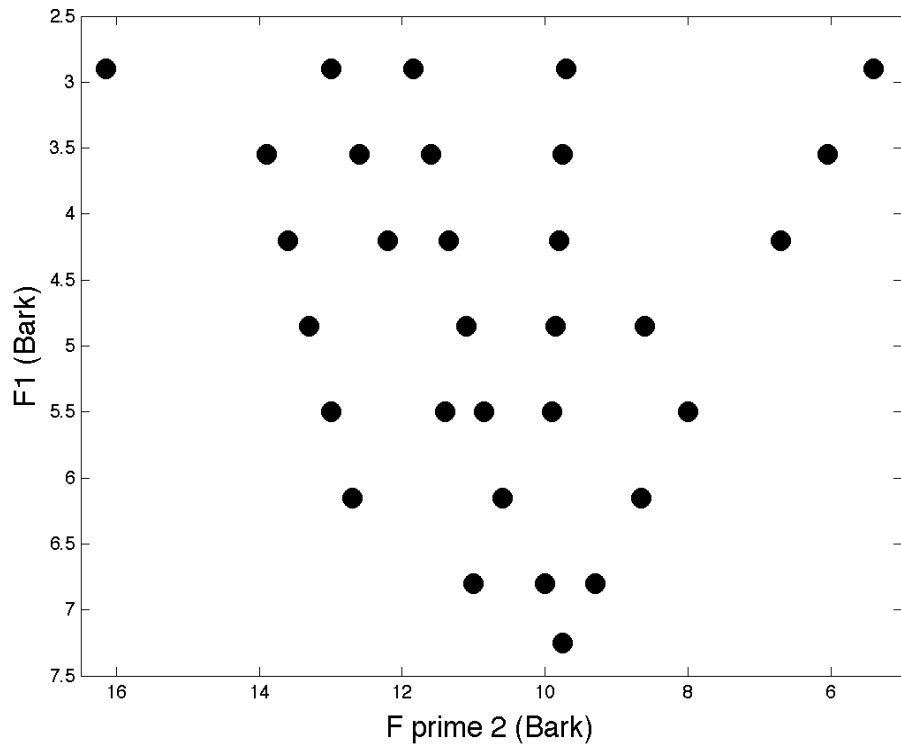


Figure 22. Espace vocalique de 31 qualités représentées en termes de F1 et F'2 établis d'après l'inventaire d'UPSID, voix masculine.

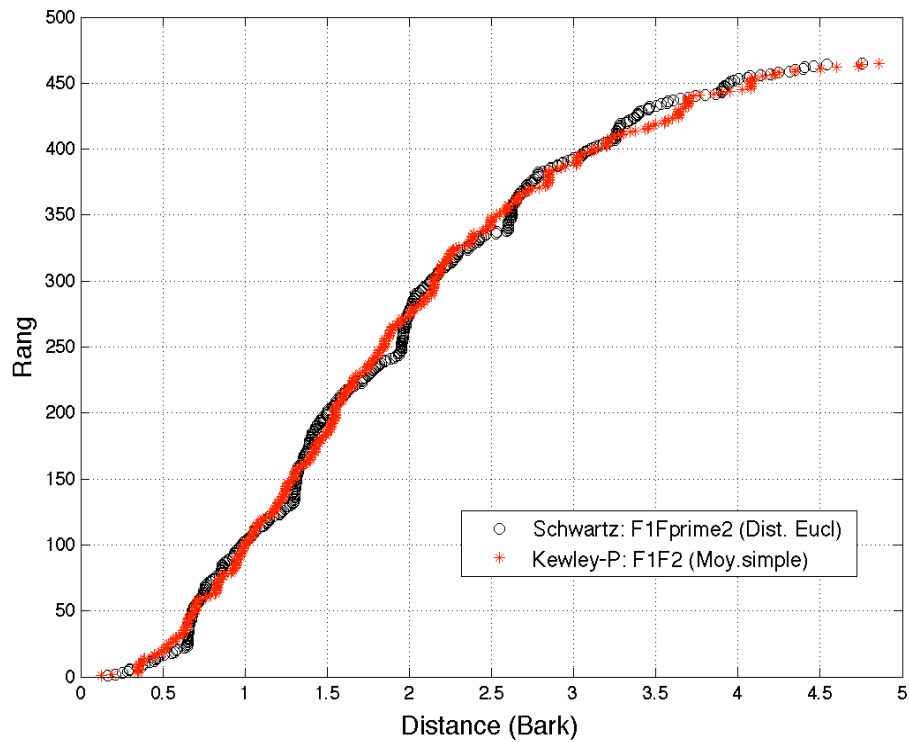


Figure 23. Mesure des 465 distances possibles entre les 31 qualités présentées dans la figure 22 selon deux méthodes: Schwartz et al. (1997b) et Kewley-Port & Zheng (1999).

I.1.4.2. Comparaisons entre des distances dans plusieurs systèmes

Notre but étant de trouver la plus petite distance robuste dans l'espace vocalique, nous devons travailler sur des espaces établis d'après des voix masculines, puisqu'il est établi que l'espace vocalique observé auprès des locuteurs hommes présente, entre ses voyelles cardinales, des distances inférieures à celles de l'espace correspondant obtenu auprès des locuteurs féminins. Cette réalité était confirmée par l'étude de Hillenbrand *et al.* (1995), qui a établi l'espace vocalique de l'anglais américain pour des locuteurs enfants, des locuteurs femmes et des locuteurs hommes. Nous représentons dans la figure 24 les trois espaces maximaux pour les trois catégories de locuteurs, enfant, femme et homme à partir des résultats de l'étude de Hillenbrand *et al.* (1995), sur laquelle nous pouvons observer les différentes tailles que l'espace peut prendre selon l'âge et le sexe.

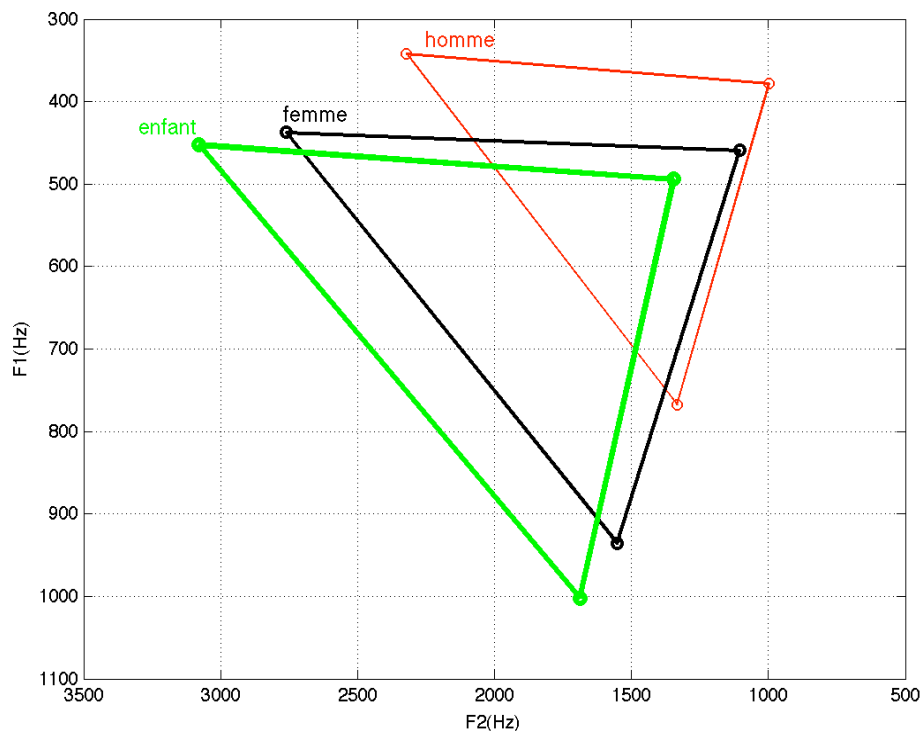


Figure 24. Espaces vocaux pour les trois voyelles cardinales (i, a, u) de l'anglais américain pour trois catégories de locuteurs, enfant, femme et homme. D'après l'étude de Hillenbrand *et al.* (1995).

1.1.4.2.1. L'anglais

Nous avons pris l'espace vocalique de l'anglais établi par Hillenbrand *et al.* (1995, figure 25), et nous avons effectué un calcul de toutes les distances possibles comme le proposaient Kewley-Port & Zheng (1999, figure 26).

L'anglais montre des distances proches de 0,3 Bark. D'après les résultats de nos recherches, détaillés en paragraphe I.1.3.2.3., ce sont des voyelles qui s'opposent par leurs durées, et non pas seulement par leur différence spectrale. L'opposition qualitative pour l'anglais, d'après l'étude de Hillenbrand *et al.* (1995), commence à la valeur de 0,68 Bark. Nous sommes quasiment à 0,1 Bark de ce que Kewley-Port & Zheng (1999) ont trouvé comme distance minimale en anglais. La distance maximale est de 4,1 Bark. Parmi les 12 voyelles de l'anglais, nous constatons la présence de 10 qualités séparées par une distance minimale de 0,68 Bark. Nous disons 10 parce que nous pensons que les distances qui séparent les deux couples de voyelles, que nous avons entourés sur la figure 25, ne sont pas suffisantes pour créer une opposition qualitative.

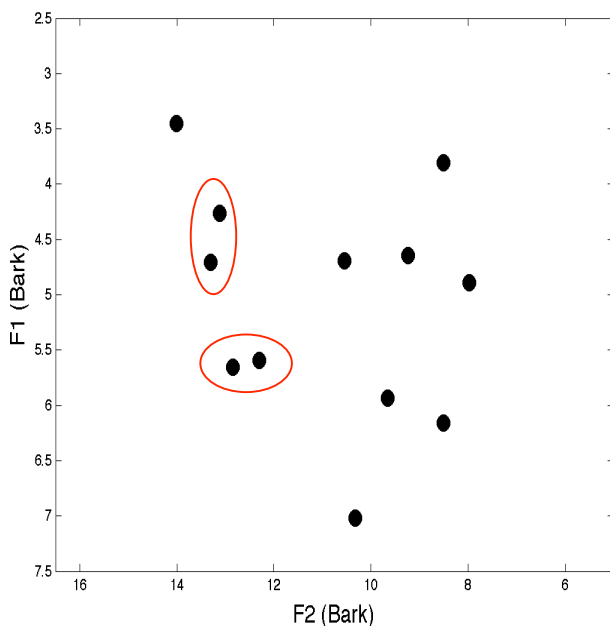


Figure 25. Système phonologique des 12 voyelles de l'anglais américain établi par Hillenbrand *et al.* (1995). Locuteurs masculins.

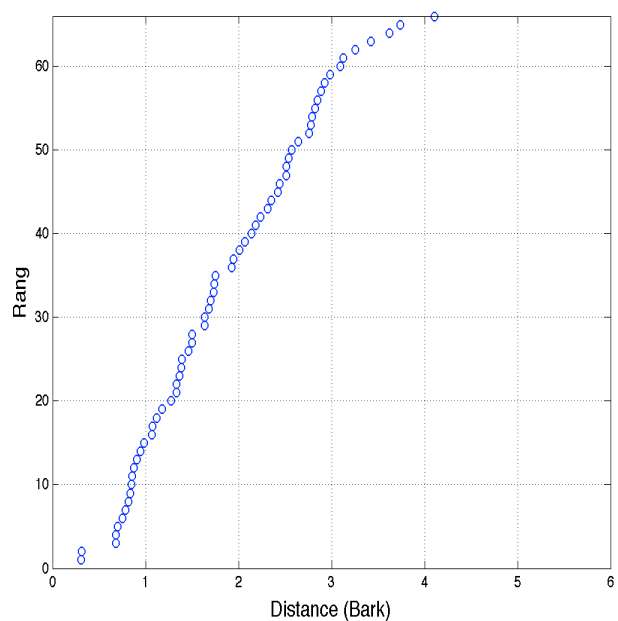


Figure 26. Calcul des 66 distances possibles entre les 12 voyelles de l'anglais américain présentées dans la figure 25. Moyenne sur F1 et F2.

1.1.4.2.2. Le français

Considérant les voyelles orales, le français, qui base toutes ses oppositions sur la qualité acoustique, ne présente pas des distances inférieures à 0,6 Bark. En effet, d'après le système du français établi par Tubach (1989), la plus petite distance observée est exactement égale à 0,6140 Bark et la plus grande est égale à 4,5 Bark. Dans la limite de cette étude, nous pouvons rejoindre la proposition de Kewley-Port & Zheng (1999) sur la valeur de 0,56 Bark pour un système, comme l'anglais américain et le français, à 10 voyelles orales. Nous pensons, effectivement, que 0,6 Bark peut constituer la valeur de la borne inférieure de l'intervalle où le seuil de la distance se situera, parce que avec 10 qualités, le français et l'anglais américain se trouvent parmi les systèmes les plus denses.

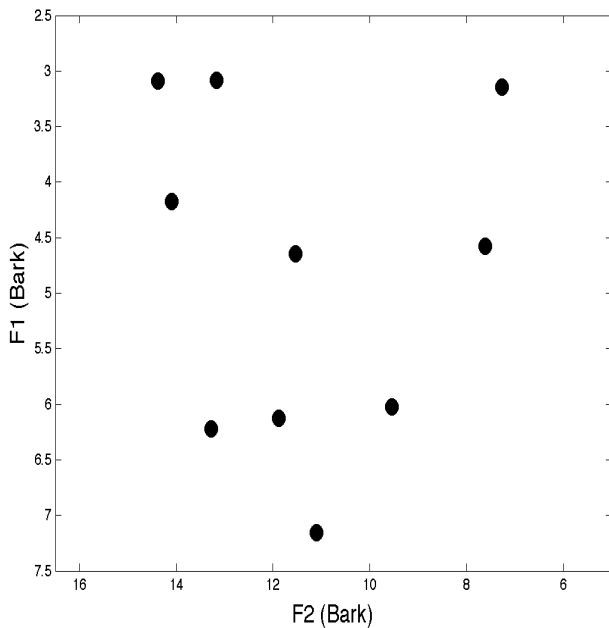


Figure 27. Système phonologique des 10 voyelles du français établi par Tubach (1989). Locuteurs masculins.

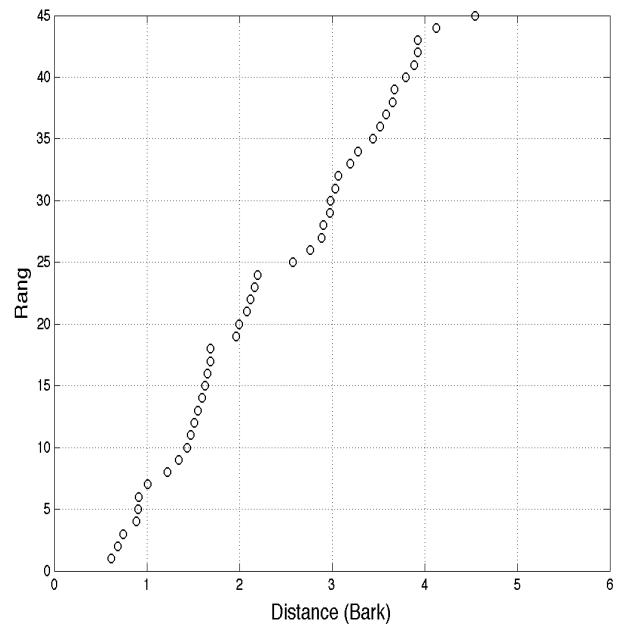


Figure 28. Calcul des 45 distances possibles entre les 10 voyelles du français présentées dans la figure 27. Moyenne sur F1 et F2.

I.1.4.2.3. L'arabe

L'arabe, qui est une langue à quantité vocalique, présente, comme l'anglais, des distances très faibles qui ne supportent vraisemblablement pas une opposition catégorielle. La distance entre les deux [a], long et bref, est égale à 0,24 Bark, et d'après l'étude d'Abou-Haidar (1994), l'opposition qualitative pour l'arabe commence avec des distances supérieures à 1 Bark. En effet, la plus petite distance opposant deux qualités est égale à 1,1 Bark, et la distance maximale est égale à 3,5 Bark. Cela rejoint l'idée que moins le système a de voyelles (moins le système est dense) et plus grandes sont les distances minimales requises qui permettent l'opposition catégorielle entre les voyelles, comme nous remarquons que la dispersion maximale du système, elle aussi, est affectée par le nombre de voyelles dans le système.

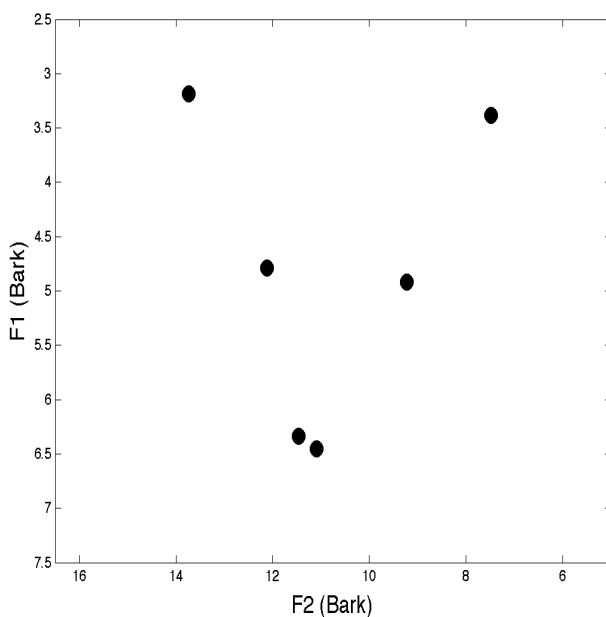


Figure 29. Système vocalique des 6 voyelles de l'arabe [i: , i , u: , u , a: , a], établi par Abou-Haidar (1994), d'après l'étude des parlars de 8 pays arabophones. Locuteurs masculins.

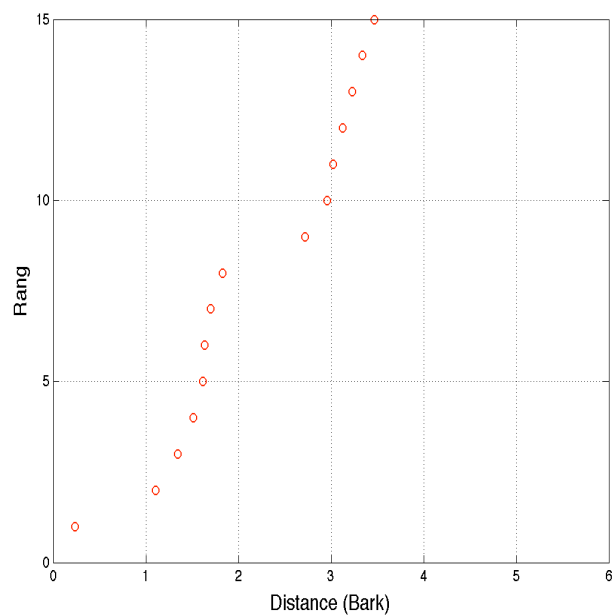


Figure 30. Calcul des 15 distances possibles entre les 6 voyelles de l'arabe présentées dans la figure 29. Moyenne sur F1 et F2.

Nos données sur l'arabe dialectal de Syrie, Allatif & Abry (2004), ne sont pas en total accord avec les données d'Abou-Haidar (1994), puisque nous trouvons deux distances qui sont égales à 0,3 Bark environ (Abou-Haidar, 1994 en trouve une seule). La plus grande est égale à 3,1 Bark. Nous avons pu démontrer, par ailleurs, Allatif (2007), que la distance

séparant les deux [a], long et bref, n'était pas suffisante pour supporter une opposition catégorielle entre les deux [a].

Finalement, nous pensons que le seuil catégoriel concernant la distance spectrale se trouve dans un intervalle et qu'il n'y a pas une seule valeur précise qui vaille pour tous les systèmes. La valeur du seuil se rapproche de la borne inférieure de l'intervalle quand le système est dense. L'inverse est aussi vrai, c'est-à-dire que moins le système est dense et plus la valeur du seuil se rapproche de la borne supérieure de l'intervalle. Les deux valeurs qui délimitent cet intervalle, d'après nos observations et notre étude comparative, seront environ 0,6 Bark pour la borne inférieure (système très dense) pouvant légèrement baisser en cas de système excessivement dense, et environ 1 Bark pour des systèmes très peu denses, comme l'arabe.

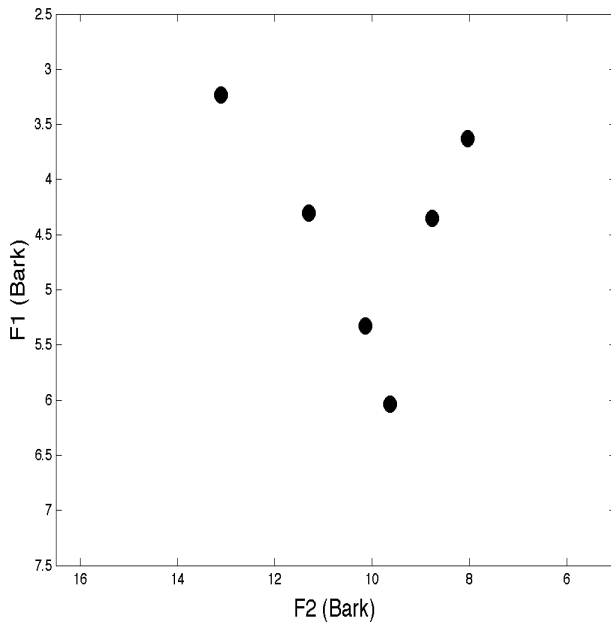


Figure 31. Système vocalique des 6 voyelles de l'arabe dialectal de Syrie [i: , i , u: , u , a: , a], établi par Allatif & Abry (2004). Locuteurs masculins.

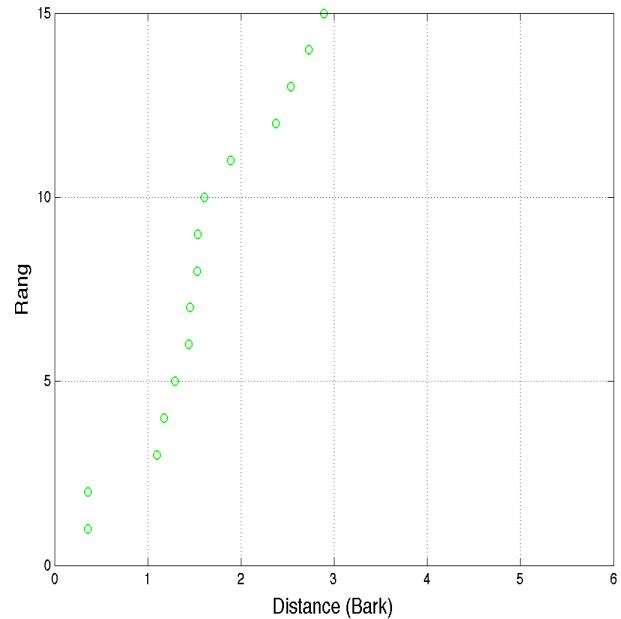


Figure 32. Calcul des 15 distances possibles entre les 6 voyelles de l'arabe dialectal de Syrie, présentées dans la figure 31. Moyenne sur F1 et F2 pour /i/ vs. /i:/, et sur F1 seulement pour /a/ vs. /a:/ et /u/ vs. /u:/, F2 non significatif.

I.1.4.3. Conclusion

La lecture graphique des résultats des mesures des distances pour les quatre systèmes que nous avons examinés plus haut, et que nous avons voulu présenter sur une même figure, figure 33 pour mieux les comparer, ne soutient pas tout à fait la thèse de Meunier *et al.* (2003) selon laquelle le nombre de voyelles, au sens de qualité vocalique, n'a pas d'influence sur la taille de l'espace vocalique, puisque nous observons que l'arabe qui n'a que 4 qualités (mais 6 voyelles) a une aire plus petite que l'anglais avec 10 qualités (et 12 voyelles).

Il est évident qu'avec un si petit nombre de systèmes, nous ne pouvons pas prétendre donner des valeurs très solides pour l'intervalle du seuil, mais une approximation très réaliste de l'étendue de l'intervalle où le seuil fluctuera en fonction du nombre de voyelles dans le système considéré. Cependant si notre démarche est juste, elle s'avèrera particulièrement efficace si nous l'appliquons sur un très grand nombre de systèmes décrits avec beaucoup de précision, tels les systèmes que nous avons étudiés dans ce paragraphe.

Le but principal de ce chapitre, qui est de trouver le seuil de la qualité vocalique

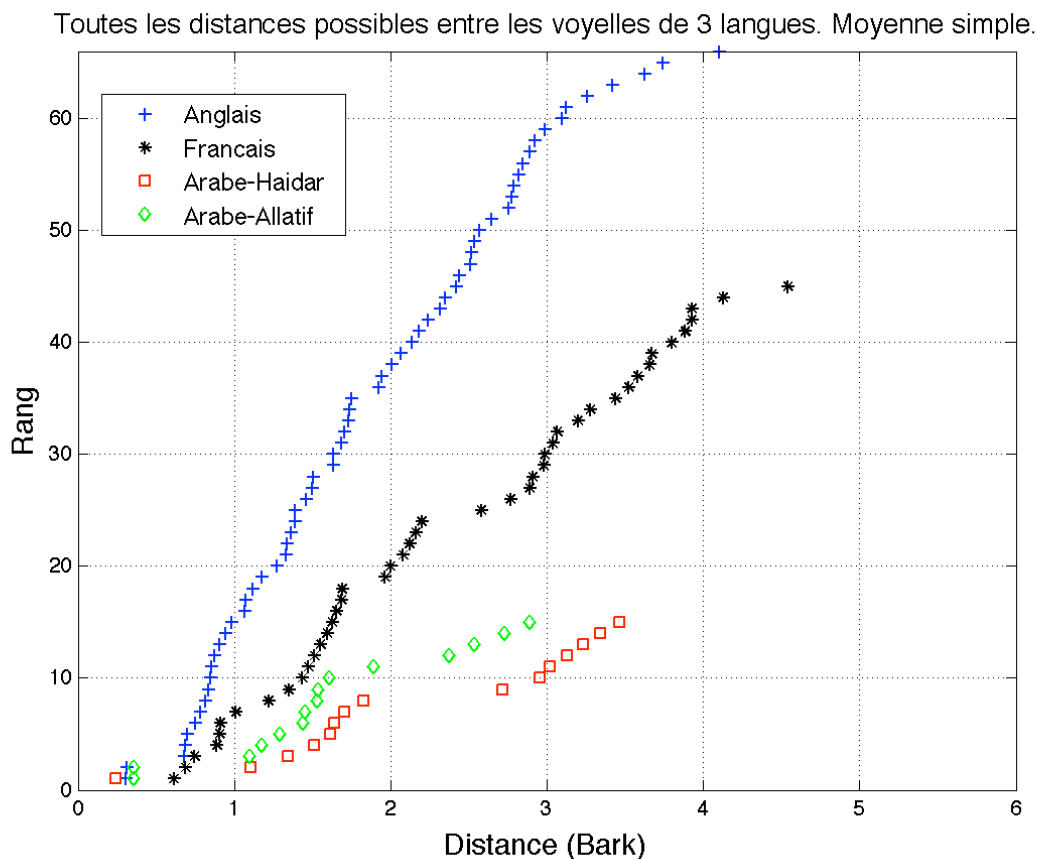


Figure 33. Comparaison entre les distances de trois systèmes vocaliques des 4 études présentées en figures (26,28,30 et 32).

(perceptible vs. catégoriel) est ici atteint. Les données convergent vers une conclusion globale selon laquelle le seuil est mieux représenté en termes de valeur absolue et non pas en termes de valeur relative, et qu'il n'y a pas de frontière critique, mais plutôt un intervalle critique pour le seuil catégoriel où toutes les langues peuvent se situer.

Concernant les deux niveaux de seuil de la qualité vocalique, nous avons un seuil pour la perception d'un changement vocalique (Just Noticeable Difference, JND), et un seuil catégoriel de la qualité vocalique. Le premier niveau permet de savoir si une modification dans la qualité de la voyelle est perceptible, le deuxième indique si la distance entre deux entités vocaliques est suffisante pour permettre leur indépendance phonologique. En nous basant sur les propositions de Kewley-Port & Watson (1994), Kewley-Port & Zheng (1996, 1999) et Kewley-Port (2001), concernant cette question, nous pouvons admettre qu'une distance spectrale combinée ($F1 \times F2$) d'une valeur avoisinant les 0,30 Bark peut franchir le premier niveau (JND), alors que pour franchir le deuxième niveau, cette valeur doit être double, soit 0,60 Bark.

Nous concluons ce grand chapitre en disant que nos observations des distances calculées dans les systèmes présentés ci-dessus, nous permettent d'avancer un intervalle pour le seuil, qui sera inclus entre 0,6 (systèmes très denses) et 1 Bark (systèmes peu denses).

1.2. Seuil différentiel de la perception de la durée vocalique.

La question de la différence temporelle perceptible entre deux durées a beaucoup d'importance pour notre problématique concernant la différenciation des voyelles longues et les voyelles brèves. Nous voudrions savoir ici, de combien de ms doit être l'écart temporel entre deux voyelles pour que cet écart temporel, à lui seul, puisse constituer un indice acoustique sur lequel un contraste catégoriel peut reposer.

La littérature sur la plus petite hausse ou baisse de durée perceptible par l'oreille humaine est loin d'être unanime. Pour la description du même phénomène nous trouvons des valeurs très hétérogènes. Cela pourrait provenir de: la nature des stimuli utilisés dans les différentes études portant sur la question, stimuli de types différents d'une expérience à une autre (son pur, bruit, sons périodiques, etc.); des sujets (entraînés ou non, jeunes ou âgés); ou encore du protocole expérimental, la frontière catégorielle n'étant pas toujours fixée de la même manière (50%, 70%, etc).

En effet, à titre d'exemple, Henry (1948), Small & Campbell (1962) trouvaient que le seuil était de 7 à 10 ms pour une durée de 30 à 40 ms, ce qui était assez différent des résultats de l'étude de Ruhm *et al.* (1966), qui trouvaient que le seuil était de 2 à 3 ms pour un stimulus de 40 ms.

Ces divergences nous conduisent à explorer plus amplement le champ de la perception de la durée afin de trouver une réponse satisfaisante à notre question concernant la différence entre les voyelles longues et brèves. Ainsi, nous allons examiner, parfois en détail, des études qui ont traité de ce problème. Les objectifs finaux de ces études-là ne sont pas toujours la durée en soi. La durée est utilisée dans ces études en tant qu'outil qui permettait de découvrir d'autres phénomènes, l'effet de l'âge sur la perception du temps par exemple, la différence entre l'homme et le singe dans le traitement de l'opposition temporelle des sons, l'impact de certaines pathologies auditives sur la perception de la durée des sons de la parole, etc. Nous, ici, nous extrayons de ces études la partie intéressante pour notre question, c'est-à-dire le seuil différentiel de la perception de la durée des voyelles.

I.2.1. Facteurs affectant le seuil de la durée

I.2.1.1. Facteurs extra-stimulus

Dans ce paragraphe, nous discuterons de 2 facteurs. 1) l'âge: nous cherchons à savoir si la perception des contrastes temporels concernant les voyelles est directement liée à l'ouïe. Autrement dit, est ce que nous sommes capables de différencier les durées des voyelles dès que nous sommes capables de les entendre? Ou est-ce que cette capacité n'est liée qu'à l'apprentissage? La réponse a de l'importance pour nous parce qu'elle nous permet d'exclure ou d'intégrer les résultats portant sur le seuil de la durée provenant de sujets appartenant à telle ou telle tranche d'âge.

2) le sexe de la personne: hommes et femmes produisent-ils les mêmes voyelles avec des durées différentes? Ou est ce que nous pouvons faire abstraction du sexe de la personne lors du traitement du seuil de la durée des voyelles?

I.2.1.1.1. le facteur âge

La perception de la différence entre les durées des voyelles est une des facultés auditives que l'homme acquière, relativement, très tôt, mais pas à la naissance, contrairement à l'acquisition des qualités vocaliques où les différences entre qualités disparaissent au cours de l'enfance en fonction de la disposition des éléments vocaliques de la langue maternelle

(surdit  phonologique), Werker *et al.* (1981).

L'exploitation de la dur e vocalique continue son acquisition au cours de l'enfance, la repr sentation de la dur e, et son exploitation dans la langue, peut changer en fonction de l' ge au cours de l'enfance, comme elle peut  tre affect e par les troubles du langage, telle la dyslexie. Carr  *et al.* (2001) ont montr  que la diphtongue synth tique [ai] s'entendait [ai] quand la dur e de la transition qui liait les deux voyelles  tait inf rieure   200 ms, et s'entendaient comme une r alisation de [aei] si la dur e de cette m me transition  tait sup rieure   200 ms, Messaoud *et al.* (2003) pensaient que ce seuil de dur e, 200 ms, d pendait de l' ge. Selon ces derniers, plus le locuteur est jeune et plus la transition doit  tre longue pour  tre per ue, comme la perception du stimulus  tant en partie conditionn e par l'exp rience linguistique du sujet.

Messaoud *et al.* tirent ces r sultats de leur exp rience qui a port  sur 3 groupes d'enfants: 36 personnes  g es de 13 ans. 34 suivis longitudinalement de 6   7 ans. 19  g s de 13 ans, et suivis pendant 2 ans. 17 autres personnes ont particip    l'exp rience contr le. Les stimuli  taient 6 s quences de: [a] + transition + [i]. Les parties vocaliques de la pr  et de la post transition  taient des segments vocaliques stables, alors que la dur e de la transition variait de 50   300 ms par pas de 50 ms.

Le [a] avait une dur e de 100 ms, F1, F2 et F3 commen aient respectivement   743, 1543 et 2796 Hz. Le [i] avait une dur e de 150 ms et ses 3 premiers formants terminaient la transition   240, 2500 et 3140 Hz. Les formants F4, F5 et F6  taient fix s respectivement   280, 4500 et 5500 Hz. Dans cette exp rience, chacun des six stimuli  tait pr sent  10 fois dans un ordre al atoire, les sujets avaient   appuyer sur un bouton  tiquet  2 ou un autre  tiquet  3 selon qu'ils entendaient 2 ou 3 voyelles.

Le traitement statistique des r sultats a montr  que plus les sujets  taient jeunes et plus ils avaient besoin que la transition soit plus longue pour passer de [ai]   [aei]. Les auteurs de l' tude en d duisaient que la cat gorisation de la dur e de la transition  taient fonction de l' ge, les plus  g s (les 13 et 7 ans) percevaient le passage d'une cat gorie   une autre plus rapidement que les plus jeunes (les 7 et 6 ans), la transition  tait plus  vidente pour les plus  g s. Ceux qui avaient 6 ans ne comptaient pas du tout sur le param tre de la syllabe, et entendaient toujours la r alisation [ai]. L'effet de l' ge  tait statistiquement significatif ($p < 0,05$).

Ces données mettent en évidence une description développementale de la détection du seuil de 200 ms proposée par Carré *et al.* (2001). La maturation linguistique aide les locuteurs à l'identification du troisième son. La transition devient un indice plus consistant à mesure que l'expérience linguistique s'enrichit.

L'effet de l'âge du locuteur sur la perception de la durée du stimulus était confirmé aussi par une étude menée par Eliers *et al.* (1984). Les auteurs de cette étude ont démontré que le bébé, alors âgé de quelques mois seulement, était capable de discriminer les durées des voyelles en anglais. Cette étude est en rapport direct avec notre question à propos de la perception de la durée, parce qu'elle nous fournit des renseignements sur le seuil de la perception de la durée. C'est pour cela que nous tenons à en détailler la procédure et les résultats.

Les chercheurs ont examiné 27 bébés, âgés de 5 à 11 mois répartis en trois groupes de 9 bébés, et 4 adultes âgés de 20-30 ans. Leurs résultats ont montré que ces bébés étaient capables de faire la différence entre des voyelles qui ne se contrastaient que par leur durée. L'expérience était effectuée avec trois rapports où la longue des deux voyelles était 33% soit $[(400-300)/300]$, 67% soit $[(500-300)/300]$ et 100% soit $[(600-300)/300]$ plus longue que la brève. Pour ces trois rapports, la discrimination était significative.

Le corpus était formé de 3 séquences, toutes synthétisées: une séquence d'une syllabe [mad = 540 ms], une séquence de deux syllabes [sa-mad = 797 ms] et une séquence de trois syllabes [ma-sa-mad = 1045 ms]. Pour chacune des séquences, Eliers *et al.* augmentaient uniquement la durée de la voyelle de la dernière syllabe (le [a] dans [mad] = 300 ms) de 100 ms en 3 fois. Ainsi la voyelle [a] durait 400, 500 et 600 ms, les autres unités de la séquence ne subissant aucun changement. Par conséquent, la durée de la séquence accusait une hausse de 100, de 200 et de 300 ms et était comparée à chaque fois avec la séquence qui avait la durée la moins longue. Cela produisait une condition de 9 comparaisons structurées comme suit (les durées ci-dessous sont toutes en ms):

Groupe 1: [mad = 540 (a = 300)] est comparé à [mad = 640 (a = 400)], à [mad = 740 ms (a = 500)] et à [mad = 840 (a = 600)].

Groupe 2: [sa- mad = 797 (a = 300)] est comparé à [sa- mad = 897 (a = 400)], à [sa- mad = 997 (a = 500)] et à [sa- mad = 1097 (a = 600)].

Groupe 3: [ma-sa- mad = 1045 (a = 300)] est comparé à [ma-sa- mad = 1145 (a = 400)], à [ma-sa- mad = 1245 (a = 500)] et à [ma-sa- mad = 1345 (a = 600)].

Les études statistiques ont montré que les 9 discriminations, chez les adultes et chez les bébés, étaient significatives, même si la capacité du bébé était plus limitée que celle de l'adulte. Il y avait par ailleurs un effet de la durée, c'est-à-dire qu'au sein d'un même groupe, chaque comparaison était différente de l'autre; en effet, la discrimination était d'autant meilleure que l'écart temporel était plus important. Par exemple, la différence [ma-sa-mad = 1045 (a = 300)] vs. [ma-sa- mad = 1345 (a = 600)] où la hausse ($\Delta T/T$) est de 100 %, était mieux discriminée que la différence [ma-sa-mad = 1045 (a = 300)] vs. [ma-sa- mad = 1145 (a = 400)] où la hausse ($\Delta T/T$) est de 0,33 %.

Le fait que la performance du bébé s'améliorait plus l'écart temporel entre les deux voyelles de la comparaison augmentait, atteste de la possession du bébé d'un pouvoir régulier (lawful) à discriminer les variations de la durée des voyelles. En effet, l'écart (ΔT) de 300 ms était mieux discriminé que l'écart de 100 ms.

L'analyse de la variance n'a pas montré un effet du groupe, c'est-à-dire que la durée initiale du stimulus, 540, 797 et 1045 ms, n'influçait pas la capacité à la détection des variations de la durée. Ce dernier résultat ne se situe pas dans nos attentes, parce que nous pensons que le seuil différentiel de la perception de la durée est proportionnel à la durée initiale du stimulus; plus le stimulus de départ est long et plus le seuil serait élevé. Ce n'est pas le cas ici, ni pour les bébés ni pour les adultes, puisque le seuil reste constant avec la hausse de la durée du stimulus. Nous pensons que la raison de ce résultat réside dans le fait que le rapport ($\Delta T/T$) est toujours très élevé, et largement au-dessus de la limite du perceptible (le seuil).

En effet, les 9 contrastes étaient très aisément discriminés par les sujets adultes, touchant à la perfection pour certains contrastes, ce qui suggèrent fortement que le seuil différentiel de la perception de la durée pour les signaux de la pseudo parole (logatomes) se situe bien au-dessous du plus petit rapport, ici $\Delta T/\text{Durée Voyelle} = 0,33$, qu'ils ont réussi à discriminer dans cette expérience. Ce rapport ($\Delta T/T$) nous pouvons l'envisager de trois manières:

1) ($\Delta T/\text{Durée de la totalité de la séquence (les trois syllabes)}$). Le plus petit rapport est (100/1045) légèrement inférieur à 0,1.

2) (ΔT /Durée de la syllabe qui contient la voyelle objet de la modification). Le plus petit rapport est (100/540) légèrement inférieur à 0,2.

3) (ΔT /Durée voyelle objet de la modification) le plus petit rapport (100/300) est égal à 0,33.

Les auteurs de l'étude se basent sur le dernier rapport pour estimer le pourcentage de la hausse.

Nous pensons qu'il faut plutôt se baser sur la deuxième procédure, c'est-à-dire rapporter la variation sur la syllabe, soit (ΔT /Durée de la syllabe qui contient la voyelle objet de la modification). De notre point de vue, c'est le rapport le mieux adapté pour rendre compte de la valeur du seuil. En effet, dans cette étude, ainsi que nous l'avons vu plus haut, augmenter le nombre de syllabes n'a pas eu d'effet sur le seuil. Quant aux bébés, ils n'ont pas eu tendance à mieux discriminer la différence de durée selon que la voyelle est insérée dans une séquence d'une, deux ou trois syllabes. L'idée que la segmentation temporelle de la chaîne parlée se fait en durées de syllabes était proposée par Carré *et al.* (2001), où les auteurs de l'étude ont montré que la diphtongue synthétique [ai] s'entendait [ai] si la durée de la transition qui liait les deux voyelles était inférieure à 200 ms, et s'entendait comme une réalisation de [aei] si la durée de cette même transition était supérieure à 200 ms. Carré *et al.* (2001) pensent que la naissance de ce seuil est influencée par la durée de la syllabe, parce que cette durée, 200 ms, est la durée moyenne d'une syllabe en français.

Les auteurs expliquent ce phénomène comme suit: 1) en passant de [a] à [i] on passe par [e], ce qui explique que la voyelle perçue est de timbre [e]. 2) la durée moyenne de la syllabe en français est de 200 ms, ce qui correspond au seuil est de 200 ms. Carré *et al.* expliquent cet effet de la durée de la transition sur la perception de deux ou de trois voyelles par le fait que le locuteur segmente le continuum vocalique en syllabes et qu'à chaque durée de 200 ms il assigne une syllabe. Quand la durée est inférieure à 200 ms, la réalisation est intégrée en tant qu'une seule syllabe, mais quand elle dépasse cette même durée, elle est intégrée en tant que 2 syllabes, et par conséquent la troisième voyelle constituera la deuxième syllabe.

D'autres données vont dans le sens de notre proposition (rapporter la hausse sur la durée de la syllabe et non pas sur la durée de la voyelle ou sur la durée de la phrase), des données selon lesquelles si la structure phonétique se complexifie au-delà de la syllabe, cette complexité n'a pas d'effet majeur sur la perception des indices acoustiques. Kewley-port & Zheng (1999) ont mesuré le seuil de la qualité vocalique, ils affirment que la taille du contexte

phonétique influence significativement le pouvoir discriminatoire auditif à résoudre les variations acoustiques qualitatives quand ils considéraient des structures inférieures à la syllabe, mais presque pas au-delà de la syllabe. Ce qui signifie que le seuil est différent selon que nous considérons un formant isolé, une voyelle (plusieurs formants simultanément) ou une syllabe; mais quasiment pas au-delà. La capacité à la résolution de la fréquence du formant diminue de 36 % en passant de la voyelle à la syllabe, alors que cette capacité diminue de 8 % seulement en passant de la syllabe à une phrase où il y a 4 syllabes (3 mots), par ailleurs, au-delà de cette taille (4 syllabes) la capacité à la discrimination reste la même. Ce qui démontre qu'au-delà de la syllabe, la taille du contexte phonétique a un effet négligeable.

C'est pour cela que nous pensons que le seuil est largement inférieur non à 0,33% mais inférieur à 20 %, puisque ce rapport (20 %) est très facilement repérable. Nous nous rapprochons de ce que Lehiste (1970) trouve, à savoir le seuil différentiel de la perception de la durée de la voyelle est de 14 % pour les voyelles supérieures à 400 ms.

Cette étude montre que les bébés, même de 5 mois, perçoivent la différence de durée, donc il y avait une faculté innée à discriminer les durées des sons. Elle prouve que jusqu'à l'âge de 1 an, au moins, la capacité à discriminer les différences de durées entre les unités sonores de la pseudo parole, et vraisemblablement de la parole, n'est pas encore définitivement établie, mais elle est en évolution, puisque les performances des adultes étaient significativement meilleures que celles des bébés. Ce qui permet de dire que cette faculté innée est finalement façonnée par l'entraînement qui sera exercé par le système phonologique de l'environnement linguistique. Cette étude a pu aussi mettre en évidence que des bébés âgés de 5 à 11 mois étaient capables de discriminer une hausse (de 20% pour nous) de 33% (pour les auteurs).

D'autres études se sont intéressées à d'autres facteurs extra-stimulus affectant la perception de la durée de la voyelle, tels l'espèce primate d'appartenance (Sinnott *et al.*, 1987a), ou le sexe de la personne (Simpson & Ericsson, 2003).

1.2.1.1.2. Sexe

Dans une recherche qui a porté sur la description de la qualité vocalique et de la durée des voyelles chez des enfants, des femmes et des hommes natifs de l'anglais américain, Hillenbrand *et al.* (1995), les locuteurs féminins ont présenté des durées plus importantes que les durées observées auprès des locuteurs masculins. Cela était vrai pour toutes les voyelles de

l'anglais. Pour les deux sexes, les voyelles étaient émises exactement dans les mêmes contextes, enregistrées et analysées avec un même protocole expérimental. Cela laisse penser que les femmes natives de l'anglais des Etats-Unis produisent des voyelles plus longues que celles des hommes, et par conséquent que la durée de la voyelle est tributaire du sexe du locuteur.

Une étude de Simpson (1998), qui a porté sur 60 000 voyelles de l'allemand (25 femmes et 29 hommes), a montré aussi que toutes les voyelles des femmes étaient plus longues que celles de leurs homologues masculins. Pour le français québécois (Martin, 1998; 2001), la réponse n'était pas catégorique, parce que la différence était significative pour certaines voyelles mais elle ne l'était pas pour certaines autres.

Les données recueillies, selon Martin (1998; 2001), mettent en évidence deux faits:

1) la durée des voyelles chez les femmes est plus importante que celle des voyelles chez les hommes.

2) la différence de durée entre les voyelles longues et les voyelles brèves est plus importante chez les femmes que chez les hommes. Selon l'auteur de l'étude, il peut y avoir 2 explications à cette différence de durée: (i) sociophonétique: faire des voyelles plus longues équivaut à parler plus clairement, ce qui est une qualité attribuée aux femmes. (ii) physiologique: les dimensions du conduit vocal chez la femme sont différentes de celles de l'homme.

Cependant, Simpson (2002) trouvaient que la différence de durée en ce qui concerne les énoncés n'était pas significative. Suite à ces considérations concernant la différence due à l'appartenance sexuelle pour les durées des voyelles dans plusieurs langues, Simpson & Ericsson (2003) ont mené une étude à partir d'un corpus réalisé par 48 locuteurs de l'anglais des Etats-Unis et 9 locuteurs du suédois pour la différence des durées due à la différence des sexes. Cette étude a montré qu'il y avait un ensemble complexe de patrons de durées. La durée la plus longue chez les femmes se trouvait dans des endroits contrastifs au niveau de l'information, les endroits de prééminence, comme l'accent en anglais qui a un rôle contrastif. Alors que la durée la plus longue chez les hommes se trouvait sur les consonnes. Il y avait peu de différences au niveau de la durée de la totalité de l'énoncé, alors qu'il y avait une différence en considérant les différentes parties de l'énoncé.

Cette étude était un examen préliminaire de la distribution des patrons des durées des segments (phonèmes) dans tous les énoncés de l'anglais et du suédois de deux bases de

données, où la différence des durées des voyelles, une différence due au sexe, était déjà mesurée.

Le calcul de la moyenne des durées a mis en évidence qu'il n'y avait pas de différence de durée entre les énoncés des hommes et ceux des femmes. Cela implique que toute différence produite, le sera au niveau du segment.

Les données de l'anglais mettent en évidence deux patrons: (i) dans la syllabe accentuée les voyelles des femmes tendent à être plus longues que celles des hommes (les voyelles postérieures arrondies font exception, puisque soit elles sont plus longues chez l'homme, soit il n'y a pas de différence entre homme et femme). (ii) les hommes ont leurs durées les plus longues avec les segments consonantiques.

Le corpus du suédois était spécialement choisi pour examiner la différence des durées des voyelles dans différents environnements pour les voyelles accentuées. Les 3 voyelles suivantes [i:, a:, u:] lorsqu'elles étaient sous l'accent dans les mots cibles, étaient significativement plus longues chez les femmes que chez les hommes; alors que les différences entre les durées de ces voyelles lorsqu'elles étaient non accentuées, étaient non significatives.

Simpson & Ericsson (2003) pensent que la description des patrons que nous venons de voir est insatisfaisante, parce qu'ils croient que la soumission des résultats à un examen plus minutieux, montrerait beaucoup trop d'exceptions. Simpson & Ericsson (2003) avancent que malgré la complexité des patrons qui sortent de leur étude, ils peuvent en tirer certaines conclusions: les différences dans les longueurs des énoncés selon le sexe seraient en fait des exceptions. Cela voudra-t-il dire que si nous considérons 2 segments, la différence de durée pour chacun est compensée par ailleurs dans l'énoncé, c'est-à-dire que s'il y en a un qui est relativement long, il y en aura un autre qui sera relativement court. Ce dernier point est en contradiction avec les résultats de Simpson (1998) pour l'allemand, où parce qu'il trouvait qu'il y avait une différence significative entre la durée de l'énoncé produit par des hommes et la durée de ce même énoncé lorsqu'il était produit par des femmes.

Les patrons mis en évidence par ces études indiquent que les différences entre les durées des patrons, selon Simpson & Ericsson (2003), peuvent être dues à la place de la syllabe dans la phrase, à la place de la syllabe dans la structure rythmique, aussi bien qu'à la place grammaticale et lexicale de la syllabe et non pas à la différence homme/femme. Donc la question du rapport entre le sexe et la durée des sons de la parole reste entièrement ouverte.

Nous pouvons conclure qu'en l'état actuel des recherches sur cette question, nous ne pouvons pas affirmer que les voyelles sont significativement plus, ou moins, longues en fonction du sexe du locuteur malgré la présence de quelques indices qui nous laissent penser que les locuteurs féminins ont tendance à réaliser des voyelles plus longues que celles de leurs homologues masculins. Ainsi, en cherchant à savoir quelle doit être la valeur du seuil différentiel de la perception de la durée de la voyelle, nous pourrions faire abstraction, en production et en perception, du sexe de la personne.

I.2.1.2. Facteurs intra-stimulus

I.2.1.2.1. Fréquence, Intensité et durée initiale

Concernant le rôle de la fréquence du stimulus, nous allons admettre qu'elle n'a pas d'effet sur le seuil de la durée pour une plage de fréquences pertinente pour la perception des voyelles, c'est-à-dire entre 100 et 4000 Hz, Hz (Moore, 1973; cf. figures 1 et 2), comme l'affirmaient aussi Tyler *et al.* dans leur étude de 1982.

Dans une étude de Sinnott *et al.* (1987a), delta T était étudié en fonction de l'intensité de 30 à 70 dB dans le sens de la hausse, avec $T = 25$ ms et $T = 200$ ms, le signal était centré autour de 2000 Hz sans harmonique (1/3 d'octave au-dessus et 1/3 d'octave au-dessous). Les résultats de cette étude sont en accord avec les autres études qui ont testé l'effet de l'intensité sur la perception du seuil qualitatif d'un son (Nelson *et al.*, 1983; Wier *et al.*, 1977). Sinnott *et al.* (1987a) montrent que la hausse de l'intensité de 30 à 70 dB par pas de 10 dB pour $T = 25$ ms occasionne une baisse graduelle du seuil de 7 ms (à 30 dB) à 4 ms (à 70 dB). Alors que lorsque la durée du stimulus est de 200 ms, les auteurs de l'étude remarquent une baisse en passant de 30 dB (42 ms) à 40 dB (31 ms) comme c'est montré dans la figure 35.

L'intérêt de cette étude réside dans le fait qu'elle nous éclaire sur trois points importants.

(1) Pour la durée initiale: le seuil est significativement différent selon (T) la durée initiale du stimulus; en effet, nous observons très clairement la différence entre les seuils, selon que la durée initiale est de 25 ou de 200 ms. (2) Pour l'intensité: au-delà de 40 dB, la hausse de l'intensité n'a plus d'effet sur le seuil; il y a une interaction entre l'intensité et la durée initiale, puisque la hausse de l'intensité n'a pas le même effet selon la durée initiale du stimulus. (3) Enfin, il n'existe pas une valeur en pourcentage qui vaudrait pour toutes les

durées: puisque nous trouvons que le rapport de Weber $\Delta T/T$ est de 0,20 pour $T = 25$ ms, alors qu'il est de 0,15 quand $T = 200$ ms.

L'étude de Tyler *et al.* (1982) présentent des résultats qui ne vont pas tout à fait dans le même sens que ceux de Sinnott *et al.* (1987a). Tyler *et al.* ont essayé de déterminer pour 16 sujets normo-entendants et 16 autres avec une perte auditive, le seuil différentiel de la durée pour le bruit (noise bursts), le seuil différentiel du silence dans le bruit (la durée minimale du silence perceptible) et le seuil différentiel pour le VOT. Ici, nous nous intéressons uniquement aux résultats des normo-entendants. Pour le seuil différentiel de la durée du bruit, la durée du stimulus était de 30 ms, laquelle augmentait par pas de 2 ms. Le stimulus était émis selon deux fréquences 500 et 4000 Hz combinées avec deux niveaux d'intensité de 90 et 50 dB. Tyler *et al.* affirmaient que la fréquence n'avait pas d'influence sur le seuil de la durée alors que l'intensité avait de l'influence. En effet, le seuil était plus élevé pour des niveaux d'intensité relativement bas. Cette hausse d'intensité n'influait pas le seuil indéfiniment. D'autres études (Creelman, 1962; Abel, 1972a), ont montré qu'avec un niveau d'intensité assez élevé la variation de l'intensité n'influait pas le seuil. Pour Sinnott *et al.* rappelons-le, ce niveau était de 40 dB, au-delà la hausse de l'intensité n'a produit aucun effet. Ainsi, nous nous intéressons seulement aux résultats pour le niveau d'intensité élevé dans l'étude de Tyler *et al.*

Tyler *et al.* ont trouvé finalement, avec un niveau d'intensité élevé, que le seuil pour un bruit de 30 ms était de 19 ms, soit un rapport de Weber est de 0,63, ce qui est très différent de ce qu'ont trouvé Sinnott *et al.* (1987a).

Pour le seuil concernant le silence, les signaux entourant le silence étaient des bruits qui avaient une durée de 1000 ms. Le silence commençait 500 ms après le début du stimulus, ce silence avait une durée initiale soit de 30 ms soit de 100 ms, et les durées augmentaient par pas de 2 ms. Le seuil trouvé était de 51 ms pour le silence de 30 ms, et de 73 ms pour le silence de 100 ms. Pour la détection de la durée minimale d'un silence, les chercheurs demandaient aux sujets de leurs indiquer le stimulus, constitué de bruit, qui contenait du silence. Les stimuli avaient une durée de 1000 ms, dont 500 ms avant le début du silence. Les expérimentateurs aménageaient un silence de 80 ms, lequel se réduisait par pas de 2 ms. Les résultats ont montré que cette durée minimale était de 10 ms.

En ce qui concerne le seuil du VOT, le travail a porté sur deux continua de syllabes synthétisées, de [ba] à [pa] et de [bi] à [pi]. Les valeurs théoriques des VOT de ces deux syllabes sont respectivement 0 et 60 ms. Il y avait deux durées initiales de VOT, soit 15 ms,

soit 30 ms, lesquelles augmentaient par pas de 2 ms. Pour le continuum [ba] à [pa], lorsque la valeur initiale du VOT était de 15 ms, le seuil était de 11 ms (rapport de Weber est de 0,73), et lorsque la valeur était 30 ms le seuil était de 30 ms (rapport de Weber est de 0,73). Tandis que pour le continuum [bi] à [pi], lorsque la valeur initiale du VOT était de 15 ms le seuil était de 23 ms, et lorsque la valeur était de 30 ms, le seuil était de 35 ms.

Notre lecture des résultats de cette étude nous permet d'observer deux phénomènes: 1) les seuils qui jouent sur deux durées initiales différentes (silence et VOT) mettent en évidence le fait qu'il n'y a pas de rapport (seuil/durée initiale) qui vaut pour les deux durées initiales, mais à chaque durée correspondra un rapport. La valeur de ce seuil est d'autant plus importante que la durée initiale est importante, donc pas de valeur absolue, mais une valeur relative.

Silence: 51/30, et 73/100.

VOT: 11/15, et 30/30.

2) plus la structure acoustique du signal est complexe et plus la valeur du seuil est basse:

Seuil différentiel du silence > seuil différentiel du VOT > seuil différentiel du bruit, respectivement 51, 30 et 19 ms.

En ce qui concerne le seuil différentiel de la durée d'un stimulus, cette étude, comme celle de Sinnott *et al.* (1987a), ne nous renseigne pas sur le seuil pour d'autres durées que 15 et 30 ms. Elle ne nous fournit pas une fonction qui nous permet de prédire le seuil à partir de la durée. C'est pour cela que nous poursuivons notre recherche avec d'autres études qui ont essayé d'établir le seuil de la durée en fonction de la durée initiale.

Fitzgibbons & Gordon-Salant (1994) avaient pour objectif d'examiner un sujet controversé qui est l'effet de l'âge sur le déficit de la sensibilité de l'audition en ce qui concerne le traitement du temps. Ainsi l'étude a consisté en la comparaison des facultés des personnes jeunes adultes (20-40 ans) et des personnes âgées (65-76 ans) en ce qui concerne la discrimination de la durée. Les sujets étaient soit avec, soit sans perte auditive. Les stimuli étaient des bruits qui comportaient deux fréquences 500 et 4000 Hz. Le test a porté sur le seuil différentiel du bruit, pour deux durées: 250 et 6,4 ms, d'une part; et sur le seuil différentiel de la durée du silence ménagé à l'intérieur de ces bruits, d'autre part. Ici nous nous intéressons aux seuils chez les jeunes normo-entendants seulement.

Leurs résultats montrent que chez les normo-entendants l'âge a un effet sur la perception de la durée, en ce sens que le seuil est significativement plus élevé chez les personnes les plus âgées de l'étude. En effet, ils trouvent que le rapport de Weber est de 0,19 chez les jeunes alors qu'il est de 0,24 chez les plus âgés. Ils attestent par ailleurs que la fréquence du stimulus a un effet chez les personnes les plus âgées mais pas chez les jeunes. Toujours chez les jeunes, et contrairement à Tyler *et al.* (1982), ils trouvent qu'il n'y a pas de différence de seuil entre bruit et silence. Le seuil pour le bruit et pour le silence est de 48 ms pour une durée de 250 ms chez les jeunes normo-entendants, avec un rapport de Weber de 0,19. Concernant le silence, ces résultats sont en accord avec Creelman (1962), Abel (1972a) qui a examiné la capacité de l'oreille à juger la différence entre les durées de deux intervalles de silence, a trouvé que le rapport de Weber était de 0,20 pour les durées supérieures à 40 ms. Elle affirmait l'existence d'une différence entre le seuil pour le bruit et celui pour le silence pour des durées identiques, alors que, rappelons-le, Fitzgibbons & Gordon-Salant (1994) ne trouvent pas de différence.

Contrairement aux études qui traitent du seuil de la durée que nous avons examinées plus haut (Fitzgibbons & Gordon-Salant, 1994; Sinnott *et al.*, 1987a; Tyler *et al.*, 1982), lesquelles ne mettent en évidence le rapport entre la durée initiale (T) et le seuil (ΔT) que pour une seule valeur à chaque fois, et non pas pour un intervalle de durées possibles, les travaux d'Abel (1972a) pour le seuil du silence et d'Abel (1972b) pour le bruit introduisent la valeur de T comme une valeur déterminante pour le seuil. Nous examinons cette étude, non pas pour les valeurs exactes de ses résultats concernant le silence, mais pour le fait que ces résultats mettent en évidence l'effet de la durée initiale sur ΔT .

Dans l'étude d'Abel (1972a) qui a porté sur le silence, les durées des silences allaient de 0,63 à 630 ms en 11 pas où la valeur suivante était le double de celle qui la précédait. Nous ne nous intéressons pas à des seuils, que ce soit pour le bruit ou pour le silence, pour des durées très petites (inférieures à 40 ms), parce que les petites valeurs ne sont pas observées dans la parole et par conséquent n'intéressent guère notre recherche (c'est pour cela que dans nos représentations graphiques nous ne représentons que les valeurs à partir de 40 ms).

Les intervalles de silence étaient bordés par un bruit dont les fréquences s'étalaient de 0 à 20 kHz, avec des durées de 300 ms et des intensités de 70 dB. L'auteur de l'étude a trouvé que pour des durées comprises entre 10 et 160 ms, $\Delta T = a \cdot (T)^b$; où $a = 7,85$ et $b = 0,35$. En changeant la durée et l'intensité du bruit qui délimitaient ces intervalles de silence, ΔT

changeait aussi. Les résultats de l'étude montraient que le seuil était d'autant plus petit que l'intensité du bruit des marqueurs était importante et que la durée des marqueurs diminuait. Parmi les conditions étudiées, celle où la durée était de 10 ms et l'intensité de 80 dB produisait le seuil le plus bas, et c'est celle-ci que nous avons choisi de représenter dans la figure 34. Cette représentation montre clairement l'effet de la durée initiale T sur le seuil (ΔT).

Une autre étude d'Abel (1972b) avait pour but de déterminer la capacité de l'oreille à discriminer la hausse de la durée. Elle a consisté en la comparaison de deux durées: (T) et ($\Delta T + T$), lesquelles étaient émises avec 500 ms d'intervalle inter-stimulus. Les durées des stimuli variaient entre 0,16 et 960 ms en 14 pas. Les fréquences des stimuli variaient entre 20 et 3500 Hz, et leurs intensités variaient entre 65 et 85 dB. Les résultats ont montré que pour toute nature de stimuli, ΔT était relatif à la racine carrée de T. Ce dernier résultat nous semble trop éloigné de ce que nous avons pu observer dans d'autres études traitant de la même question, pour que nous lui accordions beaucoup de considération. C'est pour cela que nous avons abandonné l'idée de la représentation graphique de ΔT en fonction de T selon Abel (1972b). Par ailleurs, l'analyse de la variance a révélé que ΔT ne dépendait que de T. En effet, si l'effet de T était significatif, les effets de la largeur de la bande, de la forme de l'onde, de l'intensité (dans la limite de 65 et 85 dB) et du sujet n'avaient pas d'effets significatifs sur la valeurs de ΔT .

De même qu'Abel (1972b) avait tenté d'établir ΔT en fonction de T, Sinnott *et al.* (1987a) ont essayé d'établir cette fonction avec des sinusoïdes dont la plage de fréquence n'excédait pas un octave. Les sinusoïdes étaient centrées autour de 2000 Hz (1/3 au-dessus et 1/3 au-dessous) sans qu'il y ait d'harmoniques. Leur but était de comparer ΔT chez l'homme et chez le singe afin de savoir si l'homme avait perdu, au cours de l'évolution, certains de ses pouvoirs auditifs, tel le seuil différentiel de la perception de la durée des sons. ΔT était mesuré avec des signaux de 2000 Hz chez le singe et chez l'homme. Cette étude examinait les effets des trois facteurs suivants sur la perception du ΔT :

- 1) variation de la durée du stimulus.
- 2) variation de l'intensité du stimulus.
- 3) comparer la hausse vs. la baisse de la durée.

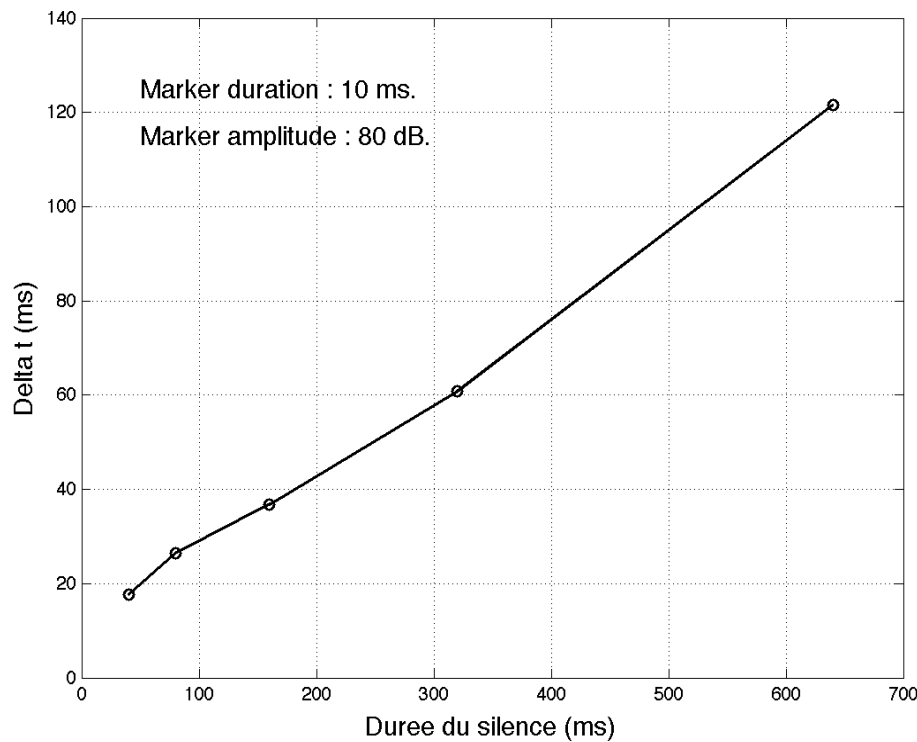


Figure 34. Seuil différentiel de la perception du silence en fonction de la durée du silence, d'après Abel (1972a).

L'expérience avait trois conditions:

1) intensité constante à 70 dB et delta T étant étudié dans le sens de la hausse pour T = 25; 50; 100; 200; et 400 ms.

2) delta T était étudié comme une fonction de l'intensité de 30 à 70 dB dans le sens de la hausse, avec T = 25 et 200 ms (figure 29, plus haut).

3) delta T était étudié dans le sens de la baisse pour T = 100; 200; 400 et 800 ms.

Nous allons suivre ci-après seulement l'étude de la première condition pour l'effet du premier facteur (T), chez les humains. Nous obtiendrons ainsi delta T en fonction de T.

Les résultats de la première condition sont représentés dans la figure 35. Delta T est d'autant plus important que T est important. Le seuil chez l'homme est plus fin que chez le singe. Le rapport de Weber est, en moyenne, égal à 0,15 chez l'homme. Pour les mêmes données ce rapport chez le singe est de 0,37. Le seuil chez le singe est de 2 à 2,5 fois celui de l'homme. Cependant, chez les deux espèces, ce rapport décroît plus T augmente (le rapport le plus bas était enregistré avec les durées de 200 et 400 ms). Ces résultats restent comparables à d'autres résultats (Henry, 1948). Small & Campbell (1962), dont l'étude a porté sur la même

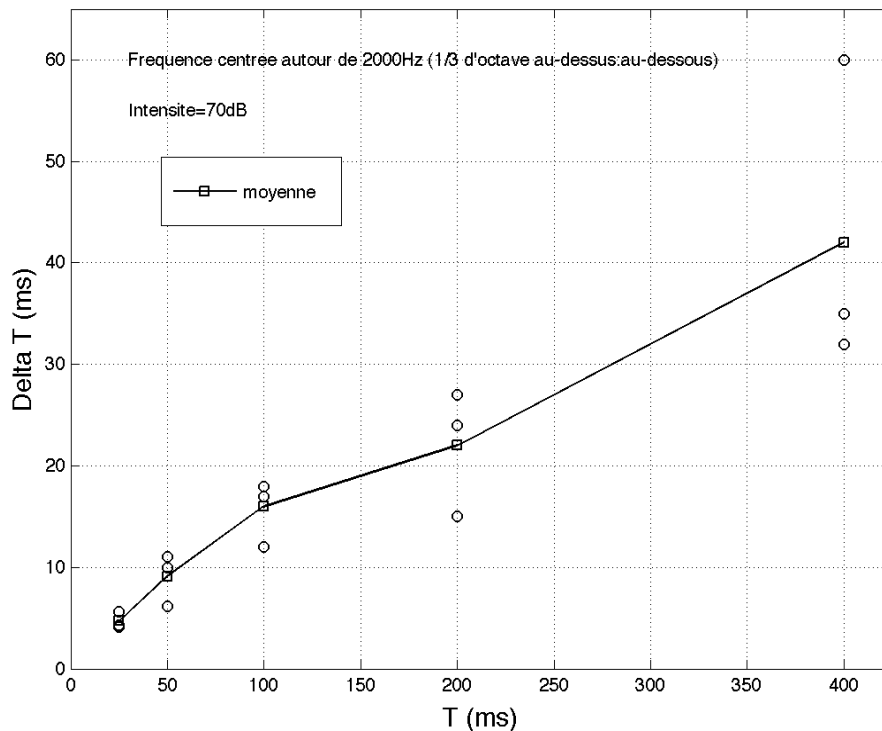


Figure 35. Seuil différentiel de la perception de la durée en fonction de la durée du stimulus (d'après Sinnott et al., 1987a).

question donnent un rapport de Weber de 0,10 à 0,15 pour T variant de 200 à 400 ms, et un rapport de Weber de 0,15 à 0,20 pour T variant de 25 à 50 ms. Ou encore, comme vu plus haut, avec les résultats de Fitzgibbons & Gordon-Salant (1994) pour les jeunes normo-entendants où le rapport de Weber est de 0,19 pour T= 250 ms.

I.2.2. Seuil pour les segments vocaliques.

I.2.2.1. Fréquence variable

Nous cherchons à savoir si la variation de la fréquence au cours de l'émission d'un segment vocalique influence le seuil de perception de la durée du segment vocalique en question. La littérature sur les seuils en ce qui concerne la transition vocalique - contrairement à celle concernant les émissions vocaliques stables - est abondante. Elle nous permettra, par la suite, au cas où il n'y aurait pas de différence entre ces deux phases des segments vocaliques, la transition et la partie stable, de généraliser ces résultats, et d'avoir ainsi plus de réponses concernant les émissions vocaliques stables.

Abondante, parce que la durée de la transition revêt de l'importance pour l'effet qu'elle a sur la perception des unités phonétiques adjacentes. Effectivement, la durée de la transition

formantique peut changer la perception de la consonne qui la précède. Liberman *et al.* (1956) ont montré qu'avec un continuum synthétique de [bɛ] à [wɛ], les locuteurs percevaient une consonne occlusive voisée quand la durée de la transition était inférieure à 40 ms et que cette consonne était perçue comme un glide quand la transition était supérieure à 40 ms. Miller & Liberman (1979) montrent que les transitions formantiques sont identifiées en tant que [b] ou en tant que [w] en fonction de la durée de la voyelle post – consonne. Ils affirment, en effet, qu'une même transition de 40 ms sera perçue [b] si la durée de la voyelle est 300 ms alors qu'elle sera perçue [w] quand la durée de la voyelle est de 80 ms. Ainsi que nous l'avons vu plus haut (paragraphe I.2.1.1.1.), Carré *et al.* (2001) ont pu montrer que la perception de la diphtongue synthétique [ai] s'entendait [ai] quand la durée de la transition qui liait les deux voyelles était inférieure à 200 ms, et s'entendaient comme une réalisation de [aei] lorsque la durée de cette même transition était supérieure à 200 ms.

Ainsi, nous nous sommes intéressés à un travail de recherche de l'institut des Sciences Phonétiques de l'université d'Amsterdam, mené par Wieringen & Pols (1994), qui a porté sur le seuil différentiel de la perception de la durée dans les transitions formantiques montantes et descendantes pour un seul formant et pour deux régions de fréquence (entre 200 et 700 Hz et entre 500 et 1000 Hz).

La discrimination de la durée d'un glide, qui change de fréquence, dépend du centre de la région de la fréquence, de l'étendue de la fréquence, de la durée du glide, de la présence ou de l'absence d'une zone stable et de la direction de la transition. À cause de l'interdépendance de la fréquence, de la durée et de la pente de la transition déterminée par le rapport Hz/s (Hertz par seconde), il est difficile d'attribuer la raison de la discrimination à la valeur fréquentielle de la fin de la transition, à la pente de la transition ou à la différence entre la valeur finale et la valeur initiale de la fréquence.

Wieringen & Pols (1994) ont déterminé le seuil de la transition (i) en variant l'étendue de la fréquence avec une même durée et (ii) en variant la durée de la transition avec une même étendue de fréquence. La variation de la pente de la transition était obtenue par le changement de la durée de la transition (20, 30 et 50 ms), en gardant une même étendue de fréquence.

Résultats. Quand la durée augmente, le seuil augmente (tableau 7). Les résultats ne dépendent pas de la région de fréquence, comme il n'y a pas de différence significative entre les sens des transitions (montante/descendante).

Durée de la transition		20 ms	30 ms	50 ms
		Débit		
étendues de fréquence	220 => 550 Hz	16,5 Hz/ms	11,0 Hz/ms	6,6 Hz/ms
	520 => 850 Hz			
	220 => 650 Hz	21,5 Hz/ms	14,3 Hz/ms	8,6 Hz/ms
	520 => 950 Hz			
	220 => 750 Hz	26,5 Hz/ms	17,6 Hz/ms	10,6 Hz/ms
	520 => 1050 Hz			
	Seuil:	SD-débit (Hz/ms)	3,4 Hz/ms	2,2 Hz/ms
La moyenne toutes régions de fréquence.	SD-durée (dynamique)	2,7 ms	4,5 ms	4,9 ms
	SD-durée (stationnaire)	—	—	3 ms

* SD=Seuil différentiel

Tableau 7. Seuil différentiel du débit (Hz/ms) et de la durée de la transition vocalique, pour trois durées différentes 20, 30, et 50 ms. Nous avons, en outre, le seuil pour un segment stable de 50 ms. D'après Wieringen & Pols (1994).

L'Anova a montré qu'il y avait un effet significatif de la durée de la transition sur le seuil de la perception de la durée de la transition (figure 36), et sur le seuil de la pente de la transition. Alors que les effets du sujet, de la région de la fréquence, de la direction de la transition (montante *vs.* descendante) et de l'étendue de la fréquence (30, 400 ou 500 Hz) ne sont pas significatifs.

Les résultats de cette étude nous permettent de constater que plus la durée de la transition est importante et plus le seuil est important, sans qu'il y ait toutefois un rapport évident, comme cela a été remarqué pour des sons non harmoniques et qui n'appartenaient pas à la parole (Fitzgibbons & Gordon-Salant, 1994; Sinnott *et al.*, 1987a; Tyler *et al.*, 1982). Ainsi notre question concernant la durée minimale requise pour donner naissance à une nouvelle sensation jugée plus longue, n'a pas de réponse finie, parce que les deux facteurs, temporel et spectral (la pente de la transition, en Hz/ms), interviennent dans la détermination du seuil de

la durée de la transition. Et il n'est pas possible de donner une réponse à la question: lequel des deux facteurs est le facteur responsable de cette discrimination.

Pour essayer de trouver une réponse à cette question, Wieringen & Pols (1994) ont posé l'hypothèse suivante: si la variation de la fréquence dans la transition contribue à la perception du seuil de la durée, alors le seuil de la durée du stimulus stationnaire doit être plus important (moins fin) que le seuil de la transition; sinon, le seuil de la durée sera le même pour les deux stimuli, dynamique et stationnaire.

Les auteurs de l'étude ont établi le seuil de la durée d'une transition stationnaire où la durée du stimulus était de 50 ms. Leurs résultats ont montré, contrairement aux attentes, que le seuil était plus fin quand le stimulus était stationnaire. Ils ont trouvé qu'il était de 3 ms, soit un rapport de Weber de 0,06; alors que pour la même durée, rappelons-le, ce rapport pour la transition valait 0,1. Cela signifie que la variation de la fréquence au cours du temps n'aide pas à affiner le seuil de la durée, mais qu'elle a plutôt un rôle négatif, puisqu'elle le rend plus important.

Nous concluons en disant que la discrimination est basée essentiellement sur le débit de la transition (Hz/ms) plutôt que sur sa durée. En effet, deux transitions ayant le même débit mais deux durées différentes, ont le même seuil. Le seuil était le même pour toutes les fréquences, pour toutes les régions de fréquence et pour les deux sens de variation (hausse vs. baisse).

Ce qui nous intéresse concrètement dans cette étude, c'est le seuil de la durée d'une voyelle stationnaire, c'est-à-dire lorsque la fréquence de départ et celle d'arrivée sont les mêmes (cas des voyelles monophthongues), et c'est parce que le rôle de la variation de la fréquence dans la perception du seuil est éliminé. Nous retenons alors que le seuil pour des segments vocaliques dynamiques et des segments vocaliques stables n'est pas le même, et que pour ces derniers le seuil (non linguistique) pour une durée de 50 ms est de 3 ms.

Les résultats de Wieringen & Pols (1994), quant au rôle négatif de variation de la fréquence sur le seuil de la durée, ne sont pas unanimes. En effet, dans une étude de Dooley & Moore (1988), les chercheurs comparaient le seuil différentiel de la durée (750 ms) pour des sinusoïdes stables (fréquence = 2000 Hz et intensité = 65 dB ne changeant pas entre le début et la fin) et des sinusoïdes qui avaient soit la fréquence (d'une amplitude de 100 Hz), soit l'intensité (d'une amplitude de 10 dB), soit les deux à la fois qui variaient uniformément du

début du stimulus à sa fin. 11 conditions d'expérimentation au total, 1 où tout était stable et 10 autres où fréquence et intensité subissaient des modifications.

Les résultats montraient que le seuil pour les sinusoïdes qui variaient était inférieur à celui du son stable, ce qui permettait de penser que la variation peut être utilisée par les sujets comme un indice de durée. Le seuil pour la condition stable était de 47 ms pour un stimulus de 750 ms, le rapport de Weber est de 0,06, le même que celui trouvé par Wieringen & Pols (1994), alors que le seuil était de 24 à 34 ms pour les 10 autres conditions, donc inférieur au seuil de la condition (1). L'analyse de la variance a montré que le seuil pour la condition (1) était significativement différent du seuil des 10 autres conditions. Cette étude montrait finalement que la variation de l'intensité et de la fréquence au cours de l'émission du stimulus pouvaient être intégrés sous forme de temps. Ainsi la perception de la durée n'était pas véhiculée que par le temps mais par la variation aussi.

Les résultats de l'étude de Wieringen & Pols (1994) et ceux de l'étude de Dooley & Moore (1988) ne vont pas du tout dans le même sens. La première affirme que la variation de la fréquence au cours de l'émission d'une transition vocalique augmente la valeur du seuil. Tandis que la deuxième montre que la variation de la fréquence et de l'intensité d'un son au cours de son émission baisse la valeur du seuil. Le rapide examen des études ci-dessus nous dévoile que, quoique divergentes quant au rôle de la variation sur le seuil, ces études s'accordent sur le fait que le seuil pour un stimulus dynamique et le seuil pour un stimulus stationnaire ne peuvent être les mêmes. Finalement, la variation de la fréquence au cours de l'émission d'un segment vocalique influence le seuil de perception de la durée du segment vocalique, et ne permet pas de généraliser les résultats à partir de ce qui est variable pour ce qui est stable. C'est pour cela que nous abandonnons la recherche du seuil sur des stimuli dynamiques, et nous cherchons par conséquent une réponse avec des signaux stationnaires.

I.2.2.2. Fréquence stable

Dans ce paragraphe nous poursuivons notre recherche avec des études qui se sont basées sur des stimuli où il n'y a que la dimension temporelle qui change. La question du seuil reste une question difficile en raison de l'interdépendance des facteurs qui en sont responsables. Watson & Foyle (1985) ont montré que la discrimination de la variation de la durée du stimulus était plus faible avec des stimuli de longue durée, par rapport à des stimuli simples de plus petite durée. Et ils ont montré que cette discrimination était fonction de plusieurs facteurs tels l'entraînement des sujets, le degré de complexité du stimulus, la

familiarité du sujet avec ce genre de tâches et le niveau d'ambiguïté en ce qui concerne les caractéristiques acoustiques du stimulus.

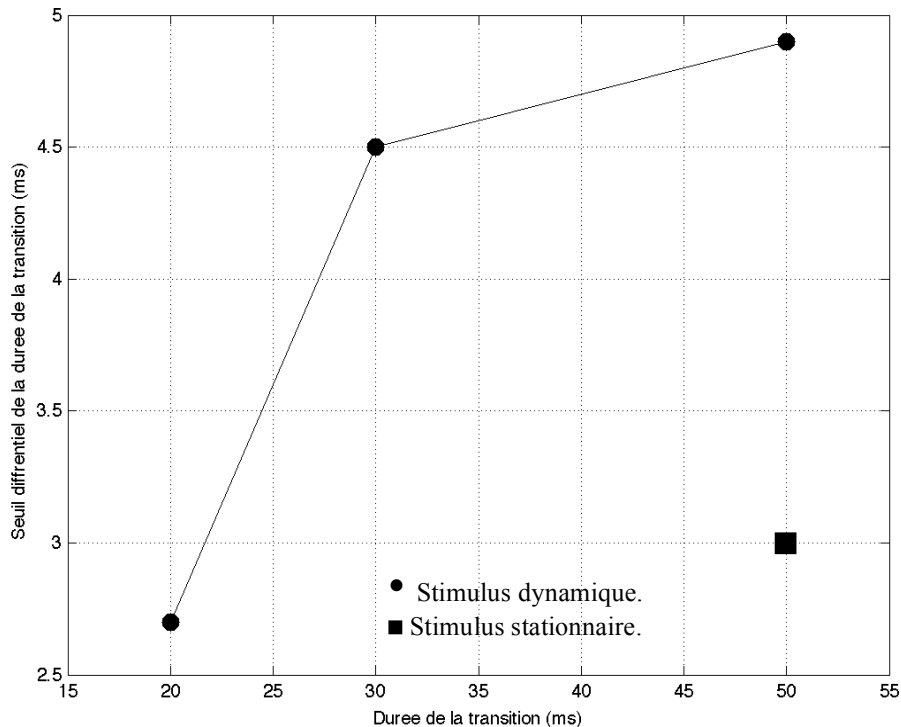


Figure 36. Seuil différentiel de la durée de la transition vocalique en fonction de la durée. D'après Wieringen & Pols, (1994).

Une étude qui nous a semblé pertinente pour notre recherche, est celle de Fitzgibbons & Gordon-Salant (1995), dont l'objectif n'était pas l'étude du seuil en soi, mais la détermination, pour le seuil différentiel de la perception de la durée, du rôle de l'âge, de la perte auditive et de la complexité des stimuli (présentation en isolation vs. présentation dans une séquence). Les sujets de cette étude étaient divisés en 4 groupes, 10 sujets par groupe: 1) Jeunes personnes (20-40 ans) normo-entendants; 2) jeunes personnes avec une perte auditive; 3) Personnes âgées (65-76 ans) normo-entendants et 4) Personnes âgées avec une perte auditive. Nous suivrons seulement le premier groupe, les jeunes normo-entendants, notre but étant de trouver le seuil pour la perception de la durée, sans les perturbations occasionnées par l'âge ou par la perte auditive.

Le premier groupe était amené à identifier les changements temporels pour: 1) les bruits de 250 ms; 2) les intervalles de silence qui étaient présentés dans des contextes de

bruits qui n'avaient pas la même durée. Le contexte de bruit était formé de 5 séquences de 250 ms chacune.

Dans l'étude de Fitzgibbons & Gordon-Salant (1995) les seuils du bruit étaient mesurés avec:

(1) *des stimuli simples (condition de base)*: constitués de bruits d'une fréquence de 4000 Hz et d'une durée de 250 ms. Les seuils pour le silence étaient mesurés pour des intervalles qui se trouvaient entre une paire de bruits d'une fréquence de 4000 Hz et d'une durée de 250 ms.

(2) *des stimuli complexes*: chacun étant une concaténation de 5 stimuli simples non séparés, soit un stimulus de 1250 ms. Un seul stimulus simple au sein du stimulus complexe subissait un changement. La fréquence était d'un tiers d'octave centrée géométriquement autour de 4000 Hz.

La tâche de l'identification était rendue plus complexe par plusieurs procédés:

1) en donnant une fréquence différente à chacune des 5 séquences au lieu d'une seule, la séquence cible restant la troisième. 2) en changeant l'arrangement des séquences à chaque présentation et en gardant la séquence cible en troisième position. 3) en changeant l'ordre de toutes les séquences y compris la séquence cible. Ces mêmes conditions étaient dupliquées pour le seuil de la durée du silence.

Nous nous intéressons seulement à la condition de base, parce que l'Anova a révélé, en ce qui concerne le seuil du bruit, que les jeunes normo-entendants (la population que nous suivons) n'étaient pas influencés par le niveau de complexité des stimuli. Par conséquent l'effet du niveau de complexité ne nous intéressera pas. En effet, le fait que le seuil différentiel de la durée du son chez les jeunes sujets avec une audition normale ne change pas avec la complexité de la condition expérimentale, ce fait nous permet de dire que si le seuil change en fonction de la complexité du stimulus, cela proviendrait soit de l'âge soit de la perte auditive. La complexité avait par ailleurs un rôle significatif sur la durée du silence. Le seuil de la durée du silence, contrairement à celui du bruit, était effectivement affecté par la complexité du stimulus, ce qui laisse penser que, même si ces deux seuils ont des valeurs très proches, le seuil du silence est moins robuste que celui du bruit. L'analyse de la variance sur toutes les conditions a montré que le seuil différentiel de la durée du silence était plus important que son correspondant pour le bruit, qu'il y avait une différence significative entre les conditions, et que l'âge avait un effet significatif sur le seuil.

Pour les stimuli présentés en isolation (condition de base) chez les jeunes normo-entendants, le seuil différentiel de la perception de la durée du bruit était de 46,8 ms, le rapport de Weber ($\Delta T/T$) était de 0,19. Pour le silence, le seuil est de 48,5 ms, le rapport de Weber est de 0,19 aussi. D'autres études (Creelman, 1962; Small & Campbell, 1962) pour le bruit, Abel (1972b) pour le silence, rapportent une valeur similaire pour le rapport de Weber, autour de 0,2.

Contrairement au seuil différentiel de la durée du son, chez les jeunes sujets avec une audition normale, le seuil différentiel de la durée du silence change avec la complexité de la condition expérimentale. Effectivement, le rapport de Weber variait entre 0,2 (pour la condition de base) et 0,45 (pour les conditions plus complexes). Chez les jeunes, moins la condition est complexe et moins le rapport de Weber est élevé.

Les résultats de cette étude montre que, globalement, les jeunes ont un seuil plus fin (plus petit) que les plus âgés, le seuil chez les plus âgés est plus facilement affecté que chez les jeunes par la complexité du stimulus.

Généralement le seuil du son est plus robuste que le seuil du silence.

L'étude de Fitzgibbons & Gordon-Salant (1995), qui était menée sur un large échantillon, et qui a mis en évidence l'effet de l'âge sur le seuil, ne fournit le seuil que pour une valeur bien précise, 250 ms, et non pas pour un intervalle de durées possibles, et par conséquent ne peut donner pour constituer une réponse à notre question. En revanche, Rossi (1972) a essayé d'établir le seuil différentiel de la perception de la durée vocalique pour un intervalle allant de 60 à 290 ms.

I.2.3. Étude linguistique du seuil

Nous abordons dans ce paragraphe des études qui ont essayé de rapporter le seuil de la durée pour des indices acoustiques directement liés à la perception de la parole. L'étude de Rossi (1972: 445) a établi des données sur un intervalle de durées pour les voyelles naturelles isolées dans le domaine de la parole. Il admet qu'entre 60 ms et 140 ms, le seuil est à peu près constant en valeur absolue et vaut approximativement 30 ms. Il affirme qu'entre 130 et 290 ms, la loi de Weber ($\Delta T/T$) est vérifiée, le seuil est alors de 22,5% en moyenne. Pour des durées supérieures à 300 ms, Rossi montre que $\Delta T/T$ augmente très rapidement, mais il accorde peu de considération à cette catégorie de durées, il justifie cela par le fait que ces durées existent rarement pour des voyelles de la parole naturelle. Les résultats de Nishinuma (1977) sur le seuil, où il utilisait des syllabes naturelles CV qu'il présentait à des

locuteurs japonais, sont comparables à ceux de Rossi. Il y a ici tout de même quelques points qui ne nous semblent pas clairs dans l'étude de Rossi (1972).

- Concernant le seuil. Nous ne savons pas quel est le seuil pour les voyelles qui ont une durée entre 130 et 140 ms: est-ce que c'est la valeur absolue de 30 ms, ou c'est la loi de Weber qui le détermine?

- Concernant le calcul. Si nous calculons le seuil pour un son d'une durée de 120 ms, cette durée appartient à la catégorie où le seuil est absolu et d'une valeur de 30 ms. Ainsi cette durée devra être de 150 ms pour qu'une nouvelle sensation de durée naisse. Faisons le calcul pour cette durée que nous venons d'obtenir pour savoir de combien il faudra la baisser pour qu'une nouvelle sensation naisse. 150 ms: cette durée rentre dans la catégorie des durées dont le seuil se calcule selon la loi de Weber $\Delta T/T = 22,5\%$, la durée devient $(150 - x)/150 = 0,225$, et $x = 116$, cela implique qu'il faut effectuer une baisse de 33,75 ms, alors que lorsque nous l'avons calculée, dans le sens de la hausse, rappelons-le, c'était 30 ms. Donc nous remarquons qu'il y a un décalage pour un même écart entre deux mêmes durées, un décalage qui ne doit pas exister normalement.

La première partie des résultats de l'étude de Rossi (seuil constant = 30 ms pour les durées entre 60 et 140 ms) reste éloignée de ce que nous avons pu observer dans la littérature des seuils. Nous rappelons que Wieringen & Pols (1994) ont trouvé que le rapport de Weber pour un segment vocalique de 50 ms de pseudo parole était de 0,06. Quant à la deuxième partie, elle est raisonnablement comparable aux résultats de Eliers *et al.* (1984) pour les adultes, Sinnott *et al.* (1987a) pour les courtes durées, et à ceux de Fitzgibbons & Gordon-Salant (1994, 1995) pour les jeunes normo-entendants.

Une autre étude dans le domaine de la parole était effectuée par Dorman *et al.* (1985). Les auteurs de l'étude se sont basés sur le continuum "SHOP-CHOP" qui fonde son contraste uniquement sur la durée de la fricative [ʃ]. Dorman *et al.* variaient la durée de la fricative [ʃ] entre 60 ms (perçu [tʃ]) et 140 ms (perçu [ʃ]), en réduisant le bruit du [ʃ] de "SHOP", prélevé au milieu de la consonne, par pas de 10 ms, produisant ainsi 9 stimuli qui ne différaient que par la durée de la friction. Les signaux manipulés étaient répétés 10 fois par stimulus et présentés dans un ordre aléatoire, Dorman *et al.* ont trouvé que la frontière catégorielle, à 50 % de réponses "SHOP", se situait à 94 ms. Il a donc fallu 34 ms pour passer d'une catégorie à une autre, en obtenant le rapport de Weber (34/60) de 0,57. Pour nous, comme nous l'avons argumenté plus haut (en I.2.1.1.1.), nous obtenons ce rapport lorsque nous rapportons cette

valeur (34 ms) sur la durée de la syllabe " CHOP " et non pas sur la durée de la consonne [ʃ], la durée de cette syllabe n'étant pas fournie dans l'étude.

Cette étude montre que les seuils des indices temporels de la parole sont en accord avec les seuils obtenus pour des signaux non parole, quand les conditions d'expérimentation sont similaires.

Nous avons cherché une réponse à notre question, concernant la durée requise pour le basculement catégoriel, dans l'étude effectuée par Gordon-Salant *et al.* (2006). Cette étude examinait l'effet de l'âge sur la perception des indices temporels dans la parole naturelle. Son objectif était de séparer les effets de la perte auditive par rapport aux effets de l'âge dans la perception des différences temporelles dans les signaux de parole. Cette étude nous a particulièrement intéressée parce que les auteurs manipulaient des indices acoustiques qui variaient uniquement sur la dimension temporelle dans des continua séparés. Ces indices étaient: (1) le VOT (2) la durée de la voyelle (3) la durée du silence et (4) la durée de la transition formantique. Ainsi, nous examinerons cette étude de très près, en décrivant en détail la procédure expérimentale et les résultats.

Le but étant de déterminer l'effet de l'âge, les sujets de cette étude appartenaient à deux tranches d'âge. Les jeunes (moyenne d'âge 21,4 ans) identifiaient des signaux auxquels les expérimentateurs avaient superposé un faible bruit de 5 dB, ce bruit était présenté tout le long de l'expérience. Ayant trouvé au préalable que le seuil d'audition chez les jeunes était plus fin que celui des personnes plus âgées, cela avait pour effet d'égaliser les seuils d'audition chez les jeunes et les âgés, et ne permettre qu'à une seule dimension, le temps, d'être derrière la performance des sujets. Le deuxième groupe était formé de personnes âgées (71 ans, en moyenne) avec une audition normale, alors que le troisième groupe était formé de personnes âgées (72,5 ans, en moyenne) avec une perte auditive.

Une analyse de la variance sur les groupes a montré que l'identification chez les personnes âgées était différente de celle des jeunes pour le traitement de la durée du silence et celle de la transition, mais pas pour les continua du VOT et de la voyelle. Pour les continua où il y avait de la différence entre jeunes et vieux, les jeunes basculaient dans la catégorie cible plus vite que les personnes âgées. Mais comme nous ne nous intéressons pas à l'effet de l'âge ou à celui de la perte auditive, nous ne suivrons que le travail de recherche effectué sur les jeunes normo-entendants

Nous décrivons tout d'abord les procédés expérimentaux qui ont permis l'établissement des stimuli compris entre les deux extrémités du continuum.

(1) VOT: En ce qui concerne la manipulation du VOT, les expérimentateurs ont choisi le continuum BUY vs. PIE [baɪ vs. paɪ] où les deux extrêmes du continuum ne différaient qu'au niveau du VOT. Le VOT était de 0 ms pour BUY (le relâchement du [b] coïncide avec la mise en vibration des cordes vocales) et de 60 ms pour PIE. Le stimulus de départ était la prononciation naturelle de BUY où le VOT est nul. Le bruit d'aspiration était importé de l'émission naturelle du stimulus PIE, et était inséré ensuite par pas de 10 ms au passage par zéro entre le bruit du relâchement du [b] et l'attaque du voisement de la voyelle du stimulus BUY. Cette opération était répétée 6 fois, ainsi le septième stimulus avait un VOT d'une durée de 60 ms, lequel était perçu PIE. Donc le VOT augmentait de 0 ms (BUY) à 60 ms (PIE). La durée totale du premier stimulus était de 310 ms et la durée totale du dernier stimulus était de 310 + 60, soit 370 ms.

(2) Durées des voyelles: l'opposition WHEAT vs. WEED en anglais étant basée sur la durée de la voyelle [i]. Le stimulus de départ était l'émission naturelle du mot WEED, le bruit de relâchement du [d] était coupé et remplacé par le bruit de relâchement du [t] provenant de l'émission naturelle du mot WHEAT. Le stimulus obtenu était un stimulus hybride. Ce stimulus était le plus long du continuum, et il restait perçu comme WEED alors qu'il était phonétiquement "WEET ". Ceci a prouvé que l'opposition WHEAT vs. WEED reposait sur la durée de la voyelle, [i] vs. [i:] et non pas sur l'opposition [t] vs. [d]. Le continuum était créé en plusieurs étapes, le stimulus qui était phonétiquement "WEET " était raccourci progressivement en coupant un cycle (T=7-8 ms) entre deux passages par zéro de la zone stable de la voyelle à chaque étape. Le raccourcissement se faisait en 9 étapes discrètes où la voyelle suivante avait un cycle de moins que sa précédente. Le dernier stimulus, la voyelle la plus courte, était perçu comme WHEAT, la durée totale de ce stimulus était égale à 249 ms avec une zone stable de la voyelle de 93 ms. Le stimulus de départ, phonétiquement "WEET " mais perçu WEED, avait une durée égale à 311 ms avec une durée de la zone stable de la voyelle de 155 ms, soit un raccourcissement de 62 ms (de 155 à 93, par pas de 7-8 ms).

(3) La durée du silence: l'opposition DICH vs. DITCH soit [dɪʃ] vs. [dɪtʃ] est basée sur la durée du silence entre la voyelle [ɪ] et la fricative [ʃ] ou l'affriquée [tʃ]. Il y a plus de silence entre [ɪ] et [tʃ] qu'entre [ɪ] et [ʃ]. Le stimulus manipulé était l'émission naturelle du mot DITCH, dont la consonne [d] et la voyelle [ɪ] ont été gardées, alors que la consonne affriquée [tʃ] était coupée (bruit d'explosion + bruit de friction) et remplacée par la consonne [ʃ] du mot

naturel DISH. Le stimulus hybride final reste perçu en tant que DITCH, ce dernier était le plus long stimulus du continuum avec une durée de la phase d'occlusion dans [tʃ] égale à 60 ms. Les autres stimuli étaient créés en réduisant la durée de l'occlusion du [tʃ] par pas de 10 ms. Il y a eu 6 étapes de raccourcissement, pour un total de 7 stimuli pour ce continuum. Le dernier stimulus, le plus court, perçu alors DISH, avait une durée d'occlusion de 0 ms (stimulus initial DITCH = 543 ms et stimulus final DISH = 483 ms). Les 5 dB de bruit masquant présenté aux jeunes n'avait pas d'effet sur la perception du silence, parce qu'une expérience secondaire, parallèle aux expériences principales, a montré que le seuil différentiel de la perception de la durée du silence dans ce mot était le même sans le bruit masquant.

(4) La durée de la transition: les mots BEAT/WHEAT fondent leur opposition sur la durée de la transition entre la consonne initiale et la zone stable de la voyelle; elle est plus longue dans WHEAT que dans BEAT. Pour WHEAT, la durée de la transition est de 51 ms, laquelle était réduite par la suppression d'un cycle de 7-8 ms à chaque étape. Il y a eu 6 étapes de raccourcissement de la durée de la transition et un total de 7 stimuli. Les transitions de BEAT et de WHEAT étaient égales à 7 et à 51 ms respectivement, pour une durée totale du stimulus de 247 et 289 ms respectivement.

Nous tenons à présenter ici comment les stimuli étaient présentés et comment l'identification s'effectuait, parce que les résultats n'étaient pas les mêmes selon que l'identification suivait telle ou telle condition expérimentale. Pour les 4 continua, il y a eu 3 conditions expérimentales pour déterminer les seuils de changement catégoriels.

Condition (1): identification du mot. Dans cette condition il y a eu dix séries d'identification. Pour chaque série, les stimuli étaient présentés dans un ordre aléatoire, pour un total de 10 présentations par stimulus, le nombre des stimuli dépendait du continuum considéré. Après chaque écoute, une paire de mots apparaissait à l'écran de l'ordinateur. Les sujets étaient amenés à choisir l'un des deux mots qui apparaissaient: BUY-PIE, DISH-DICH, BEAT-WHEAT ou WHEAT-WEED. Les résultats obtenus selon cette condition nous semblent être ceux qui répondent le mieux à notre question. En effet, le fait que la comparaison se fait par rapport à un stimulus standard nous permet de savoir de combien il faudrait augmenter ce dernier pour que le basculement catégoriel se produise, ce qui nous donne, contrairement à la condition (2), un seuil différentiel pour un changement catégoriel pour une durée particulière, celle du stimulus standard.

Condition (2): Deux stimuli (même/différent). Le procédé d'identification a consisté en la présentation de deux stimuli, le premier, dans l'ordre, était le stimulus standard, le plus

court du continuum, et l'autre était un stimulus de comparaison, plus long que le standard. Après présentation, les sujets devaient dire si les deux stimuli sont les mêmes ou s'ils sont différents. Le premier stimulus de comparaison était le plus long du continuum. Le procédé a suivi la règle adaptative qui stipulait une baisse dans la durée du stimulus après deux bonnes réponses, et une hausse dans la durée du stimulus après une réponse incorrecte (Levitt, 1971), toujours par 1 pas à la fois (la longueur des pas dépendait du continuum).

Condition (3): Trois stimuli. Gordon-Salant *et al.* présentaient (1) le standard, ensuite (2) le standard (le même) et (3) le stimulus de comparaison. Après présentation, les sujets devaient répondre à la question: Est-ce que c'est (2) ou (3) qui est différent de (1)? Cette condition est intéressante parce qu'il fournit un seuil différentiel, non pour un changement catégoriel comme la précédente condition, mais pour la durée minimale perçue à partir de celle du stimulus standard.

Il était prévu que la condition (2), à deux intervalles, produirait des résultats similaires à ceux de la condition (1) de l'identification, alors que la condition (3), à trois stimuli, était conçue pour des expériences en psychoacoustique avec des signaux qui ne relèvent pas du domaine de la parole.

En ce qui concerne l'effet des conditions expérimentales sur les résultats, l'analyse de la variance a montré que la valeur du seuil était significativement plus basse avec la condition à trois stimuli qu'avec la condition à 2 stimuli, et cela était vrai pour les quatre continua. En effet, dans cette expérience, le seuil pour le silence avec la condition à 3 intervalles était de 2,77 ms. Des valeurs proches de cette valeur ont été obtenues dans des expériences en psychoacoustique par Plomp (1964), avec des stimuli non-parole, pour lesquels il cherchait à savoir quelle était la durée minimale d'un silence introduit entre deux bruits pour que celui-ci soit perçu. Il a trouvé que le seuil critique de cette durée était de 2-3 ms. Cela est en accord aussi avec les résultats Wieringen & Pols (1994) qui trouvaient que le seuil pour un son vocalique de la non-parole est de 3 ms pour une durée de départ de 50 ms.

Nous présentons maintenant les résultats de la deuxième condition, qui est comparable aux procédés standards utilisés pour l'identification des continua de la parole.

Le point de changement catégoriel était comme suit. Gordon-Salant *et al.* ont considéré la portion linéaire de la fonction (% de réponses correctes/ms), la partie de la fonction comprise entre 20 % et 80 % de réponses correctes. Par une régression linéaire sur cette

portion, ils ont établi le coefficient directeur (a) et l'ordonnée à l'origine (b) de cette fonction. Le point de basculement dans la catégorie suivante (x) était calculé selon l'équation suivante:

$$[x = (50 - b)/a]. \text{ équation (1)}$$

Application: pour le continuum WHEAT vs. WEED, qui démarre avec un [i:] de 155 ms pour terminer avec un [i] de 93 ms, sur la figure, entre 20 % et 80 % de correctes réponses, nous avons pu relever les valeurs suivantes: x (Durée en ms) = [109 115 123 132]; y (% de bonnes réponses) = [90 50 26 10]; un calcul de la régression linéaire nous a donné les valeurs suivantes pour (a) et (b): $a = -3.3406$; $b = 444.0351$.

En remplaçant a et b dans équation (1): Point catégoriel = $(50 - 444.0351)/(-3.3406) = 117.9534$ ms, les auteurs de l'étude donnent la valeur de 118,78 ms, un décalage mineur dû sûrement au fait que nous avons relevé les valeurs sur la figure à vue d'oeil.

L'analyse de la variance montre un effet significatif du groupe pour les deux continua DISH/DITCH et BEAT/WHEAT mais pas pour les deux autres continua BUY/PIE et WHEAT/WEED.

Résultats chez les jeunes normo-entendants, les seuils sont en ms:

		Continuum			
		Silence	VOT	Voyelle (93-155)	Transition formantique
	Durée continuum	(483-543)	(310-370)	(249-311)	(247-289)
	Durée stimulus	60	60	62	42
		Seuil			
Condition	Identification	20	21	26	29
	Deux intervalles	19	29	38	28
	Trois intervalles	3	11	9	11

Tableau 8. Seuil différentiel de la perception de la durée pour 4 continua où dans chaque continuum il y a un stimulus qui subissait la variation de la durée selon 3 conditions différentes (toutes valeurs en ms). Les 2 chiffres qui sont entre parenthèses, ligne "durée continuum", sont les durées en ms des deux stimuli situés aux extrémités du continuum considéré. D'après Gordon-Salant et al. (2006).

Les seuils avec la condition à trois stimuli, condition (3) utilisée traditionnellement dans des expériences psychoacoustiques, étaient statistiquement significativement plus bas que ceux qui étaient obtenus avec les conditions (1), utilisée traditionnellement dans des expériences où les stimuli font partie de la parole, et (2). Ces résultats sont exposés pour les 4

indices acoustiques dans la figure 37. Ces différences entre résultats nous permettent d'avancer que les sujets sont capables de discerner des différences acoustiques très fines (psychoacoustiques) qui ne sont pas perçues dans la condition de l'identification (linguistique) où les sujets avaient besoin, pour la voyelle par exemple, de trois fois plus de temps pour percevoir la différence au niveau de la durée. Cette acuité perceptive en condition 3 est peut-être due aussi en partie au fait que les locuteurs savent que l'un des deux stimuli doit être différent du stimulus qui est la référence.

Pour le VOT, les résultats sont très différents de ceux de Tyler *et al.* (1982) pour qui le rapport de Weber est entre 0,75 et 1. Quant à la transition vocalique, il n'y a pas non plus d'accord, puisque dans l'étude de Wieringen & Pols (1994), le rapport de Weber est entre 0,1 et 0,14, alors qu'ici, avec la condition 3, il est de 0,05 environ.

En résumé; nous nous centrons sur les résultats concernant la durée de la voyelle qui est notre objectif principal. Les résultats concernant les autres indices auraient pu nous servir de cadre dans la recherche de notre objectif principal s'il y avait un consensus entre études, mais nous constatons que ce n'est pas le cas.

Le seuil différentiel pour la durée de la voyelle nous a été fourni pour une seule durée

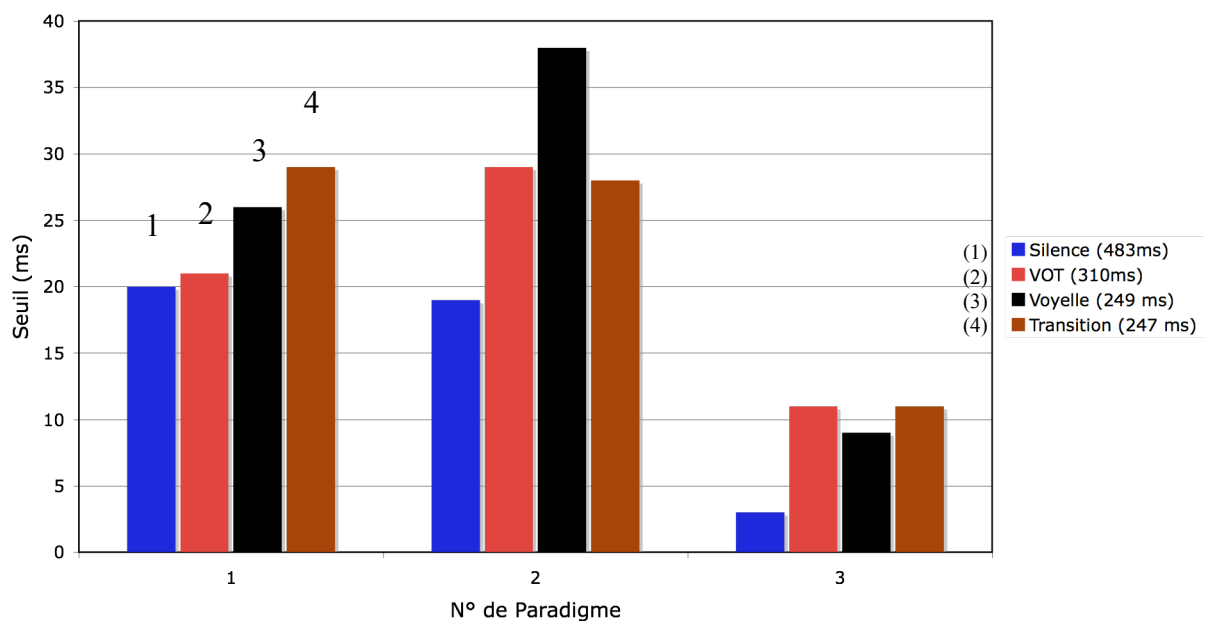


Figure 37. Seuil différentiel de la durée, 3 conditions (1)=identification, (2)=deux intervalles et (3)=trois intervalles, pour 4 indices acoustiques de la parole. D'après Gordon-Salant et al. (2006).

de voyelle de 93 ms, insérée dans une syllabe de 249 ms. Selon la condition 3, pour qu'un changement soit perceptible, il a fallu 9 ms pour 249 ms. Le rapport de Weber est de 0,04 (nous calculons toujours la hausse par rapport à la durée de la syllabe qui contient la voyelle et non pas par rapport à la durée de la voyelle). Ce résultat est très proche du résultat de Wieringen & Pols (1994) pour le segment vocalique de 50 ms de pseudo parole, où le rapport de Weber est de 0,06. C'est parce que la tâche est la même, à savoir, le discernement de la plus petite hausse perceptible, indépendamment du rôle linguistique de cette hausse. Autrement dit, la proximité des résultats est due au fait que la condition utilisée par Wieringen & Pols (1994) soit similaire à la condition 3 de l'étude de Gordon-Salant *et al.* (2006).

Pour que ce changement soit opérationnel dans la langue, soit la condition 2, il a fallu 26 ms, et le rapport de Weber est de 0,11. Selon Rossi (1972) le rapport de Weber pour cette durée-là (249 ms) est de 0,23, donc 2 fois supérieur à ce que Gordon-Salant *et al.* (2006) ont trouvé.

I.2.4. Conclusion

Dans notre recherche du seuil de la durée, nous avons examiné les résultats de 13 études différentes (tableau disponible en annexe n° VI). Nous observons que le comportement du seuil en fonction de la durée initiale est beaucoup trop compliqué pour qu'il puisse être exprimé par un seul pourcentage qui vaille pour toutes les durées initiales qui couvrent les durées des sons de la parole.

Nous n'avons pas trouvé de réponses satisfaisantes dans les études que nous avons examinées dans la deuxième partie du premier chapitre. Nous remarquons cependant les points suivants:

1) Aucune étude ne permet de prévoir le seuil pour l'intervalle qui couvre les durées des sons de la parole. Elles donnent toutes, exceptée celle de Rossi (1972) pour l'intervalle 130-290 ms, le seuil pour une durée donnée.

2) Bruits et sons vocaliques confondus, les résultats varient considérablement d'une étude à une autre, la majorité varie entre 4% et 20% avec une tendance autour de 18% (Annexe n° VI). Cette variation est due à plusieurs facteurs. Nous pouvons en identifier certains: les stimuli ne sont pas de la même nature (son pur, voyelle, bruit); les sujets n'ont pas le même niveau d'entraînement; la frontière catégorielle n'est pas fixée de la même manière

(50% pour certains; 71% pour d'autres, selon que leur approche est linguistique vs. psychoacoustique).

En nous basant sur ce que nous avons pu observer dans les études que nous avons examinées dans la deuxième partie du premier chapitre, nous pouvons avancer que pour les sons de la parole, avec une intensité suffisamment élevée, supérieure à 50 dB (Sinnott *et al.*, 1987a), et pour des durées situées entre 50 et 250 ms, pour qu'une variation de la durée puisse être tout simplement perçue par des personnes jeunes adultes normo-entendantes, le rapport de Weber devrait être autour de 0,05. Pour que cette variation de la durée puisse être exploitée linguistiquement, il faudrait que la durée initiale accuse une hausse supérieure à 10%, soit un rapport de Weber supérieur à 0,1. Une hausse de la durée initiale, entre 10% et 20%, peut être suffisante pour créer une opposition temporelle linguistiquement exploitable. Au-delà de 20%, il est fort probable que la hausse soit suffisante pour permettre aux deux durées, initiale et initiale+hausse, d'opérer indépendamment l'une par rapport à l'autre dans un système phonologique.

II. La quantité vocalique

L'examen des systèmes phonologiques des langues du monde a mis en évidence que l'opposition de durée des voyelles à des fins phonologiques (la quantité) est la seconde des stratégies de multiplication des éléments vocaliques, à partir des qualités de base, juste après la nasalité. Elle est adoptée par environ 11% (51 langues, des 451 langues répertoriées dans la base UPSID Maddieson, 1986).

II.1. Aspects des langues à quantité vocalique

Nous discutons ci-dessous de certains des comportements des langues qui opposent les durées de leurs réalisations vocaliques. Une discussion qui ne se veut pas exhaustive, mais concise, se réduisant à quelques points remarquables qui ont attiré notre attention dans notre consultation de la littérature sur la quantité vocalique.

Les langues à quantité vocalique semblent être insensibles à l'opposition temporelle en dehors des oppositions qui sont répertoriées phonologiques dans la langue considérée. En effet, une partie de l'étude de Eerola *et al.* (2003) qui a porté sur la production et la perception, par 17 locuteurs, des [i] long et bref en finnois - une langue où le contraste de quantité est phonologique - a examiné si la variation de la durée des segments perturbait la catégorisation des qualités bien établies. Ainsi l'examen a porté sur un continuum de [y] à [i], où les auteurs de l'étude faisaient varier uniquement le F2, alors qu'ils maintenaient fixes tous les autres formants: F1 = 250, F3 = 3010, F4 = 3300 et F5 = 3850. Le F2 changeait de 1520 Hz pour [y] à 2968 Hz pour [i] par 20 paliers. Chacune des deux qualités était émise avec 4 durées différentes: 50 ms, 100 ms, 250 ms et 500 ms.

Eerola *et al.* ont trouvé que la désignation de la frontière qualitative entre [y] et [i] ne dépendait pas de la durée, puisqu'elle se trouvait entre 2058-2069 Hz pour les 4 durées (tableau 9). Ils concluent que pour des voyelles qui ne s'opposent pas par leurs durées, la durée, même dans des langues à contraste quantitatif, semble avoir un effet qui reste négligeable en ce qui concerne la catégorisation et le classement des voyelles. Cette étude laisse penser, par ailleurs, que même si la durée de la partie stable de la voyelle est brève (ici 50 ms), elle est capable de porter l'essentiel de l'information acoustique requise pour

l'identification de la voyelle.

Durée (ms)	50	100	250	500
F2 (Hz)	2069	2060	2068	2058

Tableau 9. Valeurs de F2 à laquelle la voyelle passait de [y] à [i]. L'identification de la qualité, [y] ou [i], est indépendante de la durée de la voyelle. D'après Eerola et al. (2003).

Des informations importantes sur le comportement de la quantité vocalique et sur ses particularités en Japonais - son acquisition par contact avec des natifs ou par entraînement et le rôle de la position du contraste dans le mot - ont été fournies dans l'étude de Tajima *et al.* (2003), où ils testaient la capacité des anglophones natifs à percevoir la différence de durée, une différence phonologique en japonais. Ils ont utilisé la technique des paires minimales pour des voyelles et pour des consonnes. Le premier objectif de leur travail était l'étude de la perception de différents types de contrastes de durée en japonais auprès des natifs et des non natifs du Japonais. Ils cherchaient à savoir, auprès des non natifs, quel était le rôle que jouerait le contact avec les natifs et le rôle de l'entraînement dans la perception du contraste de durée. Ils cherchaient à savoir aussi si l'entraînement à un type de contraste augmentait la capacité à mieux discerner les autres types de contrastes, et finalement si la perception de la durée variait en fonction de l'occurrence du contraste dans le mot (position initiale, médiane, etc.). Ils ont divisé leurs locuteurs en trois groupes en fonction de l'origine linguistique:

Origine	Âge	Nombre
Anglais (aucune expérience avec le Japonais « ENG »)	18-30	7
Anglais (ont passé de 1 à 6 mois au Japon « JRE »)	18-32	7
Japonais natifs « JPN ».	19-22	7

L'étude a consisté en 3 phases, pré-test, entraînement et post-test. Les types de phonèmes dont ils testaient la différence de durée dans des paires minimales étaient (1) l'une des 5 voyelles du japonais; (2) une obstruente; (3) une des deux consonnes nasales [m] ou [n]; (4) l'approximante palatale [ɺ]. Toutes les longueurs des autres phonèmes du contexte étaient égalisées, restait seule différente la longueur du phonème dont la différence de la durée fondait la paire minimale. Le contraste apparaissait en plusieurs positions (initiale, médiane ou finale).

Les phrases étaient prononcées par des professionnels de la lecture. Chaque item connaissait deux contextes, en isolation ou à l'intérieur d'une phrase porteuse. Les deux contextes étaient prononcés avec 3 débits d'élocution (lent, normal et rapide). Les sujets entendaient un seul stimulus à la fois par un casque audio et choisissaient leurs réponses sur l'écran d'un ordinateur. Le pré-test et le post-test se sont soldés chacun par 1296 réponses pour un corpus de 2,5 heures.

Après le pré-test, le groupe JRE a suivi 5 jours d'entraînement à l'identification, pas le groupe ENG.

Les résultats ont révélé que tous contextes, tous débits et tous types de contrastes confondus, les comparaisons entre les réponses du pré-test et les réponses du post-test montraient, quoique faible, un effet significatif ($p < 0,01$) de l'entraînement sur l'amélioration de la perception du contraste temporel. Malgré l'amélioration de la performance des non natifs, la précision de leurs réponses restait significativement inférieure à celles des natifs ($p < 0,01$).

La comparaison des pré-tests entre les deux groupes de ENG et JRE ne montrait aucun effet significatif. Alors que les JRE étaient entraînés pour le contraste des voyelles seulement, après l'entraînement, la performance s'est améliorée pour les autres contrastes aussi. Les comparaisons individuelles montraient que la perception des contrastes des durées des voyelles s'est améliorée ($p < 0,05$), pour les obstruents aussi ($p < 0,01$), mais pas pour les nasales et la palatale.

Quant à l'effet de la position du contraste dans le mot, pour les natifs, le contexte n'avait pas d'effet, ils identifiaient le contraste à plus de 98 % dans les deux conditions. Quant aux JRE, ils identifiaient mieux le contraste des nasales ($p < 0,01$), et des voyelles ($p < 0,05$). L'interaction position - contexte est significative à ($p < 0,05$) quand le contraste apparaissait en position initiale qu'ailleurs dans le mot. L'existence du contexte n'avait pas d'effet sur l'identification de ce contraste des nasales, alors que l'existence du contexte avait un effet sur les voyelles, le contraste était mieux identifié en isolation qu'en contexte.

Les auteurs de l'étude concluent que le contact des locuteurs anglophones avec des japonais natifs augmente leur capacité à la perception du contraste de la durée. L'entraînement a le même effet que le contact. Malgré l'entraînement, une importante différence reste entre les natifs et les non natifs de la langue japonaise. La position du contraste n'a pas d'effet pour les natifs. L'entraînement au contraste temporel auprès d'une

catégorie particulière de phonèmes améliore la performance pour les autres catégories aussi. Ce dernier résultat est particulièrement intéressant, parce qu'il met en évidence l'existence d'un processus global de l'interprétation du temps, et non seulement une performance linguistique pour un contraste local à une fin finie.

II.2. Différence temporelle et quantité vocalique

II.2.1. La durée est un trait secondaire

Qualités de base et voies de l'apparition des traits secondaires: dans la description des systèmes phonologiques nous distinguons deux types de voyelles, les premières sont des voyelles dites de base, et les secondes sont des voyelles (de base) avec des traits secondaires. Ces traits secondaires contribuent à la multiplication des éléments vocaliques d'un système phonologique. Les langues du monde ont recours, en général, à ce procédé d'enrichissement par traits secondaires, lorsque le système phonologique des voyelles d'une langue exploite tout l'espace vocalique¹⁰, et de ce fait, s'il veut continuer à apporter d'autres qualités vocaliques, le système en question sera contraint de modifier les qualités de base qu'il a pu produire.

Nous trouvons que ces traits secondaires, qui entraînent la modification partielle de la voyelle de base, peuvent être effectués:

1) soit par la modification de la durée vocalique: les modifications appartenant à ce procédé consistent en l'augmentation ou la réduction de la durée de la réalisation vocalique tout en gardant par ailleurs le geste articulatoire.

2) soit par la modification de la qualité acoustique de la voyelle: la modification de la qualité acoustique de la voyelle peut être effectuée de différentes manières, tels le changement du régime phonatoire (la manière dont nous faisons vibrer les cordes vocales), l'abaissement ou le relèvement de la luette, avancer ou reculer la racine de la langue et le retournement de la pointe de la langue et son accolement contre le palais

3) soit par la modification de ces deux facteurs à la fois, et dans ce cas-là, il s'agira de deux modifications, l'une est inhérente à l'autre et forment, de ce fait, un seul trait. Il est à

¹⁰ Les diphtongues et les semi-voyelles sont des éléments vocaliques qui enrichissent le système vocalique d'une langue, parce qu'elles sont susceptibles d'être des syllabèmes, mais ici nous ne considérons que les qualités acoustiques avec des zones formantiques stables, donc les diphtongues et les semi-voyelles ne sont pas considérées dans notre discussion, d'autant plus qu'elles ne présentent pas d'opposition de durée entre elles.

noter que dans un cas de figure de ce type, il est particulièrement difficile d'attribuer à chacun de ces facteurs (temporel et qualitatif) sa part dans le poids perceptif de la forme phonétique résultante. La difficulté d'apporter des qualités de base commence avant la saturation de l'espace vocalique, pour la majorité des langues du monde, cette difficulté apparaît avec la neuvième qualité. Ainsi pour produire un certain nombre de voyelles, les langues ont plus tendance à avoir des qualités de base qui seront dédoublées par des traits secondaires.

Pourquoi les traits secondaires?

1- Les Contraintes: l'espace acoustique des voyelles de base est limité, la gymnastique des articulateurs étant tributaire des contraintes physiologiques inhérentes à l'appareil phonatoire. Effectivement, les extrémités des mouvements articulatoires sur les deux axes horizontal (le mouvement Avant Arrière de la langue) et vertical (le mouvement Baisse Remontée de la mandibule) coïncident, du point de vue acoustique, avec les voyelles cardinales du triangle vocalique.

L'existence de ces contraintes motrices, qui favorisent l'apparition ou la disparition de telle ou telle forme phonétique, se manifeste par certains phénomènes au cours de la production de la parole. (i) des formes que nous ne voulons pas produire, s'actualisent, telle la transmission d'une unité phonétique de l'une de ses caractéristiques à une autre unité voisine, ce qu'est communément appelée "Assimilation", progressive ou régressive, selon que la caractéristique est transmise vers une unité avant ou vers une unité après. Ce phénomène est particulièrement vérifié quand le débit d'élocution est rapide. (ii) le phénomène de l'harmonie vocalique, qui se manifeste en morphologie dérivationnelle par le choix des suffixes en fonction de la structure phonétique de la partie radicale de l'unité lexicale, ou encore (iii) l'épenthèse qui prouve l'existence de certaines formes qui naissent suite à un contact entre deux formes phonétiques, un contact qui est plutôt un chevauchement des caractéristiques articulatoires des unités phonétiques qui, pour des raisons que la phonétique articulatoire saurait expliquer, fait émerger une nouvelle forme phonétique. Ces phénomènes témoignent, enfin, de l'existence des limites de la capacité de la production de la parole.

2- Le système surnuméraire: il y a sûrement une autre raison qui expliquerait l'utilisation de cette technique (traits secondaires), puisque nous remarquons que certains systèmes ont recours à ce procédé alors que leur espace vocalique est très loin d'être entièrement exploité. Cette tendance à économiser des qualités de base peut être expliquée par la réalité suivante: économiser des voyelles de base revient à économiser des configurations qui nécessitent des contrôles globaux qui mobilisent tout le conduit vocal, des gestes

articulatoires relativement coûteux en énergie par rapport aux consonnes qui exigent des contrôles locaux. Chaque configuration est alors responsable d'une qualité acoustique, et d'une seule, mais l'inverse n'est pas vrai, c'est-à-dire que deux configurations peuvent produire la même qualité acoustique, parce que la parole est un système surnuméraire. Des recherches qui ont été menées dans le cadre de la théorie de l'inversion articulatoire-acoustique l'ont bien montré, Abry *et al.* (2000: 257) présentent ce phénomène comme suit: «Cette propriété des systèmes, dits surnuméraires (...) pose de sérieux problèmes pour l'inversion puisque, sans contrainte, il existe tout un ensemble de commandes possibles qui correspondent au même signal de sortie. Même intrinsèquement, bien contraint, un modèle articulatoire peut donc générer, s'il est numéraire des voyelles identiques au niveau acoustique mais présentant des dispositions articulatoires différentes en termes de classification phonétique». Ainsi la création d'une nouvelle configuration du conduit vocal ne garantit pas l'apport d'une nouvelle qualité vocalique, d'où, peut-être, un recours prématuré aux traits secondaires.

La fréquence des traits secondaires

Les traits secondaires n'ont pas la même fréquence dans les langues du monde, nous présentons ces possibilités secondaires selon le pourcentage de leur occurrence (Vallée, 1994), dans les systèmes phonologiques des langues du monde: 1) la nasalisation. 2) la longueur: celles qui sont réalisées plus longues que les réalisations normales sont des longues, alors que celles qui sont réalisées moins longues que les réalisations normales sont des « ultra brèves ». 3) la pharyngalisation; 4) l'assourdissement; 5) la laryngalisation; 6) l'aspiration; 7) la dimension rétroflexe.

Le choix de ces dimensions est conditionné par une faisabilité motrice au préalable, et par une rentabilité entre coût en énergie et effet perceptif. Nous observons que la durée, du point de vue de la fréquence de son occurrence dans les langues du monde, est le deuxième procédé utilisé dans le mécanisme de la multiplication des éléments vocaliques des systèmes phonologiques. Cela s'expliquerait en partie par le fait qu'elle soit une dimension souple, en ce sens qu'elle admet de se greffer en même temps que d'autres traits secondaires sur la même qualité de base. Théoriquement, elle peut accompagner toutes les autres dimensions, une propriété qui n'est bio-mécaniquement pas possible pour tous les traits relatifs à la qualité. C'est simplement parce que certains de ces traits sont réalisés avec le même organe et par conséquent les deux gestes sont incompatibles, puisque la présence de l'un implique l'absence de l'autre, tels la laryngalisation, le murmure (la voix soufflée) et l'assourdissement

qui sont réalisés avec les cordes vocales où la présence de l'un implique l'absence de l'autre. Dédoubler la voyelle par la durée appelle moins de contraintes qui empêcheraient sa réalisation. Ce procédé est cependant plus favorisée par certaines qualités vocaliques plutôt que par d'autres.

Les différentes formes de la durée « trait secondaire »

En ce qui concerne les langues qui utilisent la durée des voyelles comme un trait phonologiquement pertinent (UPSID 451, Maddieson, 1986), nous pouvons les diviser en 4 groupes selon qu'elles opposent: 1) des voyelles longues à des voyelles brèves (seulement avec la durée). 2) des voyelles brèves à des voyelles ultra brèves (seulement avec la durée). 3) des voyelles longues à des voyelles brèves avec un seul autre trait qui porte sur la qualité, et 4) des voyelles longues à des voyelles brèves avec deux autres traits qui portent sur la qualité.

1) Langues qui opposent les voyelles brèves aux longues: elles sont 51 langues, soit 11,31% des 451 langues de la base de données UPSID (Maddieson, 1986). Comme les langues ont moins tendance à superposer des traits secondaires qui relèvent de la dimension temporelle avec des traits qui relèvent de la dimension spectrale, ainsi les langues qui n'associent pas à la modification de la durée une modification de la qualité sont les plus nombreuses parmi toutes les langues qui opposent des durées.

1- Irlandais	2- Allemand	3- Norvégien	4- Lithuanien	5- Français
6- Pashto	7- Kurde	8- Hindi-ourdou	9- Cachemirien	10- Cinghalais
11- Hongrois	12- Selkoup	13- Cambodgien	14- Khmu?	15- Lakkia
16- Yay	17- Atayal	18- Adzera	19- lai	20- Lai
21- Mien	22- Kabardien	23- Lak	24- Télougou	25- Toulou
26- Brahoui	27- Wolof	28- Noni	29- Fur	30- Arabe
31- Tigré	32- Néo-araméen	33- Tamasheq	34- Irakou	35- Angas
36- Ngizim	37- Kanakuru	38- !xu~	39- Navajo	40- Hupa
41- Chipewyan	42-Eyak	43- Totonaque	44- Karok	45- Diegueno
46- Ojibwa	47- Tonkawa	48- Wichita	49- Bardi	50- Gadsup
51- Angaatiha				

Tableau 10. Langues qui opposent des voyelles brèves à des voyelles longues.

2) Langues qui opposent les voyelles brèves aux ultra brèves: nous en trouvons 9, soit 2% des 451 langues de la base. Les ultra brèves sont beaucoup moins fréquentes que les longues dans les systèmes phonologiques des langues du monde. Elles n'accompagnent pas d'autres traits secondaires qui portent sur la qualité.

1- Khanty	2- Nenets	3- Tchouvache	4- Po-Ai	5- Lungchou
6- Géorgien	7- Tchouktchi	8- Sebei	9- Angaatiha	

Tableau 11. Langues qui opposent des voyelles brèves à des voyelles ultra brèves.

3) Langues qui opposent les voyelles nasales brèves aux nasales longues: il y en a 8, soit 1,77% des 451 langues de la base UPSID.

1- Irlandais	2- Hindi-ourdou	3- Cachemirien	4- Lakkia	5- !xu~
6- Navajo	7- Chipewyan	8- Eyak		

Tableau 12. Langues qui opposent des voyelles nasales brèves à des voyelles nasales longues.

4) Langues qui opposent les voyelles nasales pharyngalisées brèves aux nasales pharyngalisées longues, il n'y a qu'une seule langue, le !xu~, qui représente 0.22% des 451 langues de la base de donnée UPSID.

II.2.2. Spécification linguistique de la durée

L'opposition temporelle est interprétée différemment en fonction du système phonologique où cette opposition existe, c'est-à-dire que l'écart entre les durées de deux stimuli n'a pas la même valeur linguistique pour toutes les langues du monde, chaque langue pouvant avoir une manière particulière d'exploiter cet écart. Nenonen *et al.* (2003), en comparant la catégorisation de l'écart temporel entre les voyelles longues et les voyelles brèves auprès des natifs (L1) et des non natifs (L2) du finnois, ont observé que les L2 avaient des difficultés dans l'acquisition du contraste quantitatif, qui est phonologique, des difficultés tant en production qu'en perception.

Le but de leur travail étant de déterminer si la catégorisation de la longueur était différente entre des natifs et des non natifs en raison de la différence des langues maternelles respectives des 2 groupes. Les sujets de leur étude étaient 11 russophones à l'école en Finlande, ils avaient passé 5 ans et 2 mois en Finlande (moyenne d'âge 12 ans et 7 mois), et 14 enfants natifs du finnois (moyenne d'âge 11 ans et 5 mois).

Selon les auteurs de l'étude, dans certaines langues, comme le russe, la durée est un indice de l'accent au lieu d'être un indice de la quantité. L'accent, en finnois, est un accent fixe qui frappe la première syllabe du mot, alors que le russe est une langue à accent mobile où l'accent joue un rôle phonologique basé essentiellement, entre autres facteurs, sur la durée. Les locuteurs de langue maternelle russe et qui ne parlent pas le finnois, peuvent percevoir les voyelles longues non accentuées du finnois comme des voyelles accentuées, une accentuation basée sur la durée. Pour les russes qui parlent le finnois, L2, la différence entre durées est perçue en tant qu'indice de la quantité. Cependant, cette catégorisation peut être affectée par la position de la voyelle dans le mot pour les L2. Ainsi, Nenonen *et al.* cherchaient à déterminer si la position de l'opposition affectait vraiment la catégorisation de l'opposition temporelle auprès des russes L2.

Ainsi que nous l'avons vu plus haut, Tajima *et al.* (2003) ont démontré que, pour le japonais, la position du contraste quantitatif dans le mot n'avait pas d'effet pour les natifs, alors que les non natifs identifiaient mieux le contraste des nasales et des voyelles quand il était à l'initial qu'ailleurs dans le mot.

Nenonen *et al.* se sont basés sur 2 pseudos continua de durées vocaliques, de 70 à 194 ms et de 117 à 249 ms, la durée était manipulée par la concaténation des cycles vocaliques sur le signal de parole. Nenonen *et al.* manipulent la durée uniquement, parce qu'ils considèrent que le facteur acoustique le plus saillant dans l'opposition de la quantité vocalique en finnois, même si des différences au niveau spectral existent (les brèves sont un peu plus centrales que les longues), est la durée de la voyelle.

Les résultats montraient que les fonctions de catégorisation étaient significativement différentes ($p < 0,05$), entre les groupes quand la syllabe considérée était en position finale. Le basculement de la catégorie des brèves dans celle des longues advenait plus tôt (avec moins de ms) chez les non natifs que chez les natifs. Alors qu'en syllabe initiale, il n'y avait pas de différence significative entre les deux groupes. Quant au temps de réaction, des différences significatives ont été enregistrées entre les deux groupes dans les deux conditions, en position initiale ($p < 0,01$), et en position finale ($p < 0,01$). Les natifs réagissaient plus vite

pour catégoriser la voyelle, en tant que longue ou brève, que les non natifs. La différence en ce qui concerne la syllabe finale (les L2 ont leur frontière avant les L1), est peut-être due au fait que cette syllabe est inaccentuée, et la voyelle brève de cette position syllabique est relativement longue, cette longueur intrinsèque n'a pas été intégrée en tant qu'intrinsèque mais en tant qu'allongement. La catégorisation du continuum temporel semble être mieux établie chez les natifs que chez les L2 puisque leur temps de réaction est inférieur à celui des L2. Ces derniers résultats attestent que la différence temporelle peut-être exploitée différemment, elle est indice d'accent en russe, alors qu'elle est indice de quantité en finnois, et que cette différence dans l'interprétation peut avoir un impact significatif sur la pratique d'une langue seconde.

Les résultats d'une expérience perceptive, effectuée par Mora & Mott (2003), soutiennent l'idée que la durée peut-être exploitée pour servir d'indice acoustique autre que celui de la quantité. Mora & Mott montrent que la durée peut être un indice pour la perception de la syllabation des consonnes en anglais.

Leur expérience, qui visait à déterminer le rôle de la durée, testait la capacité des natifs de l'anglais à distinguer les consonnes syllabiques des consonnes non syllabiques. Les consonnes qui étaient testées étaient [n] *lightning-lightening*, [l] *finely-finally*, [r] *hungry-Hungary* et [s] *sport-support*. Selon les auteurs, la consonne du deuxième mot est syllabique, celle du premier mot est non syllabique.

L'idée traditionnelle est que les consonnes sont syllabiques grâce à leur longueur relativement importante, elles sont plus longues que leurs homologues non syllabiques, toutes choses égales par ailleurs. La supposition selon laquelle la durée de la consonne est un indice de syllabation ne fait pas l'unanimité. En effet, Cruttenden (2001) affirmait que certains phonèmes avec une durée de 50-60 ms pouvaient être syllabiques. Barry (2000) trouvait que c'était plutôt le contexte qui jouait le rôle déterminant. Selon ce dernier, ce qui permet d'identifier un [l] comme syllabique, c'est le fait qu'il soit précédé par une occlusive sourde non pré-glottalisée (*duckling vs. chuckling*: avant le [k] de *duckling* il y a un coup de glotte).

Afin de trouver une réponse à ce dilemme, Mora & Mott (2003) ont élaboré une étude qui a porté sur 21 locuteurs anglophones natifs, âgés de 21 à 68 ans. La tâche consistait en l'identification comme syllabique ou non syllabique des consonnes [n], [l], [r] et [s] dans le corpus suivant « *lightning vs. lightening, finely vs. finally, hungry vs. Hungary et sport vs.*

support ». D'après Mora & Mott, les premiers mots des 3 premières paires illustrent les mots avec un contexte dans lequel les consonnes vocaliques syllabiques pouvaient alterner avec leurs homologues non syllabiques précédées par [ə] précédé par un segment d'un rang de sonorité inférieur dans la première des deux syllabes post-toniques inaccentuées. La phrase porteuse était "say ... again". Les consonnes syllabiques ont été raccourcies, et les consonnes non syllabiques ont été rallongées, le schwa était enlevé quand il était avant une sonore. Le rallongement et le raccourcissement des durées étaient proportionnels aux durées initiales des segments. Les consonnes non syllabiques étaient rallongées en copiant un segment de la consonne considérée et en le collant à l'intérieur, cela était répété autant de fois qu'il le fallait jusqu'à la création de deux sons différents. Il y avait deux degrés d'allongement, A1: 40-50% et A2: 60-75% de la durée initiale de la consonne. Le raccourcissement des consonnes syllabiques était fait par la suppression d'une partie du centre de la consonne, deux degrés de raccourcissement, R1: 30-50% et R2: 50-75%. Le test a consisté en 320 stimuli.

Le plan de l'expérience de Mora & Mott (2003) est le suivant: 4 types de mots pour les 4 consonnes de l'expérience, des mots où (1) la consonne est non-syllabique (**P**). (2) la consonne est syllabique (**S**). (3) la consonne est non syllabique précédée par un schwa (**VP**). (4) la consonne est non-syllabique et le schwa qui le précède était enlevé (**D**). Il y a eu 2 degrés de raccourcissement pour les syllabiques (**R1**) et (**R2**), et 2 degrés d'allongement pour les non syllabiques (**A1**) et (**A2**). Il y a eu 2 contextes (**C1**) et (**C2**).

Selon les résultats obtenus par Mora & Mott, l'allongement des consonnes vocaliques non syllabiques dans les mots *finely* [l], *hungry* [r] et *lightning* [n], a produit un changement significatif où ces mots étaient identifiés comme *finally*, *Hungary* and *lightening*. Chacune des trois consonnes citées ci-dessus est syllabique. Ainsi, le fait que le premier mot de la paire minimale passe pour le deuxième, signifie que les non syllabiques passent pour des syllabiques. Cela confirme l'assertion selon laquelle la durée est un indice acoustique pour la syllabation des consonnes en anglais, même si elle n'est pas le seul indice.

Les études statistiques appliquées sur les résultats confirment en partie l'hypothèse selon laquelle les non-syllabiques vocaliques [n, l, r] sont souvent perçues comme syllabiques quand elles sont allongées. Parce que la perception était différente avec **A2** ($p < 0,001$), alors que la différence n'était significative avec **A1** que pour le [l]. Dans tous les cas, une différence significative ($p < 0,01$) était trouvée entre **P** et **S**. Comme prévu, pour le [s], ni l'allongement ni le raccourcissement n'a eu d'effet sur le statu de la syllabation de la consonne. Un grand nombre de consonnes de la condition **D** était identifié comme appartenant

à la condition **S**. Les consonnes de la condition **VP** sont plus longues que celles de la condition **P**.

Les auteurs de l'étude soulignent que les résultats du raccourcissement des syllabiques présentent deux types d'asymétries par rapport aux non-syllabiques allongées. D'un côté, allongement et raccourcissement, 2 différents moyens pour modifier la durée d'une consonne, n'ont pas exactement le même effet sur la perception de la syllabation de la consonne. L'effet du raccourcissement de la syllabique se traduit par son identification en tant que non-syllabique, mais cela n'est vrai que pour *finely-finally* et *hungry-Hungary* ($p < 0,01$), soit pour [l] et [r].

Ainsi, le [n] non-syllabique est perçu comme syllabique quand il est allongé, alors que le [n] syllabique n'est pas perçu en tant que non-syllabique quand il est raccourci. La deuxième asymétrie réside dans (1) le comportement du [l] et du [n] par rapport au nombre des identifications en tant que syllabiques, ce nombre est plus élevé pour le [l]. Mora & Mott attribuent la raison au fait que le [l] soit réalisé plus couramment en tant que syllabique que le [n]. (2) le fait que le raccourcissement affecte la syllabation du [l] mais pas celle du [n].

Les résultats laissent penser que la longueur fonctionne comme un indice acoustique dans la perception de la syllabation des consonnes, et qu'il y a une asymétrie entre allongement et raccourcissement.

Ci-dessous nous présenterons et discuterons d'une étude qui a essayé de montrer, non pas que le rôle du contraste des durées de certaines unités phonétiques est relatif à la langue (quantité, accent, syllabation, etc.) mais que l'unité d'intégration temporelle, elle-même, peut être spécifique à la langue où nous l'observons. Par ce biais, et lorsqu'un sujet passe de sa langue maternelle à une deuxième langue acquise ultérieurement, un décalage dans le traitement métrique du temps peut naître, puisque avec les règles de la première langue, le sujet essaiera d'analyser la seconde. Une telle "naissance" constituera la preuve que le temps n'est pas traité avec les mêmes unités.

En effet, les locuteurs japonais analyseraient la chaîne de la parole des langues qui leur sont étrangères en mores, comme ils utiliseraient la more, plutôt que la syllabe, pour le contrôle de la durée quand ils parlent en français. Les affirmations de Kondo & Shinohara, dans leur étude publiée en 2003, vont dans ce sens. Les auteurs ont cherché à savoir à quel point le timing de la langue maternelle influençait la production d'une langue seconde. Ils se

sont intéressés au phénomène de la compensation (ajustement) temporelle comme un outil de l'organisation rythmique.

L'organisation des paramètres temporels est relative à la langue où ces paramètres sont relevés. Des études psycholinguistiques, Otake *et al.* (1993) ont montré que la segmentation de la chaîne parlée en japonais se faisait en termes de *mores*, même si les durées des mots peuvent coïncider avec des durées de mots à structures syllabiques, la durée d'un mot est proportionnelle au nombre de *mores* qui le composent.

En français, la segmentation de la chaîne parlée se fait en termes de syllabes, où les syllabes ont plus ou moins les mêmes durées (Wenk & Wioland, 1982). Alors qu'en anglais la segmentation se fait en termes de pied trochaïque accentué (Cutler & Otake, 1994). Ces nombreuses recherches se sont intéressées à cette segmentation parce qu'elle était supposée préalable à l'analyse lexicale, un thème majeur dans l'apprentissage langagier.

En japonais, la *more*¹¹ est l'unité de base de l'organisation temporelle, alors qu'en français c'est la syllabe. Yamasaki (1996), a montré que les nouveaux-nés et les enfants japonais ne procédaient pas par segmentation moraïque. Kondo & Shinohara (2003) cherchent à savoir quel type de segmentation phonétique les locuteurs japonais adultes utilisent - la segmentation moraïque ou la segmentation syllabique - lorsqu'ils écoutent une langue non moraïque, tels le français ou l'anglais.

Cette recherche a de l'intérêt pour notre problématique, parce qu'elle s'intéresse à la représentation phonologique de la durée selon la langue où le contraste temporel est relevé. Elle met en évidence, par de nombreux exemples, le rôle que jouent les règles phonotactiques de la langue maternelle dans la perception des durées des unités phonétiques des autres langues. Cela constitue une preuve en faveur de "la phonologisation" (adaptation linguistique) de la dimension temps. En effet, les auteurs rapportent les résultats de Mochizuki-Sudo & Kiritani (1991) concernant l'apprentissage de l'anglais par les japonais, où, ces derniers, montrent que lorsque les durées des voyelles de l'anglais sont modifiées sous l'effet de l'accent, les locuteurs japonais ne peuvent détecter le changement que lorsqu'il est assez important pour créer un passage d'une voyelle longue à une voyelle brève et inversement. Cela montre qu'il y a une adaptation de la part des locuteurs japonais des voyelles de l'anglais, puisque la distinction se base sur une segmentation moraïque.

¹¹ Une unité de quantité ou de durée, la *more* peut être considéré comme une composante de la syllabe. Une voyelle longue, une diphtongue, une voyelle brève plus une coda peuvent constituer deux *mores*, alors qu'une voyelle brève ne constitue qu'une seule *more*.

Kondo & Shinohara (2003) soulignent que certaines voyelles de l'anglais correspondent, après adaptation des locuteurs japonais, à 2 mores et d'autres à une seule. Ils présentent l'exemple suivant: "*pita*" [pɪtə] devient [pita] vs. "*peter*" [pi:tə] devient [piita].

Une autre manifestation de la réalité de l'adaptation des mots étrangers, c'est la syllabation des syllabes à structures complexes. Le japonais ne permet que les syllabes à attaques simples, lesquelles peuvent se terminer à la moitié d'une gémignée ou avec une nasale, ce qui donne naissance à des voyelles épenthétiques, ex.: "*extra*" devient [ekusutora]. Ceci laisse penser, encore une fois, que lorsque les locuteurs japonais traitent les mots étrangers, ils le font en utilisant la more.

Permettons-nous ici de rapporter certaines de ces contraintes, fournies dans l'étude de Kondo & Shinohara (2003), concernant l'adaptation en japonais. La première moitié d'une consonne gémignée, la 2e moitié d'une voyelle longue ou d'une diphtongue sont chacune une more. Quand une syllabe en français se termine par une consonne simple, alors, dans la forme adaptée, la rime s'allonge. La syllabe pré finale peut être lourde dans les cas suivants:

1. Elle se termine avec la première moitié d'une consonne gémignée "*archevêque*" devient [arusjubekku].

2. Elle contient une voyelle longue "*madeleine*" devient [madureenu].

Le type d'allongement dépend de la nature de la consonne. Lorsque les consonnes sourdes avec obstacle, dite obstruente (occlusives, constrictives ou mi-occlusives), se rencontrent en finale du mot français, elles sont perçues comme des gémignées.

(a) les obstruents sourdes précédées par une voyelle rallongée, la voyelle peut être optionnellement rallongée à la place de la consonne. Exemple le mot "*haute*" donne [ooto] ou [otto], le mot "*fête*" donne [Φeeto] ou [Φetto].

(b) les fricatives voisées: quand la consonne finale est une fricative voisée, c'est toujours la voyelle qui est rallongée: "*rose*" donne [roozu], "*terre*" donne [teeru].

(c) les patrons de la variation des occlusives orales et nasales voisées: "*robe*" donne [robba] ou [roobu]. "*Aide*" donne [eddo] ou [eedo]. "*Pomme*" donne [pommu] ou [poomu], "*reine*" donne [rennu] ou [reenu].

Suite à ces observations, Kondo & Shinohara (2003) émettent l'hypothèse selon laquelle les effets de la compensation temporelle peuvent se manifester à de différents niveaux dans

les deux langues. Ainsi, les japonais analysent les sons des autres langues en termes de mores, et en production, ils utilisent la more pour le contrôle de la durée des mots français.

Afin de savoir si la compensation de la durée par more existe dans des phrases françaises prononcées par des japonais, les auteurs ont élaboré une expérience avec un corpus qui a consisté en 6 paires de phrases en français, lesquelles étaient prononcées par des locuteurs francophones et japonais: 8 locuteurs natifs, 5 français et 3 japonais, avec 3 répétitions chacune. Les phrases de chaque paire avaient le même nombre de syllabes, alors qu'elles pouvaient potentiellement être adaptées avec des nombres différents de mores. Cette différence en nombre de mores peut influencer la durée de la phrase française.

Les analyses statistiques ont montré qu'il n'y avait pas de différence de durée entre les deux phrases de chaque paire. Les auteurs attribuent l'absence de la significativité à la "petite" taille de l'échantillon. Ils remarquent que le nombre de mores n'a pas d'influence sur les énoncés des locuteurs francophones. Souvent, dans une paire de phrases, celle qui avait le plus grand nombre de mores avait la durée la plus courte. Par contre, auprès des locuteurs japonais la phrase qui avait le plus grand nombre de mores avait la durée la plus longue, alors que les deux phrases pouvaient avoir le même nombre de mores, cela dépendait de l'interprétation des locuteurs des énoncés.

Finalement, les résultats expérimentaux laissent penser que les locuteurs japonais analysent la chaîne de la parole des langues qui leur sont étrangères en mores, comme ils utilisent la more, plutôt que la syllabe, pour le contrôle de la durée quand ils parlent en français, les voyelles du français étant analysées comme bimoraïques. Il semble que les voyelles épenthétiques peuvent être éliminées par les sujets qui parlent bien le français, mais il est clair que l'acquisition du contrôle rythmique des unités du timing d'une deuxième langue par les locuteurs japonais reste difficile.

L'étude de Nenonen *et al.* (2003) sur la représentation de l'écart temporel en finnois pour des natifs et des non natifs, le travail de Mora & Mott (2003) sur la syllabation des consonnes en anglais et l'étude de Kondo & Shinohara (2003) sur l'influence du timing du japonais langue maternelle sur la production d'une langue seconde, sont des exemples qui démontrent que la notion du temps est relative à la langue: une différence temporelle n'est pas exploitée pour une même fin et n'est pas représenté de la même manière en fonction de la langue où elle opère. Cela nous conduit à dire que ce n'est pas parce que nous observons des sons brefs avec des partenaires longs que nous devons penser forcément à un contraste

quantitatif, l'écart temporel pouvant servir d'autres procédés phonétiques dans la langue considérée.

II.3. Quantité vocalique en arabe

Présente en arabe standard et dans ses aboutissants dialectaux, la quantité vocalique a été l'objet de plusieurs études acoustiques (Al-Ani, 1970; Ghazeli, 1979; Belkaid, 1984; Abou-Haidar, 1994; Newman, 2002; Barkat-Defradas, 2003; Allatif & Abry, 2004). Ces études rapportent que les voyelles longues ont une durée supérieure à celle des brèves correspondantes, le rapport Brève/Longue fluctuant, pour la plupart des résultats, entre 40% et 60%. L'opposition de durées des consonnes en arabe a fait aussi l'objet de plusieurs études, Bonnot (1979) a établi une description expérimentale de la gémination consonantique en arabe. Mais ici nous nous concentrons sur les études qui ont porté sur la quantité vocalique que sur la quantité consonantique.

En effet, Al-Ani (1970), un des pionniers de la question de la quantité vocalique en arabe, prétendait qu'en arabe irakien toutes les brèves [i, u, a] avaient la même durée, soit 300 ms, alors que les longues correspondantes [i:, u:, a:] avaient une durée qui était le double de celle des premières, c'est-à-dire 600 ms. Le rapport Brève/Longue était ainsi de 0,5.

Dans une étude de Barkat-Defradas (2001) a mesuré les durées des voyelles et leurs qualités acoustiques pour 9 sujets du Maghreb (Maroc, Algérie et Tunisie) et pour 9 sujets du Moyen-Orient (Syrie, Jordanie et Liban), avec un corpus constitué de la prononciation de la traduction spontanée, en arabe dialectal, de 12 phrases du texte " *La Bise et le Soleil* " où chaque phrase était répétée 12 fois. Les résultats, en ce qui concerne la durée, tableau 13, sont en total accord avec les descriptions de la durée des voyelles de l'arabe présentes dans les études citées plus haut, exceptées les études d'Embarki (2002, 2003 et 2004) concernant la présence de l'opposition temporelle entre voyelles brèves et longues en arabe dialectal de la ville marocaine de Ksar El Kébir.

	Durée (ms)	
	Brève	Longue
Maghreb		
Maroc	58	106
Algérie	52	103
Tunisie	51	102
Moyenne	54	104
Moyen-Orient		
Syrie	61	140
Liban	58	145
Jordanie	54	131
Moyenne	58	139

Tableau 13. Durées en ms des voyelles longues et brèves de 6 parlars dialectaux de 6 pays arabophones (3 du Maghreb et 3 du Moyen-Orient). D'après Barkat-Defradas (2001).

Elle montre que le rapport Brève/Longue est, approximativement, de 0,52 pour les parlars du Maghreb, alors qu'il est de 0,42 pour les parlars du Moyen-Orient. Le fait que l'auteur mélange des données provenant de locuteurs féminins et des données qui proviennent de locuteurs masculins n'a pas d'impact sur la valeur moyenne de la durée, comme nous l'avons vu dans le paragraphe I.2.1.1.2.

Jomaa (1991) pour l'arabe tunisien et Rhardisse (1989) pour l'arabe marocain montrent des résultats comparables à ceux de Barkat-Defradas (2001) pour les parlars du Maghreb, en effet ils rapportent que le rapport [a/a:] est de 59% et de 77% respectivement. Tandis que Norlin (1987), pour l'arabe égyptien, rapporte des valeurs proches de celles de Barkat-Defradas (2001), le rapport Brève/Longue (85/152) étant de 56%. Port *et al.* (1980) affirment que ce même rapport observé auprès des locuteurs irakiens, koweïtiens et égyptien, qui lisent un texte de l'arabe standard, est de 39% (60/155). Les résultats de Mitleb (1984) pour l'arabe jordanien sont très différents de ceux de Port *et al.* (1980), puisque le premier trouve que le rapport est de 62%.

Cette unanimité a été ébranlée par les études d'Embarki (2002, 2003, 2004). En effet ce dernier, en étudiant l'acquisition de la quantité vocalique en arabe marocain (arabe dialectal de la ville de Ksar El Kébir), montre que les enfants en âge préscolaire (de 4 à 6 ans) ne

possèdent pas ce procédé phonétique, pourtant largement attesté dans les parlers arabes. Nous observons une attention particulière à l'égard de ces études qui sont les seuls que nous connaissions qui vont en quelque sorte à l'encontre des études qui ont porté sur la quantité vocalique dans les dialectes arabes, en effet, les études d'Embarki réfutent la vérité selon laquelle la présence de différences temporelles entre voyelles longues et brèves qui supporte le trait de la quantité vocalique dans les dialectes arabes soit une évidence. Embarki (2002) a mesuré les durées de la voyelle [a] en syllabe ouverte en arabe dialectal marocain en production normale et en production hyperarticulée de 3 enfants âgés de 4 à 6 ans afin de savoir si l'opposition temporelle existe vraiment en arabe dialectal du Maroc, puisque à cet âge les locuteurs ne sont pas encore influencés par la représentation scolaire de l'opposition quantitative. Il soumet alors cette opposition de durée à l'effet perturbateur de l'hyperarticulation afin d'en tester la robustesse. Selon lui, si de précédentes études (Fennan, 1986; laabi, 1975), ont rapporté la présence d'un contraste temporel entre voyelles longues et voyelles brèves en arabe marocain, c'est parce que les deux termes de chacune des paires minimales utilisées dans les corpus de ces études s'opposent plutôt par la qualité acoustique de la voyelle que par la durée. Le corpus d'Embarki (2002) était formé de 10 mots bisyllabiques [CV/CV] de l'arabe dialectal de la ville de Ksar El Kébir (Maroc), les équivalents en arabe standard ont la structure syllabique [CV:/CV]. Les résultats ont montré que la condition expérimentale de la parole hyperarticulée a significativement affecté la durée des voyelles. Cependant aucun des trois enfants ne faisait la différence au niveau temporel entre la voyelle de la première syllabe et celle de la deuxième syllabe ni en parole normale ni en parole hyperarticulée. Autrement dit, le procédé de la quantité vocalique basée sur la différence de durée n'existe pas chez les locuteurs de l'arabe dialectal de Ksar El Kébir avant scolarisation. Cela prouve que le parler de l'arabe dialectal de cette ville n'a pas cette opposition de manière substantielle, et que si ce contraste quantitatif basé sur la différence de durée existe à un âge postérieur, ce qu'il est dû à la scolarisation, qui met le sujet de l'arabe dialectal en contact avec l'arabe standard.

Dans une étude ultérieure (Embarki, 2003), l'auteur a répliqué son travail de recherche de 2002 sur un échantillon de taille plus importante (32 enfants), et sur une tranche d'âge plus importante aussi, de 6 à 16 ans, les résultats étaient les mêmes, c'est-à-dire qu'il n'y avait pas de différences significatives de durée entre les voyelles 'supposées' longues et les voyelles 'supposées' brèves.

Toujours dans le but de déterminer la présence, en production et en perception, de la

quantité vocalique en arabe marocain, Embarki (2004) intègre dans ses locuteurs 8 adultes âgés de 32 à 42 ans de la ville de Ksar El Kébir. Les sujets de son étude portant sur la production étaient 32 enfants âgés de 6 à 16 ans, tous scolarisés, et 4 hommes âgés de 33 à 38 ans sans connaissance élaborée de l'arabe standard. Ils prononçaient des mots bisyllabiques selon deux débits de parole, normal et rapide. Les résultats du débit normal montrent que ni les enfants ni les adultes ne produisent une différence de durée entre voyelles longues et voyelles brèves. Le débit rapide a réduit significativement les durées des voyelles, les résultats demeurent les mêmes que ceux qui étaient observés en débit normal, c'est-à-dire l'absence de la différence de durée entre voyelles longues et voyelles brèves. En perception, les sujets étaient 32 enfants âgés de 7 à 15 ans, et 3 hommes et une femme âgés de 32 à 42 ans. Contrairement aux adultes de la partie production de l'étude, les adultes de la partie perception de l'étude avaient une très bonne connaissance de l'arabe standard. Les stimuli étaient des mots naturels puisés dans le lexique de l'arabe dialectal de la ville de Ksar El Kébir et prononcés par un locuteur natif. Les mesures ont montré que les voyelles du corpus n'avaient pas des durées différentes. Les résultats étaient surprenants et moins uniformes que ceux de la partie production. En effet, les résultats montraient que pour 36,5% des réponses, les sujets percevaient une différence de durée entre les deux voyelles noyaux des deux syllabes des mots bisyllabiques du corpus. La perception erronée de la durée s'est présentée d'une manière quasi symétrique par rapport à la position de la syllabe dans le mot, c'est-à-dire qu'il y avait autant de voyelles perçues longues en première syllabe (la voyelle de la deuxième syllabe est alors perçue brève) que de voyelles perçues longues en deuxième syllabe (la voyelle de la première syllabe est alors perçue brève). Cette perception erronée est en lien directe avec le niveau de scolarité et la connaissance de l'arabe standard. Selon Embarki (2004) plus le niveau scolaire est élevé et plus les sujets perçoivent une différence de durée (qui n'existe pas) entre les deux voyelles du mot bisyllabique.

Ces résultats montrent que l'opposition temporelle entre voyelles longues et brèves n'est pas une caractéristique évidente en arabe dialectal, les patrons de durées des voyelles de la ville Ksar El Kébir pouvant se produire dans d'autres lieux géographiques du monde arabophone. Ainsi, pour décrire un parler arabe comme possédant un système vocalique disposant de voyelles longues et brèves des mesures instrumentales des durées des voyelles en question semblent, grâce aux travaux d'Embarki (2002, 2003, 2004), tout à fait nécessaires.

Les divergences au sujet de la particularité de la phonétique de la langue arabe persistent. Nous voudrions mentionner ici quelques divergences concernant l'effet du

contexte consonantique sur la voyelle, notamment sur sa durée, avant d'exposer les résultats de certaines études qui rapportent la présence d'un écart temporel et d'un écart qualitatif entre voyelles longues et brèves en arabe.

Dans une revue de questions, Jomaa (1991) admettait que la durée de la voyelle en arabe était en rapport direct avec le lieu d'articulation de la consonne post-vocalique, c'est-à-dire avec l'endroit où se situait le resserrement le plus étroit du conduit vocal. Plus le lieu d'articulation était reculé, et plus la durée de la voyelle était longue. Cette constatation était valable, selon lui, pour les consonnes occlusives, constrictives et nasales d'une part, et pour les consonnes voisées et non voisées d'autre part. Ainsi une voyelle serait plus longue devant les consonnes vélaires et post-alvéolaires [k, g, ŋ, ʒ, ʝ] que devant les consonnes alvéolaires [t, d, n, s, z]; elle est plus longue devant les alvéolaires que devant les labiodentales ou les bilabiales [p, b, m, f].

L'étude d'Alioua (2003) mettait en question ces résultats, l'auteur essayait de montrer que ces postulats n'étaient pas généralisables à toutes les langues du monde et il a choisi l'arabe pour montrer le contre exemple. Ainsi, il a cherché à vérifier l'effet du contexte (lieu et mode d'articulation), et le régime phonatoire sur la durée des voyelles. Il a aussi examiné la corrélation durée/aperture.

Les sujets de son étude étaient 4 locuteurs de l'arabe marocain, 2 femmes et 2 hommes qui lisaient des paires minimales inscrites sur des fiches indépendantes, lesquelles étaient présentées en ordre aléatoire. Pour la segmentation, l'auteur de l'étude délimitait les voyelles à partir des tracés spectrographiques et oscillographiques.

Les résultats de son étude vont à l'encontre des études qui, depuis Lehiste (1970), avançaient que plus la consonne adjacente (antéposée ou postposée) était reculée et plus la voyelle est longue. Pour le rapport entre la durée de la voyelle et le lieu d'articulation de l'entourage consonantique, les données sur l'arabe marocain de l'étude d'Alioua (2003) ne vont pas dans ce sens, parce que les mesures des voyelles [a:, i:, u:] dans 4 lieux d'articulation: laryngal, pharyngal, uvulaire et vélaire, avec deux consonnes par lieu, montraient qu'il n'y avait pas de corrélats entre la durée de la voyelle et le lieu d'articulation de la consonne adjacente. En effet, les voyelles n'avaient pas une durée plus importante dans le voisinage des consonnes arrière par rapport à des consonnes moins arrière. Cela a permis à l'auteur de l'étude de confirmer que le lieu d'articulation des consonnes adjacentes en arabe marocain n'avait pas d'impact majeur et stable sur les durées des voyelles.

Alioua a vérifié aussi l'hypothèse selon laquelle la durée de la voyelle est déterminée par la consonne postposée, alors que la consonne antéposée n'a pas d'effet. Son expérience qui a consisté en l'insertion de voyelles longues en 2 contextes $C_1V:C_2$ et $C_2V:C_1$, montrait que la voyelle [a:] était affectée, légèrement, par la consonne antéposée. Alors que les voyelles [i:] et [u:] confirmaient l'hypothèse de départ. Le rapport entre la durée de la voyelle et la distribution de la consonne voisine en arabe est conforme à ce qui était déjà trouvé pour d'autres langues (Peterson & Lehiste, 1960; Fintofit, 1961). Alioua ne rapporte pas cependant une validation de ses résultats par les tests statistiques appropriés.

Il est commun d'admettre que les consonnes voisées d'un côté et constrictives de l'autre allongent la durée de la voyelle qui la précède. Peterson & Lehiste (1960) ont rapporté que dans la parole naturelle en anglais la durée de la voyelle quand elle est produite devant une consonne occlusive non voisée est, en moyenne, 197 ms, et de 297 devant une consonne occlusive voisée. Le rapport: [Durée de la voyelle devant une consonne non voisée/Durée de la voyelle devant une consonne voisée] est de 2/3. Concernant l'effet du voisement et du mode articulaire de l'entourage consonantique sur la durée de la voyelle en arabe, Alioua montre que l'arabe se conforme aux tendances générales selon lesquelles les voyelles ont une durée plus importante dans l'entourage des consonnes voisées par rapport aux consonnes non voisées (Di Cristo, 1985). De même les voyelles sont plus longues dans un entourage constrictif que dans un entourage occlusif (Ladefoged & Maddieson, 1996). Effectivement, Alioua trouvait que la durée augmentait significativement lorsque la consonne antéposée ou postposée était voisée. Il en était de même pour les consonnes constrictives par rapport aux consonnes occlusives.

Il est établi que plus la voyelle est ouverte et plus elle est longue (Malmberg, 1975; Rostolland *et al.*, 1985). Les données de l'arabe marocain ne vont pas dans le même sens que la théorie, puisque les résultats montraient qu'il n'y avait pas de corrélat proportionnel et systématique entre l'aperture d'une voyelle et sa durée, le [a:] n'était pas systématiquement la voyelle la plus longue, se faisant dépasser par le [u:] ou encore le [i:], alors qu'elle était la plus ouverte. Il était démontré, par ailleurs que les voyelles longues de petite aperture sont moins ouvertes que leurs correspondantes brèves.

L'intérêt fondamental des résultats de cette étude réside dans la controverse jetée sur certaines généralités, dans la négation de certains postulats linguistiques, ce qui nous conduit à ne pas nous conformer à certains acquis qui pourraient être vrais pour certaines langues alors qu'ils ne le sont pas pour certaines d'autres, et à tenir ainsi compte de la spécificité de la

langue qui fait l'objet de notre étude.

II.4. Espace vocalique de l'arabe

Nous discutons maintenant de certains travaux de recherche qui ont étudié et décrit l'espace vocalique de certains dialectes arabes parlés au travers le monde arabophone, ce qui nous permet de comparer nos résultats à propos de l'arabe de Mayadin (Syrie) pour la durée et pour l'espace vocalique, avec d'autres résultats obtenus pour autres dialectes.

Barkat-Defradas (2001), rapporte que les parlers arabes du Moyen-Orient ont des espaces vocaliques plus larges, en ce sens que les voyelles sont plus périphériques que celles des parlers arabes du Maghreb. Ceci est clairement observable sur la figure 38. Nous trouvons dans cette étude un point discordant par rapport aux études précédentes. Bien que la tendance à la centralisation soit bien marquée dans les parlers arabes du Maghreb, notre examen des données de Barkat-Defradas (2001) montre que cette tendance est pratiquement absente dans les parlers du Moyen-Orient (figure 38), ce qui ne va pas de pair avec les résultats qui ont examiné l'espace vocalique de l'arabe, et avec nos propre résultats (Allatif & Abry, 2004), concernant l'arabe de Mayadin de la vallée de l'Euphrate de l'Est de la Syrie. Cela est peut-être dû au fait que l'auteur de l'étude mélange des résultats de mesures provenant de sujets féminins et d'autres provenant de sujets masculins, deux espaces qui n'ont pas les mêmes dimensions (Hillenbrand *et al.*, 1995, figure 24). La description de l'étude de Barkat-Defradas (2001) est assez détaillée, ce qui nous a permis de savoir que l'auteur avait en effet mis ensemble des données provenant d'un homme et d'autres provenant d'une femme.

Dans un travail préliminaire sur la variabilité interlocuteur et intra-locuteur dans la production et la perception de l'espace vocalique selon différents systèmes phonologiques, Barkat-Defradas *et al.* (2003) ont étudié la variabilité en perception et en production de 20 locuteurs de deux dialectes arabes, l'arabe jordanien (AJ) et l'arabe marocain (AM).

D'après l'auteur, l'AM dispose des 5 voyelles suivantes [i:, u, u:, ə, a], et l'AJ a les 8 voyelles suivantes [i:, i, u, u:, e, o:, a, a:]. Barkat-Defradas *et al.* demandaient à leurs sujets de prononcer 51 items pour l'AJ et 35 pour l'AM, chaque item était présenté 5 fois dans un ordre aléatoire à 10 locuteurs hommes natifs du dialecte en question. Quant à la tâche de perception, les locuteurs devaient déterminer le meilleur prototype de la voyelle qui leur était présentée, pour l'AJ il y'avait 16 prototypes et pour l'AM il y avait 10 prototypes. Ces prototypes ont été manipulés par un logiciel de synthèse à formants. Chaque item était présenté 10 fois dans un ordre aléatoire.

Les valeurs de la production ont été obtenues en prélevant les valeurs de F1 et de F2 dans la partie stable de la voyelle. L'étude de la distribution des voyelles longues et brèves a montré que les voyelles brèves étaient effectivement centralisées par rapport aux voyelles longues, en production comme en perception dans les deux dialectes arabes, AJ et AM.

Dans une autre étude sur l'une des variétés de l'arabe, Newman *et al.* (2002) ont décrit l'espace vocalique de l'arabe en termes de ses voyelles longues et brèves. Les auteurs ont procédé à l'analyse d'une récitation coranique (voix chantée) de 30 minutes par *shaykh* Muhammad Sadiq al-Minshawi (édité par Abou Hamza Recordings, Caire, 1994). Toutes les

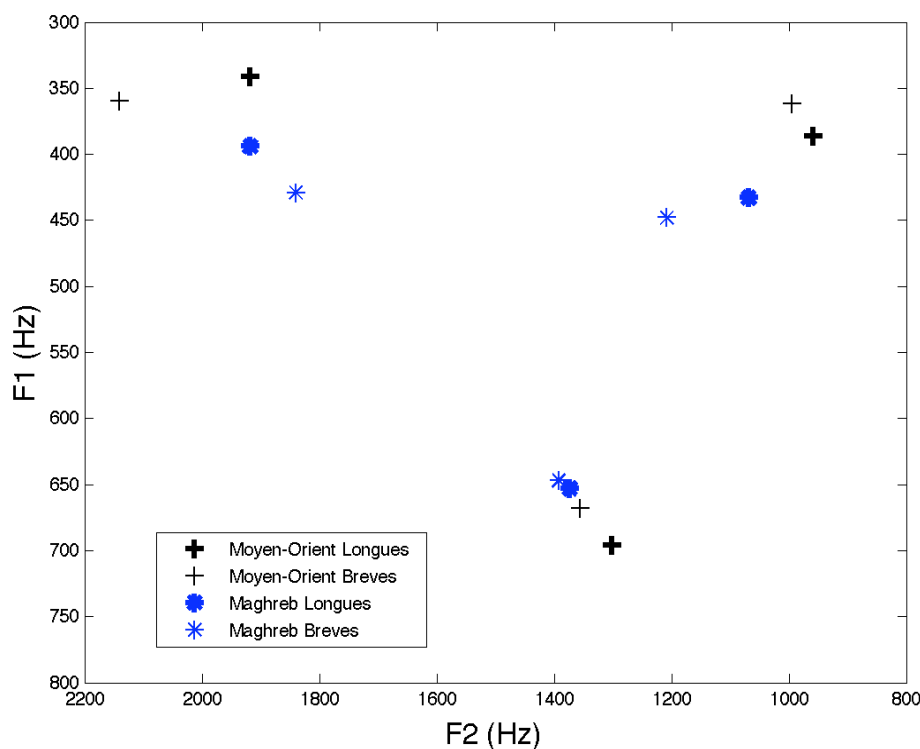


Figure 38. Espace vocalique des voyelles longues et brèves [i:, i, u:, u, a: et a] de 6 pays arabophones, (3 du Maghreb et 3 du Moyen-Orient). D'après Barkat-Defradas (2001).

voyelles de ce texte étaient segmentées manuellement et analysées par le logiciel Praat. Un total de 400 observations distribuées presque uniformément sur les 6 voyelles [i:, i, u:, u, a:, a]. La figure 39 montre que les voyelles brèves sont situées dans une zone relativement centrale. Newman *et al.* notent que cette tendance à la centralisation est plus claire pour les voyelles hautes par rapport aux voyelles basses. Cependant, les auteurs affirment que l'analyse statistique montre que ces différences (de centralisation) entre longues et brèves ne sont pas significatives. Les auteurs notent que la différence de durée, qui est phonologique en arabe standard et en arabe dialectal, n'est pas une différence significative en arabe de récitation coranique. Un résultat qui ne va pas de pair avec les résultats des études faites sur l'arabe non chanté (Al-Ani, 1970; Abou-Haidar, 1994; Ghazeli, 1979).

Ce résultat soulignerait un point d'une importance capitale, puisqu'il montre que la différence de durée est liée à une différence de timbre, parce que, sous la contrainte de la voix chantée, une contrainte qui a égalisé les durées des voyelles longues et brèves, les différences qualitatives ont disparu.

Dans la même étude, Newman *et al.* ont analysé les voyelles de l'arabe classique. Ces voyelles ont été obtenues par la lecture de la traduction d'un passage du texte "*La Bise et le Soleil*" par un natif de l'arabe du Caire. Les analyses ont été menées de la même manière que pour les voyelles du Coran.

Dans la figure 39, nous voyons la représentation graphique¹² des résultats des mesures de l'arabe standard en forme de récitation (chanté), et de l'arabe standard obtenu par la lecture. Les valeurs des voyelles dans la parole continue de l'arabe standard du Caire se rapprochent de celles de la parole non continue (corpus) des études précédentes, citées plus haut. Les auteurs soulignent notamment la centralisation des brèves par rapport aux longues, une centralisation beaucoup plus marquée que celle observée avec la récitation coranique où elle n'est pas significative.

Les données de Newman *et al.* concernant l'arabe classique, au moins en ce qui concerne la parole provenant de la lecture, semblent être en accord avec l'existence d'une structure stable de l'espace vocalique à travers les dialectes arabes, à savoir, les longues sont plus périphériques que leurs partenaires brèves.

Cette structure stable de l'espace vocalique de l'arabe était rapportée par de nombreuses

¹² La représentation graphique est approximative: Newman *et al.* ne fournissent pas les valeurs de F1 et de F2 dans leur étude. Nous avons prélevé manuellement les valeurs des deux premiers formants sur le graphique paru dans l'article, ce qui pourrait entraîner un léger décalage par rapport aux vraies valeurs.

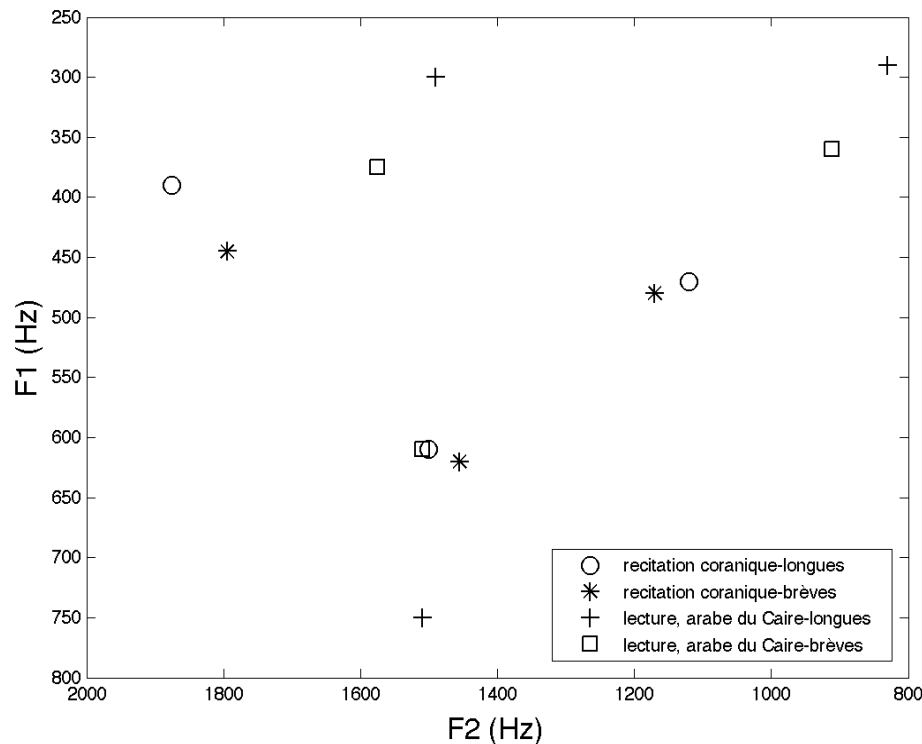


Figure 39. Comparaison de l'espace vocalique de l'arabe égyptien sous deux conditions d'expérimentation: 1) récitation coranique (voix chantée) et 2) lecture d'un texte. D'après Newman et al. (2002).

études. Nous avons choisi, comme Newman *et al.* (2002), d'en représenter 5 sur une même figure (figure 40), et d'en donner les valeurs des deux premiers formants telles qu'elles sont apparues dans les études citées dans le tableau 14.

À quelques exceptions près¹³, toutes ces études montrent que l'arabe dispose d'un espace vocalique où les brèves sont dans une zone centrale de l'espace par rapport aux longues. En effet, malgré les divergences en ce qui concerne les valeurs absolues des voyelles de l'arabe, plusieurs raisons pouvant expliquer ces différences, le sexe et l'âge du locuteur (Hillenbrand *et al.*, 1995), la région (Walter, 1982), le contexte consonantique (Barkat-Defradas *et al.*, 2003), nous pouvons observer deux triangles, un premier pour les longues et un deuxième, inclus dans le premier, pour les brèves. De même nous pouvons observer que les distances entretenues entre une longue et sa partenaire brève n'est pas la même selon la qualité de la voyelle, en effet, la distance [i:] - [i] est plus importante que la distance [u:] - [u], qui est à son tour plus importante que la distance [a:] - [a].

Cette tendance, selon laquelle les brèves ont un triangle réduit se trouvant dans le centre de l'espace vocalique, un espace délimité par les voyelles cardinales longues, était confirmé

¹³ Le F1 du /a:/ dans Belkaid (1984), le F2 du /i/ dans Al-Ani (1970).

par le travail de Abou-Haidar (1994). Dans cette dernière étude, l'auteur a réussi à décrire l'espace vocalique pour 8 locuteurs, tous de sexe masculins, tous natifs de l'arabe (parlant le standard et le dialectal) originaires des 8 pays arabophones suivants: le Qatar, le Liban, l'Arabie Saoudite, la Tunisie, la Syrie, le Soudan, les Émirats arabes unis et la Jordanie. La figure 41 est la représentation des données disposées dans le tableau 15. Nous remarquons, comme sur la figure 38, 39 et 40), la centralisation des voyelles brèves par rapport aux longues, l'ordre des distances entre les termes de chaque paire longue/brève étant aussi préservé.

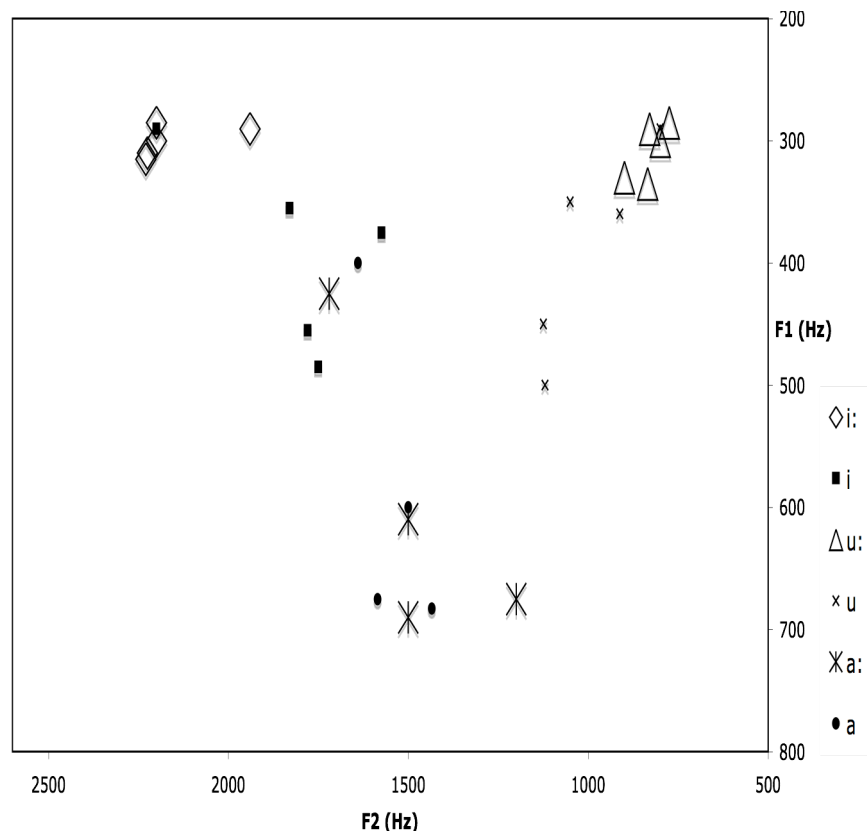


Figure 40. Espace vocalique des voyelles cardinales longues et brèves [i:, i, u:, u, a: et a] de plusieurs parlars arabes rapportés dans 5 études différentes (Al-Ani, 1970; Ghazeli, 1979; Belkaid, 1984; Abou-Haidar, 1994 et Newman, 2002)

	i:		i		u:		u		a:		a	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
(Al-Ani-70)	285	2200	290	2200	285	775	290	800	675	1200	600	1500
(Ghazeli-79)	310	2225	455	1780	330	900	450	1125	----	----	----	----
(Belkaid-84)	285	2195	355	1830	310	790	340	995	425	1720	400	1640
(Haidar-94)	315	2230	485	1750	335	835	500	1120	690	1500	675	1585
(Newman-02)	390	1725	440	1770	470	1120	480	1170	620	1455	616	1460

Tableau 14. Valeurs des deux premiers formants des voyelles cardinales longues et brèves [i:, i, u:, u, a: et a] de plusieurs parlars arabes rapportés dans les 5 études suivantes: (Al-Ani, 1970; Ghazeli, 1979; Belkaid, 1984; Abou-Haidar, 1994 et Newman, 2002). Ces valeurs sont représentées dans la figure 40.

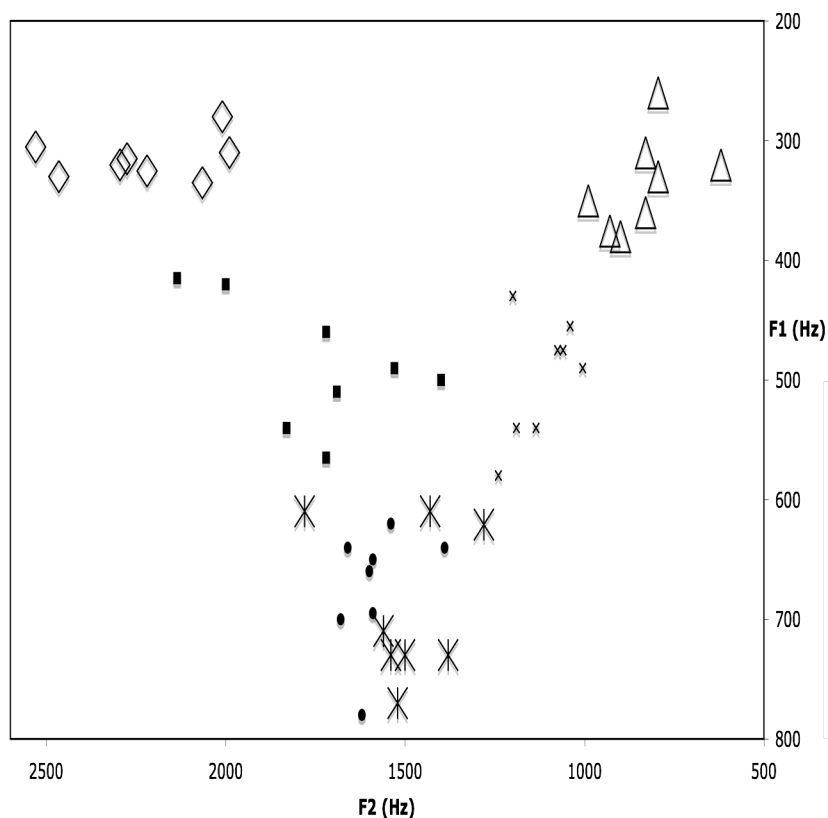


Figure 41. Espace vocalique de l'arabe pour 8 locuteurs masculins, âgés de 21 à 30 ans, originaires des 8 pays arabophones suivants: le Qatar, le Liban, l'Arabie Saoudite, la Tunisie, la Syrie, le Soudan, les Émirats arabes unis et la Jordanie. D'après Abou-Haidar (1994).

	i:		i		u:		u		a:		a	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
Qatari	310	1990	500	1400	310	830	490	1005	621	1280	620	1540
Lebanon	280	2010	490	1530	330	795	475	1060	610	1430	640	1390
Saudi	305	2530	540	1830	375	930	540	1190	730	1540	695	1590
Tunisia	315	2275	510	1690	360	830	540	1135	610	1780	650	1590
Syria	330	2465	415	2135	320	620	430	1200	710	1560	700	1680
Sudan	325	2220	420	2000	380	900	455	1040	730	1500	660	1600
UAE	335	2065	460	1720	350	990	475	1075	730	1380	640	1660
Jordan	320	2295	565	1720	260	795	580	1240	770	1521	780	1620

Tableau 15. Valeurs des deux premiers formants des voyelles cardinales longues et brèves [i:, i, u:, u, a: et a] de 8 pays arabes, mesurées par Abou-Haidar (1994). Ces valeurs sont représentées dans la figure 41.

Les études précédentes se sont contentées de constater l'existence d'une différence qualitative entre la voyelle longue et sa correspondante brève, cela en observant la non-superposition dans l'espace vocalique des deux termes du contraste, et/ou l'existence d'une différence quantitative, en rapportant l'écart temporel entre la durée de la longue et celle de la brève dans des conditions de l'émission de la parole naturelle, *a priori*, sans contrainte. La robustesse des écarts n'était pas testée. Aucune des études, exposées plus haut dans ce paragraphe, n'a posé la question de l'origine de l'information acoustique sur laquelle l'opposition Longue *vs.* Brève repose. Autrement dit, quel est réellement le trait phonétique pertinent qui permet la distinction Longue *vs.* Brève? Où la source de l'information réside-t-elle: dans l'écart quantitatif? Dans l'écart qualitatif? Dans les deux à la fois? Quelle est la part de chacun des écarts dans la création du contraste?

Pour essayer de répondre aux questions que nous venons de poser, dans le chapitre III, nous mènerons notre propre expérience, sur le dialecte arabe de Mayadin (Est de la Syrie), pour lequel nous établirons l'espace vocalique des voyelles longues et brèves, /i:/, /i/, /u:/, /u/, /a:/ et /a/, sous différentes conditions expérimentales dans le but de tester la robustesse des écarts qualitatifs, représentés par les distances euclidiennes entre les deux termes de chaque paire. De même nous mesurerons les écarts temporels pour ces mêmes voyelles sous les mêmes conditions d'expérimentation. Nous essayerons ensuite de donner des réponses à la lumière des résultats que nous avons exposés dans le premier chapitre de ce manuscrit, des résultats concernant le seuil de la qualité et le seuil différentiel de la perception de la durée des voyelles.

III. Expérience sur l'arabe dialectal de Mayadin (Syrie)

III.1. Introduction en repère proximal: quantité et qualité dans le dialecte arabe d'Amman (Jordanie).

Nous avons indiqué dans le chapitre précédent que la quantité vocalique en arabe a fait l'objet de plusieurs études acoustiques (Al-Ani, 1970; Ghazeli, 1979; Belkaid, 1984; Abou-Haidar, 2002; Barkat-Defradas, 2003). Toutes rapportent une opposition temporelle dans les conditions normales de l'émission de la parole, en formes de citation, les voyelles longues ont une durée supérieure à celle des brèves correspondantes. Certaines rapportent une différence qualitative dans des conditions normales. Dans notre étude, nous posons la question de savoir si ces différences, temporelles et spectrales, se maintiennent ou non lorsque le locuteur doit adapter sa performance phonétique à des variations de débit d'élocution ou d'intonation. Nous appliquerons ces contraintes sur le dialecte arabe de la ville de Mayadin (Syrie).

Notre hypothèse pour l'arabe de Mayadin est que si les oppositions temporelles entre brèves et longues sont robustes, et par conséquent source de l'information acoustique concernant le contraste Brève *vs.* Longue, alors ces oppositions se réorganisent sous des contraintes naturelles de la communication linguistique pour se maintenir, de telle sorte que leur rôle initial ne disparaisse pas.

Dans une étude précédente, où nous avons examiné le système vocalique des voyelles du dialecte de Mayadin (Allatif, 2003), nous avons remarqué que les voyelles brèves et longues n'avaient pas les mêmes coordonnées dans l'espace vocalique $F1 \times F2$, les brèves se situant dans une région centrale de l'espace vocalique. Cette caractéristique est commune aux dialectes arabes (Abou-Haidar, 1994). Alors, si ces différences qualitatives observées sont linguistiquement spécifiées et jouent un rôle dans l'opposition Brève/Longue, ces différences dites de qualité vocalique, autrement dit spectrales, accompagnées des conséquences des contraintes biomécaniques dues, ici, à l'adaptation aux conditions rythmiques et intonatives, devraient se maintenir pour préserver l'information linguistique. Enfin cette différence spectrale est-elle contingente ou indispensable au fonctionnement du système de 6 voyelles (3 brèves et 3 longues) de l'arabe classique? Nous examinerons ces différences du point de vue statistique et du point de vue perceptif.

Les contraintes de variation prosodique et de débit d'élocution sont des éléments significativement perturbateurs. En effet, nous rappelons que Newman *et al.* (2002) ont trouvé que la centralisation des voyelles brèves dans l'espace vocalique de l'arabe (par rapport aux longues) devenait non significative lorsque les voyelles sont émises avec la récitation coranique (voix chantée). Cette contrainte affecte aussi l'opposition quantitative, Newman *et al.* (2002) notent que la différence de durée devient non significative lorsque les voyelles sont émises avec la voix chantée.

D'autres études (Minagawa *et al.*, 2003) ont exploré les effets de la variation prosodique sur la durée des voyelles longues et brèves. En fait, il est établi que le japonais est une langue qui connaît un contraste quantitatif systématique. Minagawa *et al.* (2003) ont étudié les variations des durées des voyelles en fonction de leurs conditions prosodiques d'accent, de débit d'élocution et de position de la syllabe dans la phrase. Les analyses étaient faites à partir du CSJ "Corpus of Spontaneous Japanese". Cette étude, qui a examiné certaines caractéristiques temporelles des voyelles longues et brèves en japonais spontané, a montré que les conditions accentuelles avaient un faible, mais significatif, effet sur les durées des voyelles et des mores et que le contraste temporel entre voyelles longues et brèves était robuste sous trois niveaux de débits différents, lent, rapide et très rapide. En effet, les analyses ont montré un allongement de la durée de la voyelle dû à la position finale de la syllabe dans la phrase, en comparaison avec la position non finale mono-moraïque. Le rapport d'allongement était de 1,3 à 1,5. Ces rapports baissent quand le débit augmente. Ils notent que ce rapport, quoique significatif, ne reflète pas une tendance universelle. Pour l'anglais c'est tout à fait le contraire: ce rapport augmente avec l'augmentation du débit de la parole. Les auteurs pensent qu'une telle différence entre les patrons temporels du japonais et de l'anglais est probablement due à la spécificité phonologique de la langue dans son traitement du contrôle du timing. L'accent a eu un effet significatif sur les durées des voyelles brèves (les différences étaient entre 3 et 14 ms). Concernant la comparaison entre voyelles longues et brèves, l'analyse de la durée comme une fonction du débit de la parole a montré un effet plus important sur la voyelle longue (syllabe bi-moraïque) que sur la voyelle brève, comme il a montré que le rapport longue/brève (syllabe bi-moraïque/syllabe mono-moraïque) avait tendance à baisser plus le débit augmentait. Cette étude prouve finalement que soumettre les oppositions entre voyelles longues et brèves à des conditions expérimentales, telles que le débit et l'accent est une procédure phonétique qui est en mesure de perturber significativement les durées des voyelles.

La prosodie peut donc révéler les propriétés distinctives des systèmes vocaliques: est-ce le cas de l'arabe?

Pour l'arabe, l'étude de de Jong & Zawaydeh (2002) avait sondé en profondeur le rapport entre les variations de certaines caractéristiques prosodiques et les oppositions quantitatives et qualitatives de la voyelle /a/ afin de mieux caractériser le comportement de certains procédés prosodiques de l'arabe dialectal. Elle a décrit comment l'opposition de quantité (longue *vs.* brève) et l'état de voisement de la consonne post-vocalique (occlusive voisée *vs.* occlusive non voisée) agissent sur les durées des voyelles et les valeurs de leurs formants en arabe dialectal. Les auteurs ont décrit aussi l'effet des procédés prosodiques d'accent (décrit comme une magnification de l'activité articulatoire: "hyperarticulation"), de focus lexical et de focus segmental sur la variation de la durée des voyelles, une variation due aux effets magnifiants de la quantité (les brèves augmentent leurs durées en passant pour des longues) et de l'état de voisement de la consonne post-vocalique (les voyelles sont plus longues lorsque la consonne post-vocalique est voisée).

En examinant les effets de l'accent, du focus lexical et du focus segmental sur l'usage des contrastes segmentaux de la quantité et du voisement, de Jong & Zawaydeh (2002) ont trouvé que l'accent et le focus segmental sont des outils utiles pour déterminer si, et à quel degré, un effet phonétique particulier est linguistiquement spécifié. Dans leur étude, l'accent et le focus segmental indiquent que, en arabe, la quantité conditionne (spécifie) les durées relatives aux voyelles, alors que le voisement ne fait pas cela. Par conséquent l'accent et le focus segmental tendent à augmenter l'expression des spécifications de la quantité, alors qu'ils laissent tel qu'il est, ou ils éliminent, les différences temporelles de durées dues au voisement. Les effets du focus lexical sont moins clairs à cet égard, indiquant que dans certaines circonstances les effets du voisement peuvent être augmentés alors qu'ils n'indiquent aucune hausse dans les différences contrastives de la quantité. Il semble qu'un focus lexical plus marqué tende à créer une hausse de la production vocale sans créer forcément des hausses plus ciblées dans des aspects contrastifs spécifiques du signal.

L'accent et les deux types de focus se sont montrés différents dans leurs conséquences. Quelques effets, tel ceux de l'accent et de la quantité, sont bien conventionnels et semblent être les mêmes à travers les sujets et à travers les différentes conditions expérimentales. D'autres effets, comme le focus, sont différemment employés par les différents locuteurs. Il semble raisonnable de penser qu'effectuer une tâche comme le focus lexical, ou encore le

focus segmental, ne se pratique pas d'une manière commune. Ces résultats soulignent que tous les phénomènes de la parole n'ont pas le même statut concernant le code linguistique conventionnel que locuteurs et auditeurs partagent. Utiliser différentes tâches qui testent explicitement les compétences linguistiques des locuteurs est particulièrement utile pour comprendre la part de la parole dans le langage.

Selon de Jong & Zawaydeh (2002), la notion d'accent fait partie d'une notion linguistique plus générale, c'est la notion de Focus. Le focus (ou plus communément *mise en relief*) couvre en effet un plus grand nombre de phénomènes linguistiques: phonétique, morphologique, syntaxique et pragmatique. Au niveau phonétique, le rapport entre l'accent et le focus peut être décrit comme suit: l'accent est un focus qui se limite au domaine de la syllabe. Le focus est relatif à des items lexicaux multisyllabiques, ce qui fait que le focus est un procédé qui relève plutôt de la morphologie et de la syntaxe. Ainsi le focus qui porte sur un item lexical implique une hyperarticulation couvrant tout le mot concerné par le focus, le but du focus étant de contraster un mot avec les autres mots de la phrase. L'accent porte cependant sur une seule syllabe de l'item lexical. de Jong & Zawaydeh (2002) affirment que, dans certains cas, accent et focus lexical sont non distinguables.

Les auteurs plaident pour l'existence d'un focus qui porte sur une unité plus petite que la syllabe, ils rapportent que van Heuven (1994) a trouvé que les locuteurs étaient capables d'hyperarticuler un contraste phonologique pour attirer l'attention sur un contraste d'ordre segmental. de Jong & Zawaydeh (2002) appelle ce dernier type de focus le focus segmental pour le différencier du focus qui porte sur un mot entier (focus lexical).

Les résultats de l'étude de de Jong & Zawaydeh (2002) montrent que la quantité et le voisement, tous deux, affectent la durée, ils sont cependant différents dans leurs interactions avec l'accent et le focus. La différence de quantité est magnifiée par l'accent, alors que la différence de durée due au voisement est insensible à l'accent et au focus segmental.

Au plan quantitatif, l'accent augmente la durée de la voyelle. Au niveau qualitatif, l'accent déplace la voyelle vers la périphérie de l'espace vocalique, alors que l'état de voisement n'affecte que la durée des voyelles. Ils notent que les différents types d'allongements des voyelles n'ont pas tous les mêmes caractéristiques dynamiques. Ils rapportent que Summers (1987), en examinant les effets de l'accent et du voisement des consonnes sur la durée des voyelles en anglais, a trouvé que l'allongement dû au voisement se localisait dans la partie finale de la voyelle, au moment où le geste de fermeture est en train de s'exécuter, alors que, selon Edwards *et al.* (1991), l'accent a un effet global allongeant la

partie médiane de la voyelle, cette partie étant produite pendant que le conduit vocal est largement ouvert. D'autres études articulatoires (Harrington *et al.*, 1996) suggèrent que ce patron de l'expansion temporelle de la partie médiane de la voyelle frappée par l'accent est du au geste de rephasage, tel que le geste de fermeture commence à un moment postérieur au geste d'ouverture. Un effet commun de ce rephasage est la magnification du mouvement d'ouverture du à l'élimination de l'undershoot. Si le geste de fermeture commence plus tard, le geste d'ouverture est ainsi autorisé à évoluer plus longtemps avant d'être interrompu.

Dans une autre étude (de Jong & Zawaydeh, 1999), les auteurs ont examiné les F1 pour les voyelles basses de l'arabe. Ils ont détecté une différence de mécanisme entre les effets de l'accent et l'allongement dû à la position (pénultième) de la syllabe. Dans cette étude, ils trouvent aussi que les syllabes sous l'accent ont un F1 plus élevé pour les voyelles basses, alors que l'allongement de la syllabe pénultième n'a pas d'effet sur le F1. Ainsi, alors que l'accent engendre une durée plus importante et des valeurs pour F1 plus extrêmes, le mécanisme associé au voisement et à l'allongement dû à la position de la syllabe, produit seulement des durées plus importantes.

L'étude que nous considérons ici (de Jong & Zawaydeh, 2002), examine les durées et les qualités des voyelles en arabe dialectal de Jordanie dans le but de comparer et de contraster l'accent et le focus. L'arabe était choisi pour cette étude parce que c'est une langue à accent comme l'anglais. de Jong & Zawaydeh (1999), en étudiant les durées des voyelles et leurs qualités, ainsi que les patrons de la fréquence fondamentale de l'arabe dialectal d'Amman (Jordanie), ont trouvé que l'accent augmentait systématiquement les durées et les valeurs des F1 des voyelles basses. Ils ont constaté que l'arabe d'Amman, comme l'anglais, faisait usage de l'accent haut et bas de la F0 associé aux syllabes sous accent. Cet accent dépend du schéma mélodique de la phrase, interrogative ou déclarative respectivement. D'autres études plus récentes montrent que les syllabes sous accent en anglais et en arabe sont sélectionnées pour recevoir plus d'attention en production. Les auteurs de l'étude rapportent que Tajima *et al.* (1999) trouvent que l'arabe est plus proche de l'anglais que du japonais du point de vue rythmique. Cela a été mis en évidence par l'élaboration d'un patron rythmique basé sur la tâche qui consiste en la répétition des phrases. Une autre raison pour le choix de l'arabe pour cette étude (de Jong & Zawaydeh, 2002), est que cette langue manifeste deux effets sur les durées des voyelles, lesquels sont différents concernant leurs rôles phonologiques. Le premier effet est que la hausse de la durée des voyelles due à la quantité est phonologique en arabe, la quantité est la première base de contraste, selon eux, pour les

consonnes et pour les voyelles. Ils mentionnent que Mitleb (1984b) trouve que la quantité cause une hausse de 65% des durées des brèves. Le second effet, c'est que la hausse de la durée de la voyelle due au voisement de la consonne post-vocalique n'est pas phonologique. Ils rapportent que, concernant l'arabe, l'effet de l'état de voisement de la consonne post-vocalique sur la durée de la voyelle préconsonantique ne fait pas l'unanimité. En effet, ils signalent que Port *et al.* (1980) et Munro (1993) trouvent classiquement que les voyelles suivies par des consonnes sourdes sont significativement moins longues que lorsqu'elles sont suivies par des consonnes sonores. Alors que Flege & Port (1981) et Mitleb (1984a) trouvent que l'état de voisement de la consonne post-vocalique n'a pas d'effet significatif sur la durée de la voyelle qui la précède. Ce résultat pousse Mitleb (1984a) à penser que la hausse de la durée due au voisement en arabe ne constitue pas un indice sur l'état de voisement de la consonne, alors que c'est le cas en anglais. L'anglais est en effet une langue distinguée par le fait que l'allongement dû au voisement est exploité linguistiquement.

de Jong & Zawaydeh (2002) examinent comment l'accent, le focus lexical et le focus segmental affectent les deux effets temporels segmentaux de la quantité et du voisement dans l'arabe dialectal. Les auteurs posent la réflexion suivante: dans la mesure où accent et focus se servent des mêmes mécanismes et que ces mécanismes font partie de l'usage des locuteurs arabes, on devrait s'attendre à ce qu'accent, focus lexical et focus segmental produisent les mêmes effets dans leurs domaines d'application (syllabe, item lexical et environnement local segmental respectivement). On doit s'attendre aussi à ce qu'ils interagissent de la même manière avec les deux effets segmentaux (quantité et voisement). Si la description de Mitleb (1984a), à propos de la différence entre les deux effets comme un indice de contraste, est correcte, alors il faut que les trois mécanismes augmentent la durée des voyelles au-delà de l'augmentation due à la quantité, et aussi qu'aucun n'augmente la durée au-delà de l'augmentation due au voisement. Cette étude examine aussi les différences de qualité vocalique, en s'attendant à ce que si deux mécanismes sont les mêmes, ils produiront alors le même effet sur la durée et la qualité.

Les sujets de leur étude étaient 4 locuteurs, 2 femmes et 2 hommes, locuteurs natifs de l'arabe d'Amman (Jordanie), dont l'âge variait entre 17 et 28 ans. Les sujets de l'expérience étaient tous polyglottes, mais seul l'arabe était utilisé au moment de l'enregistrement qui était effectué à Amman.

Le plan expérimental était composé de deux parties. La première était conçue pour tester les items sans focus et les items sous focus lexical. La deuxième partie était conçue

pour tester les contrastes de quantité et de voisement des items qui sont sous focus segmental. Vu la difficulté de trouver des paires minimales, nécessaires au test du focus segmental, une partie du corpus était commune aux deux parties du plan expérimental.

Chaque mot du corpus avait un des quatre segments cibles, le segment cible était composé de [b], [a] et d'une consonne occlusive. Les voyelles sous l'accent et sans accent étaient des longues et des brèves, suivies d'une consonne coronale qui était soit sourde soit sonore. Ce qui produit 8 formes au total, soit les 4 formes [bad] vs. [bat] et [baad] vs. [baat], une fois sous l'accent et une fois sans accent.

Afin de faire correspondre le corpus à celui de de Jong (1991), la consonne finale était suivie par une voyelle. Cette caractéristique signifie que la consonne qui suit la voyelle formera une syllabe avec la voyelle qui la suivra et non pas avec la voyelle qui la précédera (la voyelle cible). Des études sur les effets du voisement, comme celle de Chen (1970), avaient inclus les effets internes (à la syllabe) et les effets trans-syllabiques, les deux types d'effets semblaient être de la même envergure. C'est ainsi que les auteurs s'attendaient à ce que la consonne augmente la durée de la voyelle qui la précède, mais cette caractéristique complique l'interprétation des effets de l'accent.

À part le contraste de longueur, l'identité de la voyelle reste fixée afin de permettre l'examen de la corrélation entre la variation de la durée et celle de la qualité de la voyelle. Les auteurs soutiennent que l'arabe d'Amman dispose de 5 voyelles contrastées systématiquement en quantité. Ils poursuivent que les voyelles longues mi-ouvertes sont dérivées historiquement des diphtongues /ay/ et /aw/ (selon la notation des auteurs). Ils assurent qu'en arabe, quand les voyelles ouvertes se rencontrent dans un environnement phonétique qui contient une consonne emphatique ou uvulaire, elles deviennent uvularisées, et ainsi la voyelle basse a plus d'allophones arrières dans ces conditions. Aucun des mots cibles ne contenait une consonne pharyngale ou emphatique à cause de leur effet trans-segmental (longue portée) sur la qualité de la voyelle (Zawaydeh, 1999).

Les 4 mots cibles (voyelle brève ou longue croisée avec une consonne finale qui peut être voisée ou non voisée) étaient placés dans des mots trisyllabiques, où ils étaient soit sous accent soit sans accent; cela dépendait du fait que la coda de la syllabe suivante contenait ou non une consonne. L'accent dans ce dialecte, selon les descriptions impressionnistes, tombe sur la dernière syllabe ultralourde, la pénultième (avant dernière) syllabe lourde ou l'antépénultième (avant avant dernière) syllabe. Le nombre total de mots de la première partie de l'expérience était 8. 5 mots étaient ajoutés en outre pour obtenir des paires minimales

nécessaires à la deuxième partie. Pour obtenir ces paires minimales, l'exigence sur la position du mot était relâchée.

Les mots apparaissent dans différentes conditions phrastiques et discursives afin de varier le focus. La première partie a nécessité deux contextes, post noyau et focus lexical. Dans le cadre post noyau, les locuteurs plaçaient le focus sur un item qui précédait la cible:

ʔana ʔult *badawi* mi ʃ huwwa. "j'ai dit *bédouin* pas lui" (le focus est placé sur le mot souligné, le mot cible est en italique).

Avec le focus lexical, les locuteurs plaçaient le focus sur le mot cible:

ʔana ʔult *badawi* mi ʃ fallaħ. "j'ai dit *bédouin*, pas paysan" (le focus est placé sur le mot souligné, le mot cible est en italique).

Dans chaque cas de focus lexical, le mot n'avait pas de similarité phonologique.

La deuxième partie du plan nécessitait une condition phrastique semblable à celle du focus lexical, sauf que les locuteurs corrigeaient à des auditeurs qui étaient censés avoir mal perçu un contraste particulier sur un segment du mot cible. Cette portion nécessitait que chaque mot cible soit minimalement contrasté en quantité ou en voisement avec une autre forme avec laquelle il était juxtaposé. Pour mettre le focus sur la quantité, les locuteurs lisaient la phrase suivante:

ʔana ʔult *baada* mi ʃ bada "j'ai dit *il a exterminé*, pas il a commencé" (le focus est placé sur le mot souligné et en italique).

Pour mettre le focus sur le voisement, les locuteurs lisaient la phrase suivante:

ʔana ʔult *baada* mi ʃ baata "j'ai dit *il a exterminé*, pas il resté debout toute la nuit" (le focus est placé sur le mot souligné et en italique).

Hypothèses:

Les auteurs observent l'effet de l'accent, du focus lexical et du focus segmental sur la durée et la qualité vocaliques, et leurs interactions avec la quantité et le voisement en subdivisant le corpus et en appliquant des analyses de la variance sur les mesures de la durée de la voyelle. Il est commun qu'une durée plus longue accompagne l'accent et le focus. Toutes choses égales par ailleurs, ils s'attendent à ce qu'il y ait une hausse de durée avec chacun des trois procédés. Ils prévoient évidemment que la durée augmente en passant des

voyelles brèves vers les longues, et en faisant suivre les voyelles par des consonnes voisées. La quantité devrait avoir aussi une interaction avec l'accent et avec les deux types de focus, parce qu'une plus haute attention se traduit par une magnification de la taille de l'effet temporel. Le contraste de la quantité devrait être nettement plus important dans les syllabes sous l'accent, partout dans le mot frappé par un focus lexical et quand les locuteurs placent segmentalement le focus sur la quantité. Le voisement ne devrait pas avoir une interaction avec l'accent ou le focus, cette absence d'interaction proviendrait du fait que la hausse de la durée due au voisement de la consonne n'est pas considérée comme un changement à valeur phonologique en arabe. Accent, focus lexical et focus segmental ont *a priori* différents domaines d'application. Les effets et les interactions de l'accent doivent être restreints aux syllabes sous l'accent et ne doivent pas apparaître dans des zones sans accent. Les effets du focus segmental et du focus lexical peuvent se trouver dans n'importe quel endroit du mot, bien qu'il soit possible que la taille de l'effet puisse être différente selon la place de l'accent.

Pour examiner statistiquement ces prédictions, les auteurs ont partitionné en séries équilibrées toute la base de données, ce qui permettait de faire varier le facteur considéré. Les mesures de durée ont été testées par des Anova multivariées.

Des Anova étaient appliquées aussi aux mesures faites sur F1 et F2. En effet, des études précédentes ont montré que l'accent et le voisement sont différents dans leurs manières d'entraîner une hausse de F1 corrélée à l'allongement temporel de la voyelle. En se basant sur de précédents résultats concernant l'accent, les auteurs s'attendent ainsi à ce que l'accent, le focus lexical et le focus segmental causent une hausse de F1. En outre, ils pensent qu'ils doivent s'attendre à ce que l'allongement dû à la quantité entraîne aussi à son tour une hausse de F1, à cause de la baisse de l'undershoot. En se basant sur des résultats, tels ceux de Summers (1987), ils ne s'attendent pas à ce que le voisement affecte le F1.

Les analyses de F2 ont été effectuées spécifiquement pour tester les différences de qualité entre les voyelles longues et brèves. Une différence qualitative sur la dimension avant-arrière des voyelles basses longues et brèves a été observée dans différents dialectes arabes. Si une telle différence existait dans ce dialecte, examiner F2 en plus permettrait d'examiner comment de telles différences dans la qualité pourraient être impliquées dans la magnification du contraste de la quantité causée par l'accent ou par le focus. En l'absence de larges différences quantitatives, il est difficile de prévoir les effets de l'accent ou les différents types de focus sur F2. Par exemple, les effets de l'undershoot pourraient produire un F2 plus bas à cause de la consonne labiale prévocale, ou ils pourraient aussi générer un F2 plus élevé à

cause de la consonne coronale post-vocalique.

Nous exposons dans ce paragraphe une liste de certains résultats des calculs statistiques à propos de l'accent et du focus lexical de l'étude de de Jong & Zawaydeh (2002).

Concernant l'accent, les auteurs notent que les syllabes sous l'accent sont plus longues de 20 ms par rapport à leurs correspondantes sans accent, une différence significative ($p < 0,01$). Quant au focus, curieusement, ce focus n'a pas eu d'effet majeur sur la durée; il n'y a pas eu non plus d'interaction entre accent et focus.

Quantité: de Jong & Zawaydeh (2002) obtiennent pour la quantité vocalique classiquement un effet significatif sur les durées des voyelles ($p < 0,01$): avec pour la longue 132 ms vs. la brève 58 ms.

Voisement: les auteurs attestent que l'effet de l'état du voisement de la consonne post-vocalique sur la durée de la voyelle n'est pas significatif; ils confirment ainsi les propositions de Flege & Port (1981) et Mitleb (1984a) concernant l'absence d'effet de l'état de voisement de la consonne post-vocalique sur la durée de la voyelle qui la précède en arabe.

Concernant les interactions:

L'interaction accent X quantité est significative à ($p < 0,05$). L'écart temporel entre la longue et la brève est plus important sous l'accent que sans accent.

L'interaction accent X voisement n'est pas significative, le voisement qui n'a pas eu d'effet sans accent n'a pas réussi à avoir de l'effet sous l'accent.

Focus lexical X quantité n'est pas une interaction significative: la différence de durées entre la longue et la brève n'est pas affectée par ce type de focus.

L'interaction Focus lexical X voisement n'est pas significative non plus: l'état de voisement (comme vu plus haut) n'a pas eu d'effet ni sous l'accent ni sans accent, ici l'état de voisement n'a pas eu d'effet sous le focus non plus.

L'interaction Focus X accent X voisement est significative ($P < 0,05$). Les autres interactions ne sont pas significatives, à part celle de quantité X voisement ($p < 0,05$).

Pour résumer: l'accent augmente la durée absolue des voyelles toutes longueurs confondues, mais aussi l'écart entre longue et brève alors qu'il n'affecte pas la différence de durée due au voisement. Le focus lexical n'a eu aucune incidence sur la durée en quantité, ni sur la durée absolue ni sur l'écart entre longue et brève, alors qu'il a eu un léger effet

significatif sur le voisement.

Résultats à propos du focus segmental:

Les analyses des deux corpus du focus segmental (focus sur la quantité et focus sur le voisement) donnent des résultats similaires à ceux qui sont observés avec l'accent et le focus lexical. Une exception est celle de l'effet significatif du voisement ($p < 0,01$). Sous le focus segmental, il y a eu des effets majeurs de l'accent sur la différence de durée des voyelles due à la quantité ($p < 0,05$) et sur la différence due au voisement ($p < 0,01$). La quantité et le voisement ont eu finalement tous deux des effets significatifs ($p < 0,01$) sur la durée des voyelles.

Il y a eu une différence entre les résultats du focus sur la quantité et les résultats du focus sur le voisement; le type de focus avait un effet significatif dans le corpus du voisement mais pas dans le corpus de la quantité. L'interaction accent X focus n'était pas significative ni avec la quantité ni avec le voisement. On note une hausse moyenne de 13 ms environ avec le focus segmental sur le voisement, mais pas de hausse qui accompagne le focus segmental sur la quantité.

Interaction du focus segmental avec ce qui est sous focus: les interactions sont similaires à ce qui était observé avec le focus lexical. Sous l'accent, les voyelles sont plus longues que leurs correspondantes (sans accent) de 10% environ, les longues sont 120% plus longues que les brèves.

L'accent tend à allonger les brèves alors que le focus segmental tend à les raccourcir.

L'interaction accent X focus segmental X quantité est non significative.

Les syllabes qui sont sous l'accent sont plus longues, mais pas d'interaction significative entre l'accent et le voisement. Pas d'interaction significative non plus entre le focus segmental et le voisement. L'interaction des trois facteurs: focus segmental X accent X voisement n'est pas significative non plus.

Pour résumer: le focus segmental produit des résultats similaires à ceux de l'accent, hausse de la durée mais pas de hausse de l'écart temporel dû à la quantité. Il n'y a pas de différence due au voisement de la consonne post-vocalique. Le focus segmental diffère de l'accent dans son impact sur les voyelles brèves, en effet, alors que l'accent augmente la durée des voyelles brèves, l'application du focus segmental sur l'opposition de la quantité a montré une tendance des voyelles brèves à être réduites.

Nous rapportons en ce qui suit certains des résultats concernant les tests statistiques relatifs à la réaction de la qualité vocalique aux effets prosodiques d'accent et de focus lexical et segmental:

En examinant le F1, les auteurs trouvent un large effet de l'accent ($p < 0,05$) et du focus ($p < 0,05$). Comme ils l'ont prévu, le F1 pour les voyelles sous l'accent et sous le focus est plus élevé et se déplace par conséquent vers la périphérie de l'espace vocalique. Les deux effets sont de l'ordre de 50 Hz. Les voyelles longues avaient à la base un F1 plus élevé que celui de la brève ($p < 0,01$). L'effet de la quantité sur F1 était clairement plus important que les effets des variations prosodiques, la différence étant de 160 Hz. Ils ont trouvé aussi un effet du voisement ($p < 0,05$); c'est l'opposé de ce qu'on aurait pu prévoir en se basant sur le mécanisme de l'undershoot; en effet, les voyelles suivies de consonnes sourdes avaient en moyenne un F1 plus élevé que celles qui étaient suivies par des consonnes voisées et qui étaient plus longues. Alors, l'allongement dû à la quantité (accompagné par une hausse de F1) semble être plus proche de celui qui est produit par l'accent que celui qui est produit par le voisement (accompagné par une baisse de F1). Il y a eu une interaction significative, accent X quantité ($p < 0,05$) et focus X quantité ($p < 0,05$). Aucune des autres interactions n'était significative.

L'analyse de F2 n'a révélé aucun effet significatif (quantité, voisement, accent et focus sont tous non significatifs). Pourtant l'analyse de l'effet fixé, qui tient compte des différences proportionnelles des valeurs des formants pour les différents locuteurs, a montré un effet significatif de la quantité, du voisement et de l'accent, tel que les effets qui causaient une hausse de durée produisaient en même temps une légère hausse de F2. Ainsi les voyelles longues, celles qui sont suivies par des consonnes voisées et celles qui sont frappées par l'accent avaient en moyenne un F2 plus élevé de 15 Hz. Il y a eu une interaction significative quantité X accent X focus. Cette interaction semble due à la différence de 40 Hz entre les voyelles brèves sous l'accent et les voyelles brèves sans accent dans la condition où elles étaient sous le focus. Aucune autre interaction ne s'est révélée significative. Les auteurs concluent, concernant F2, qu'il n'y a pas eu de distinction systématique avant-arrière entre les voyelles longues et brèves, et arrêtent la recherche sur ce point (F2) à ce stade.

L'analyse des formants dans les deux corpus du focus segmental a produit des résultats pour F1 semblables à ceux qui sont trouvés avec le focus lexical. Il y a eu un effet variable de l'accent (corpus de la quantité $p < 0,001$; corpus du voisement $0,1 < p < 0,05$), un effet de la quantité ($p < 0,01$) et pas d'effet pour le voisement. Ici, il n'y a pas eu d'effet du focus ni pour

la quantité ni pour le voisement. Les seules interactions significatives étaient quantité X focus ($p < 0,05$). voisement X accent X focus ($p < 0,05$).

Les résultats à propos de F2 révèlent 2 interactions significatives de l'accent dans le corpus quantité-focus, accent X quantité ($p < 0,01$); accent X focus ($p < 0,05$), et une interaction marginale, accent X voisement ($0,1 < p < 0,05$) dans le corpus de voisement. Chacun de ces effets semble être dû à une réduction de 30 Hz en moyenne de la forme lexicale [baada]. Aucune autre interaction n'était significative. Comme ces effets sont faibles et ne reflètent aucune tendance systématique concernant la distinction de la quantité, les auteurs arrêtent ici aussi leur examen.

Afin de comprendre les effets de la qualité et comment ces effets sont liés aux effets temporels, nous avons procédé à une lecture graphique de F1 en fonction de la durée dans l'étude de de Jong & Zawaydeh (2002). Cette lecture nous révèle que, concernant la relation entre accent et quantité, l'effet de l'accent sur les voyelles longues est semblable en général à celui de la quantité. Alors que l'accent n'agit pas comme la quantité sur les voyelles brèves, dans la mesure où l'accent affecte F1 même si les différences temporelles ne sont pas importantes. Ainsi, pour les voyelles longues, l'effet de l'accent est similaire à celui de la quantité, légèrement plus important. Pour les voyelles brèves, où les effets de l'undershoot sont probablement plus évidents, l'accent est manifesté par une hausse de F1.

Les effets du focus lexical sont différents des effets de l'accent et de la quantité dans la mesure où ils affectent principalement le F1, alors qu'ils laissent la durée essentiellement non affectée. Lorsque le focus segmental est appliqué sur la quantité, il n'augmente pas le F1, il montre plutôt un allongement des voyelles longues et un léger raccourcissement des voyelles brèves.

L'accent affecte la durée et le F1. Le focus segmental appliqué sur le voisement se montre différent du focus segmental appliqué sur la quantité, puisque ce dernier crée un allongement général et une baisse de F1. Le focus segmental sur le voisement produit un allongement général des voyelles.

Les auteurs concluent sur ces résultats en soulignant le fait qu'ils trouvent une certaine régularité dans la manière avec laquelle les locuteurs emploient le contraste de la quantité: une hausse de la durée corrélée à une hausse de F1. Ils trouvent aussi une certaine régularité dans la manière avec laquelle les locuteurs mettent l'accent en pratique (chez 3 locuteurs sur 4), il est corrélé à une hausse de F1 chez les longues et les brèves. Les focus lexical et

segmental sont employés cependant de différentes manières. Pour le focus lexical, 2 des 4 locuteurs augmentent F1 pour les voyelles longues et brèves, un locuteur baisse la durée des voyelles longues en baissant F1. Pour le focus segmental, deux locuteurs ont montré une hausse de durée et de F1, un locuteur montre seulement une baisse au niveau de F1, et le 4^e ne présente aucun effet.

Discussion

Concernant la question générale: comment l'accent interagit-il avec les effets temporels dus à la quantité et au voisement, les résultats indiquent la présence d'une différence entre la quantité et le voisement. L'accent augmente l'allongement de la durée de la voyelle dû à la quantité, mais pas l'allongement de la durée dû au voisement. Les résultats pour le voisement diffèrent de ce qui était trouvé pour l'anglais, la proposition de Mitleb (1984a) semble ainsi vraisemblable. Cette différence est due au fait que l'anglais avait spécifié la durée de la voyelle comme un indice acoustique sur la distinction voisé/non voisé, alors que cette spécification est absente en arabe. D'où la hausse de l'attention due au fait que l'accent augmente les différences temporelles alors que ces différences sont spécifiées comme une partie d'un contraste paradigmatique.

Une complication, qui doit être considérée par plus de recherches, concerne les effets de la quantité et du voisement par rapport à la structure de la syllabe. Alors que les effets de la durée de la quantité indiquent un contraste à l'intérieur de la voyelle, et par conséquent à l'intérieur de la syllabe, les effets de la durée du voisement indiquent un contraste qui réside, au moins partiellement, dans la syllabe qui suit. Des études précédentes, notamment Chen (1970), sur l'effet du voisement sur la durée de la voyelle en anglais ont montré que le voisement peut agir au-delà de la frontière de la syllabe, les résultats de cette étude (de Jong & Zawaydeh, 2002) le montrent aussi. Cette étude trouve qu'il y a un effet, lequel n'interagit pas avec l'accent. Si, conformément aux suppositions traditionnelles, l'accent opère sur toute la syllabe, alors soit la consonne, soit la voyelle qui la précèdent sont sous l'accent dans le corpus de cette étude, mais pas les deux.

Ces résultats en ce qui concerne l'accent et le focus sont équivoques. Même si le focus lexical n'augmente pas l'effet du voisement pour les voyelles sans accent, il n'est pas sûr que le focus segmental ne le fait pas.

La plus grande conclusion que les auteurs peuvent tirer de cette étude est que l'effet du voisement sur la durée de la voyelle est réalisé comme un effet secondaire en l'absence de

l'accent. Pourtant, il n'est pas sûr que les locuteurs comptent sur de telles différences quand ils essaient de montrer un contraste de voisement à leurs auditeurs. Cela diffère des effets de la quantité que les locuteurs essaient de magnifier quand ils veulent communiquer un contraste de quantité à l'auditeur.

Une question peut se poser alors pour savoir si ces deux effets temporels, quantité et voisement, sont en compétition. Comme l'arabe a un contraste vocalique très robuste et systématique qui dépend de la durée de la voyelle, les différences temporelles dues au voisement pourraient être supprimées afin de ne pas compromettre le contraste de la quantité. Des arguments similaires pour le maintien de la quantité comme un facteur dans le développement de l'accent en arabe étaient avancés par Ahn (2000). Cela devrait être différent du cas de l'anglais qui ne présente pas d'opposition de quantité de manière systématique. Certes, les dialectes de l'anglais utilisent la durée de la voyelle comme un indicateur de l'identité de la voyelle, plusieurs voyelles se contrastant typiquement par la durée; dans certains dialectes, il semble qu'il y a une opposition minimale pour les voyelles avant, non hautes, comme dans "bed" et "bad". Munro (1993) a montré que les locuteurs arabophones sur-articulaient les différences temporelles entre les voyelles de l'anglais quand ils apprenaient l'anglais, cela laisse penser que le contraste de quantité en arabe dépend plus de la différence temporelle que les voyelles de l'anglais, même celles qui sont contrastées surtout par leurs durées, n'en dépendent.

Il est nécessaire d'effectuer plus de travail sur d'autres langues pour évaluer mieux cette proposition. Il y a deux types d'évidence qui laissent penser que cette proposition ne peut pas être soutenue. En premier, des travaux précédents sur les effets du voisement (Chen, 1970; Port *et al.*, 1980) n'indiquent pas que la taille des effets du voisement est clairement corrélée à l'absence systématique des effets du contraste de la quantité. Port *et al.* (1980), par exemple, trouvent que le japonais rejoint l'anglais sur le fait que tous les deux présentent de grands effets pour le voisement. Ce n'est pas le cas de l'arabe, et cela malgré le fait que le japonais et l'arabe tous deux aient des contrastes de quantités systématiques. Deuxièmement, cette étude trouve que les dynamiques du voisement sont différentes de celles de la quantité, comme de précédentes études (Summers, 1987; de Jong, 1991) ont trouvé une différence entre les dynamiques du voisement et de l'accent. Finalement, en comparant l'accent et les deux types de focus, les auteurs trouvent quelques similarités. Premièrement, ils trouvent une stratégie générale de hausse de la durée des voyelles sous l'accent et des voyelles sous le focus segmental par rapport au voisement. En outre, accent et focus lexical tendent tous deux à

augmenter les valeurs des F1 des voyelles basses examinées dans cette étude (de Jong & Zawaydeh 2002). Finalement, ils trouvent une similarité entre l'accent et le focus segmental: tous les deux interagissent avec les effets de la quantité et du voisement, les deux augmentant la taille de l'effet de la quantité mais pas la taille de l'effet du voisement. Cette étude trouve aussi des preuves de quelques différences nettes entre l'accent, le focus lexical et le focus segmental, qui ne peuvent pas être expliquées comme des différences de domaine d'application. La première: le focus lexical dans cette étude n'allonge pas les voyelles, son effet sur F1 diffère de l'effet de l'accent; alors que l'effet de l'accent sur F1 peut être considéré comme une réduction de l'undershoot, l'effet du focus lexical est une hausse clairement ciblée (contrôlée) et indépendante de l'allongement temporel. La seconde différence est comment accent et focus segmental interagissent avec la quantité. Alors que l'effet de l'accent est d'allonger les voyelles brèves, le focus segmental sur la quantité tend à avoir un effet de raccourcissement des voyelles brèves. Ces différences rappellent l'effet de l'accent sur les voyelles hautes (discuté en de Jong, 1995). Harrington *et al.* (2000) conçoivent un conflit entre deux mécanismes possibles d'hyperarticulation qui peuvent s'appliquer aux voyelles hautes. Le premier augmenterait le degré d'aperture pour rendre la voyelle plus sonorisante (vocalique). Le second augmenterait le degré de constriction palatale afin de rendre le timbre de la voyelle plus contrastif. Harrington *et al.* (2000) trouvent que les sujets présentent des résultats quelque peu différents, certains ont des ouvertures plus importantes et certains ont des constriction plus importantes. Les résultats de de Jong & Zawaydeh (2002) à propos du focus lexical semblent correspondre au premier mécanisme d'Harrington *et al.* (1996), indiquant une stratégie générale de la hausse du niveau général de l'intensité du son produit. Les résultats en ce qui concerne l'accent semblent combiner les deux types de mécanismes, une hausse générale de la durée et de F1, conjuguée avec une importante hausse de la durée des voyelles longues. Le cas des voyelles brèves de l'arabe est quelque part semblable au cas des voyelles hautes dans le sens où l'accent peut allonger une voyelle pour la rendre plus sonorisante et par conséquent plus claire en ce qui concerne le contraste vocalique, mais cet allongement se fait au prix du contraste (très robuste) de la quantité. Selon les auteurs de cette étude, quand l'attention du locuteur est focalisée loin du contraste vocalique et sur le contraste de quantité, les locuteurs ont tendance à raccourcir ces voyelles au lieu de les allonger.

Le point supplémentaire qui doit être fait dans cette comparaison entre les résultats de l'accent et du focus, c'est qu'il y a peu d'évidence d'une interaction en domaines d'application. Focus lexical et focus segmental ont tous deux les mêmes effets, indépendamment du fait que

le matériel affecté soit ou non sous l'accent. Pour le focus segmental, l'absence de cette relation n'est pas surprenante, puisqu'il a un domaine d'application plus petit que celui de l'accent.

Une différence finale entre l'accent et le focus est que les sujets sont souvent très différents dans leur usage des deux types de focus (segmental et lexical), alors qu'ils sont très similaires dans leur manière de mettre en pratique l'accent. Il semble que l'accent est plus conventionnel que le focus. Les locuteurs sont en quelque sorte avertis du fait que des contrastes particuliers sont exprimés et sont mis spécialement en relief différemment selon le locuteur.

Les effets prosodiques se sont finalement révélés des moyens efficaces de perturbation pour permettre la caractérisation de certains phénomènes phonétiques d'un système vocalique donné. La différence entre mécanismes par exemple: l'accent n'agissant pas de la même manière sur l'allongement de la durée de la voyelle dû à la quantité, et sur l'allongement dû au voisement de la consonne post-vocalique, ce qui met en évidence la différence entre ces deux mécanismes d'allongement.

Ces perturbations prosodiques ont permis de montrer que certains impacts peuvent devenir non significatifs sous certaines contraintes alors qu'ils étaient significatifs sans contraintes (L'effet significatif du voisement sur F1 dans des conditions normales devient non significatif sous le focus segmental). Mais elles peuvent montrer aussi que certains impacts peuvent devenir significatifs sous certaines contraintes alors qu'ils ne l'étaient pas sans contraintes. En effet, le rôle du voisement qui n'était pas significatif sans effet prosodique particulier, ou sous l'accent, devient significatif sous l'effet du focus segmental. Ces perturbations permettent aussi de distinguer les procédés prosodiques entre eux: l'accent agit sur la durée des voyelles, pas le focus lexical: par conséquent les deux procédés sont différents.

Quant à notre étude maintenant, pour tester la robustesse des écarts quantitatifs et qualitatifs entre voyelles longues et voyelles brève en arabe dialectal de Mayadin, nous procéderons à deux types de contraintes indépendantes et compatibles qui se sont révélées efficaces dans la perturbation des systèmes vocaliques et ainsi l'extraction de leurs propriétés fondamentales.

La première contrainte est appliquée par une variation rythmique, le débit de la parole, et nommée *perturbation prioritaire*, utilisée par Abry *et al.* (1990). Partant des résultats de

ces derniers, nous nous attendons à ce qu'un débit rapide produise deux types d'effets. Le premier est temporel et consiste à réduire les durées des voyelles. Le deuxième est spectral, conséquence du premier, et consiste à accentuer l'effet de l'undershoot (la réduction vocalique) en déplaçant les voyelles vers le centre de l'espace vocalique.

La deuxième contrainte est une variation intonative, où une interrogation totale, traduite par une intonation montante, frappe la voyelle cible. Première conséquence logique: c'est la hausse de la fréquence fondamentale de la voyelle ciblée. Nous nous attendons à ce que l'interrogation produise des effets –temporellement et spectralement - opposés aux effets de la parole rapide, c'est-à-dire une durée plus importante et par conséquent – si l'undershoot est corrélé à une baisse de la durée de la voyelle - une réduction de l'undershoot, c'est-à-dire un déplacement des voyelles vers la périphérie de l'espace vocalique. En effet, de Jong & Zawaydeh (2002) ont étudié les réactions des voyelles basses de l'arabe dialectal d'Amman (Jordanie) à l'accent et au focus, ils ont pu démontrer que les durées des voyelles, brèves et longues, augmentaient significativement quand elles sont frappées par l'accent. Ils ont pu démontrer aussi que le F1 avait des valeurs significativement plus élevées (déplacement vers la périphérie de l'espace vocalique) sous l'accent et sous le focus. Ces deux procédés phonétiques ont des effets largement inférieurs à ceux de l'interrogation par l'intonation, qui est connue par ses effets extrêmes sur la durée, la F0 et l'intensité des voyelles comparée aux autres procédés prosodiques.

Au final, les écarts qui résistent malgré ces deux effets perturbateurs de la communication orale naturelle seraient des écarts à valeurs phonologiques dans le système, et ainsi linguistiquement spécifiés.

III.2. Notre dialecte de Mayadin

Le dialecte de la ville de Mayadin partage, comme c'est le cas de la plupart des dialectes arabes du Moyen-Orient, avec l'arabe classique beaucoup de ses caractéristiques, aussi bien au niveau lexical, syntaxique que phonétique. Son système phonétique est caractérisé par un consonantisme riche – il compte 28 consonnes – et un vocalisme relativement pauvre avec *a priori* 4 timbres. La pauvreté du système vocalique serait peut-être à l'origine des réalisations de ces trois voyelles sous des durées différentes pour enrichir son inventaire de voyelles.

Nous dressons en ce qui suit une description brève de son système phonologique:

Les consonnes. Le système phonologique des consonnes, tableau 16, est le même que celui de l'arabe classique. Il a ajouté deux consonnes: 1) la mi-occlusive post-alvéolaire sourde [tʃ]. 2) l'occlusive vélaire sonore [g]. Il utilise une grande étendue du conduit vocal, du point le plus en arrière, la glotte, jusqu'au point le plus en avant, les lèvres.

			Régimes phonatoires	LIEU D'ARTICULATION											
				bilabial	labio-dental	dental	alvéolaire	post-alvéolaire	palatales	vélaires	uvulaire	pharyngal	glotale		
MODE D'ARTICULATION (Régimes aéro acoustiques)	Occlusif	oral	sourd				t			k	q		ʔ		
			sonore	b			d			g					
		pharyngalisé	sonore				tʕ						ʕ		
			sourd				dʕ								
	nasal		sonore	m			n								
	mi-occlusif		sourd					tʃ							
			sonore					dʒ							
	constrictif		médian	sourd		f	θ	s	ʃ		x		ħ	h	
				sonore			ð	z			ɣ				
			à battements	sourd											
				sonore				r							
			pharyngalisé	sourd					sʕ						
				sonore				ðʕ							
	latéral		sonore				l								

Tableau 16. Système phonologique des consonnes du dialecte arabe de la ville de Mayadin (Syrie).

Les voyelles

Quant aux voyelles, le système phonétique de ce dialecte dispose de 4 timbres stables [i, a, u, o] alors que le système phonologique "a priori" dispose de 7 voyelles, tableau 17.

/i:/	palatal	antérieur	non-arrondi	fermé
/a:/	central	antérieur	non-arrondi	ouvert
/u:/	vélaire	postérieur	arrondi	fermé
/i/	palatal	antérieur	non-arrondi	fermé
/a/	central	antérieur	non-arrondi	ouvert
/u/	vélaire	postérieur	arrondi	fermé
/o:/	vélaire	postérieur	arrondi	semi-fermé

Tableau 17. Système phonologique des voyelles du dialecte arabe de la ville de Mayadin (Syrie).

Les voyelles longue et brève de la qualité [a], connaissent chacune une variante combinatoire réalisée plus en arrière [ɒ]. La voyelle mi-ouverte [o], présente dans le dialecte, est dérivée historiquement de la diphtongue [au] de l'arabe standard, la qualité stable a définitivement remplacé la diphtongue dans des syllabes à coda pleine. Cette tendance à remplacer les diphtongues [au] et [æi] par des qualités stables intermédiaires (du point de vue de l'aperture) dans des syllabes à coda pleine, [o] et [e] respectivement, était observée dans d'autres dialectes arabes (de Jong & Zawaydeh, 2002 pour l'arabe jordanien d'Amman). Le processus qui consiste à remplacer la diphtongue par la qualité stable intermédiaire n'est pas encore accompli concernant la diphtongue [æi] dans le dialecte de Mayadin quand la syllabe est en finale de mot. En effet, cette diphtongue est devenue [ie] dans un processus qui la rapproche de la forme statique de qualité vocalique [e], cette forme statique a définitivement remplacé la diphtongue quand la syllabe n'est pas en finale de mot. Bien qu'elle connaisse une variante (sociophonétique) libre, d'une qualité stable [e], qui, dans beaucoup de dialectes arabes a remplacé la diphtongue de manière définitive, cette diphtongue, [ie], demeure la forme standard dans ce dialecte quand la syllabe est en finale de mot. Nos mesures confirment ce que nous venons d'affirmer à propos de la diphtongue [ie], figure 42, une hausse pour F1 et une baisse pour F2.

Le système phonologique des voyelles de l'arabe standard ne dispose pas de la qualité stable [o], nous la décrivons ici de façon explicite en termes des valeurs de ses trois premiers formants. Nous l'avons enregistrée auprès des locuteurs natifs du dialecte, le résultat confirme l'existence d'un [o] caractéristique dont le sonagramme est présenté dans la figure 43.

Cette voyelle était émise dans le contexte d'une consonne pharyngale constrictive sourde [ħ] en position prévocale, et d'une consonne alvéolaire à plusieurs battements [r] en position post-vocalique. La qualité est finalement stable. La voyelle [o] ne présente pas une opposition de quantité avec une autre voyelle qui soit plus longue ou plus brève.

Quant aux voyelles brèves, elles sont 3 /i/, /a/ et /u/. Toutes les voyelles de qualités stables, à part la voyelle /o/ qui ne présente pas de contraste quantitatif, seront décrites en détail plus loin.

Diphthongaison de /i/ à /e/.

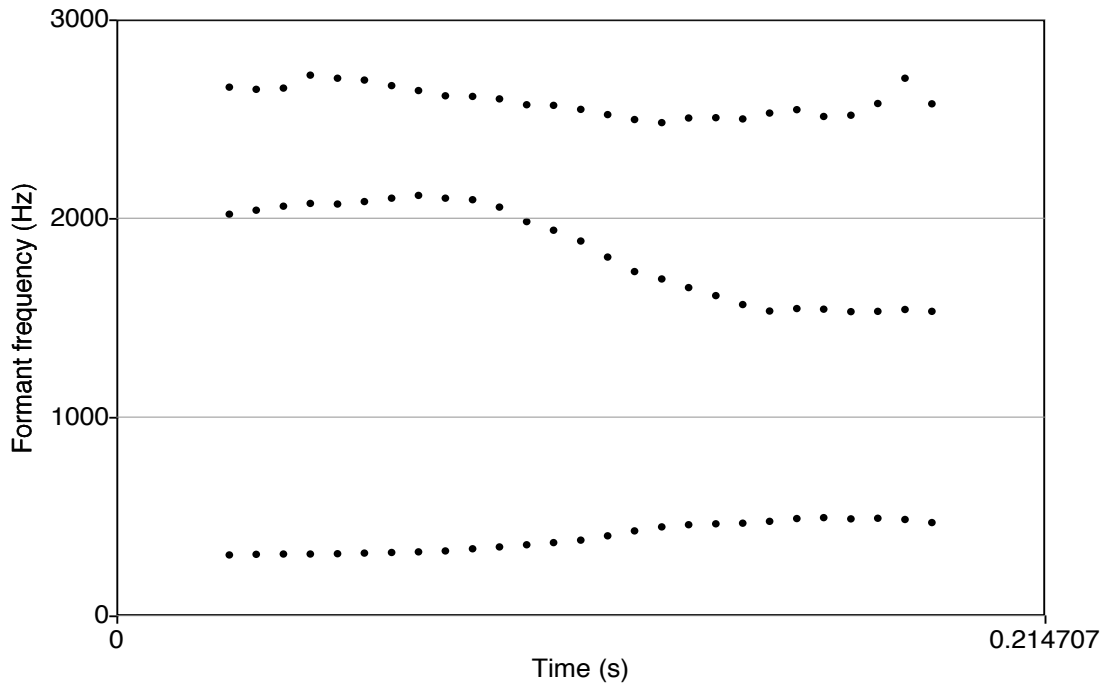


Figure 42. La diphthongue [ie] du mot [ziet] "huile" prononcé par un locuteur natif du dialecte arabe de Mayadin. L'équivalent en arabe standard est [zæit].

Voyelle longue de timbre /o/.

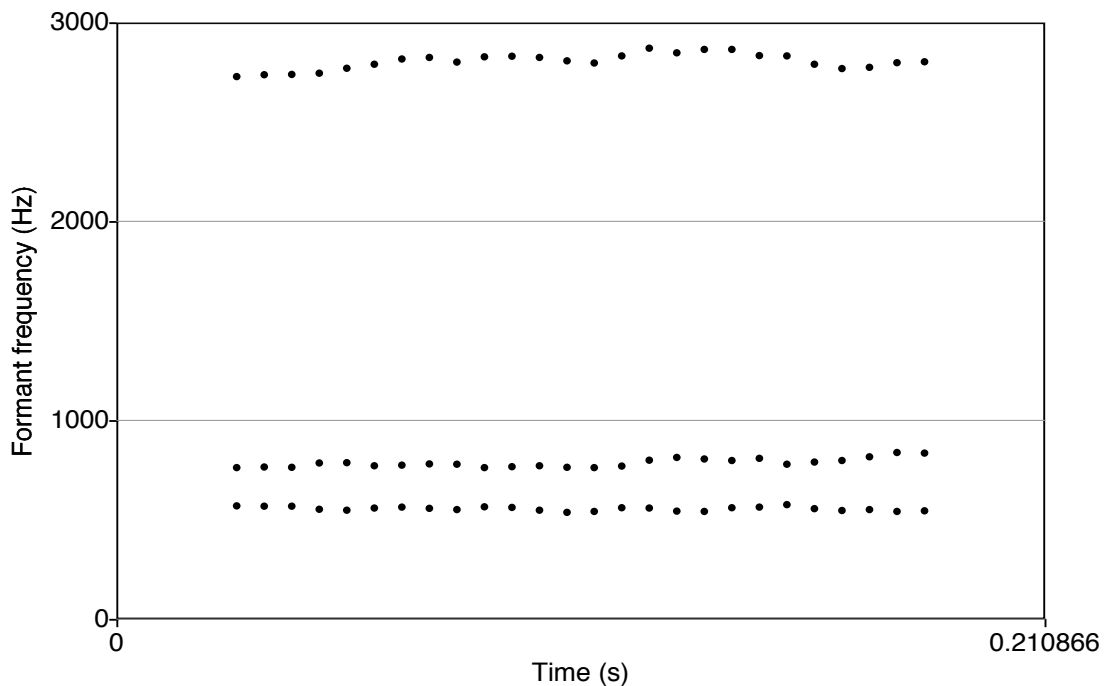


Figure 43. La voyelle /o/ dans le mot [ħor] "peuplier" prononcé par un locuteur natif du dialecte arabe de Mayadin. L'équivalent en arabe standard est [ħaur].

III.3. Partie expérimentale

Nous voudrions rappeler ici que les 2 conditions expérimentales qui seront appliquées à nos voyelles sont des effets frontalement opposés, et au niveau temporel et au niveau spectral, comme ce sont les deux extrêmes de la perturbation naturelle: un débit rapide réduisant et une intonation interrogative magnifiante.

Nos objectifs sont alors:

III.3.1. Caractérisation de la durée

Il nous faut pour cela:

(1) Trouver les durées des trois voyelles cardinales longues et des trois voyelles cardinales brèves, soit les 6 voyelles suivantes [i:, i, a:, a, u:, u] dans des conditions normales de l'émission de la parole (dites *a priori* "sans contraintes"): le débit de l'élocution est normal et le schéma mélodique est déclaratif (prosodie neutre). Cette condition constitue, pour nous, la base de la performance, parce que c'est par rapport à cette condition-là que nous évaluerons les effets de la variation du débit et de l'intonation.

(2) Mesurer l'écart entre les durées des deux termes de chacune des 3 paires de voyelles [i] vs. [i:], [a] vs. [a:], et [u] vs. [u:] formée d'une longue et d'une brève correspondante sous la condition de base, et chercher à savoir si cet écart est significatif: (i) au niveau statistique. (ii) au niveau perceptif, à la lumière de nos résultats de recherche sur le seuil différentiel de la perception de la durée vocalique, des résultats que nous avons obtenus en I.2.

(3) Refaire les mêmes tâches (1) et (2) pour les mêmes voyelles émises (i) avec un débit rapide et un schéma mélodique déclaratif. (ii) avec un débit rapide et un schéma mélodique interrogatif et (iii) avec un débit normal et un schéma mélodique interrogatif. Cela nous donne 4 conditions d'expérimentation (2 débits d'élocution X 2 schémas mélodiques).

Nous comparons ensuite les écarts obtenus sous la condition de base avec les écarts obtenus sous les trois autres conditions où nous changeons soit le débit soit le schéma mélodique, pour voir s'ils sont maintenus d'une condition à l'autre. Cette méthode, appelée méthode des "perturbations prioritaires" utilisées dans Abry *et al.* (1990), vise, comme la perturbation par la prosodie (de Jong & Zawaydeh, 2002) à tester la robustesse des écarts temporels entre une voyelle brève et sa correspondante longue. En effet, les oppositions linguistiquement spécifiées se maintiennent puisque leur présence est indispensable à

l'intelligibilité du message linguistique, les traits qui disparaissent relèvent alors de la stylistique et de la socio phonétique. Une telle robustesse, si elle est attestée, permet de savoir que l'écart soumis à la perturbation de l'élocution rapide de la parole ou à l'effet prosodique de l'interrogation, est contrôlé. Donc les oppositions sous différentes conditions d'expérimentation servent, en dernier lieu, à savoir quel écart est contrôlé et quel autre ne l'est pas.

III.3.2. Caractérisation de la qualité

En ce lieu aussi, nous devons:

(1) Trouver les valeurs des deux premiers formants des trois voyelles cardinales longues et des trois voyelles cardinales brèves dans des conditions normales de l'émission de la parole. Le débit de l'élocution est normal et le schéma mélodique est déclaratif, cette condition est la condition de base.

(2) Mesurer la distance entre les deux voyelles de la paire formée d'une longue et d'une brève correspondante, sous la condition de base. Chercher à savoir si cette distance est significative: (i) au niveau statistique. (ii) au niveau perceptif, à la lumière de nos résultats de recherche sur le seuil de la qualité vocalique (des résultats que nous avons obtenus en I.1).

(3) Refaire les mêmes tâches en (1) et (2) pour les mêmes voyelles émises (i) avec un débit rapide et un schéma mélodique déclaratif. (ii) avec un débit rapide et un schéma mélodique interrogatif et (iii) avec un débit normal et un schéma mélodique interrogatif. Cela nous donne 4 conditions d'expérimentation (2 débits d'élocution X 2 schémas mélodiques).

Comparer ensuite les distances obtenues sous la condition de base avec les distances obtenues sous les trois autres conditions où nous changions soit le débit soit le schéma mélodique, pour voir s'ils sont maintenus d'une condition à l'autre, "perturbations prioritaires".

III.3.3. Méthode

III.3.3.1. Stimuli

Les voyelles enregistrées, objet de l'expérience, seront, comme nous l'avons mentionné en III.3.1., [a:, i:, et u:] d'un côté et [a, i et u] de l'autre. Elles seront introduites dans 6 mots monosyllabiques où chaque mot sera inséré dans une phrase porteuse.

D'après notre expérience en matière de segmentation, nous proposons que le choix du contexte immédiat du phonème que nous cherchons à étudier doit se faire selon deux principes:

1) Ne pas choisir des phonèmes dont nous ignorons l'effet qu'ils peuvent exercer sur les phonèmes voisins. Nous avons notamment évité tout voisinage pharyngal pour les voyelles qui ont fait l'objet de notre expérience, ceci est en raison de l'important effet que cette catégorie de consonnes exerce sur la structure acoustique de la voyelle adjacente (Zawaydeh, 1999). En effet, Barkat-Defradas *et al.* (2003) ont étudié l'effet de la consonne pharyngale sur la voyelle adjacente en arabe jordanien; les voyelles de leur corpus étaient précédées une fois par [d], et une fois par [d^ʕ] (la version pharyngalisée du [d]), en production et en perception. Leurs résultats ont montré que la pharyngalisation a eu de l'effet sur la voyelle adjacente. Les analyses statistiques ont montré que la pharyngalisation affectait significativement les valeurs de F1 et de F2. Les auteurs notent que les voyelles en contexte pharyngal sont plus postérieures (baisse de la valeur de F2). En perception, les voyelles pharyngalisées ont subi un changement au niveau de leur F1 seulement, mais aucun effet sur le F2. Nous avons finalement exclu de ce contexte les semi-voyelles, toutes les consonnes (occlusives ou constrictives) articulées dans la cavité arrière¹⁴: les uvulaires [q], les pharyngales [ħ] et [ʕ], les pharyngalisées [t^ʕ], [d^ʕ], [s^ʕ] et [d^ʕ] et les glottales [ʔ].

2) Le contexte consonantique adjacent pour nos mots était finalement formé en position prévoicative de la consonne alvéolaire constrictive voisée [z] et en position post-vocalique de l'alvéolaire vibrante (à plus d'un battement) voisée [r]. Avec une consonne voisée en position prévoicative et une consonne vocalique en position post-vocalique, ce contexte n'est pas le contexte idéal, mais il répond à certaines des contraintes auxquelles nous avons voulu nous conformer lors de l'adoption du corpus, la disponibilité du lexique du dialecte n'a pas permis un choix meilleur.

Le [z] est une consonne alvéolaire consonantique voisée et le [r] est une consonne alvéolaire vocalique, donc fortement voisée. Cette dernière catégorie de consonnes présente beaucoup de ressemblance avec les voyelles, c'est dû au fait qu'elle a une structure formantique comme les voyelles. Les consonnes alvéolaires influencent beaucoup la durée et la direction des transitions formantiques, surtout dans la partie transitoire (Stevens & House, 1963). Quant à l'effet du voisement, au sein de la même syllabe, les voyelles qui sont suivies par des consonnes voisées sont plus longues que les voyelles suivies des consonnes sourdes (Peterson & Lehiste, 1960). de Jong & Zawaydeh (2002) affirment que l'état de voisement de la consonne post-vocalique n'a pas d'effet significatif sur la durée de la voyelle en arabe

¹⁴ Nous notons seulement les consonnes qui sont présentes dans le système phonologique du dialecte, et qui appartiennent à la catégorie des consonnes exclues.

dialectal d'Amman (Jordanie), confirmant ainsi les propositions de Flege & Port (1981) et Mitleb (1984a) à propos de cette question.

La disponibilité du lexique du dialecte a fait que notre contexte n'est pas le contexte idéal, car trouver trois paires minimales avec les trois voyelles cardinales avec un même contexte consonantique implique beaucoup de contraintes, peu faciles à satisfaire. C'est aussi parce que nous n'avions pas le droit d'introduire dans le corpus des logatomes, que nous aurions inventés pour servir de mots porteurs. Tous les mots devaient bien être des mots intégrés dans le lexique du dialecte. Ce contexte ne rend pas la délimitation des voyelles impossible mais plus difficile, car elle exige plus d'attention.

Les items de l'expérience

La voyelle [a] bref était portée par le mot [zar], le passé du verbe du sens "*boutonner*" avec la troisième personne du singulier, soit "*il a boutonné*". La voyelle [a] long était portée par le mot [za:r], le passé du verbe du sens "*visiter*" avec la troisième personne du singulier, soit "*il a visité*". La voyelle [i] bref était portée par le mot [zir], un substantif du sens "*bouton*". La voyelle [i] long était portée par le mot [zi:r], un nom propre¹⁵. La voyelle [u] bref était portée par le mot [zur], l'impératif du verbe du sens "*boutonner*" pour la deuxième personne du singulier, soit "*boutonne*". La voyelle [u] long était portée par le mot [zu:r], l'impératif du verbe du sens "*visiter*" pour la deuxième personne du singulier, soit "*visite*".

Ces six items ont été introduits dans une phrase porteuse formée de deux unités lexicales monosyllabiques de type CVC: la première unité était le mot [gal], le passé du verbe du sens "*dire*" avec la troisième personne du singulier, soit "*il a dit*"; la deuxième était l'un de nos six items qui alternaient aléatoirement de 1 à 6 derrière ce mot. Cet ordre était choisi de cette manière pour que les voyelles de l'expérience, placées ainsi dans la deuxième syllabe, reçoivent l'effet de l'intonation montante lors de l'émission avec un schéma interrogatif.

Le plan de notre expérience s'établit ainsi: les phrases porteuses sont émises selon deux schémas mélodiques 1) déclaratif. 2) interrogatif. Chaque schéma est émis selon deux débits d'élocution 1) normal. 2) rapide. L'association de ces deux schémas avec ces deux débits aboutira à quatre conditions d'expérimentation. Avec 6 (phrases) × 12 (répétitions) = 72

¹⁵ Il est connu la première fois pour le protagoniste d'un très vieux roman arabe du cinquième siècle. Ce roman raconte l'histoire d'une guerre qui a duré 40 ans où A-ZI:R était un grand chevalier, le roman est intitulé « a zi:r wa al bassouss ». Ce nom a perduré.

phrases par condition d'expérimentation et par locuteur.

III.3.3.2. Sujets

Les locuteurs sont 3, ils sont naïfs en ce qui concerne la phonétique. Ils n'étaient pas au courant de l'objectif de l'expérience. Ils sont tous les trois des hommes natifs du dialecte arabe de Mayadin. Le premier L_1 est âgé de 25 ans; L_2 de 37 et L_3 de 39 ans. Ils ne manifestent pas de troubles de langage ou de pathologie respiratoire. Ils parlent avec une fréquence fondamentale moyenne (en débit normal, schéma déclaratif toutes voyelles confondues) $L_1 = 156$; $L_2 = 168$ et $L_3 = 119$ Hz et une intensité de $L_1 = 75,7$; $L_2 = 77,38$ et $L_3 = 70,29$ dB.

Test des locuteurs, élocution et perception

Les locuteurs ont subi un test préalable à l'expérience. Le test était passé dans une chambre sourde où ils devaient identifier les six voyelles sous l'interrogation et sous le débit rapide. Nous avions pour but de:

- Nous assurer que les locuteurs réalisaient des paires minimales avec la durée, c'est-à-dire qu'ils accordaient une valeur phonologique à l'augmentation de la durée de la réalisation vocalique.
- Confirmer expérimentalement, avant toute mesure, la persistance de l'écart entre les longues et les brèves sous les différents effets perturbateurs. Ainsi, si les locuteurs réalisent un taux d'identification très élevé, cela vérifiera la robustesse de l'écart entre les longues et les brèves.

Dans ce test préliminaire une personne prononçait les paires minimales en ordre aléatoire et une autre notait ce qu'elle entendait. Chaque locuteur était amené à jouer les deux rôles. La personne qui prononçait ne voyait pas la personne qui notait, afin que la personne qui notait ne voie pas la personne qui prononçait. Ceci dans le but de neutraliser l'effet de la lecture labiale dans la reconnaissance des mots prononcés (Cathiard, 1994), et de s'assurer ainsi, que les locuteurs disposent d'une bonne audition, d'une bonne qualité de production acoustique, et faisant bien la différence entre les longues et les brèves.

Le locuteur entendait les 72 répétitions par condition expérimentale. Avec 4 conditions expérimentales, cela aboutit à $4 \times 72 = 288$ mots à identifier par locuteur. Ce que le premier locuteur a identifié était ce que le deuxième locuteur a prononcé et ce que le deuxième locuteur a identifié était ce que le premier locuteur a prononcé. Cela permettait de tester en même temps la capacité des locuteurs à la production et à la perception des contrastes

quantitatifs et qualitatifs.

Les résultats des (2*288) 576 mots à identifier, locuteur par locuteur, schéma par schéma, débit par débit et voyelle par voyelle sont disponibles en annexe n° VII.

Commentaire: À part le [a] bref chez le locuteur n° 1, qui était identifié à trois reprises en tant que [a:] long, tout était correctement identifié. Soit $(576-3)*100 / (576) = 99,5\%$ d'identification correcte. Le résultat de ce long test est finalement: Tous locuteurs, toutes voyelles, tous débits et tous schémas mélodiques confondus, le taux de réponses correctes est de **99,5%**, un résultat qui nous a permis d'adopter les sujets du test perceptif préliminaire comme des locuteurs pour notre expérience.

III.3.3.3. Procédure et enregistrement

Les six items ont été introduits dans la phrase porteuse et émis selon les 4 conditions expérimentales, chaque item étant prononcé 12 fois par condition expérimentale, ce qui fait que le locuteur prononçait 6 (phrases) × 12 (répétitions) = 72 phrases par condition expérimentale. Nous avons 3 locuteurs, ce qui fait que nous avons une expérience basée sur un corpus de $4 \times 72 \times 3 = 864$ phrases enregistrées pour l'analyse. Après examen des phrases enregistrées, des 12 répétitions nous avons choisi les 10 répétitions qui présentaient la meilleure qualité d'enregistrement, ce qui a fait que notre corpus s'est réduit à 720 phrases.

Nos locuteurs se sont entraînés avant l'enregistrement à prononcer le même corpus selon les 4 conditions expérimentales. Nous avons demandé aux locuteurs de simuler une situation de communication différente pour chacune des conditions. Nous leur laissions, juste avant l'enregistrement, le temps qu'ils voulaient pour imaginer et assimiler le rôle qu'ils devaient jouer. Voilà en détail ce qu'on leur demandait:

1) Schéma déclaratif avec débit normal: il était demandé au locuteur d'imaginer simplement qu'il était en train de raconter une histoire à un ou plusieurs copains, que la phrase porteuse se rencontrait dans cette histoire, et qu'il ne voulait avoir aucun effet avec sa voix sur l'interlocuteur; aucun procédé prosodique n'accompagnait donc la prononciation.

2) Schéma interrogatif avec débit normal: le locuteur imaginait qu'il posait une question à un interlocuteur, l'interrogation est supportée par l'accentuation de la voyelle noyau de la 2e syllabe, elle est manifestée par: (1) l'augmentation de la fréquence fondamentale (intonation montante); (2) l'augmentation de la durée de la voyelle, une augmentation de l'ordre de $\Delta T/T = 11\%$ de la durée de la voyelle avec un schéma déclaratif.

3) Schéma déclaratif avec débit rapide: le locuteur imaginait qu'il était dans un jeu-

concours où il devait répondre à une question posée à plusieurs personnes et que celle qui connaissait la réponse répondait en premier avant les autres.

4) Schéma interrogatif un débit rapide: le locuteur imaginait qu'il allait poser une question (avec le même procédé classique d'intonation montante) alors qu'il montait dans un bus qui partirait immédiatement: donc il avait très peu de temps pour poser sa question, ce qui l'obligeait à parler rapidement.

Toutes ces simulations ont été assez bien réussies, ce qui nous a permis d'obtenir, à la fin du dépouillement des données, des résultats assez similaires d'un locuteur à un autre.

Les locuteurs, assis devant le microphone, devaient prononcer les fiches que nous leur montrions. L'enregistrement était fait dans la chambre sourde de Gipsa-Lab (autrefois ICP) à Grenoble, avec un lecteur DAT Sony «DIGITAL AUDIO TAPE DECK-ZE700» entrée «ANALOG» avec un mode d'enregistrement STANDARD (44.100 KHz).

Nous avons utilisé un cadre de photographies qui permettait un défilement manuel, l'intérêt de l'utilisation de ce cadre est d'éviter la superposition du bruit dû au changement des fiches à lire lors de l'enregistrement sur le signal de parole, parce que nous avons remarqué dans d'autres expériences que ce bruit, quelque minime qu'il soit, est bien capté par les appareils de l'enregistrement. Les fiches défilaient dans un ordre aléatoire

Les intervalles temporels qui séparaient les apparitions des fiches sont quasi-égaux et d'une durée de 3 secondes. Les phrases selon le débit d'élocution, la voyelle prononcée, le locuteur et le schéma mélodique avaient une durée de 500 à 700 ms. Une fois l'enregistrement d'une condition commencé, le locuteur ne faisait une pause qu'après la 72e phrase.

III.3.3.4. Dépouillement des données

La délimitation des voyelles: Après l'enregistrement, la numérisation des phrases du corpus a été faite avec le logiciel Cool Edit 96. Il a fallu que nous délimitions sur le signal les segments vocaliques d'intérêt, afin que nous puissions ensuite faire des mesures manuelles et automatiques. La délimitation était surtout indispensable au traitement automatique des signaux par le logiciel Praat.

L'absence d'une définition stricte des règles qui régissent le procédé de la segmentation, a rendu préférable le fait que ce travail soit effectué entièrement par la même personne, pour éviter ainsi toute éventuelle incohérence qui pourrait être due aux différents jugements sur le début et la fin de la voyelle. C'était bien le cas pour la voyelle [i], par exemple, surtout la brève des deux, en contact avec la consonne alvéolaire voisée [z] en

position prévocale. En effet, le repérage du début de la voyelle est toujours problématique et discutable. Pour la segmentation, nous comptons essentiellement sur la forme des cycles vocaliques du signal de parole, cette forme étant corrélée à l'enveloppe du spectre acoustique du segment considéré.

Pour certaines voyelles, la délimitation par rapport à la consonne qui précède est plus facile que la délimitation par rapport à la consonne qui suit, et pour d'autres voyelles, c'est l'inverse.

Avec le schéma mélodique interrogatif, les voyelles sont plus facilement distinguables. C'est probablement dû à l'intensité qui est relativement élevée, par rapport au schéma mélodique déclaratif.

III.4. Résultats

Nous présentons ci-dessous les durées des 6 voyelles avec les valeurs de leurs formants. Nous nous contentons, dans un premier temps, d'exposer les valeurs que nous avons obtenues, et de les commenter simplement sans discuter de leur significativité. Nous présenterons dans les parties suivantes une analyse statistique et perceptive. Les durées, elles, seront présentées dans des tableaux montrant quels rapports les durées des voyelles entretiennent entre elles. Quant aux valeurs formantiques, leurs valeurs numériques seront présentées dans des tableaux; ces valeurs numériques seront traduites en des figures qui expriment F1 en fonction de F2 (espaces vocaliques).

Tableau des résultats. Le tableau 18 contient les valeurs moyennes de la totalité des résultats de l'expérience, pour les 6 voyelles avec les deux débits et les deux schémas mélodiques.

Voyelle	Durée (ms)	F0	F1		F2		F3	
		Hz	Hz	Bark	Hz	Bark	Hz	Bark
AND	143	146	548	5,33	1295	10,14	2567	14,67
AAND	237	137	636	6,04	1196	9,63	2486	14,46
IND	120	153	431	4,30	1549	11,30	2554	14,64
IIND	196	150	320	3,23	2030	13,11	2642	14,86
UND	125	148	436	4,35	1040	8,76	2509	14,52
UUND	209	151	360	3,63	921	8,04	2691	14,98
ANI	161	194	593	5,70	1313	10,23	2543	14,61
AANI	277	165	671	6,31	1245	9,88	2606	14,77
INI	127	196	430	4,29	1592	11,49	2502	14,50
IINI	211	183	328	3,31	2055	13,19	2619	14,80
UNI	130	218	438	4,37	1066	8,91	2484	14,46
UUNI	233	222	367	3,70	856	7,62	2485	14,46
ARD	119	161	565	5,47	1257	9,95	2469	14,42
AARD	201	164	626	5,96	1237	9,84	2467	14,41
IRD	95	175	422	4,22	1576	11,42	2560	14,65
IIRD	148	175	335	3,38	2001	13,01	2633	14,84
URD	104	172	430	4,29	1026	8,68	2521	14,55
UURD	168	175	356	3,59	999	8,52	2545	14,62
ARI	122	228	575	5,55	1356	10,43	2520	14,55
AARI	212	199	657	6,20	1330	10,31	2456	14,38
IRI	102	262	453	4,50	1603	11,53	2504	14,51
IIRI	161	232	354	3,57	2016	13,06	2588	14,73
URI	112	268	465	4,61	1136	9,31	2477	14,44
UURI	183	260	392	3,94	903	7,93	2566	14,67

Tableau 18. Résultats pour la durée et les valeurs formantiques des six voyelles [I, II, A, AA, U et UU] de l'arabe de Mayadin (Syrie) pour les 4 conditions expérimentales. Moyenne des 3 locuteurs masculins.

Légende: I = /i/ bref; II = /i/ long; A = /a/ bref; AA = /a/ long; U = /u/ bref; UU = /u/ long. N = le débit est normal; R = le débit est rapide; D = le schéma mélodique est déclaratif et I = le schéma mélodique est interrogatif.

La conversion de Hz en Bark était effectuée comme suivant: F en Bark = $7 \sinh(F \text{ en Hz}/650)$. Le formant F4 est fixé à 3560 Hz pour toutes les voyelles, soit 16,81 Bark.

III.4.1. Durée

Les durées seront représentées dans le tableau 19 avec le pourcentage que la brève représente par rapport à la longue, et avec $\Delta D/D$, où D est la durée de la plus longue des deux: $[(D_{\text{longue}} - D_{\text{brève}})/D_{\text{longue}}]$ Cela nous permet de calculer en % l'écart de durée entre longue et brève par rapport à la longue; nous cherchons à savoir si cet écart est maintenu d'une condition expérimentale à une autre.

Normal Déclaratif (ND)					Normal Interrogatif (NI)				
voyelle	Brève	Longue	%	$\Delta D/D$	voyelle	Brève	Longue	%	$\Delta D/D$ %
A	143	237	60,3	39,7	A	161	277	58,1	41,9
I	120	196	61,2	38,8	I	127	211	60,2	39,8
U	125	209	59,8	40,2	U	130	233	55,8	44,2
Rapide Déclaratif (RD)					Rapide Interrogatif (RI)				
voyelle	Brève	Longue	%	$\Delta D/D$	voyelle	Brève	Longue	%	$\Delta D/D$ %
A	119	201	59,2	40,8	A	122	212	57,6	42,5
I	95	148	64,19	35,8	I	102	232	44	56
U	104	168	61,91	38,1	U	112	183	61,2	38,8

Tableau 19. Variation des rapports de la durée entre les voyelles longues et brèves de l'arabe de Mayadin en fonction des 4 conditions expérimentales. Le [%] dénote ce que représente la brève par rapport à la longue. [$\Delta D/D$ %] dénote ce que représente l'écart entre la longue et la brève de la durée de la longue.

En ce qui concerne la durée des voyelles, nous constatons que:

- En ND les voyelles brèves constituent 60%, environ, de la durée de la longue (condition de base "sans contrainte"). C'est par rapport à cette condition que nous jugeons l'effet de la condition expérimentale sur la durée de la voyelle, c'est-à-dire que lorsque nous disons "le débit rapide baisse les durées des voyelles", nous sous-entendons "le débit rapide baisse les durées des voyelles de la condition de base".
- Sans aucune exception, le débit rapide de la parole baisse la durée des voyelles du débit normal (longues et brèves, schéma déclaratif et schéma interrogatif confondus) de $\Delta D/D = (D_{\text{normal}} - D_{\text{rapide}})/D_{\text{normal}} = 20\%$.

La vitesse de l'élocution est mesurée par l'unité Syllabe/Seconde (soit l'unité syll./s). La moyenne des durées des 398 syllabes [zVr] du débit normal est de 409 ms (soit 2,45 syll./s), en débit rapide la durée moyenne des 403 syllabes mesurées est de 332 ms (soit 3 syll./s).

Le rapport Syllabe débit normal/Syllabe débit rapide est alors de 1,23.

- L'interrogation: aucune exception non plus, l'interrogation augmente la durée des voyelles du schéma déclaratif (longues et brèves, débit normal et débit rapide confondus) de $\Delta D/D = (D_{\text{interrogatif}} - D_{\text{déclaratif}}) / D_{\text{interrogatif}} = 9\%$.
- La condition RI est intéressante parce qu'elle soumet les voyelles à 2 contraintes qui ont deux effets opposés. Le débit rapide baisse la durée de la voyelle et l'interrogation l'augmente, mais c'est l'effet du débit rapide qui l'emporte, parce que, sans exception, par rapport à ND, nous observons une baisse de la durée des voyelles (longues et brèves confondues) $\Delta D/D = (D_{\text{ND}} - D_{\text{RI}}) / D_{\text{ND}} = 13\%$.
- Nous pouvons classer ainsi nos voyelles des moins longues aux plus longues en fonction de leur condition d'expérimentation $RD < RI < ND < NI$.

En ce qui concerne maintenant l'écart entre les voyelles, alors que les valeurs des durées des voyelles changent sous l'effet des conditions expérimentales, l'écart est presque maintenu d'un débit à un autre et d'un schéma mélodique à un autre, il est perturbé lorsque les voyelles sont soumises à 2 conditions expérimentales à la fois.

Durée des voyelles de l'arabe de Mayadin selon 4 conditions expérimentales

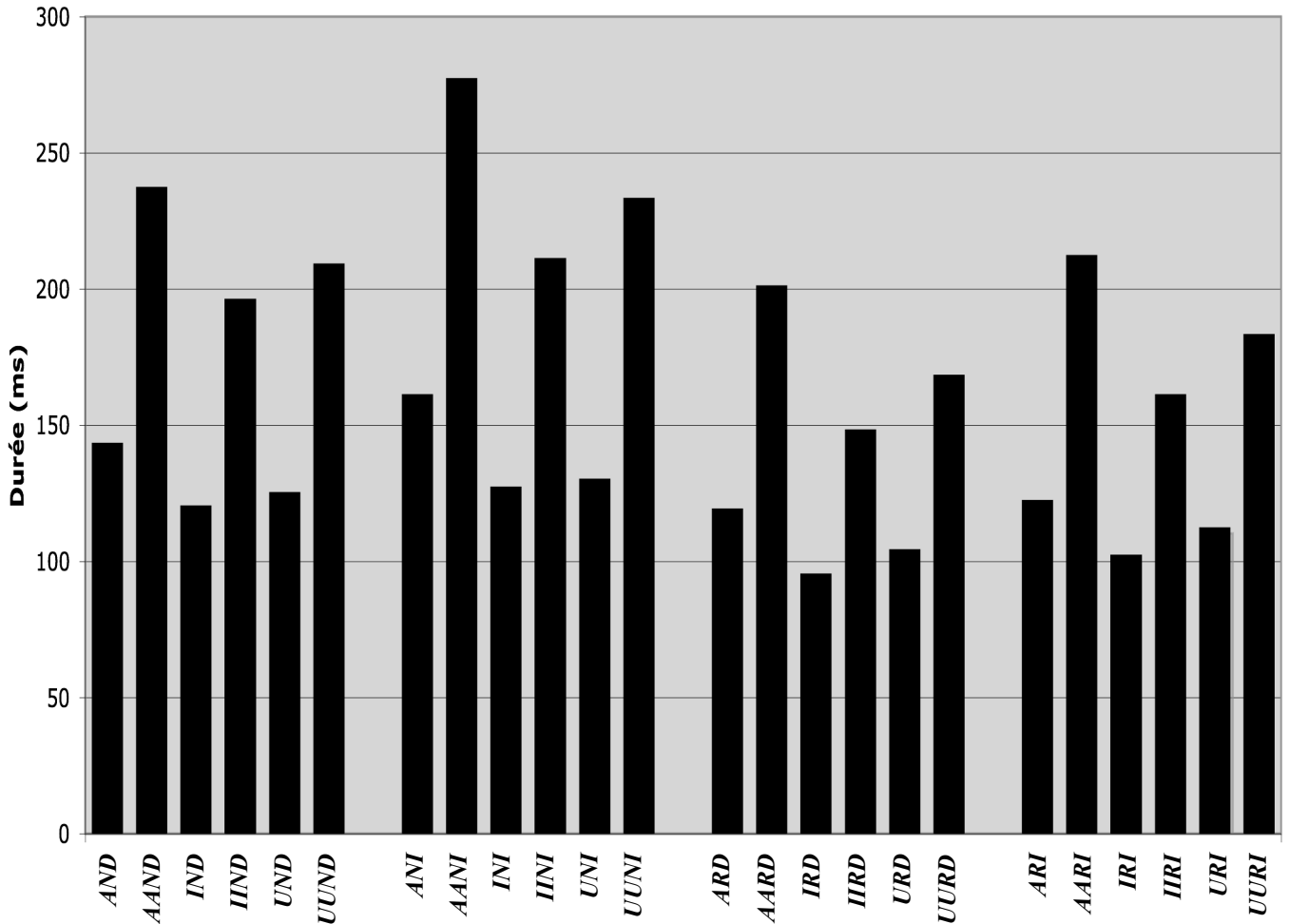


Figure 44. Durée des voyelles longues et brèves de l'arabe de Mayadin [I, II, A, AA, U et UU] selon deux débits de parole combinés avec deux schémas mélodiques. Moyennes de 3 locuteurs masculins. Légende: A = /a/ bref; AA = /a/ long; I = /i/ bref; II = /i/ long; U = /u/ bref; UU = /u/ long. N = le débit est normal; R = le débit est rapide; D = le schéma mélodique est déclaratif et I = le schéma mélodique est interrogatif.

III.4.2. Les formants

III.4.2.1. Présentation numérique

Les formants sont présentés dans les 2 tableaux 20 et 21. Le 14 pour les valeurs des F1, le 15 pour les valeurs des F2, avec le pourcentage que le F1 de la brève représente par rapport au F1 de la longue (idem pour F2), et avec $\Delta F1/F1$, où F1 est le premier formant de la plus longue des deux $(F1_{longue} - F1_{brève})/F1_{longue}$. Cela nous permettra de calculer en % l'écart entre F1 de la longue et F1 de la brève par rapport à la valeur de F1 de la longue. Nous cherchons à savoir si cet écart est maintenu d'une condition expérimentale à l'autre par la suite.

F1 (les valeurs sont en Hz).

Normal Déclaratif (ND)					Normal Interrogatif (NI)				
voyelle	Brève	Longue	%	$\Delta F1/F1$ %	voyelle	Brève	Longue	%	$\Delta F1/F1$ %
A	548	636	86,2	13,8	A	593	671	88,4	11,6
I	431	320	134,7	34,7	I	430	328	131,1	31,1
U	436	360	121	21	U	438	367	119,3	19,3

Rapide Déclaratif (RD)					Rapide Interrogatif (RI)				
voyelle	Brève	Longue	%	$\Delta F1/F1$ %	voyelle	Brève	Longue	%	$\Delta F1/F1$ %
A	565	626	90,3	9,7	A	575	657	87,5	12,5
I	422	335	126	26	I	453	354	128	28
U	430	356	120,8	20,8	U	465	392	118,6	18,6

Tableau 20. Variation des rapports du premier formant entre les voyelles longues et brèves de l'arabe de Mayadin en fonction des 4 conditions expérimentales. Le [%] dénote ce que représente F1 de la brève par rapport à F1 de la longue. [$\Delta F1/F1$] % dénote ce que représente l'écart entre F1 de la longue et F1 de la brève de F1 de la longue.

Nous observons que:

- En ND, les F1 des voyelles brèves n'ont pas le même pourcentage par rapport à ceux des longues, comme nous avons vu avec la durée, mais ce pourcentage change selon la condition expérimentale.

- Toutes voyelles confondues, le débit rapide de la parole réduit la distance entre les F1 des voyelles longues et brèves.
- L'interrogation: sans exception, l'interrogation réduit les distances entre les F1 des longues et brèves, son effet de rapprochement des F1 sur le triangle vocalique est, en général, moins fort que celui du débit rapide.
- Le RI, par rapport au ND, a un effet de rapprochement des F1.

F2 (les valeurs sont en Hz).

Normal Déclaratif (ND)					Normal Interrogatif (NI)				
voyelle	Brève	Longue	%	$\Delta F2/F2$ %	voyelle	Brève	Longue	%	$\Delta F2/F2$ %
A	1295	1196	108,3	8,3	A	1313	1245	105,5	5,5
I	1549	2030	76,3	23,7	I	1592	2055	77,5	22,5
U	1040	921	112,9	12,9	U	1066	856	124,5	24,5

Rapide Déclaratif (RD)					Rapide Interrogatif (RI)				
voyelle	Brève	Longue	%	$\Delta F2/F2$ %	voyelle	Brève	Longue	%	$\Delta F2/F2$ %
A	1257	1237	101,6	1,6	A	1356	1330	102	1,2
I	1576	2001	78,8	21,2	I	1603	2016	79,5	20,5
U	1026	999	102,7	2,7	U	1136	903	125,8	25,8

Tableau 21. Variation des rapports du deuxième formant entre les voyelles longues et brèves de l'arabe de Mayadin en fonction des 4 conditions expérimentales. Le [%] dénote ce que représente F2 de la brève par rapport à F2 de la longue. [$\Delta F2/F2$ %] dénote ce que représente l'écart entre F2 de la longue et F2 de la brève de F2 de la longue.

- En ND, les F2 des voyelles brèves n'ont pas le même pourcentage par rapport à ceux des longues sous toutes les conditions expérimentales, ce qui signifie que le changement de la condition expérimentale a un impact sur la réorganisation des qualités brèves par rapport aux longues.
- Toutes voyelles confondues, le débit rapide de la parole réduit la distance entre les F2 des voyelles longues et les F2 des voyelles brèves. Cette réduction est plus importante que celle que nous avons observée auprès de F1.
- L'interrogation augmente les distances entre les F2 des longues et des brèves pour certaines

voyelles et la réduit pour certaines autres. Son effet de rapprochement des F2 est, au total, moins fort que celui du débit rapide.

- Le RI, par rapport au ND, a un effet de rapprochement des F2, avec un effet moins fort que le débit rapide tout seul, et moins fort que l'interrogation toute seule.

III.4.2.2. Présentation graphique

Nous représentons les voyelles de notre expérience dans le plan F1 X F2 selon la condition de base, figure 45. Cet espace vocalique est l'espace de référence, parce que c'est l'espace obtenu *a priori* sans contrainte, ainsi, c'est par rapport à cet espace-là que nous estimerons les effets du débit rapide et du schéma mélodique interrogatif sur la répartition des voyelles dans le plan.

Concernant une zone de points de dispersion, nous nous intéressons à deux caractéristiques: (1) l'emplacement de la zone par rapport aux autres zones sur le triangle. (2) la surface de la zone. Le premier point traduit les valeurs de ces points en termes de F1 et de F2 et donne à cette zone son identité. Alors que le deuxième est corrélé à la variation de ces valeurs par rapport à leur moyenne. Plus les différences entre les valeurs sont importantes plus la surface de la zone est grande. La variation sur l'axe horizontal reflète la variation de la position avant arrière de la langue, alors que la variation sur l'axe vertical reflète la position haut-bas de la mandibule.

Nous tenons à préciser que nous sommes conscients que le débit normal n'est pas synonyme de parole normale, parce que dans la parole normale l'espace vocalique est plus réduit, comme les ellipses qui couvrent leurs points de dispersion sur le triangle vocalique présentent beaucoup de chevauchement par rapport à la parole prononcée pour un enregistrement. Parce que la personne qui prononce les mots à enregistrer prend beaucoup soin de sa manière de prononcer, c'est le cas pour les voyelles de notre expérience. À ce propos, Lindblom (1967) a écrit, après avoir effectué une expérience qui a consisté à enregistrer des voyelles dans différents contextes consonantiques pour tester l'influence du contexte sur les transitions formantiques: «*There is every reason to believe that for normal conversational speech similar data would show even greater overlap*». Dans tous les cas, les valeurs du triangle, figure 45, nous serviront de valeurs de référence pour estimer l'effet de la condition expérimentale sur la localisation des formants.

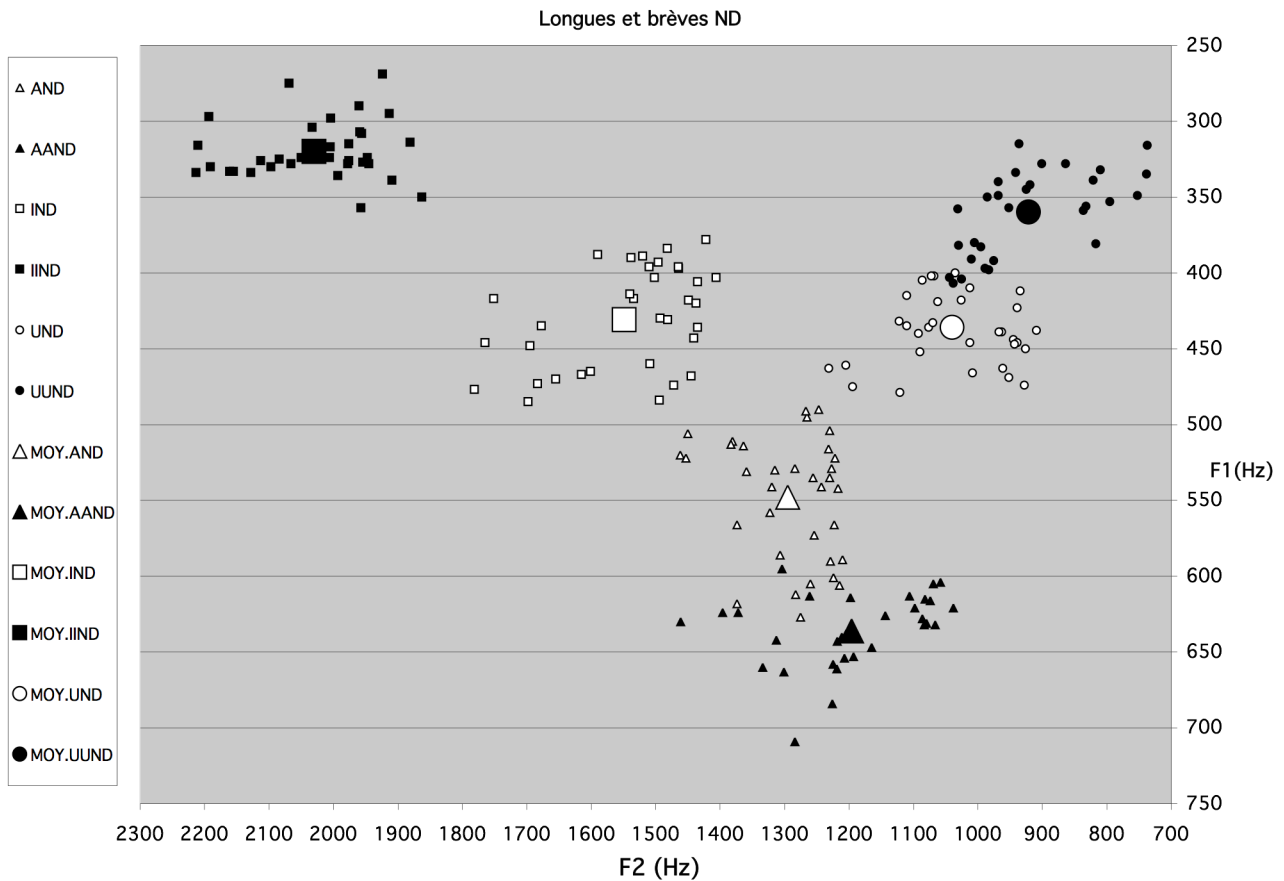


Figure 45. Moyennes et points de dispersion des voyelles longues et brèves de l'arabe de Mayadin [I, II, A, AA, U et UU], en fonction de F1 et de F2 avec un débit normal et un schéma mélodique déclaratif, 3 locuteurs, 10 répétitions par locuteur et par voyelle.

Condition ND: nous notons que les voyelles brèves sont dans une zone centrale du triangle par rapport aux longues. Conformément aux paramètres de l'établissement graphique des triangles vocaliques, nous avons trouvé que les brèves forment un triangle inscrit dans le triangle des longues et aucun sommet de celui des brèves n'est confondu avec l'un des sommets de celui des longues. De par leur proximité, les brèves sont plus sujettes à la confusion entre elles que les longues entre elles. Les brèves ne se comportent pas toutes de la même manière vis-à-vis de leurs longues correspondantes. En effet, nous observons que les ellipses des deux voyelles de la paire [a] vs. [a:] présentent beaucoup de chevauchement, c'est-à-dire que l'aire de l'intersection des ellipses de dispersion est relativement importante, ce chevauchement est moins important pour la paire [u] vs. [u:], et il est complètement absent pour la paire [i] vs. [i:].

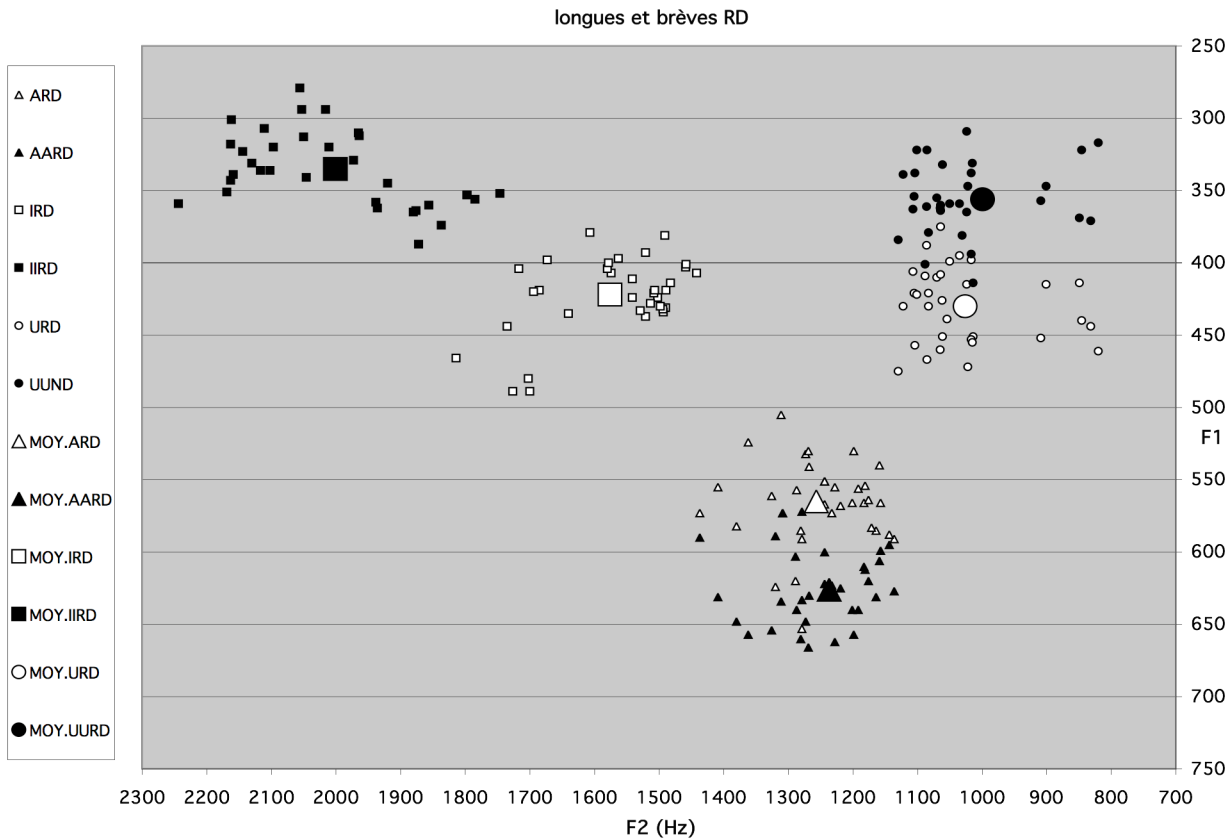


Figure 46. Moyennes et points de dispersion des voyelles longues et brèves de l'arabe de Mayadin [I, II, A, AA, U et UU], en fonction de F1 et de F2 avec un débit rapide et un schéma mélodique déclaratif, 3 locuteurs, 10 répétitions par locuteur et par voyelle.

Condition RD: les voyelles brèves demeurent dans le centre de l'espace par rapport aux longues qui en constituent la périphérie. Nous avons remarqué dans cette condition que la distance Brève/Longue d'une paire diminue, alors que la distance entre les 3 paires, deux à deux, augmente. En effet, pour les deux paires [a] vs. [a:] et [u] vs. [u:], brèves et longues tendent à s'aligner sur l'axe de F2 ce qui réduit la distinction entre les deux termes de chaque paire. Ainsi la confusion entre la brève et la longue correspondante augmente, et c'est dû à la fois à une décentralisation des brèves et à une centralisation des longues (excepté le [u] long). Nous remarquons par ailleurs que les zones des points de dispersion ont des surfaces moins importantes que celles du débit normal, ce qui veut dire qu'il y a moins de variation d'une répétition à une autre.

Comme en condition ND, la paire [a] vs. [a:] présente beaucoup de chevauchement, suivie par la paire [u] vs. [u:], suivies enfin par la paire [i] vs. [i:] où il n'y a pas de chevauchement.

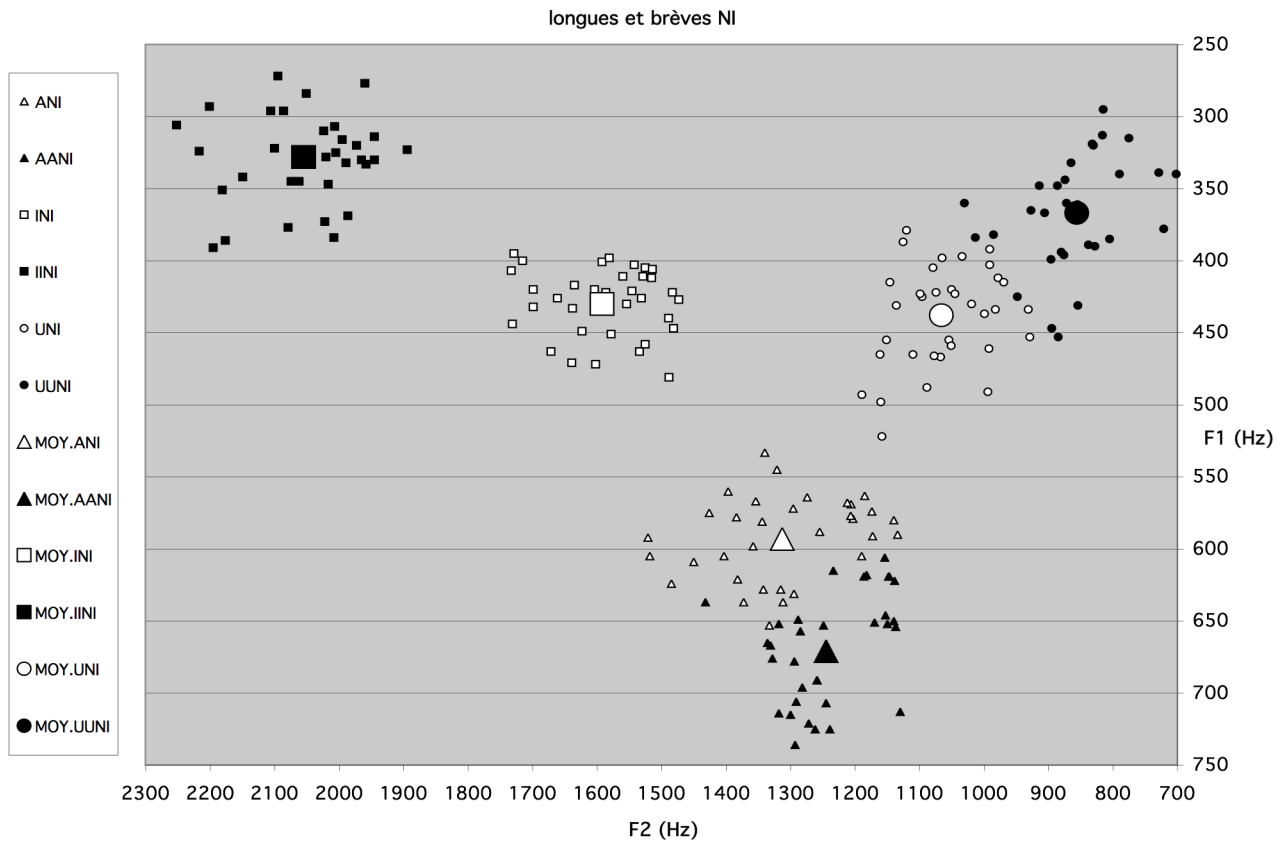


Figure 47. Moyennes et points de dispersion des voyelles longues et brèves de l'arabe de Mayadin [I, II, A, AA, U et UU], en fonction de F1 et de F2 avec un débit normal et un schéma mélodique interrogatif, 3 locuteurs, 10 répétitions par locuteur et par voyelle.

Condition NI: il nous semble que cette condition d'élocution est la meilleure du point de vue de la distance entre les voyelles, parce qu'à la fois les tailles des ellipses sont plus petites par rapport à la condition expérimentale de référence (ND), ce qui signifie qu'il y a moins de variation d'une répétition à une autre, et parce que les emplacements des ellipses sont plus éloignés, en effet, la différence au niveau de F2 pour la paire [a] vs. [a:] et pour la paire [u] vs. [u:] qui a disparu sous le débit rapide de la parole semble rétablie par l'effet de l'interrogation, ainsi les confusions sont minimisées par rapport à la condition précédente (RD). Comme dans les deux conditions précédentes, c'est-à-dire ND et RD, le chevauchement des ellipses de dispersion est présent pour la paire [a] vs. [a:] et pour la paire [u] vs. [u:], mais pas pour la paire [i] vs. [i:]. La configuration des brèves par rapport aux longues reste la même que les deux autres conditions ND et RD, c'est-à-dire que les voyelles brèves demeurent dans le centre de l'espace par rapport aux longues qui en constituent la périphérie.

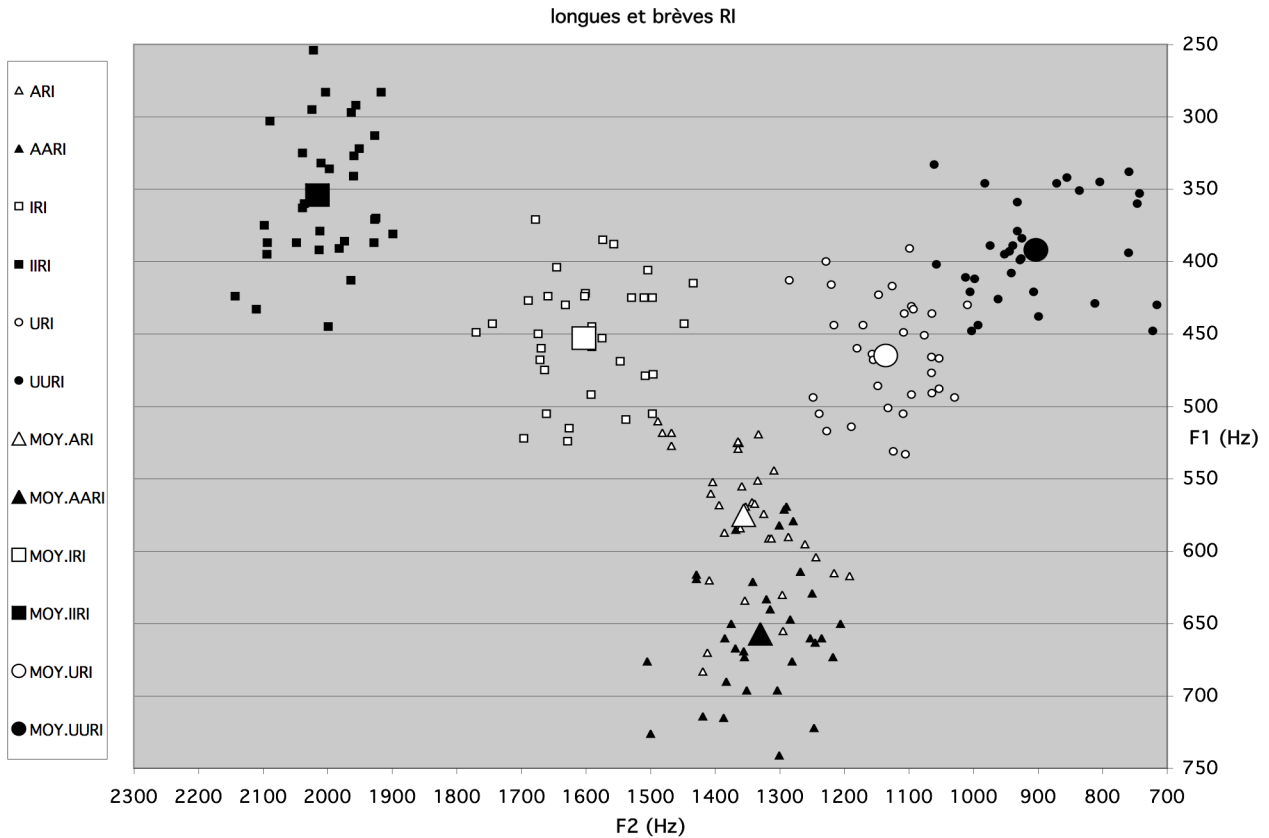


Figure 48. Moyennes et points de dispersion des voyelles longues et brèves de l'arabe de Mayadin [I, II, A, AA, U et UU], en fonction de F1 et de F2 avec un débit rapide et un schéma mélodique interrogatif, 3 locuteurs, 10 répétitions par locuteur et par voyelle.

Ce mode d'émission des voyelles semble être le plus perturbateur, parce que les tailles des ellipses augmentent d'une manière remarquable, ce qui signifie qu'il y a plus de variation d'une répétition à une autre par rapport à la condition expérimentale de référence (ND). Les emplacements des ellipses sont par ailleurs plus rapprochés

Le chevauchement des ellipses de dispersion est présent pour la paire [a] vs. [a:] et pour la paire [u] vs. [u:], mais toujours pas pour la paire [i] vs. [i:]. Les emplacements des brèves par rapport aux longues dans l'espace vocalique restent les mêmes que ce que nous avons pu observer dans les 3 autres conditions ND, RD et NI, c'est-à-dire que les voyelles brèves demeurent dans le centre de l'espace par rapport aux longues qui se trouvent sur la périphérie.

III.5. Significativité des écarts

Nous voudrions tout d'abord mentionner que le schéma mélodique interrogatif (l'intonation montante) s'est révélé efficace dans sa manière d'affecter les trois indices principaux qui signalent l'interrogation, la durée, l'intensité et la F0. Tous locuteurs, toutes qualités vocaliques, toutes longueurs et tous débits de parole confondus, l'interrogation a eu un effet significatif ($p < 0,01$) sur la F0, l'intensité et la durée des voyelles.

En effet, nous avons testé premièrement l'effet du schéma mélodique sur la F0 en confondant toutes les réalisations (400 réalisations par schéma) afin d'éliminer les éventuels effets des autres variables indépendantes (qualité, débit et longueur), et ne laissant en jeu que l'effet du schéma mélodique. L'Anova a montré à ce stade-là que la hausse de la F0 due à l'intonation montante était significative, figure 49.

Deuxièmement, nous avons fait des analyses qualité par qualité, débit par débit et longueur par longueur ($3*2*2$ respectivement) tous locuteurs confondus, les 12 analyses se sont montrées toutes significatives ($p < 0,01$).

Pour tester l'effet du schéma mélodique sur l'intensité, nous avons procédé de la même manière que pour la F0, c'est-à-dire que nous avons effectué le test une première fois en confondant toutes les réalisations, le schéma avait un effet significatif sur l'intensité (figure 50), et une deuxième fois pour chaque qualité, chaque débit et chaque longueur à part ($3*2*2$ respectivement), les 12 analyses se sont montrées toutes significatives ($p < 0,01$).

En ce qui concerne la durée, en confondant les 400 réalisations par schéma, le schéma mélodique a eu un effet significatif sur la durée de la voyelle (figure 51), alors qu'en testant cet effet pour chacune des variables indépendantes à part, les 12 analyses ne se sont pas montrées toutes significatives.

En prélude de nos analyses statistiques, nous voudrions annoncer que les tests statistiques portant sur les trois éléments que nous venons d'observer sous l'interrogation (la F0, l'intensité et la durée), ont montré que, tous locuteurs, toutes qualités vocaliques, toutes longueurs et tous débits de parole confondus, le débit rapide affecte significativement ($p < 0,01$) ces trois éléments.

Hausse de la F0 en fonction du schéma mélodique

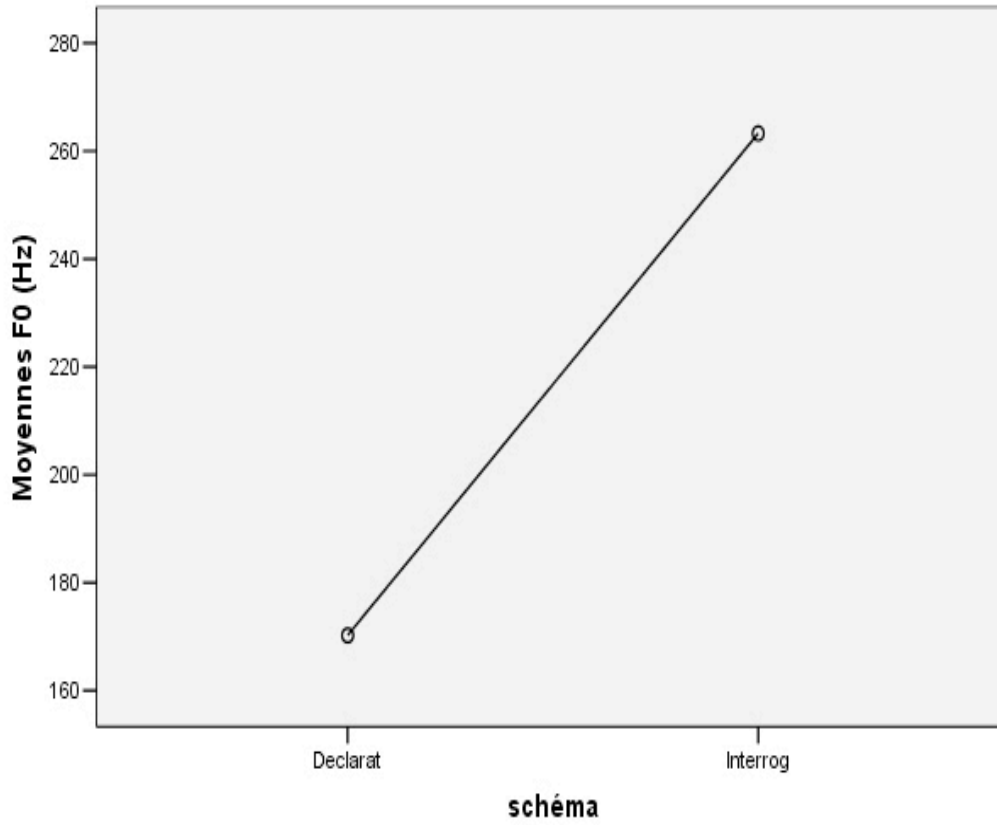


Figure 49. Effet du schéma mélodique (déclaratif/interrogatif) sur la F0 en arabe de Mayadin. Tous locuteurs (3 locuteurs masculins), toutes qualités [i, a, u], toutes longueurs (Brève/Longue) et tous débits de parole (normal/rapide) confondus. Effet significatif ($p < 0,01$).

Hausse de l'intensité en fonction du schéma mélodique

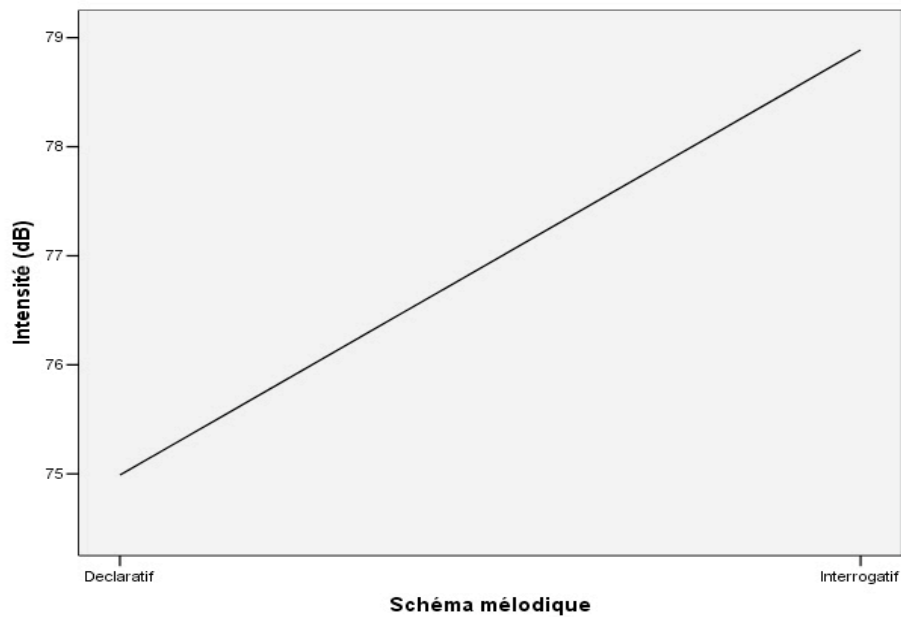


Figure 50. Effet du schéma mélodique (déclaratif/interrogatif) sur l'intensité en arabe de Mayadin. Tous locuteurs (3 locuteurs masculins), toutes qualités [i, a, u], toutes longueurs (Brève/Longue) et tous débits de parole (normal/rapide) confondus. Effet significatif ($p < 0,01$).

Hausse de la durée en fonction du schéma mélodique

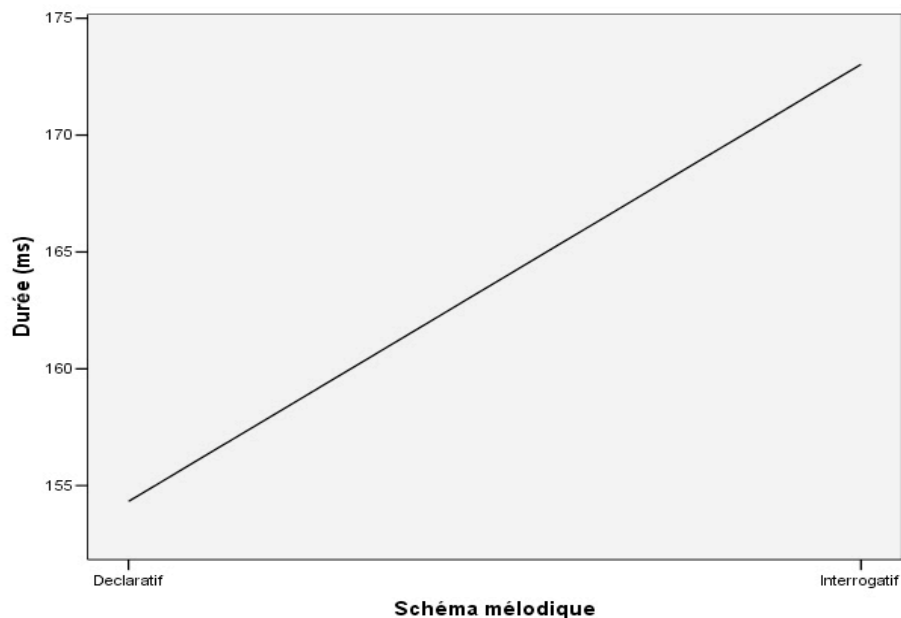


Figure 51. Effet du schéma mélodique (déclaratif/interrogatif) sur la durée en arabe de Mayadin. Tous locuteurs (3 locuteurs masculins), toutes qualités [i, a, u], toutes longueurs (Brève/Longue) et tous débits de parole (normal/rapide) confondus. Effet significatif ($p < 0,01$).

III.5.1. Significativité statistique

Le logiciel qui nous a permis d'établir ces calculs statistiques est SPSS.

Le plan d'expérience est le suivant:

$St_{30} * Q_3 * L_3 * D_2$: (avec St pour stimulus; Q pour qualité; L pour longueur et D pour débit d'élocution). Il y a 30 stimuli par qualité, par longueur et par débit.

Les différences dont nous cherchons la significativité portent sur les variables dépendantes suivantes: (1) Durée, (2) F1 et (3) F2.

III.5.1.1. L'écart entre les durées

Nous cherchons à obtenir et à comparer les résultats des deux points suivants:

- La différence de durée entre la voyelle brève et sa correspondante longue en débit normal.
- La différence de durée entre la voyelle brève et sa correspondante longue en débit rapide.

Nous fixons le seuil de significativité dans cette étude à 0,01, c'est-à-dire que lorsque nous disons "significatif" c'est que $p < 0,01$.

Nous avons fait une analyse de variance par schéma (déclaratif ou interrogatif). L'Anova avait 3 variables indépendantes: qualité [i, a, u], longueur (brève, longue) et débit (normal, rapide). La 4e variable indépendante (locuteur) n'était pas testée, nous avons mélangé les données de nos 3 locuteurs. En effet, le test post hoc de Scheffé a montré que les durées des voyelles ne sont pas différentes entre les locuteurs comparés deux à deux.

III.5.1.1.1. Schéma déclaratif

L'Anova a montré un effet significatif de la qualité, c'est-à-dire que les voyelles longues étaient différentes entre elles et que les voyelles brèves étaient différentes entre elles du point de vue de la durée. Le test de Scheffé montre que les [i] et [u] brefs ne sont pas différents entre eux et tous deux sont différents du [a] bref, ceci est vrai sous les deux débits. Ce test montre aussi que les [i] et [u] longs ne sont pas différents entre eux et tous deux sont différents du [a] long en débit normal, alors qu'ils sont deux à deux différents sous le débit rapide. Donc l'effet de la qualité n'est pas uniforme, parce qu'il dépend de la qualité et du débit.

Nous avons testé l'effet du débit pour chacune des 6 voyelles, l'Anova a montré un effet significatif du débit sur la durée pour les 6 voyelles testées.

Enfin nous avons testé l'effet de la longueur sur la durée des voyelles. C'est ce dernier

point que nous cherchons à déterminer dans cette étude, car elle permet de savoir si les voyelles brèves ont une différence temporelle statistiquement significative par rapport aux longues. L'Anova nous permet de confirmer la présence d'une différence temporelle entre les 2 termes d'une paire brève vs. longue pour les trois paires étudiées.

Les figures 52 et 53 représentant l'effet de la longueur sur la durée sous le débit normal et rapide respectivement, avec un schéma mélodique déclaratif, nous indiquent que:

- Chacune des voyelles brèves est significativement plus courte que la voyelle longue correspondante, l'écart entre brève et longue en débit normal étant significatif.
- Le débit rapide modifie significativement les valeurs des durées obtenues en débit normal, mais l'écart entre brève et longue en débit rapide demeure significatif.
- L'étendue de l'écart diminue significativement en passant du débit normal au débit rapide.

En répliquant l'analyse de la variance pour chaque locuteur à part, nous avons trouvé que chacun des locuteurs réalise une différence de durée significative entre la longue et la brève correspondante pour chacune des trois paires I/II, A/AA et U/UU avec les 2 débits de parole.

III.5.1.1.2. Schéma interrogatif

L'interrogation sous le débit normal a significativement augmenté la durée des voyelles [a] bref, [a:] long et [u:] long, les durées des trois autres voyelles sont restées inchangées.

Le test de Scheffé montre que les [i] et [u] brefs ne sont pas différents entre eux et tous deux sont différents du [a] bref en débit normal. Sous le débit rapide, deux à deux, il n'y a aucune différence significative entre ces trois voyelles brèves. Ce test montre aussi que les [i] et [u] longs ne sont pas différents entre eux et tous deux sont différents du [a] long en débit normal, alors qu'ils sont deux à deux différents sous le débit rapide. Donc l'effet de la qualité n'est pas uniforme, parce que, encore une fois, il se montre dépendant de la qualité et du débit.

Le débit n'a pas eu d'effet sur les [i] et [u] brefs, ce n'était pas le cas avec un schéma déclaratif où il a affecté les six voyelles de l'étude. Ici, le débit a affecté le [a] bref et toutes les longues.

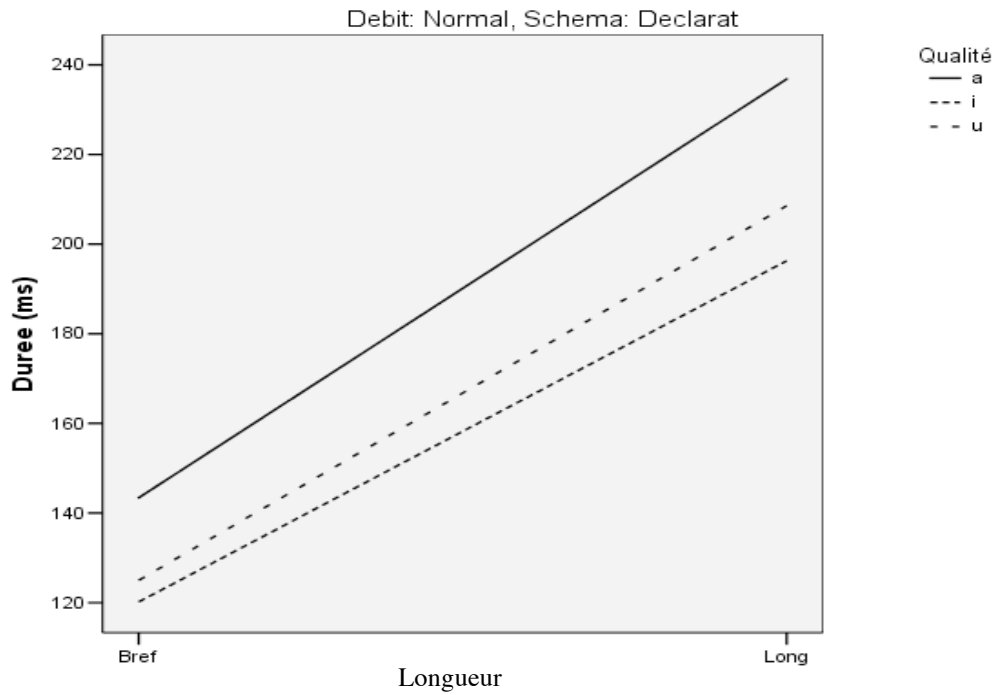


Figure 52. Différence temporelle entre les voyelles longues [i:, a:, u:] et leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit normal. Le schéma mélodique est déclaratif. 3 locuteurs masculins.

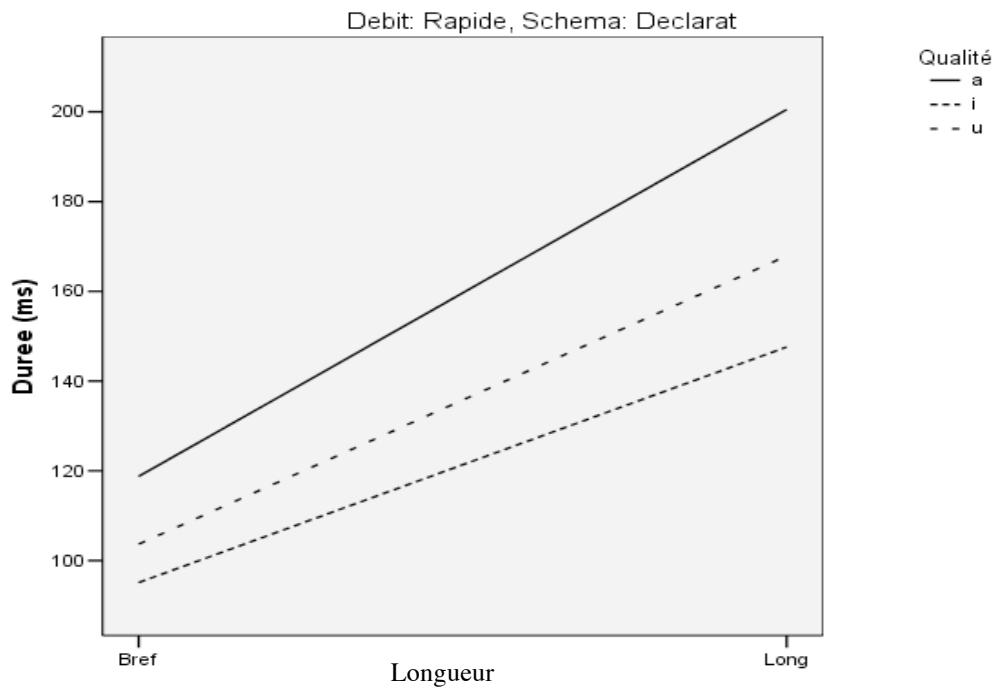


Figure 53. Différence temporelle entre les voyelles longues [i:, a:, u:] et leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit rapide. Le schéma mélodique est déclaratif. 3 locuteurs masculins.

L'effet de la longueur sur la durée se trouve au centre de notre intérêt, parce qu'il permet de savoir si les voyelles brèves ont une différence temporelle statistiquement significative par rapport aux longues sous le schémas mélodique interrogatif.

L'effet de la longueur sur la durée était significatif pour les trois paires et avec les deux débits de parole. Lorsque nous avons procédé à l'analyse de cet effet pour chaque locuteur à part, nous avons pu détecter que chez l'un des 3 locuteurs, l'opposition I/II en débit rapide n'était pas significative.

Les figures 54 et 55 représentant l'effet de la longueur sur la durée sous le débit normal et rapide respectivement avec un schéma mélodique interrogatif, nous indiquent que:

- l'écart entre brève et longue en débit normal est significatif.
- l'écart entre brève et longue en débit rapide reste significatif.
- le débit rapide modifie significativement les valeurs des durées du débit normal de toutes les voyelles longues et du [a] bref.
- l'étendue de l'écart diminue significativement en passant du débit normal au débit rapide.

En d'autres termes, nous n'observons pas le même comportement en durée des voyelles pour le schéma déclaratif et interrogatif.

Résumé

Dans ce paragraphe nous avons testé l'effet de 3 variables indépendantes (qualité, débit et longueur) sur une variable dépendante (durée) sous 2 schémas mélodiques: déclaratif vs. interrogatif. Nous pouvons ainsi résumer, tableau 22, nos résultats en affirmant que: aussi bien sous une intonation assertive qu'une intonation interrogative, le débit rapide peut changer les valeurs absolues des durées des voyelles du débit normal, mais ce changement n'efface pas l'écart temporel entre les longues et les brèves, l'écart résiste à l'effet perturbateur du débit rapide et demeure significatif, ce qui veut dire: **la robustesse des écarts des durées résiste à la perturbation du débit rapide.**

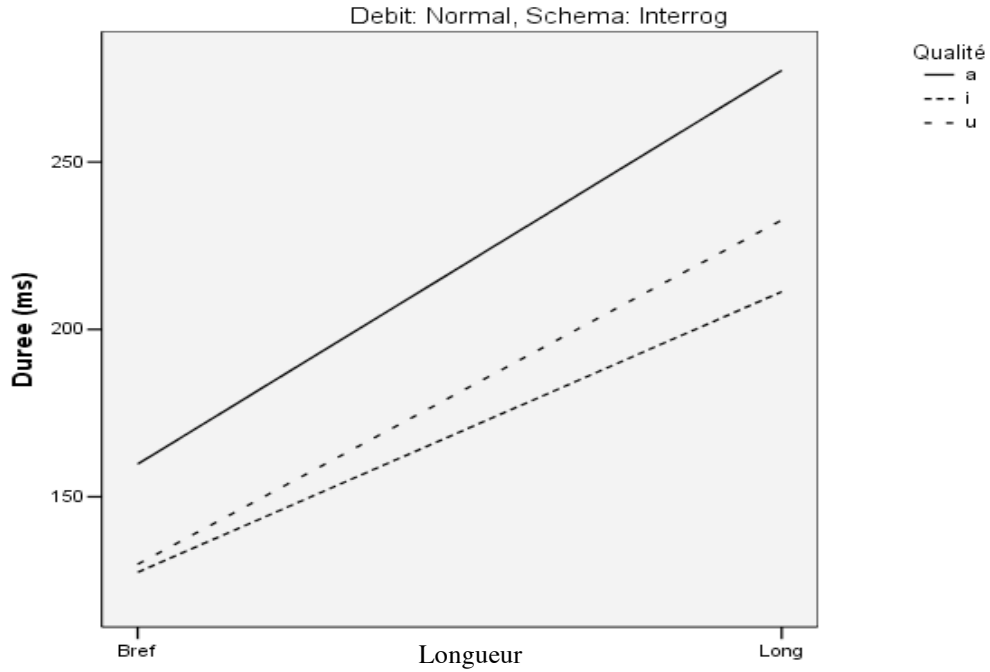


Figure 54. Différence temporelle entre les voyelles longues [i:, a:, u:] et leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit normal. Le schéma mélodique est interrogatif. 3 locuteurs masculins.

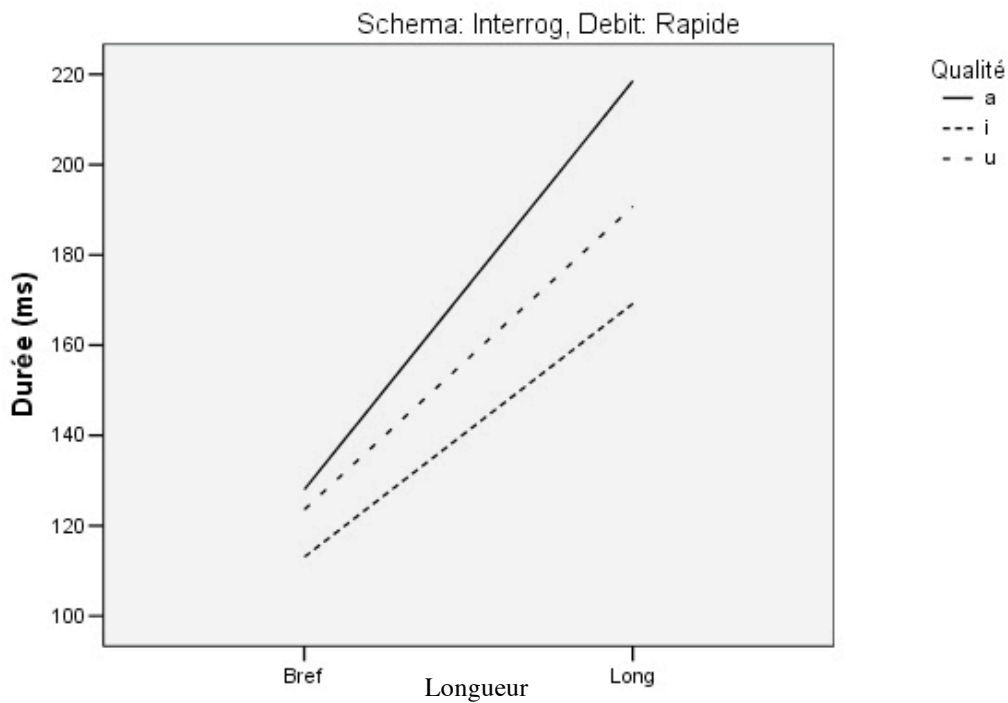


Figure 55. Différence temporelle entre les voyelles longues [i:, a:, u:] et leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit rapide. Le schéma mélodique est interrogatif. 3 locuteurs masculins.

Durée

	Déclaratif		Interrogatif	
	Normal	Rapide	Normal	Rapide
A/AA	(+)	(+)	(+)	(+)
I/II	(+)	(+)	(+)	(+)
U/UU	(+)	(+)	(+)	(+)

Tableau 22. Résumé des significativités statistiques de l'effet de la longueur sur la durée pour les 3 paires I/II, A/AA et U/UU de voyelles de l'arabe de Mayadin selon deux schémas mélodiques (déclaratif vs. interrogatif) et deux débits de parole (normal vs. rapide). Le signe "+" signifie que l'écart est significatif ($p < 0,01$), le signe "-" signifie que l'écart n'est pas significatif.

III.5.1.2. L'écart entre les valeurs des F1

Nous cherchons à obtenir et à comparer les résultats des deux points suivants:

- La différence entre les valeurs des F1 des voyelles brèves et les valeurs des F1 des voyelles longues correspondantes en débit normal.
- La différence entre les valeurs des F1 des voyelles brèves et les valeurs des F1 des voyelles longues correspondantes en débit rapide.

Nous avons fait une analyse de variance par schéma (déclaratif ou interrogatif). L'Anova avait 3 variables indépendantes: qualité [i, a, u], débit (normal, rapide) et longueur (brève, longue). Nous avons mélangé les données de nos 3 locuteurs. Avec 10 répétitions par locuteur, nous avons 30 stimuli

Le plan d'expérience est le suivant:

$$St_{30} * Q_3 * L_2 * D_2$$

III.5.1.2.1. Schéma déclaratif

Qualité: Concernant l'effet de la qualité sur F1, Le test de Scheffé montre qu'en débit normal, figure 56, les [i] et [u] brefs ne sont pas différents entre eux et tous deux sont différents du [a] bref. Cette configuration reste la même sous le débit rapide, figure 57. Ce test montre aussi que les longues sont deux à deux différentes sous le normal. En débit rapide Les [i] et [u] longs ne sont pas différents entre eux et tous deux sont différents du [a] long.

Débit: A part le F1 du [i] long, aucun F1 n'a été affecté par le débit rapide de la parole.

Longueur: L'Anova montre un effet significatif de la longueur sur les valeurs de F1.

En effet, paire par paire, nous trouvons que la valeur de F1 de la voyelle longue est significativement différente que celle de la voyelle brève, avec le débit normal, figure 56, pour les 3 paires I/II, A/AA et U/UU. Cette différence demeure significative sous le débit rapide, figure 57.

- Le F1 de chaque voyelle brève est différent de celui de sa longue correspondante sous les deux débits de parole.

- L'effet de la brièveté sur les voyelles longues est la centralisation, puisque les valeurs de F1 de chaque voyelle brève se rapprochent vers un point central dont le F1 se situe vers 500 Hz (F1 de la voyelle centrale [ə]), et cela est vrai pour les deux débits.

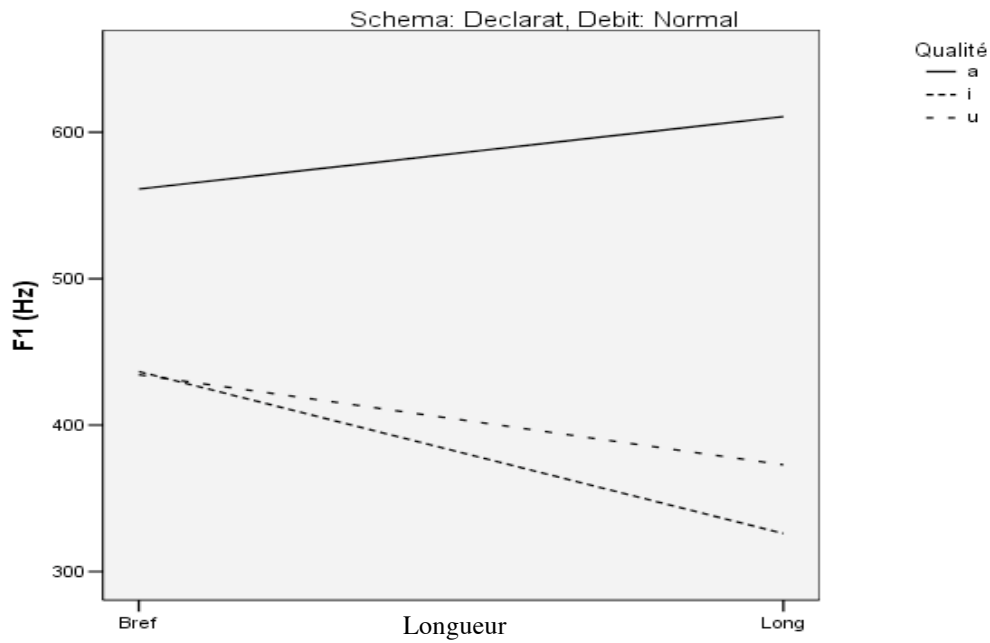


Figure 56. Différence entre les valeurs des F1 des voyelles longues [i:, a:, u:] et ceux de leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit normal. Le schéma mélodique est déclaratif. 3 locuteurs masculins.

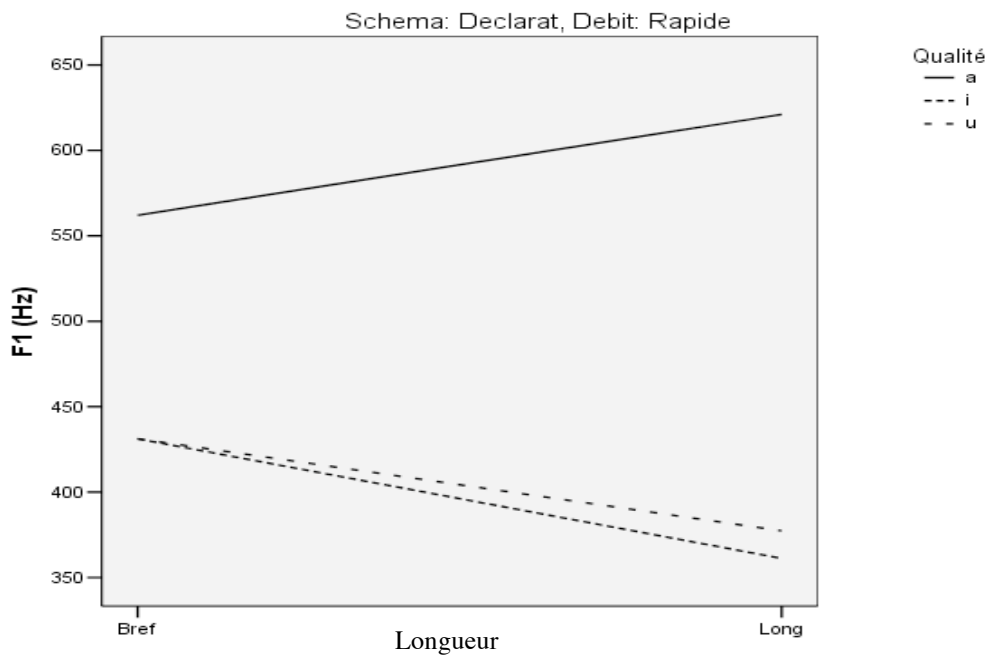


Figure 57. Différence entre les valeurs des F1 des voyelles longues [i:, a:, u:] et ceux de leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit rapide. Le schéma mélodique est déclaratif. 3 locuteurs masculins.

III.5.1.2.2. Schéma interrogatif

L'interrogation (avec un débit normal) ne change pas significativement les valeurs des F1, seule exception est le F1 du [i:] long.

Qualité: Concernant l'effet de la qualité sur F1 avec un schéma mélodique interrogatif, le test de Scheffé montre qu'en débit normal, figure 58, les [i] et [u] brefs ne sont pas différents entre eux et tous deux sont différents du [a] bref, configuration maintenue sous le débit rapide. Ce test montre aussi que les longues sont deux à deux différentes sous le débit normal. En débit rapide les [i] et [u] longs ne sont pas différents entre eux et tous deux sont différents du [a] long.

Débit: Comme le schéma déclaratif, à part le F1 du [i] long, aucun F1 n'a été affecté par le débit rapide de la parole avec un schéma interrogatif.

Longueur: Nous avons examiné chaque paire de voyelles à part. Paire par paire, nous trouvons que la valeur de F1 de la voyelle longue est significativement différente de celle de la voyelle brève, avec le débit normal, figure 58, pour les 3 paires I/II, A/AA et U/UU. Cette différence demeure significative sous le débit rapide, figure 59.

Nous pouvons dire, à partir des figures 58 et 59:

Le F1 de chaque voyelle brève est différent de celui de sa longue correspondante sous les deux débits de parole.

Nous retrouvons l'effet de centralisation déjà vu avec le schéma déclaratif.

Comme pour la déclaration, l'effet du débit est peu important, les voyelles longues et brèves maintenant leur écart en passant d'un débit à l'autre.

Les voyelles ont un comportement uniforme: En débit normal, les valeurs des voyelles brèves sont significativement différentes de celles des longues. En débit rapide, les valeurs des brèves sont significativement différentes de celles des longues. .

Résumé: Dans ce paragraphe nous avons testé l'effet de 3 variables indépendantes (qualité, débit et longueur) sur une variable dépendante (F1) sous 2 schémas mélodiques: déclaratif vs. interrogatif. Le débit rapide ne change pas significativement les valeurs absolues du premier formant des voyelles du débit normal (la seule exception est le [i] long, un changement sans impact sur l'écart), ni avec un schéma mélodique déclaratif ni avec un schéma mélodique interrogatif. L'écart entre les valeurs du premier formant des voyelles longues et les valeurs du premier formant des voyelles brèves correspondantes, tableau 23, est

significatif en débit normal et en débit rapide pour les trois paires I/II, A/AA et U/UU, ce qui veut dire: **la robustesse de l'écart des F1 résiste à la perturbation du débit rapide.**

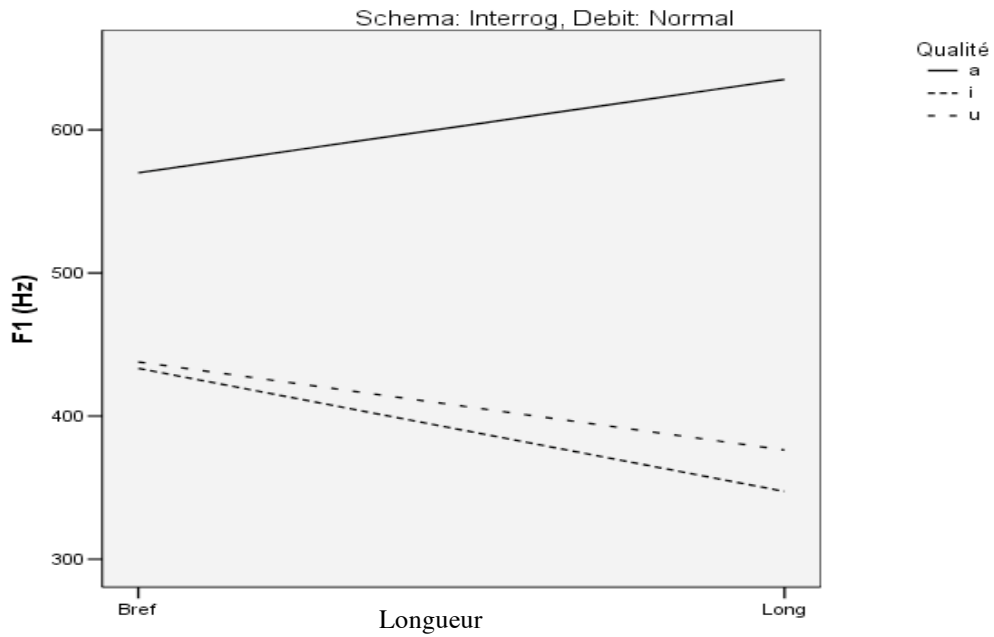


Figure 58. Différence entre les valeurs des F1 des voyelles longues [i:, a:, u:] et ceux de leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit normal. Le schéma mélodique est interrogatif. 3 locuteurs masculins.

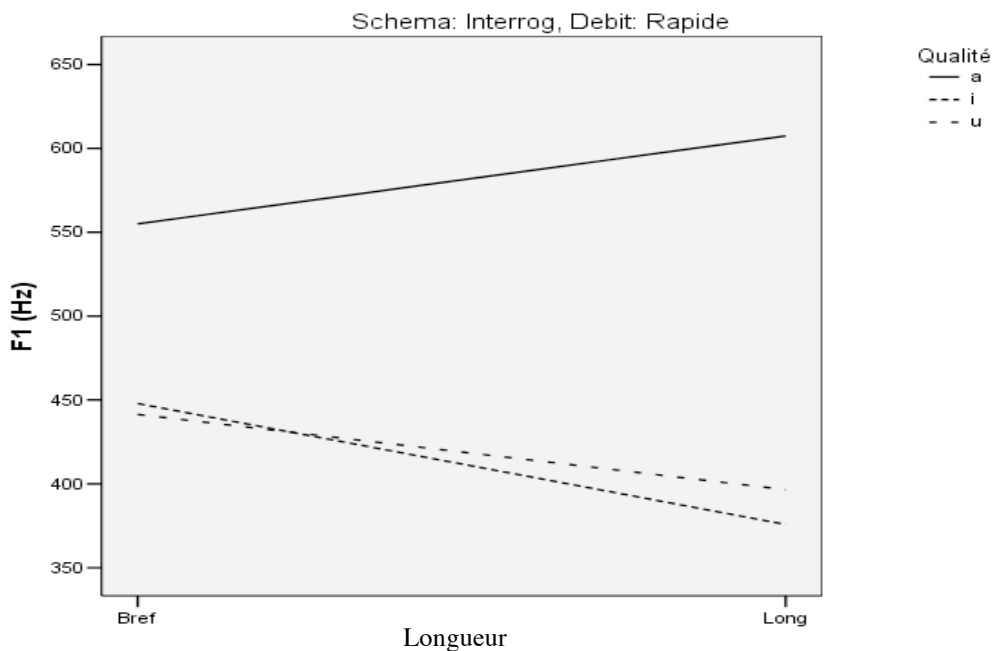


Figure 59. Différence entre les valeurs des F1 des voyelles longues [i:, a:, u:] et ceux de leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit rapide. Le schéma mélodique est interrogatif. 3 locuteurs masculins.

F1

	Déclaratif		Interrogatif	
	Normal	Rapide	Normal	Rapide
A/AA	(+)	(+)	(+)	(+)
I/II	(+)	(+)	(+)	(+)
U/UU	(+)	(+)	(+)	(+)

Tableau 23. Résumé des significativités statistiques de l'effet de la longueur sur F1 pour les 3 paires I/II, A/AA et U/UU de voyelles de F1 entre les 3 voyelles longues et les 3 voyelles brèves de l'arabe de Mayadin selon deux schémas mélodiques (déclaratif vs. interrogatif) et deux débits de parole (normal vs. rapide). Le signe "+" signifie que l'écart est significatif, le signe "-" signifie que l'écart n'est pas significatif ($p < 0,01$).

III.5.1.3. L'écart entre les valeurs des F2

Nous cherchons à obtenir et comparer les résultats des deux points suivants:

- La différence de F2 entre la voyelle brève et sa correspondante longue en débit normal.
- La différence de F2 entre la voyelle brève et sa correspondante longue en débit rapide.

Nous avons fait une analyse de variance par schéma (déclaratif ou interrogatif). L'Anova avait 3 variables indépendantes: qualité [a, i, u], débit (normal, rapide) et longueur (brève, longue). Nous avons mélangé les données de nos 3 locuteurs. Avec 10 répétitions par locuteur, nous avons 30 stimuli

Le plan d'expérience est le suivant:

$$St_{30} * Q_3 * L_2 * D_2$$

III.5.1.3.1. Schéma déclaratif

Qualité: Concernant l'effet de la qualité sur F2, le test de Scheffé montre qu'en débit normal, figure 60, et en débit rapide, figure 61, les brèves sont deux à deux différentes. Cette configuration est la même pour les voyelles longues.

Débit: les F2 des brèves et ceux des longues sont insensibles au changement du débit, à l'exception du F2 du [i] long. L'Anova ne montre pas d'effet significatif du débit sur le F2 pour les autres voyelles.

Longueur: en débit normal, l'Anova montre un effet significatif de la longueur sur les valeurs de F2 pour les 2 paires I/II et A/AA, mais pas pour la paire U/UU. En débit rapide, figure 61, seule la paire I/II présente une différence significative en ce qui concerne le F2, en effet, la différence pour la paire A/AA, qui peine à être significative en débit normal, disparaît sous l'effet de la perturbation du débit rapide.

III.5.1.3.2. Schéma interrogatif

L'interrogation (avec un débit normal) a eu un effet significatif sur le F2 du [a] bref, sur le F2 du [a:] long et sur F2 du [i:].

Qualité: Concernant l'effet de la qualité sur F2, les résultats sont identiques à ceux que nous avons obtenus avec le schéma déclaratif, puisque le test de Scheffé montre qu'en débit normal, figure 62, et en débit rapide, figure 63, les brèves sont deux à deux différentes. Le test montre aussi que cette configuration est la même pour les voyelles longues.

Longueur: en débit normal, l'Anova montre un effet significatif de la longueur sur les valeurs de F2 pour les 2 paires I/II et U/UU, mais pas pour la paire A/AA. Cette configuration demeure vraie sous le débit rapide.

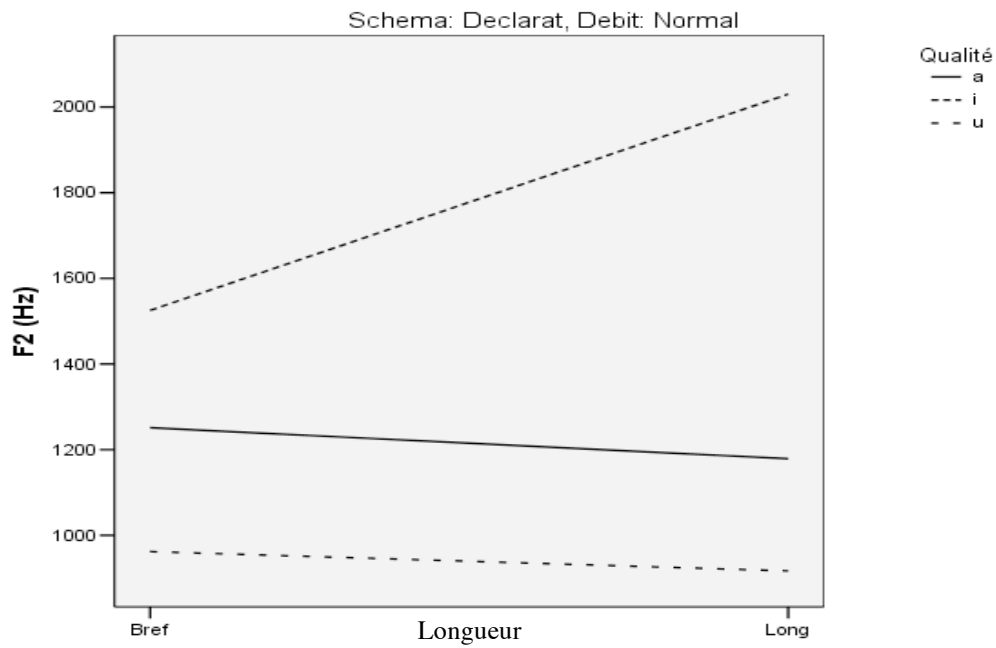


Figure 60. Différence entre les valeurs des F2 des voyelles longues [i:, a:, u:] et ceux de leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit normal. Le schéma mélodique est déclaratif. 3 locuteurs masculins.

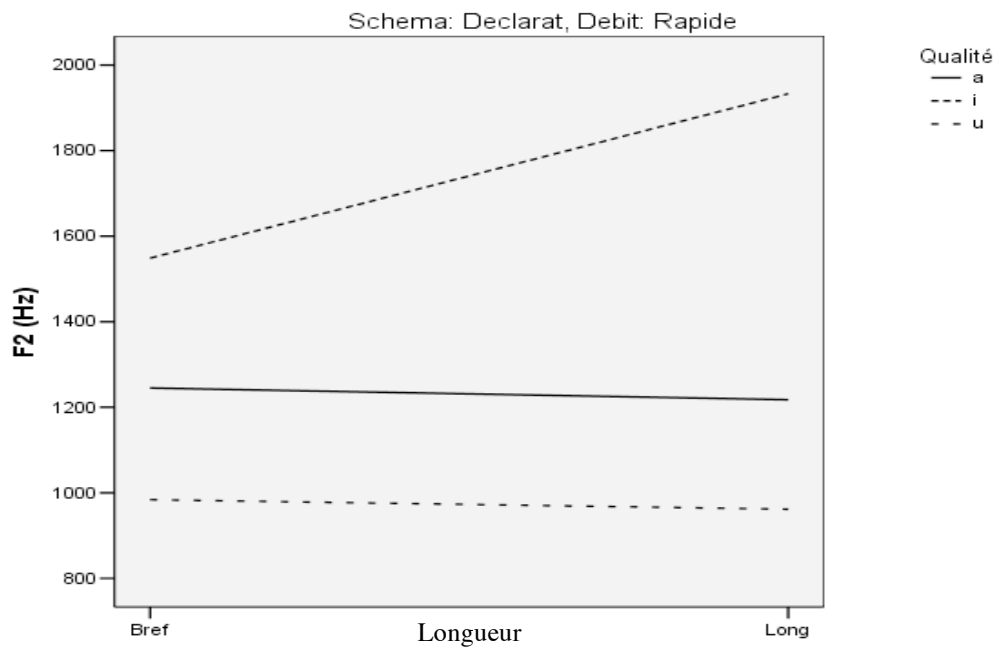


Figure 61. Différence entre les valeurs des F2 des voyelles longues [i:, a:, u:] et ceux de leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit rapide. Le schéma mélodique est déclaratif. 3 locuteurs masculins.

Résumé: Dans ce paragraphe nous avons testé l'effet de 3 variables indépendantes (qualité, débit et longueur) sur une variable dépendante (F2) sous 2 schémas mélodiques: déclaratif vs. interrogatif. Le débit rapide change significativement les valeurs absolues du deuxième formant de certaines voyelles du débit normal.

Contrairement à F1, l'écart entre les valeurs du deuxième formant des voyelles longues et les valeurs du deuxième formant des voyelles brèves n'est pas significatif pour les trois paires de voyelles avec un schéma déclaratif et un débit normal (sans perturbation). Et le schéma interrogatif et le débit rapide montrent des effets significatifs sur le F2. Le tableau 24 rend compte du comportement de la totalité des F2 de toutes les paires avec les débits et avec les deux schémas mélodiques.

Cela veut dire que la paire A/AA ne présente d'écart significatif que lorsqu'elle est émise avec un débit normal, l'écart ne résiste pas à l'effet du débit rapide. Avant de continuer sur la robustesse de A/AA, nous voudrions nous attarder sur une apparition fondamentalement importante pour notre travail: le débit rapide, qui a un effet significatif sur les valeurs de la durée, de F1 et de F2, montre pour la première fois un effet significatif sur l'écart Brève/Longue, un effet qui était jusque là, pour l'écart concernant la durée et le F1, absent.

Cette absence totale de l'effet du débit sur les écarts de la durée et de F1, nous a poussé à nous poser des questions à propos du bien fondé du choix de cette condition expérimentale (le changement de débit) pour tester la robustesse des écarts entre les différentes voyelles. Mais là nous voyons effectivement que (*Ceteris paribus* "toutes choses étant égales par ailleurs") rien que le passage d'un débit à l'autre a apporté une nouvelle donnée de grande importance.

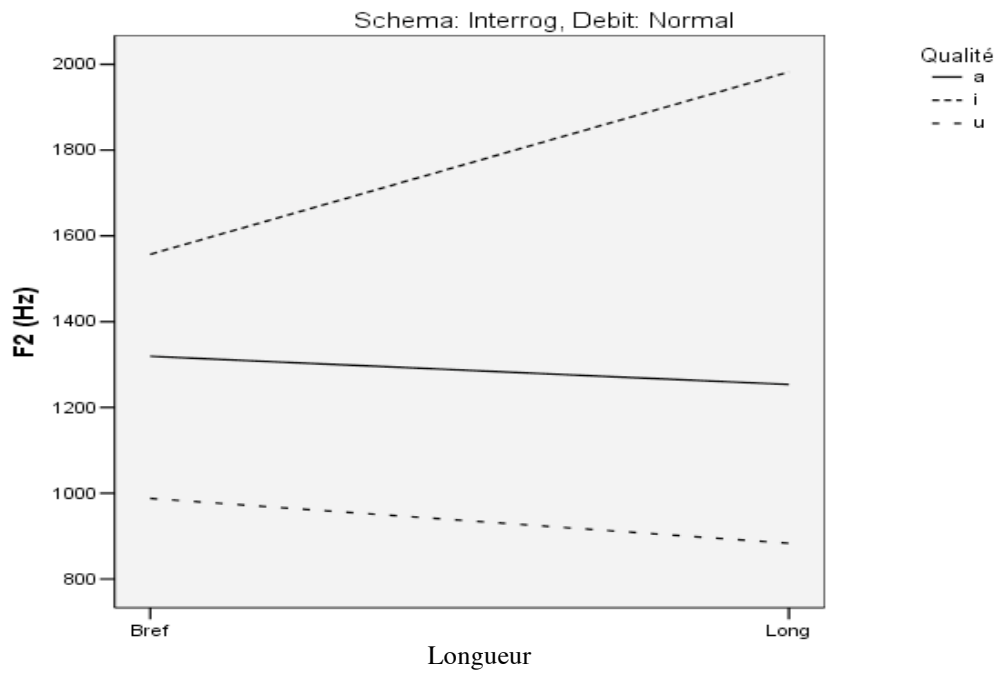


Figure 62. Différence entre les valeurs des F2 des voyelles longues [i:, a:, u:] et ceux de leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit normal. Le schéma mélodique est interrogatif. 3 locuteurs masculins.

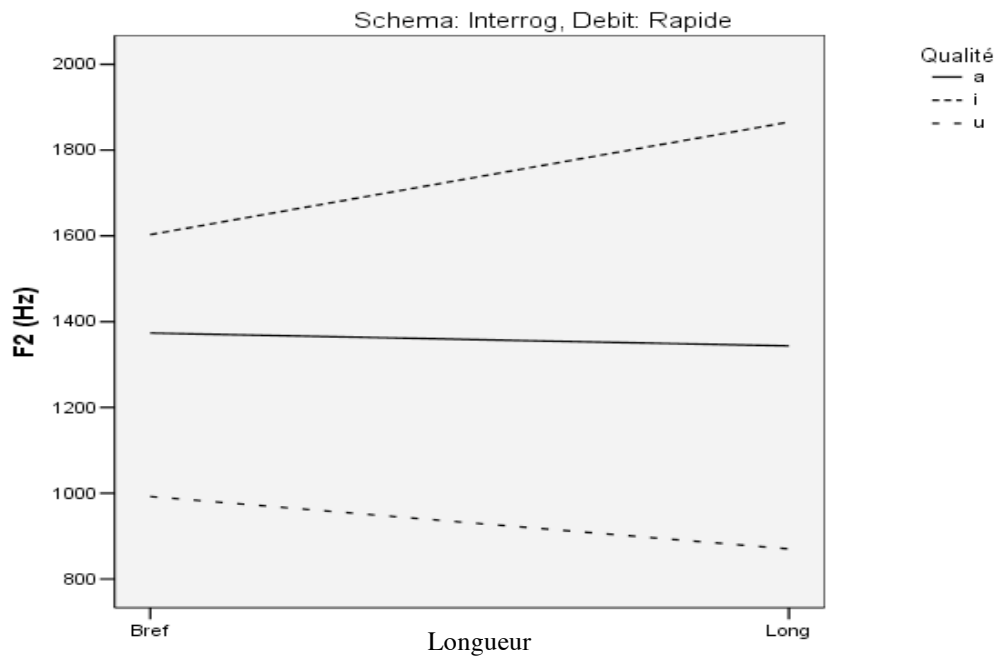


Figure 63. Différence entre les valeurs des F2 des voyelles longues [i:, a:, u:] et ceux de leurs partenaires brèves [i, a, u] de l'arabe de Mayadin sous le débit rapide. Le schéma mélodique est interrogatif. 3 locuteurs masculins.

Nous retournons au tableau 24: l'opposition A/AA ne résiste pas non plus à l'effet de l'interrogation, sur ce point, l'interrogation se confirme, de même que le débit rapide, une condition expérimentale perturbatrice, c'est-à-dire que deux formants qui ont normalement une différence qui permet leur discrimination peuvent se trouver confondus, statistiquement parlant, à cause de leur soumission à l'interrogation. Finalement, débit rapide et interrogation ont des effets perturbateurs, pourtant nous ne pouvons pas les comparer, car ils agissent différemment, si nous regardons le tableau 24, F2 U/UU est perturbé par le débit mais par l'interrogation.

F2

	Déclaratif		Interrogatif	
	Normal	Rapide	Normal	Rapide
A/AA	(+)	(-)	(-)	(-)
I/II	(+)	(+)	(+)	(+)
U/UU	(-)	(-)	(+)	(+)

Tableau 24. Résumé des significativités statistiques de l'effet de la longueur sur F2 pour les 3 paires I/II, A/AA et U/UU de voyelles de l'arabe de Mayadin selon deux schémas mélodiques (déclaratif vs. interrogatif) et deux débits de parole (normal vs. rapide). Le signe "+" signifie que l'écart est significatif ($p < 0,01$), le signe "-" signifie que l'écart n'est pas significatif.

Nous concluons cette fois-ci sur l'écart F2 A/AA, il ne résiste ni au débit rapide ni à l'interrogation: **l'écart n'est pas robuste.**

L'écart F2 I/II est l'écart le plus important parmi les 3 paires, il résiste au débit rapide et résiste aussi au schéma interrogatif: **l'écart est robuste.**

L'écart F2 U/UU, a mis en évidence la non parité des effets des deux conditions d'expérimentation qui servent à tester la robustesse, c'est parce qu'il ne résiste pas au débit rapide alors qu'il résiste à l'interrogation: **l'écart n'est pas robuste**

III.5.1.4. Résumé général des études statistiques

Le débit rapide perturbe la durée des voyelles plus que l'interrogation ne fait, alors que l'interrogation perturbe les valeurs des formants plus que le débit ne fait. Voilà un résumé des effets de ces deux conditions perturbatrices sur F1 et sur F2.

Voyelle	Effet débit			Effet interrogation		
	Durée	F1	F2	Durée	F1	F2
i	(+)	(-)	(-)	(-)	(-)	(+)
a	(+)	(-)	(-)	(+)	(-)	(-)
u	(+)	(-)	(-)	(-)	(-)	(-)
i:	(+)	(+)	(+)	(-)	(+)	(+)
a:	(+)	(-)	(-)	(+)	(-)	(+)
u:	(+)	(-)	(-)	(+)	(-)	(-)

(+): effet significatif. (-): effet non significatif.

La durée vocalique pouvant imposer sa significativité, toutes paires de voyelles et toutes conditions expérimentales confondues, **elle est robuste**.

La qualité vocalique représentée par F1 qui résiste toutes paires de voyelles et toutes conditions expérimentales confondues, et par F2 qui ne résiste pas du tout pour l'opposition A/AA, qui résiste parfaitement pour l'opposition I/II et qui, pour l'opposition U/UU résiste plus que pour A/AA et moins que pour l'opposition I/II. Cela nous permet de conclure que **la qualité vocalique n'est pas robuste pour A/AA, robuste pour I/II et à moitié robuste pour U/UU (elle résiste à deux perturbations sur 3)**.

Les calculs statistiques ne nous donnent pas d'informations sur l'importance des écarts du point de vue de la perception, c'est-à-dire que nous ne savons pas lesquels sont au-delà du seuil de la perception des valeurs formantiques (perceptibles) et lesquels sont en deçà (non perceptibles), d'où l'intérêt d'étudier ces écarts du point de vue perceptif.

III.5.2. Significativité perceptive

III.5.2.1. Les écarts entre les durées (quantité)

Afin de savoir si un écart temporel est perceptible, nous avons eu recours à deux sources d'information.

(1) Nos recherches bibliographiques concernant le seuil différentiel de la perception de la durée et dont les résultats sont exposés en I.2. Un tableau récapitulatif de 11 de ces études est disponible en annexe n° V. En effet, nous avons examiné treize études qui ont porté sur le seuil différentiel de la perception de la durée des excitations sonores. Les résultats à propos de l'étendue de l'écart temporel qui permet à deux voyelles de se démarquer au niveau perceptif uniquement par leur durée, montrent que pour des durées situées entre 50 et 250 ms, pour qu'une variation de la durée puisse être exploitée linguistiquement, il faudrait que la durée initiale accuse une hausse ou une baisse supérieure à 10%, soit un rapport de Weber supérieur à 0,1. Une hausse ou une baisse de la durée initiale se situant entre 10% et 20%, peut être, mais pas nécessairement, suffisante pour créer une opposition temporelle linguistiquement exploitable, au-delà de 20% il est fort probable que la hausse soit suffisante pour permettre aux deux durées, initiale et initiale+hausse (ou +baisse) d'opérer indépendamment l'une par rapport à l'autre dans un système phonologique.

(2) Les travaux de Rossi concernant le seuil différentiel de la perception de la durée vocalique (Rossi, 1972: 445). Nous citons ici l'essentiel du passage qui intéresse notre travail: «Nous pouvons admettre qu'entre 6 c/s et 14 c/s le seuil est à peu près constant en valeur absolue et voisin de 3 c/s. Entre 13 et 29 c/s, où la loi de Weber est vérifiée $\Delta T/T = 22,5\%$ en moyenne. Enfin au-dessus de 30 c/s, $\Delta T/T$ augmente très rapidement, mais ces durées ne nous intéressent guère dans l'étude de la durée vocalique». Nous trouvons que ces mots, utilisés par Rossi "à peu près; voisin de; en moyenne" témoignent de la grande précaution qu'il nous conseille de prendre en exploitant ces résultats. D'après ce qu'il conclut, restent quelques points qui ne nous semblent pas clairs.

- **Seuil:** Nous ne savons pas quel est le seuil pour les voyelles qui ont une durée entre 130 et 140 ms: est-ce que c'est la valeur absolue de 30 ms, ou est-ce que c'est la loi de Weber qui le détermine.

- **Calcul:** Si nous calculons le seuil pour un son d'une durée de 120 ms, cette durée appartient à la catégorie où le seuil est absolu et d'une valeur de 30 ms. Ainsi cette durée devra être de 150 ms pour qu'une nouvelle sensation de durée naisse. Mais faisons le calcul pour cette

durée que nous venons d'avoir pour savoir de combien il faudra la baisser pour qu'une nouvelle sensation naisse. 150 ms: cette durée appartient à la catégorie des durées dont le seuil se calcule selon la loi de Weber $\Delta T/T = 22,5\%$, $\Rightarrow (150 - x)/150 = 0,225$, cela implique qu'il faut effectuer une baisse de 33,75 ms, alors que lorsque nous le calculons, dans le sens de la hausse, rappelons-le, c'était 30 ms. Le seuil établi par Rossi ne rend pas compte du sens de la variation (baisse vs. hausse).

Finalement, quoique les résultats de Rossi (1972) prévoient un seuil plus important (moins fin) de ce que nous avons pu observer dans les études qui ont évalué la taille du seuil, les données de Rossi (1972) comparées à celles des autres recherches disponibles, ne sont pas radicalement différentes et restent comparables.

Pour l'étude de la significativité des variations de la durée vocalique, nous considérons que les écarts entre les longues et les brèves sont dus à une baisse des durées des longues, et que les écarts entre les normales et les rapides sont dus à une baisse des durées des normales.

Significativité perceptive au niveau de la longueur: comme toutes les voyelles longues ont des durées qui sont supérieures à 140 ms, elles seront traitées selon la loi de Weber: $[(\text{la longue} - \text{la brève}) * 100 / \text{longue}] = \text{ce que l'écart entre la longue et la brève représente en \% de la durée de la longue}$. Si cet écart est supérieur ou égal à 22,5 alors il est significatif, et s'il est inférieur à 22,5, alors il est non significatif.

	Normal	rapide
$[(i:-i)] * 100 / i:]$	$((196 - 120) * 100) / 196 = 38,77$	$((148 - 95) * 100) / 148 = 35,81$
$[(a:-a)] * 100 / a:]$	$((237 - 143) * 100) / 237 = 39,66$	$((201 - 119) * 100) / 201 = 40,80$
$[(u:-u)] * 100 / u:]$	$((209 - 125) * 100) / 209 = 40,19$	$((168 - 104) * 100) / 168 = 38,10$

Tableau 25. Rapport de Weber concernant l'écart temporel avec un schéma mélodique déclaratif pour trois paires de voyelles: I/II, A/AA et U/UU, avec 2 débits de parole, normal et rapide pour 3 locuteurs masculins de l'arabe de Mayadin.

Le tableau 25 montre qu'aussi bien en débit normal qu'en débit rapide, pour les 3 paires de voyelles, la différence entre la brève et la longue est perceptible, puisque tous les rapport sont largement supérieurs à 22,5%.

	Normal	rapide
[(i:-i)]*100/i:]	$((211 - 127)*100)/211 = 39,81$	$((161 - 102)*100)/161 = 36,65$
[(a:-a)]*100/a:]	$((277 - 161)*100)/277 = 41,88$	$((212 - 122)*100)/212 = 42,45$
[(u:-u)]*100/u:]	$((233 - 130)*100)/233 = 44,21$	$((183 - 112)*100)/183 = 38,80$

Tableau 26. Rapport de Weber concernant l'écart temporel avec un schéma mélodique interrogatif pour trois paires de voyelles: I/II, A/AA et U/UU, avec 2 débits de parole, normal et rapide pour 3 locuteurs masculins de l'arabe de Mayadin.

Le tableau 26 montre qu'aussi en débit normal qu'en débit rapide, pour les 3 paires de voyelles, la différence entre la brève et la longue est perceptible, puisque le rapport de Weber est supérieur au seuil fixé à 22,5% pour chacune des 3 paires.

Résumé: les tableaux 25 et 26 nous enseignent que, quel que soit le débit de la parole et quel que soit le schéma mélodique, la différence temporelle entre chaque voyelle brève et sa correspondante longue est perceptible.

III.5.2.2. Les écarts entre les formants (qualité)

Nous avons discuté en I.1.4.1. le choix de la méthode de mesure de la distance entre deux qualités vocaliques, Schwartz *et al.* (1997b) vs. Kewley-Port & Zheng (1999). Nous pouvons distinguer trois étapes pour établir une distance entre 2 voyelles dans l'espace vocalique. 1) Conversion à l'échelle des Bark des fréquences exprimées en Hz: Schwartz *et al.* (1997b) utilisent la formule proposée par Schroeder *et al.* (1979): $1\text{Bark} = 7 * \text{asinh}(f \text{ (en Hz)}/650)$; alors que Kewley-Port & Zheng (1999) utilisent la formule proposée par Trau Müller (1990): $1\text{Bark} = ((26,81 * f \text{ (en Hz)})/(1960 + f \text{ (en Hz)})) - 0,53$. La comparaison des deux formules de conversion ainsi que nos données converties à l'échelle des Bark selon les deux formules sont disponibles en annexe n° V. Cette comparaison a montré qu'il n'y a pratiquement pas de différences entre les 2 formules de conversion pour des fréquences inférieures à 4000 Hz. Cela nous permet de conclure que si nous obtenons deux mesures différentes pour une même distance, cette différence ne proviendra pas alors du choix de la formule de conversion, puisque les 2 fournissent des résultats sensiblement identiques.

2) Intégration: Schwartz *et al.* (1997b) proposent d'intégrer F2, F3 et F4 en F'2 (le deuxième formant effectif) qui est le centre de gravité de ces trois derniers formants. Cette intégration est basée sur l'idée qu'une différence $\leq 3,5$ Bark entre deux ou trois formants d'une même voyelle (F1 est à part), fait que l'oreille humaine intègre ces formants en un seul, ce qui

est appelé l'intégration large bande (Chistovich & Lublinskaya, 1979). Le F_2' est calculé à partir des valeurs de F2, F3 et F4 selon le modèle proposé par Mantakas *et al.* (1986), (Algorithme de calcul disponible en annexe n° II). Kewley-Port & Zheng (1999) ne tiennent pas compte de F3 et de F4 dans leur procédure, donc il n'y a pas eu d'intégration pour les formants supérieurs à F2 (F3 et F4) dans leur méthode de mesure de la distance entre les éléments vocaliques. Donc à ce niveau-là, nous observons une divergence en ce qui concerne les procédures de mesures des distances, intégration de F3 et de F4 vs. non intégration de F3 et de F4. Nous avons constaté ainsi que cette différence qui réside dans le fait que Kewley-Port & Zheng (1999) ne tiennent pas compte de F3 et de F4, contrairement à Schwartz *et al.* (1997b), avait une conséquence importante en ce qui concerne les résultats des mesures lorsque la distance a une grande valeur, pour les faibles valeurs de distances, les résultats des deux méthodes restent comparables.

3) Distance: Schwartz *et al.* (1997b) proposent une distance euclidienne pondérée (puisque'ils affectent le F_2' par coefficient inférieur à 1) comme suit:

$$d_{ViVj} = [(F_{1i}-F_{1j})^2 + 0,3^2 (F_{2i}'- F_{2j}')^2]^{1/2}.$$

Cette distance prend en compte les valeurs des 4 premiers formants. La méthode de mesure utilisée par Kewley-Port & Zheng (1999) consiste tout simplement à calculer une moyenne arithmétique qui base la distance uniquement sur F1 et F2:

$$d_{ViVj} = [(|F_{1i}V_2 - F_{1j}V_1| + |F_{2i}V_2 - F_{2j}V_1|)/2].$$

En ce lieu encore nous assistons à une divergence concernant les deux procédures de mesure des distances dans l'espace vocalique, et là encore il y a matière à penser que les résultats de mesures soient différents.

Les résultats de l'exécution de la procédure suivie par Schwartz *et al.* (1997b), **conversion** de nos six voyelles en débit normal et en débit rapide avec un schéma mélodique déclaratif, **intégration** de F2, F3 et F4 en F_2' , ainsi que les mesures des **distances** séparant les brèves des longues, sont disponibles dans le tableau 27.

Voy.	F1		F2		F3		F4		F'2	distance en Bark
	Hz	Bark	Hz	Bark	Hz	Bark	Hz	Bark		
AN	548	5,36	1295	10,08	2567	14,58	3560	16,81	10,080	0,716
AAN	636	6,06	1196	9,59	2486	14,36	3560	16,81	9,590	
IN	431	4,36	1549	11,22	2554	14,54	3560	16,81	11,490	1,309
IIN	320	3,32	2030	13,00	2642	14,77	3560	16,81	14,160	
UN	436	4,40	1040	8,74	2509	14,42	3560	16,81	8,740	0,730
UUN	360	3,70	921	8,03	2691	14,90	3560	16,81	8,030	
AR	565	5,50	1257	9,90	2469	14,31	3560	16,81	9,900	0,485
AAR	626	5,99	1237	9,80	2467	14,31	3560	16,81	9,800	
IR	422	4,27	1576	11,33	2560	14,56	3560	16,81	11,710	1,087
IIR	335	3,46	2001	12,90	2633	14,75	3560	16,81	14,130	
UR	430	4,35	1026	8,66	2521	14,45	3560	16,81	8,660	0,683
UUR	356	3,66	999	8,51	2545	14,52	3560	16,81	8,510	
ANI	593	5,73	1313	10,17	2543	14,51	3560	16,81	10,168	0,611
AANI	671	6,33	1245	9,84	2606	14,68	3560	16,81	9,836	
INI	430	4,35	1592	11,40	2502	14,40	3560	16,81	11,994	1,067
IINI	328	3,40	2055	13,08	2619	14,71	3560	16,81	13,623	
UNI	438	4,42	1066	8,89	2484	14,35	3560	16,81	8,890	0,753
UUNI	367	3,77	856	7,62	2485	14,36	3560	16,81	7,621	
ARI	575	5,58	1356	10,37	2520	14,45	3560	16,81	10,371	0,642
AARI	657	6,22	1330	10,25	2456	14,28	3560	16,81	10,249	
IRI	453	4,55	1603	11,44	2504	14,41	3560	16,81	12,067	1,004
IIRI	354	3,65	2016	12,95	2588	14,63	3560	16,81	13,511	
URI	465	4,66	1136	9,27	2477	14,33	3560	16,81	9,273	0,771
UURI	392	4,00	903	7,92	2566	14,57	3560	16,81	7,922	

Voyelle	F1		F2		distance en Bark
	Hz	Bark	Hz	Bark	
AND	548	5,33	1295	10,14	0,608
AAND	636	6,04	1196	9,63	
IND	431	4,30	1549	11,30	1,438
IIND	320	3,23	2030	13,11	
UND	436	4,35	1040	8,76	0,721
UUND	360	3,63	921	8,04	
ARD	565	5,47	1257	9,95	0,297
AARD	626	5,96	1237	9,84	
IRD	422	4,22	1576	11,42	1,215
IIRD	335	3,38	2001	13,01	
URD	430	4,29	1026	8,68	0,432
UURD	356	3,59	999	8,52	
ANI	593	5,70	1313	10,23	0,475
AANI	671	6,31	1245	9,88	
INI	430	4,29	1592	11,49	1,343
IINI	328	3,31	2055	13,19	
UNI	438	4,37	1066	8,91	0,982
UUNI	367	3,70	856	7,62	
ARI	575	5,55	1356	10,43	0,387
AARI	657	6,20	1330	10,31	
IRI	453	4,50	1603	11,53	1,232
IIRI	354	3,57	2016	13,06	
URI	465	4,61	1136	9,31	1,027
UURI	392	3,94	903	7,93	

Tableau 27. Tableau du haut, procédure utilisée par Schwartz et al. (1997b): conversions en Bark des 4 premiers formants selon la formule proposée en Schroeder et al. (1979), intégrations des 3 formants F2, F3 et F4 en F'2 (F4 est fixé) et calculs des distances perceptives entre les 3 voyelles brèves [i, a, u] et leurs correspondantes longues [i:, a:, u:] de l'arabe de Mayadin.

Tableau du bas, procédure utilisée par Kewley-Port & Zheng (1999): conversions en Bark des 2 premiers formants selon la formule proposée en Traunmüller (1990) et calculs des distances perceptives entre les 3 voyelles brèves [i, a, u] et leurs correspondantes longues [i:, a:, u:] de l'arabe de Mayadin.

Deux débits de parole, normal (N) et rapide (R). Deux schémas mélodiques, déclaratif (D) et interrogatif (I).

Nous avons calculé les distances entre les brèves et les longues correspondantes selon la procédure utilisée par Kewley-Port & Zheng (1999) pour les deux schémas mélodiques, déclaratif et interrogatif, combinés avec les deux débits de parole, normal et rapide, pour chacune des trois paires ($2 \times 2 \times 3 = 12$ distances). Ces résultats sont présentés avec les résultats que nous avons obtenus pour les mêmes distances en utilisant la procédure de Schwartz *et al.* (1997b), et dont les résultats sont disponibles en tableau 27, dans les tableau 28 et 29. Nous avons choisi de reproduire les résultats selon la procédure Schwartz *et al.* (1997b) en tableau 27 et mettre ainsi les résultats selon les deux procédures côte à côte afin de faciliter la comparaison, ce qui nous permet de nous rendre compte exactement de la taille de la différence entre ces deux façons de considérer les écarts dans l'espace vocalique.

Déclaratif	en débit normal	
	A/AA	0,61
	I/II	1,44
	U/UU	0,72
	en débit rapide	
	A/AA	0,30
I/II	1,22	
U/UU	0,43	
Interrogatif	en débit normal	
	A/AA	0,48
	I/II	1,34
	U/UU	0,98
	en débit rapide	
	A/AA	0,39
I/II	1,23	
U/UU	1,03	

Déclaratif	en débit normal	
	A/AA	0,72
	I/II	1,31
	U/UU	0,73
	en débit rapide	
	A/AA	0,49
I/II	1,09	
U/UU	0,68	
Interrogatif	en débit normal	
	A/AA	0,61
	I/II	1,07
	U/UU	0,75
	en débit rapide	
	A/AA	0,64
I/II	1,00	
U/UU	0,77	

Tableau 28. Toutes les valeurs sont en Bark. Distances perceptives calculées entre les 3 voyelles brèves [i, a, u] et leurs correspondantes longues respectives [i:, a:, u:] de l'arabe de Mayadin avec 2 schémas mélodiques, déclaratif et interrogatif et selon 2 débits de parole, normal et rapide. Le calcul est établi selon la méthode proposée par Kewley-Port & Zheng (1999): $dV_iV_j = [(|F1V2 - F1V1| + |F2V2 - F2V1|)/2]$.

Tableau 29. Toutes les valeurs sont en Bark. Distances perceptives calculées entre les 3 voyelles brèves [i, a, u] et leurs correspondantes longues respectives [i:, a:, u:] de l'arabe de Mayadin avec 2 schémas mélodiques, déclaratif et interrogatif et selon 2 débits de parole, normal et rapide. Le calcul est établi selon la méthode proposée par Schwartz *et al.* (1997b): $dV_iV_j = [(F1i - F1j)^2 + 0,3^2 (F'2i - F'2j)^2]^{1/2}$.

Un regard global sur les résultats des mesures des distances perceptives dans l'espace vocalique, obtenus selon les deux procédures, nous montre que les résultats sont du même ordre de grandeur. La procédure de Schwartz *et al.* (1997b) donne systématiquement des distances plus élevées pour A/AA et moins élevées pour I/II par rapport à la procédure de Kewley-Port & Zheng (1999). Malgré leurs différences, ces deux procédures restent généralement comparables. Elles donnent par ailleurs le même classement, à savoir la distance I/II > à la distance U/UU > à la distance A/AA.

Kewley-Port & Zheng (1999) ont établi le seuil de la qualité vocalique selon deux examens. L'un a étudié le seuil psychoacoustique expérimentalement et l'a établi à **0,28 Bark, c'est le seuil prédit**. Le second examen a consisté en l'observation de la plus petite distance qui existe dans l'anglais parlé aux Etats-Unis, les auteurs se sont rendu compte que la plus petite distance observée dans la parole ne coïncidait pas avec le seuil prédit, mais plutôt avec 2*le seuil prédit, ainsi ils ont établi **le seuil observé à 0,56 Bark** (les détails sur ces résultats et sur l'obtention de la base de données qui a permis ce résultat sont disponibles en I.1.3.2.3.).

Nous considérons ainsi la première valeur ($\approx 0,30$ Bark) comme le seuil de perception de la qualité vocalique, puisque cette distance a un impact réel sur la discrimination des voyelles sans qu'elle soit suffisante pour supporter une opposition catégorielle. La deuxième valeur ($\approx 0,60$ Bark) est suffisante pour démarquer catégoriellement deux entités vocaliques, nous la considérons comme la valeur du seuil catégoriel de la qualité vocalique.

L'étude de Schwartz *et al.* (1997b) n'établit pas de seuil, ni prédit ni observé, leur travail se cantonnant à la mesure de la distance sans se soucier de l'étendue de la distance qui permet la séparation perceptive ou catégorielle des éléments vocaliques. Les résultats de l'application de leur procédure demeurent cependant un important point de repère qui nous sauvegarde d'une éventuelle aberration due à une mauvaise estimation du poids des formants supérieurs de la part de Kewley-Port & Zheng (1999), qui les ont supprimés de leur formule établissant la distance. Ainsi pour estimer une distance, nous prendrons en compte les résultats des deux procédures, mais nous l'évaluerons à la lumière des résultats de Kewley-Port & Zheng (1999) en ce qui concerne le seuil.

Nous avons conclu en III.5.1.3. que les distances sur l'axe de F2 pour les paires A/AA et U/UU n'étaient pas robustes, c'est-à-dire que, du point de vue statistique, le F2 de la brève et celui de la longue correspondante sont égaux. Donc nous pensons qu'il faut désormais les éliminer des formules qui permettent l'établissement des distances dans l'espace vocalique,

nous considérons ainsi que le F2 du /a/ bref est le même que le F2 du /a:/ long, idem pour les /u/ bref et long.

Par conséquent, les valeurs que nous avons présentées dans le tableau 28 se réduisent, liste de valeurs ci-contre, concernant les paires A/AA et U/UU puisque nous avons éliminé la contribution des F2 de la disance séparant la brève de la longue. Ces distances restent cependant inchangées pour la paire I/II.

Considérant ces derniers résultats à propos des seuils prédit et observé et de l'absence de la distance sur l'axe de F2 pour A/AA et U/UU, avec le schéma déclaratif, nous pouvons avancer qu'en ce qui concerne la paire I/II, les résultats de l'application de la procédure de Kewley-Port & Zheng (1999) donnent la distance I/II au delà du seuil catégoriel de la qualité vocalique et cela avec les deux débits de parole. Pour la paire A/AA, les calculs donnent cette distance perceptible en débit normal alors qu'elles l'estiment à la limite du perceptible en débit rapide, cette distance affranchit le premier seuil (de perception) mais pas le deuxième (catégoriel). En ce qui concerne la distance U/UU, les calculs jugent perceptible cette distance en débit normal et en débit rapide, comme pour la paire A/AA, cette distance affranchit le premier seuil (de perception) mais pas le deuxième (catégoriel).

Le schéma interrogatif ne change pas les données obtenues avec un schéma déclaratif pour les trois paires de voyelles et sous les deux débits de parole.

Ces mesures montrent que l'effet du débit est constant d'une paire à l'autre avec un schéma déclaratif, le débit rapide réduit la distance entre les deux termes de la paire. L'effet de l'interrogation par rapport à la déclaration n'est pas constant, tantôt c'est la réduction, tantôt c'est la magnification des distances. L'effet du débit sous l'interrogation n'est pas constant non plus, puisque tantôt c'est la réduction tantôt c'est la magnification des distances.

La distance I/II est, de loin, la distance la plus importante que des voyelles d'une paire longue/brève entretiennent entre elles, elle est donnée largement perceptible. La distance U/UU, avec un schéma déclaratif, se trouve souvent légèrement au-dessus du perceptible. La

Declaratif	en débit normal	A/AA	0,36
		I/II	1,44
		U/UU	0,36
	en débit rapide	A/AA	0,25
		I/II	1,22
		U/UU	0,35
Interrogatif	en débit normal	A/AA	0,31
		I/II	1,34
		U/UU	0,33
	en débit rapide	A/AA	0,33
		I/II	1,23
		U/UU	0,34

distance A/AA est la dernière du point de vue de sa taille.

Si nous essayons d'établir une corrélation entre la portée du contraste au niveau de la quantité et celle du contraste au niveau de la qualité, nous trouvons le suivant: les voyelles qui présentent un grand contraste au niveau de la quantité comme A/AA, présentent un faible contraste au niveau de la qualité et inversement.

III.6. Conclusions

Nous rappelons sommairement ici nos résultats finaux:

Au niveau statistique:

- Les écarts de durées entre voyelles brèves et longues sont significatifs quel que soit le débit de parole et quelle que soit la nature du schéma mélodique.
- Les écarts entre les F1 des voyelles brèves et les F1 des voyelles longues sont significatifs quel que soit le débit de parole et quelle que soit la nature du schéma mélodique.
- Les écarts des deuxièmes formants brèves/longues sont significatifs pour I/II quel que soit le débit et quel que soit le schéma mélodique. Pour les deux autres paires, c'est-à-dire U/UU et A/AA, la significativité dépend du débit et du schéma (tableau 24).

Au niveau perceptif:

- Quel que soit le débit de la parole et quel que soit le schéma mélodique, la différence temporelle entre chaque voyelle brève et sa correspondante longue est perceptible, puisque le rapport de Weber [(longue-brève)/longue] est supérieur à 35% quelle que soit la condition expérimentale, alors qu'un rapport supérieur à 20% indique que l'écart temporel est suffisant pour démarquer catégoriellement la brève de la longue.

En nous basant sur les propositions de Kewley-Port & Watson (1994), Kewley-Port & Zheng (1996, 1999) et Kewley-Port (2001) concernant la distance spectrale, nous admettons qu'une distance spectrale combinée (F1 X F2) d'une valeur avoisinant le 0,30 Bark est perceptible et peut ainsi jouer un rôle dans un système vocalique donné en contribuant, avec d'autres indices acoustiques (temporel ou prosodique), à la démarcation des qualités vocaliques. Alors que pour que cette distance spectrale puisse démarquer seule (sans l'apport d'autres indices acoustiques, temporel ou prosodique), cette valeur doit être le double, soit 0,60 Bark. Partant, seule la distance séparant [i] bref de [i:] long est suffisante pour démarquer les deux qualités quel que soit le débit et quel que soit le schéma mélodique. Pour

les deux autres paires, c'est-à-dire U/UU et A/AA, cette distance est perceptible et joue un rôle dans la séparation de la brève de la longue. En effet, si nous considérons que les différences au niveau de F2 n'existent pas pour les 2 paires /a/ vs. /a:/ et /u/ vs. /u:/, parce qu'elles ne se sont pas montrées robustes, elles n'ont pas pu résister aux effets perturbateurs du débit rapide de la parole et du schéma interrogatif, ces distances spectrales calculées à partir des F1 seuls selon la méthode de Kewley-Port & Zheng (1999), toutes conditions expérimentales confondues, varient entre 0,30 et 0,35 Bark (excepté la distance entre le /a/ bref et le /a:/ long qui est égale à 0,24 Bark en débit rapide avec un schéma déclaratif). Cela signifie que la distance au niveau de F1 pour chacune des deux paires est une distance perceptible, mais elle n'est pas suffisante pour distinguer, toute seule, ces deux catégories de voyelles.

Le premier point de notre travail de recherche est réalisé. Effectivement, nous sommes parvenus à caractériser les 3 voyelles cardinales longues et les 3 voyelles cardinales brèves du dialecte arabe de la ville de Mayadin (Syrie) en termes de leurs durées et en termes de leurs deux premiers formants. Nous avons trouvé que les voyelles brèves, sans perturbation, forment trois sommets d'un triangle, lequel est inscrit dans le triangle dont les 3 voyelles longues forment les sommets. Nous avons attesté par les mesures appropriées que certains des six sommets des deux triangles, le grand et le petit, peuvent s'aligner, sans se superposer, figure 64, sous l'effet de la perturbation de certaines conditions expérimentales.

Outre la caractérisation du système phonologique des voyelles du dialecte, nous avons pour but de savoir si les voyelles brèves étaient des qualités acoustiques différentes des longues, ou si elles étaient des longues réalisées brèves.

Les brèves ayant leur propre triangle vocalique, lequel ne se confond pas avec celui des longues, elles observent alors une certaine distance par rapport aux longues. Dans des conditions normales de l'émission de la parole, *a priori* sans contrainte (débit normal et schéma mélodique déclaratif), toutes les distances observées contribuent à la démarcation perceptive (pour U/UU et A/AA) ou carrément catégorielle (pour I/II). Ainsi l'opposition Brève/Longue pour les 3 paires que nous avons examinées est basée sur deux contrastes, un contraste temporel et un second contraste, lequel est qualitatif. Ainsi, les perturbations engendrées par la variation du débit et du schéma ont révélé les réalités suivantes:

A/AA est une opposition basée au moins sur deux contrastes, temporel et spectral, le temporel est suffisant, le spectral n'est pas suffisant.

I/II est une opposition basée au moins sur deux contrastes, temporel et spectral, le temporel est suffisant, le spectral est suffisant aussi.

U/UU est une opposition basée au moins sur deux contrastes, temporel et spectral, le temporel est suffisant, le spectral n'est pas suffisant.

L'opposition qualitative A/AA est l'opposition la plus fragile parmi les trois paires étudiées, puisque le F2 de la brève de la longue fusionnent suite aux variations prosodiques et rythmiques.

L'opposition qualitative U/UU semble être fragile aussi, puisque les F2 des deux termes de l'opposition montrent un contraste vulnérable face aux variations prosodiques, mais pas face aux variations rythmiques. Le contraste qualitatif de l'opposition U/UU est moins important que le contraste qualitatif solide que nous avons observé pour I/II et plus important que le contraste qualitatif fragile observé pour l'opposition A/AA.

L'opposition qualitative I/II est certainement une opposition robuste résistant aux effets

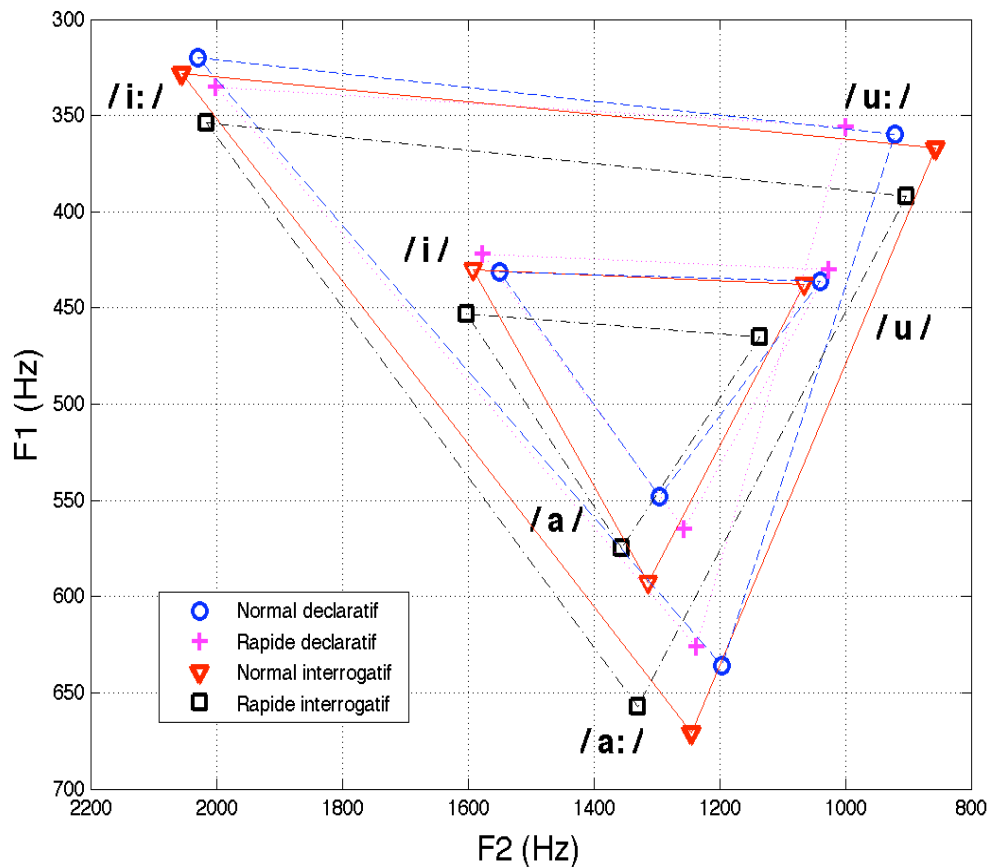


Figure 64. Moyennes des 3 voyelles brèves [i, a, u] et leurs correspondantes longues [i:, a:, u:] de l'arabe de Mayadin selon 4 conditions expérimentales, 2 schémas mélodiques (déclaratif et interrogatif) et 2 débits de parole (normal et rapide). 3 locuteurs masculins, 10 répétitions par locuteur et par voyelle.

perturbateurs des variations prosodiques et rythmiques, la distance spectrale séparant ces deux entités vocaliques est sûrement suffisante pour les démarquer catégoriellement.

Le triangle des voyelles brèves

Nous notons à ce niveau que les distances entre les 3 voyelles brèves avec un débit normal et un schéma déclaratif, calculées selon la formule proposée par Kewley-Port & Zheng (1999), sont 2 à 2 supérieures à 1 Bark.

Distance entre /a/ bref et /i/ bref = $((5,33-4,30)+(11,30-10,14))/2 = 1,20$

Distance entre /a/ bref et /u/ bref = $((5,33-4,35)+(10,14-8,76))/2 = 1,18$

Distance entre /i/ bref et /u/ bref = $((4,35-4,30)+(11,30-8,76))/2 = 1,30$

Ces distances restent supérieures à 1 Bark quelles que soient la condition expérimentale.

Au plan statistique nous trouvons que le F1 du /a/ bref est significativement ($p < 0,01$) différent de celui du /i/ et du /u/ brefs, les F1 de ces deux dernières voyelles ne sont pas différents entre eux ($p = ns$). Quant aux F2, ils sont deux à deux différents. Cette configuration statistique demeure vraie quelle que soit la condition expérimentale. Ces dernières données prouvent finalement que le triangle vocalique formé par les brèves peut être la base d'un système phonologique, puisque les voyelles de ce triangle sont deux à deux distinctes.

Nous avons pour objectif aussi dans cette étude de tester l'adaptabilité du système vocalique de l'arabe dialectal de Mayadin aux conditions naturelles de la communication linguistique en termes de réduction et de magnification de la distinctivité des oppositions sous les variations rythmiques et intonatives.

Les effets (adaptations) sur la durée et sur la qualité des voyelles confirment les prédictions sur la réduction et la magnification de la distinctivité. Le débit rapide de la parole s'est montré réducteur des durées des voyelles et, moins régulièrement, réducteur de l'espace vocalique. L'interrogation augmente les durées des voyelles et, moins régulièrement, magnifie les distances entre voyelles dans l'espace vocalique.

Sous l'effet du débit, les longues et les brèves se réduisent en durée par rapport à la condition optimale, soit en débit normal sous contour interrogatif, puisque dans cette dernière condition, les longues enregistrent leurs durées les plus longues. Sous cette dernière condition, elles voient leur espace vocalique formantique irrégulièrement se dilater (annexe n° VIII pour une illustration graphique de la réduction sous le débit et la magnification sous la

mélodie interrogative).

A ce propos deux remarques s'imposent par rapport à la théorie des systèmes vocaliques (en nous référant aux héritages de Lindblom et de Stevens, cumulés dans la TDF de Schwartz *et al.*, 1997b).

1. Le système de Mayadin, qui n'aurait pas besoin de différencier plus de trois qualités ([i a u] dans le cas de l'arabe classique) puisque les différences temporelles sont suffisantes pour démarquer l'identité de la brève de celle de la longue, et préserver ainsi leur distinctivité phonologique, réalise des exemplaires extrêmes pour les voyelles hautes ou basses. Ces dernières occupent ainsi les zones des voyelles dites focales: ce système ne présente donc pas de réduction générale vers une distinctivité "suffisante", il ne se réduit pas au triangle vocalique des brèves, significativement moins étendu que celui des longues (figure 64), même si cette réduction est conforme au principe de l'économie des gestes articulatoires. Nous rappelons que ces trois voyelles brèves sont, deux à deux, séparées par des distances supérieures 1 Bark, une valeur largement supérieure à la valeur estimée (0,60 Bark) du seuil catégoriel de la qualité vocalique.

2. La voyelle brève, en opposition avec [i] long, a fini par occuper la zone de [e] (phénomène courant en arabe dialectal). Et elle s'y maintient remarquablement, ne se confondant jamais en qualité, quelles que soient les perturbations, avec sa partenaire longue. Ces deux observations seront à considérer dans le cadre de la théorie des systèmes vocaliques, par rapport aux systèmes réduits des langues d'Australie, dont les formants ont été mesurés depuis quelques années par Andrew Butcher et son équipe. Ce point sera discuté dans la conclusion de ce manuscrit.

Alors, que nous renseignent ces derniers résultats sur les réorganisations des systèmes vocaliques sous les perturbations? En fait, cela laisse penser que nous sommes en train d'assister à deux types de changements: (1) accompli et (2) non accompli.

Accompli: après avoir occupé la périphérie de leurs espaces vocaliques, les langues tendent majoritairement à occuper le centre par de nouvelles qualités vocaliques (Vallée, 1994, Schwartz *et al.*, 1997b) conformément à la TDF. Le /i/ bref, qui est le produit d'une réduction vocalique du /i:/ long, s'est démarqué par une distance importante de ce dernier, se réalisant plutôt [e]. Cette distance s'est montrée hautement robuste, puisque le /i/ bref ne redevient jamais une voyelle de qualité [i] quel que soit le type de perturbation, cette réalisation vocalique s'impose alors comme une voyelle indépendante. Cela nous permet

d'avancer l'idée que le /i/ bref devenant [e] est le signe d'un changement phonétique accompli.

Non accompli: le /a/ bref qui est le produit de la réduction vocalique du /a:/ long, observe une certaine distance par rapport à ce dernier. Cette distance non importante, au sens perceptif du terme, s'est montrée non robuste, ne résistant pas aux variations rythmiques et prosodiques. Cette observation nous permet de penser que le /a/ bref tend à avoir une distance qualitative suffisante et suffisamment robuste qui lui garantit, comme le /i/ bref vis-à-vis du /i:/ long, de se maintenir qualitativement différent et ne pas redevenir une réalisation vocalique appartenant à la qualité /a/ sous certaines conditions perturbatrices de la parole. Le /a/ bref tendant à se réaliser avec une qualité [œ] mais restant pratiquement de qualité [a] est le signe d'un changement phonétique non accompli à venir.

De la mesure à la manipulation.

Nous avons réussi jusqu'ici à mesurer toutes les différences temporelles et spectrales entre chaque voyelle brève et sa correspondante longue. Pour désigner la source de l'information sur la catégorie de la voyelle, brève ou longue, nous nous sommes contentés d'effectuer des mesures sous 4 conditions expérimentales de la parole naturelle, en considérant que les indices acoustiques spécifiés phonologiquement doivent être constants à travers tous les types de perturbations.

Le /a/ bref se différencie du /a:/ long par la durée, une différence suffisante pour démarquer leurs identités, le /i/ bref se différencie du /i:/ long par la durée et la qualité, alors que l'une des deux suffit (vu avec le /a/). Si la différence qualitative est suffisante, alors l'absence de la différence temporelle entre les deux voyelles n'entraînera pas leur fusion perceptive. Si les différences, temporelle et spectrale, entre /i/ et /i:/ se maintiennent sous les perturbations, selon le principe des perturbations prioritaires, les deux informations sont alors à valeur phonologique dans le système de l'arabe de Mayadin. Cette redondance dans les indices nous a poussé à penser que même si au niveau de la production nous assistons à deux indices, il est possible qu'au niveau de la perception il y en ait qu'un seul. Voilà une limite pour les seuils qui se cantonnent à désigner ce qui est perceptible au sens phonétique physique du terme, laissant la question, au niveau linguistique, entière. En outre, si l'absence de l'un des deux indices n'affecte pas la perception de la catégorie de la voyelle, alors ce résultat constituera une limite pour la méthode des perturbations prioritaires qui se veulent un recours pour signaler ce qui est à valeur phonologique dans un système phonétique donné,

puisqu'elles auraient désigné un des indices comme indispensable à la communication linguistique en lui accordant une valeur phonologique, une valeur qu'il n'a pas.

C'est ainsi que, dans le chapitre suivant (n° IV), nous effectuerons une manipulation qui consiste à allonger les voyelles brèves et à raccourcir les voyelles longues afin de savoir lequel des deux indices, temporel *vs.* spectral, est à l'origine du contraste catégoriel: Brève *vs.* Longue. Cette manipulation nous permettra de savoir enfin si vraiment les deux indices, temporel et spectral, sont tous les 2 indispensables au contraste: /i/ bref *vs.* /i:/ long, et de savoir aussi s'il existe d'autres corrélats acoustiques (à part les deux corrélats classiques: temporel et spectral) qui participent à la démarcation catégorielle Brève *vs.* Longue.

IV. Le poids perceptif de la durée de la voyelle face à sa qualité acoustique dans la discrimination des voyelles longues et brèves

IV.1. Introduction

Nous sommes effectivement parvenus dans le chapitre précédent à caractériser les oppositions entre la catégorie des voyelles brèves et celle des voyelles longues sur le plan temporel et sur le plan spectral. Sans aucune contrainte, chacune des voyelles brèves est démarquée de sa partenaire longue par des différences temporelles et spectrales. Les perturbations ont montré que les différences temporelles pour les trois oppositions I/II, A/AA et U/UU sont robustes, que la différence spectrale pour I/II est robuste, qu'elle est *semi* robuste pour U/UU et pour A/AA (robuste au niveau de F1 seulement). Ces résultats indiquent-ils que la différence spectrale pour l'opposition I/II est capable toute seule de supporter le contraste entre les deux catégories? Ainsi la disparition de la différence temporelle ne devrait pas affecter la perception de ce contraste. La différence spectrale pour l'opposition U/UU sera non suffisante pour maintenir le contraste sans que la différence temporelle n'intervienne, cette opposition sera par conséquent sensible à la disparition de la différence temporelle entre /u/ et /u:/. Finalement, l'opposition A/AA, qui compte, comme l'opposition U/UU, essentiellement sur la différence de durée entre les deux termes de la paire brève/longue, ne pourra pas se maintenir en l'absence de cette différence temporelle.

Nous rappelons que les études qui se sont intéressées à la mesure de la durée en arabe ont montré une différence temporelle importante entre la voyelle brève et sa partenaire longue. Barkat-Defradas (2003) montre que le rapport longue/brève est de 1,9 pour l'arabe du Maghreb, alors qu'il est de 2,3 pour les régions du Moyen-Orient. Nous avons trouvé (chapitre III), que ce rapport était compris entre 1,6 et 1,7 (suivant le débit et le schéma mélodique) pour cette localité de l'Euphrate en Syrie. Les études portant sur l'espace vocalique conviennent que les voyelles brèves se situent dans une zone centrale de l'espace vocalique par rapport aux voyelles longues. Abou-Haidar (1994) avait établi l'espace vocalique de 8 dialectes du monde arabe; l'auteur confirme la tendance des voyelles brèves à la centralisation pour tous les dialectes qui ont fait l'objet de son étude. Les résultats de Newman (2002) sur l'arabe du Caire sont en accord avec les résultats d'Abou-Haidar (1994). Nos données sur l'arabe dialectal de Mayadin vont aussi dans le même sens que les travaux d'Abou-Haidar

(1994) et de Newman (2002) sur l'arabe du Caire.

Cette tendance à la centralisation des voyelles brèves a été démontrée pour d'autres langues que l'arabe. En effet, Fletcher & Butcher (2003) ont étudié les effets de la structure prosodique dans la mise en avant des contrastes paradigmatiques pour les voyelles de trois langues des aborigènes du nord de l'Australie (Kayardild, Dalabon et Mayali). Ils affirment que la langue Kayardild dispose d'un contraste quantitatif, ils montrent ainsi que dans l'espace vocalique de cette langue se dessinent deux triangles vocaliques, un pour les voyelles longues et un autre se trouvant dans le centre de l'espace vocalique pour les voyelles brèves.

Nenonen *et al.* (2003) ont étudié l'acquisition de l'opposition temporelle entre voyelles longues et voyelles brèves en finnois, leur étude a montré que longues et brèves avaient les mêmes propriétés distributionnelles, longues et brèves pouvant se rencontrer dans le même contexte phonétique. Ils soutiennent que la durée de la voyelle est le facteur acoustique le plus saillant dans l'opposition de la quantité vocalique en finnois, même si des différences au niveau spectral entre les longues et les brèves existent. En effet, et sans préciser l'étendue de ces différences, ils affirment que les brèves sont (un peu) plus centrales que les longues. La question que nous posons est à quel point cette centralité (cette distance entre brèves et longues dans l'espace vocalique) joue un rôle dans le contraste Brève vs. Longue.

La question du trait phonétique dominant, voire seul en jeu, des deux indices temporel vs. spectral, autrement dit "quantitatif vs. qualitatif", reste cependant mal évaluée dans les études qui ont décrit les caractéristiques du système phonétique de l'arabe citées plus haut; les distances perceptives entre les voyelles dans les espaces vocaliques considérés ne sont pas étudiées. Ainsi, aux voyelles brèves, les phonéticiens assignaient systématiquement la qualité de la longue, en considérant que la différence se situe uniquement au niveau de la durée. Cependant Alioua (2003) affirme que la différence de timbres entre longues et brèves est assez importante pour être perceptible pour les trois oppositions I/II, A/AA et U/UU en arabe marocain. L'auteur ne donne pas d'avantage de détails.

Dans ce paragraphe, et afin de vérifier les résultats de notre chapitre précédent, nous procédons à une perturbation expérimentale, qui consiste en l'élimination de la différence temporelle entre les deux termes de la paire brève/longue, ne laissant ainsi que la différence spectrale en jeu. En effet, nous procéderons à l'allongement des voyelles brèves et au raccourcissement des voyelles longues. Nous observerons ensuite les réactions de 10 sujets natifs du dialecte de Mayadin en ce qui concerne la perception de l'opposition Brève/Longue dans les signaux manipulés. Il s'agit d'une réplique d'une expérience informelle (annexe n°

IX) que nous avons effectuée de manière préliminaire afin de tester la portée de cette procédure expérimentale consistant en la manipulation de la durée du signal de parole, et dont les résultats se sont révélés prometteurs. Cette perturbation est non naturelle, par opposition au débit rapide de la parole et au schéma mélodique interrogatif qui sont des perturbations qui relèvent du domaine de la parole naturelle.

Cette question de trait phonétique dominant: temporel *vs.* spectral, a été discutée pour d'autres langues que l'arabe, la plupart rapportant un rôle déterminant de la durée dans ce contraste vis-à-vis de la différence spectrale associée. Schwartz *et al.* (1997a), en se basant sur l'inventaire de la base de données UPSID, affirment la présence de systèmes vocaliques combinant des différences temporelles et spectrales, un tel système est appelé alors "unique" par opposition à un système dit "double" où il n'y a que des différences temporelles qui sont à l'origine de l'opposition Bref/Long.

Dans une étude perceptive avec des voyelles synthétisées, Ainsworth (1972) a montré qu'en anglais britannique la durée affecte l'identité de la voyelle. Il affirme que certaines voyelles, notamment celles du centre de l'espace vocalique, qui sont proches les unes des autres en termes de F1, F2, comme /ɜ/ - /ʌ/ et /o/ - /ɔ, ɑ /, peuvent être confondues entre elles si le jugement de l'identité de la voyelle n'est basé que sur les informations acoustiques spectrales. Mermelstein (1978) avait montré qu'en anglais les deux voyelles /ɛ/ et /æ/, insérées dans des syllabes synthétisées [bVd] et [bVt], en raison de leurs durées très différentes, alors qu'elles ont des qualités acoustiques proches, sont identifiées en fonction de leurs durées. La longue est perçue /æ/ alors que la brève est perçue /ɛ/. Pour la différence de quantité, phonologique en japonais, Tajima *et al.* (2003) affirment que le premier corrélat acoustique pertinent du contraste quantitatif est la différence temporelle, et ceci pour les voyelles et les consonnes. Tajima *et al.* (2003) accordent ainsi peu de considération au rôle de la différence au niveau spectral.

Le problème est aussi bien posé pour le français où la quantité vocalique n'est pas un procédé phonétique régulier, et spécifié linguistiquement. Contrairement aux études précédentes (Ainsworth, 1972; Mermelstein, 1978; Tajima *et al.*, 2003), les locuteurs du français comptent essentiellement sur l'information spectrale, même si les deux informations, temporelle et spectrale, sont présentes dans l'opposition (V/V:). Effectivement, Gottfried & Beddor (1988) ont mené une étude qui a examiné l'impact de l'information temporelle *vs.* spectrale dans la discrimination des deux voyelles de la paire /o/ - /ɔ/. Les auteurs de l'étude rapportent que la mesure des durées et des valeurs des trois premiers formants de ces 2

voyelles dans la parole naturelle confirment la présence d'une différence aux niveaux temporel spectral:

	Durée (ms)	F1	F2	F3
/o/	111	438	1160	2608
/ɔ/	85	541	1512	2654

Nos mesures (chapitre V) du français de la région de la Savoie auprès d'une génération septuagénaire vont dans le même sens que l'affirmation de Gottfried & Beddor (1988), puisqu'elles montrent que ces deux voyelles présentent une différence temporelle combinée avec une différence spectrale (cf. V.3.1 pour la durée et V.3.2. pour la qualité).

Lorsque ces deux voyelles sont insérées dans un contexte consonantique [kVt], elles produisent une paire minimale [kot] (côte) vs. [kɔt] (cotte), basée sur la durée de la voyelle noyau de la syllabe. Ces deux voyelles présentent des différences et au niveau quantitatif et au niveau qualitatif. Des études précédentes (Gottfried, 1984) ont montré que la longue des deux /o/ est 25 % environ plus longue que la brève /ɔ/ dans un contexte [tVt]. Une telle différence temporelle, selon nos critères développés en chapitre I, et selon les critères de Rossi (1972), concernant le seuil différentiel de la perception de la durée vocalique, est assez importante pour constituer la base d'une opposition quantitative. Gottfried & Beddor (1988) ont élaboré une expérience qui a consisté à modifier, de manière synthétique, le F1 et le F2 de /o/ en 10 pas pour arriver au F1 et au F2 de la voyelle /ɔ/. Chacun des 10 stimuli était émis avec 3 durées différentes de 220, 180 et 140 ms. Les résultats de leurs tests perceptifs pour des natifs du français montrent que la durée n'est pas utilisée comme un indice pour identifier la voyelle, les sujets sont insensibles à l'information temporelle. Lorsque cette même expérience était conduite auprès des natifs de l'anglais des États-Unis, la catégorisation de ces deux voyelles auprès de ces sujets était influencée par la durée de la voyelle du français. Cela illustre deux importants points: (i) l'insensibilité chez les sujets natifs du français n'est pas due à la faiblesse de la différence temporelle entre les deux termes de la paire, puisque cette différence est perçue par des sujets d'une autre langue; (ii) la sensibilité des locuteurs natifs de l'anglais des États-Unis est due au fait que le contraste temporel est phonologiquement spécifié en anglais. En effet, lorsque l'information spectrale devient ambiguë, lors du passage de /o/ à /ɔ/ (les valeurs formantiques pouvant se situer dans une zone intermédiaire entre les deux qualités), alors les locuteurs natifs de l'anglais américain, ceux qui connaissent le français

comme ceux qui ne le connaissent pas, identifient la plus longue des deux comme /o/ et la plus brève en tant que /ɔ/.

L'importance de l'indice temporel pour l'identification de la voyelle ne varie pas seulement d'une langue à une autre comme entre le français et l'anglais, mais cet indice peut varier au sein d'une même langue. Miller & Grosjean (1997) affirment que la durée de la voyelle peut jouer un rôle concernant l'identité de la voyelle en français en Suisse (nous verrons ce qu'il en est pour une région voisine, la Savoie, cf. chapitre V), alors que ce n'est pas le cas pour le français standard de France.

La description des systèmes phonétiques a besoin de savoir quel est le trait perçu et qui est à l'origine du contraste Brève vs. Longue. En effet, dans les langues (ou les parlers) où l'opposition quantitative n'est pas systématique et n'est pas évidente, classer une langue (ou un parler) étudiée comme une langue à quantité vocalique dépend du poids perceptif que l'opposition temporelle possède vis-à-vis des autres éventuels indices prosodiques ou qualitatifs associés. En fait, Uguzzoni *et al.* (2003) ne sont pas parvenus à décrire les langues Frignanaises comme des langues qui font usage de la quantité vocalique, malgré les différences apparentes concernant les durées des voyelles, parce qu'ils n'ont pas réussi à déterminer la source de l'information dans l'opposition entre les voyelles brèves et longues. Nous présentons ci-dessous leur travail sur des variétés italo-romaines. Le travail de Uguzzoni *et al.* (2003) a porté sur quelques langues européennes où le contraste temporel entre voyelles peut jouer un rôle phonologique. Les auteurs font cependant la différence entre deux types de langues à quantité vocalique: libre vs. combinatoire. Selon Uguzzoni *et al.* dans des langues comme le finlandais et le tchèque, les contrastes quantitatifs sont indépendants de l'accent, les voyelles longues et brèves se produisant aussi bien dans des syllabes accentuées que dans des syllabes non accentuées. Alors que dans des langues, comme l'allemand standard et le hollandais, la quantité vocalique est liée à la place de l'accent dans le mot. En effet, l'opposition entre voyelles longues et brèves ne se rencontre que dans des syllabes accentuées. Les deux groupes de langues ont des différences concernant certaines de leurs propriétés distributionnelles (les occurrences des voyelles). Uguzzoni *et al.* (2003) ont remarqué que dans le premier groupe de langues, les longues et les brèves peuvent se rencontrer dans une syllabe accentuée ouverte, alors que les deux autres langues, l'allemand standard et le hollandais, sont caractérisées par une contrainte particulière qui, dans les syllabes accentuées ouvertes, ne permet que l'occurrence des voyelles longues, et non pas des brèves. Cette contrainte n'est pas la seule pour ces deux dernières. Les auteurs soulignent que

les contrastes entre les durées des voyelles ne se produisent que sous l'accent du mot, lequel est dynamique par nature. Les voyelles brèves accentuées ne sont pas permises dans les syllabes ouvertes, mais dans les fermées. Une seule consonne intervocalique, quand elle est précédée par une voyelle brève accentuée, est toujours ambisyllabique.

L'interprétation des contrastes entre les durées des voyelles dépend finalement du type de la langue considérée: en finlandais et en tchèque, il est interprété en tant que contraste de quantité, et en allemand standard et en hollandais, en tant que contraste de coupe syllabique.

Uguzzoni *et al.* (2003) ont mené une enquête sur des variétés italo-romaines parlées dans la région de Frignano, dans le nord de l'Italie, afin de savoir si elles sont des parlers à quantité vocalique. Selon les descriptions classiques, le contraste de durée n'est pas phonologique dans les langues romanes.

Les résultats des recherches phonétiques et phonologiques ne sont pas en accord avec ce qui était traditionnellement avancé. Uguzzoni & Busà (1995) trouvent un inventaire de voyelles assez important dans les dialectes de la zone Frignanese. En syllabes accentuées, ils présentent 13 voyelles, 9 longues [i:, y:, u:, e:, ø:, o:, ε:, ɔ:, a:] et 4 brèves [e, ø, ɔ, a]. Le contraste de quantité se présente dans ces dialectes à la fois en syllabe finale et en syllabe non finale. En effet, les résultats des mesures ont montré que la durée de la brève constituait 49,28% de celle de la longue. Uguzzoni *et al.* (2003) ont remarqué que ce contraste de durées de voyelles était accompagné d'autres facteurs acoustiques, telles les qualités vocaliques et les durées des consonnes *post-vocaliques*. Par exemple, dans des paires minimales telles que [pela] vs. [pe:la] et [bɔt] vs. [bɔ:t], les voyelles présentaient une différence au niveau de la qualité et une différence au niveau de la durée. En comparant les voyelles en termes de leurs F1, F2, et F3, ils ont trouvé les comportements suivants: le /e/ bref est plus bas et moins antérieur que le /e:/ long; le /ø/ bref est plus bas et moins arrondi que le /ø:/ long; le /ɔ/ est plus bas, plus en avant et plus arrondi que le /ɔ:/; le /a/ est plus haut et plus antérieur que le /a:/. Les auteurs de l'étude pensent, sans présenter des mesures spectrales dans leur étude sur les longues et les brèves, que les différences de qualités entre longues et brèves jouent un rôle moins important que celui que jouent les différences de durées. En ce qui concerne les durées des consonnes en contact avec ces voyelles, les auteurs rapportent que la consonne précédée par une voyelle brève donnait l'impression d'être plus longue que la consonne précédée par une voyelle longue. Les résultats expérimentaux des mesures de la durée des consonnes qui se rencontraient après une voyelle longue et ceux qui se rencontraient après une voyelle brève avaient sensiblement la même durée (/CVCV/ vs. /CV:CV/). Ce n'était que dans les mots

monosyllabiques (/CVC/ vs. /CV:C/) que les durées des consonnes étaient un peu plus importantes après une brève qu'après une longue (le rapport était de 81,29%). Au final, les chercheurs, après avoir éliminé l'effet de la consonne post-vocalique, ne parviennent que partiellement à déterminer la principale source de l'information dans ce contraste.

Ces résultats n'ont pas permis aux auteurs de l'étude d'assigner les variétés italo-romaines à la catégorie des langues à quantité vocalique. Les auteurs voulaient simplement souligner que les dialectes Frignanese pourraient appartenir à la catégorie des langues à quantité vocalique. Mais le rôle de l'accent est un facteur incontestable dont dépendent étroitement les contrastes des durées des voyelles en Frignanese.

La recherche de la source de l'information dans l'opposition Brève/Longue a motivé bien d'autres études. Une des langues pour laquelle la question Durée vs. Spectre a été minutieusement étudiée, est le Twi (parlé au Ghana). Les oppositions de longueur des voyelles en Twi sont spécifiées linguistiquement (Adu Manyah, 2003), cette spécification se manifeste dans plusieurs domaines: 1) lexicale: opposer deux mots de même catégorie (adjectif vs. adjectif, par exemple), ou de 2 catégories différentes (substantif vs. verbe). 2) morphologique: déclinaison verbale (conjugaison). Adu Manyah (2001) a procédé à des analyses temporelles et spectrales pour certaines des voyelles qui présentaient ce contraste. En 2003, Adu Manyah a effectué un travail de recherche similaire, mais plus étendu, ayant toujours pour objectif la détermination des éléments phonétiques qui sont à l'origine de cette différence. La méthode a consisté en l'enregistrement de paires minimales construites à partir des voyelles longues et brèves par deux locuteurs. Le Twi est une langue qui compte 9 voyelles orales, l'auteur de l'étude affirme cependant que la disposition du lexique du Twi ne permettait pas la construction d'une paire minimale pour les 2 /e/, bref et long. L'étude a porté par conséquent sur 8 paires de voyelles: /i/ vs. /i:/, /ɪ/ vs. /ɪ:/, /ɛ/ vs. /ɛ:/, /a/ vs. /a:/, /ɔ/ vs. /ɔ:/, /o/ vs. /o:/, /ʊ/ vs. /ʊ:/ et /u/ vs. /u:/. Le débit d'élocution était laissé au choix des locuteurs, l'environnement consonantique était la syllabe [pVk]. Chaque item était répété 10 fois dans un ordre aléatoire.

Les analyses des qualités acoustiques et des durées des voyelles longues et de leurs correspondantes brèves par Adu Manyah (2003) ont montré que le contraste entre les deux catégories était plutôt basé sur l'opposition temporelle que sur l'opposition spectrale. En effet, la différence de durée était significative ($p < 0,01$). L'analyse acoustique des mesures

effectuées sur les trois premiers formants des longues et des brèves a montré que la différence des valeurs formantiques n'était pas significative ($p = ns$), (excepté pour la paire /u/ vs. /u:/, où les différences temporelles et spectrales étaient significatives), par conséquent, la durée seule était à l'origine de ce contraste distinctif Brève/Longue.

Cette étude a mis en évidence le fait que la durée est le paramètre privilégié dans l'établissement du contraste entre les deux catégories de voyelles en Twi, parce que de faibles différences qualitatives ont été néanmoins observées, figure 65.

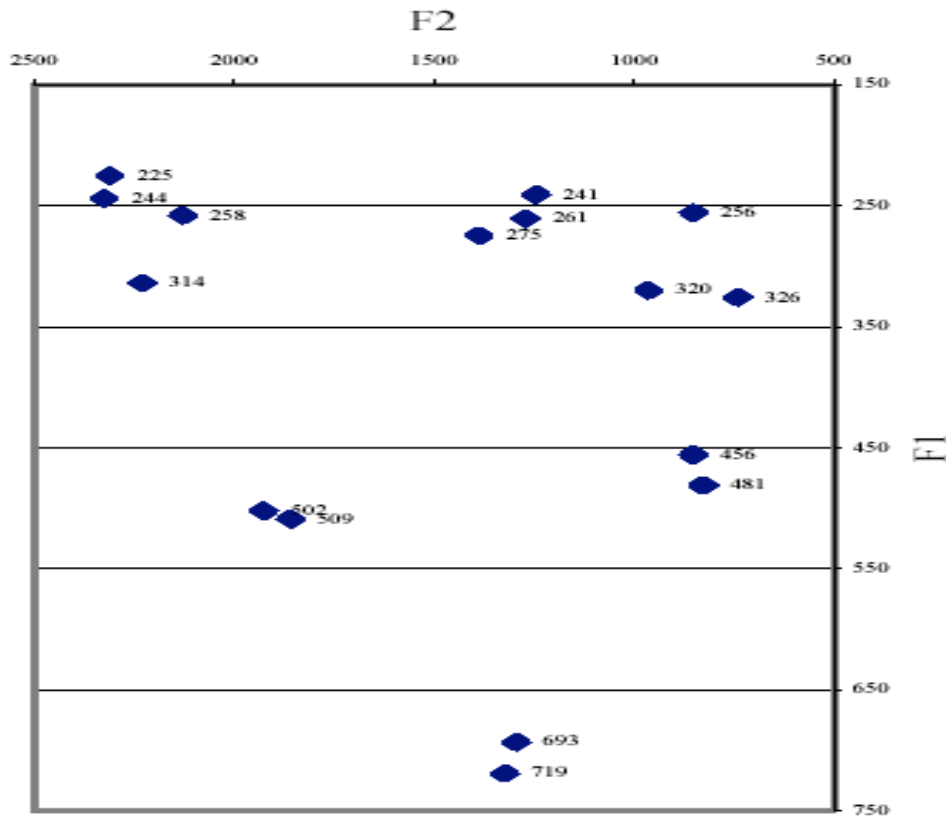


Figure 65. Voyelles longues et brèves du Twi de Ghana [i, ii, I, II, E, EE, a, aa, ɔ, ɔɔ, o, oo, ʊ, ʊʊ, u et uu], en fonction de F1 et de F2, 2 locuteurs natifs. D'après Manyah (2003).

La partie de l'expérience de l'étude d'Adu Manyah (2003), qui a porté sur l'opposition des durées des voyelles brèves et longues d'un côté, et l'opposition des syllabes porteuses VC/V:C de l'autre en Twi, a montré que cette opposition était supportée par la différence des durées des réalisations, et non pas par les différences qualitatives de timbres. L'expérience a montré que la durée de la voyelle n'était pas le seul facteur à l'origine de cette distinction, la complémentarité syllabique jouait, elle aussi, un rôle. En effet, les voyelles brèves étaient suivies d'une consonne longue et les voyelles longues étaient suivies d'une consonne brève,

ce qui suggère fortement que la durée de la consonne post-vocalique, tributaire de la durée de la voyelle, contribue à la distinction des deux termes de la paire Brève/Longue. En effet, les analyses statistiques ont montré que la différence de durée entre voyelles brèves et longues était significative, mais aussi la différence de durée entre les consonnes précédées par des voyelles longues et les consonnes précédées par des voyelles brèves était significative ($p < 0,01$). Toutes voyelles et tous locuteurs confondus, le rapport Longue/Brève = 2,4 (figure 66). La durée de la syllabe est plus ou moins la même quelle que soit la durée de la voyelle, elle est de l'ordre de 230 ms, la brève constitue 20% de la durée de la syllabe, alors que la longue constitue 50%. Pour les voyelles nasales, les brèves constituent 27% environ de la durée de la syllabe, et les longues 56%. Les voyelles nasales sont systématiquement plus longues que leurs correspondantes orales.

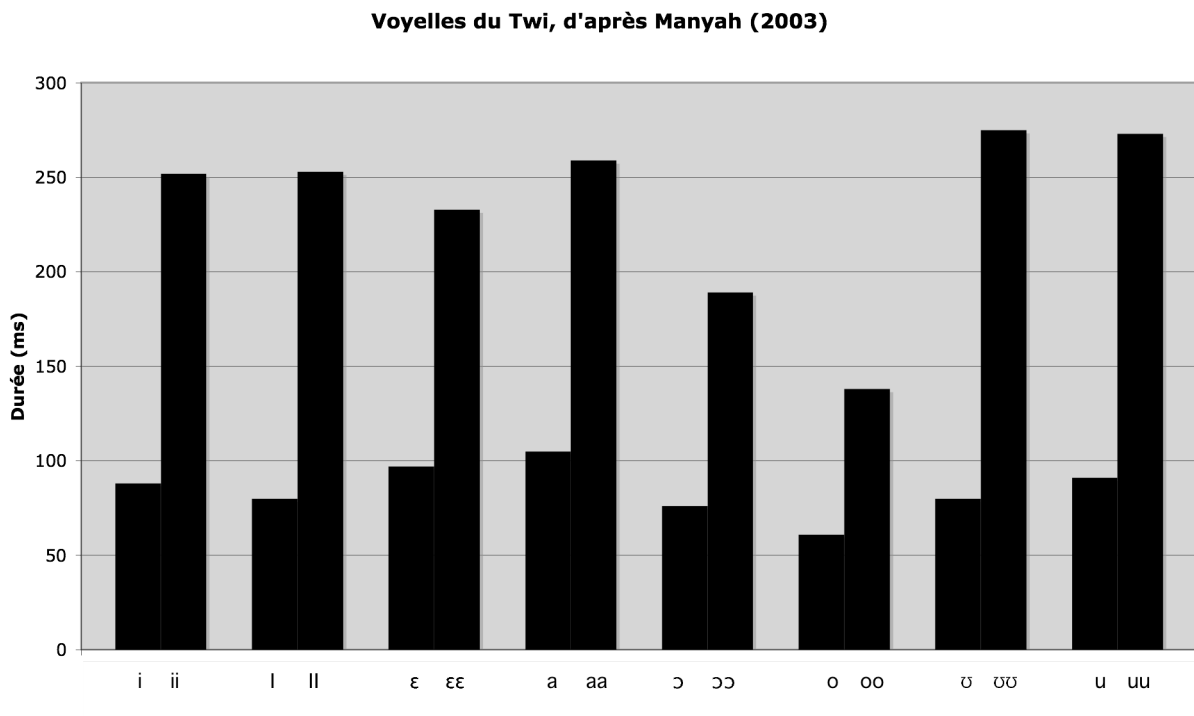


Figure 66. Durée des voyelles longues et brèves du Twi de Ghana [i, ii, l, ll, ε, εε, a, aa, ɔ, ɔɔ, o, oo, ɔ̃, ɔ̃ɔ̃, u et uu]. D'après Adu Manyah (2003).

Finalement, les résultats des études acoustiques mettent en évidence le rôle déterminant que joue la durée de la voyelle dans la création de la différence entre les deux catégories de voyelles, brèves et longues, comme ils mettent en évidence un rôle significatif de la durée de la consonne post-vocalique comme un indice acoustique sur la durée intrinsèque de la voyelle qui la précède.

D'autres études ont cherché à savoir si le trait temporel était effectivement le seul indice acoustique sur le contraste entre les deux catégories Brève/Longue. Dans ce but, elles ont essayé de contrôler les caractéristiques acoustiques en neutralisant certaines et en ne laissant que certaines en jeu afin de juger de leur poids perceptif dans ce type de contraste. Abramson (2003) avait suspecté le rôle de la différence temporelle comme étant le seul responsable de la démarcation des deux catégories de consonnes Brève/Longue. L'étude d'Abramson (2003) sur les consonnes en Pattani Malay, une langue parlée dans le sud de la Thaïlande, mentionne que les consonnes qui sont dans des syllabes initiales peuvent présenter des contrastes phonologiques de durée. La différence de durée en Pattani Malay, pour les consonnes en position initiale, est audible quels que soient le lieu et le mode d'articulation de la consonne, ce qui signifie que même les consonnes occlusives sont contrastées par leurs durées. La discrimination des consonnes brèves et longues à l'initiale des mots se faisant avec un taux qui est bien au-delà de la contribution du hasard, Abramson pose la question sur la durée de l'occlusion pour savoir si elle est le seul facteur acoustique qui est derrière ce contraste bref/long, ou s'il y a bien d'autres facteurs.

En toute logique, la différence de durée de l'occlusion ou de la constriction, ne peut être entendue que lorsqu'il y a une excitation acoustique audible qui accompagne l'occlusion ou la constriction. En position initiale, les sons occlusifs et les mi-occlusifs sourds sont dépourvus d'une excitation acoustique et ils ne peuvent être perçus comme plus longs ou moins longs que lorsqu'ils sont précédés par une voyelle. La question que l'auteur pose alors: comment les locuteurs arrivent-ils à évaluer les durées des occlusifs quand ils sont en position initiale? Parce qu'il était effectivement démontré (Abramson, 1987) que l'occlusion était plus longue pour les consonnes longues que pour les consonnes brèves. Dans une étude du même auteur en 1986, l'efficacité de la durée de l'occlusion était testée par l'allongement de l'occlusion des consonnes brèves et par le raccourcissement des durées des consonnes longues; dans les 2 sens, le changement de la durée s'est avéré efficace. L'intensité est, par ailleurs, plus importante sur les premières syllabes dans les mots qui avaient des consonnes longues dans leurs premières syllabes. D'autres expériences (Abramson, 1987) ont montré que l'interaction: durée de l'occlusion X intensité était significative, mais l'intensité toute seule n'était pas assez imposante pour permettre le passage d'une catégorie à une autre.

Dans une autre expérience, Abramson (1999), la F0 a été augmentée sur la première syllabe des mots avec des consonnes brèves et abaissée sur les consonnes longues des autres

mots des paires minimales contrastées par la durée de la consonne à l'initiale absolue. Les effets étaient très significatifs, par contre le changement de la F0 n'était pas assez important pour permettre, tout seul, le passage d'une catégorie à une autre (passage de la brève à la longue).

Dans une expérience ultérieure, Abramson (2003) a cherché à établir un lien entre l'accentuation et le contraste de durée pour des stimuli où la F0 et l'intensité co-variaient d'une manière élaborée. Pour chaque mot de chaque paire, il y avait 35 stimuli, le mot original plus 34 stimuli; donc pour 2 mots il y avait 70 stimuli.

La conception de cette expérience était basée sur le contrôle des caractères les plus pertinents des syllabes pour des paires minimales CVCV *vs.* CCVCV. Le changement affectait seulement les indices de la proéminence de la syllabe, notamment la F0 et l'intensité, alors que les autres indices, notamment la durée de l'occlusion, restaient inchangés. Le but étant de détecter les facteurs qui sont à l'origine du contraste bref/long. Pour les mots avec des consonnes brèves en syllabe initiale, la proéminence de la première syllabe était augmentée, alors que celle de la deuxième était réduite. Pour les mots avec des consonnes longues, la proéminence de la première syllabe était diminuée alors que celle de la deuxième était augmentée. L'auteur s'intéressait aux 2 indices acoustiques les plus prometteurs, la F0 et l'intensité sur la première et la deuxième syllabe. Donc, 4 syllabes, 2 paramètres pour chaque syllabe et 3 degrés de changements pour chaque paramètre, avec un programme qui permettait d'alterner les propriétés prosodiques par rapport à la phrase de base, le changement était de plus en plus important, il était nul pour le premier stimulus et maximal pour le 35^e stimulus. Les 70 stimuli ont été passés en 6 tests différents.

Pour la consonne brève /k/ dans le mot [kito], la consonne était identifiée à 100% brève avant les changements progressifs, au fur et à mesure que l'intensité et la F0 augmentaient sur la première syllabe et baissaient sur la deuxième, du même mot, la brève passait pour une longue /k:/, jusqu'au 35^e changement où la brève était identifiée à 100% en tant que sa partenaire longue. Alors que pour la longue, quand le chercheur lui attribuait petit à petit les caractéristiques acoustiques (essentiellement la F0 et l'intensité) de la brève, elle montrait le même comportement observé auprès de la brève quand le chercheur lui attribuait graduellement les caractéristiques de la longue, mais en moins robuste, c'est-à-dire que la longue était perçue en tant que la brève au fur et à mesure que l'auteur de l'étude lui attribuait la F0 et l'intensité de la brève. Les autres paires de consonnes /p/ *vs.* /p:/, /t/ *vs.* /t:/, /d/ *vs.* /d:/, /m/ *vs.* /m:/, /s/ *vs.* /s:/ présentaient des comportements similaires à ceux du /k/, mais ils

étaient moins robustes et moins bien marqués.

Les résultats mettent en évidence l'existence d'un corrélat évident entre le changement de la perception de la catégorie (brève *vs.* longue) de la consonne occlusive sourde à l'initiale de la première syllabe et le changement des deux facteurs prosodiques Intensité et F0. Les résultats montrent que l'intensité et la F0 sont les seuls facteurs qui permettent la perception de la longueur, et non pas la durée de l'occlusion. L'auteur de l'étude pense qu'il est possible que la durée, en tant que telle, disparaisse à long terme. Abramson (2003) affirme que lorsque ces consonnes sont en position post-vocalique, la durée de l'occlusion est un indice très fort. Pour les constrictives et les occlusives voisées, les termes des paires minimales sont essentiellement différenciés par leurs durées.

Cette étude nous permet de conclure qu'une différence de durée peut exister sans qu'elle constitue (toute seule) un indice acoustique pertinent, permettant la démarcation catégorielle des phonèmes, mais elle est accompagnée par d'autres indices qualitatifs ou prosodiques, qui sont eux-mêmes porteurs de l'information sur le contraste bref/long.

Les études d'Adu Manyah (2001, 2003) et d'Abramson (1987, 1999, 2003) nous ont bien montré que les informations sur l'opposition quantitative peuvent être véhiculées par des indices qui existent en dehors de la durée la voyelle. Ici, nous exécutons un procédé expérimental, "la neutralisation de trait", que nous effectuons dans le but de déterminer la source de l'information (temporelle *vs.* spectrale) dans l'opposition entre voyelles longues et brèves en arabe de Mayadin.

IV.2. manipulation instrumentale

IV.2.1. Introduction

Dans cette expérience perceptive, nous éliminerons la différence temporelle entre la voyelle longue et sa correspondante brève afin de neutraliser le rôle de l'opposition temporelle dans le contraste Brève *vs.* Longue, ce qui nous permet de déterminer par la suite la contribution de chacune des informations, temporelle *vs.* spectrale. Nous augmentons les durées des voyelles brèves et nous réduisons les durées des voyelles longues. Nous cherchons ainsi à valider nos résultats statistiques et perceptifs concernant les trois paires de voyelles étudiées. Ces résultats ont montré que la différence spectrale pour I/II est robuste et par

conséquent suffisante pour créer le contraste Brève vs. Longue, elle est *semi* robuste pour U/UU et pour A/AA, donc ces 2 dernières oppositions sont tributaires de la différence temporelle.

Des études précédentes ont argumenté en faveur d'un espace vocalique non uniforme, Polka & Bohn (2003) croient à l'asymétrie dans la perception des voyelles (cf. I.1.3.2.1.). Ils pensent que si l'espace perceptif est supposé euclidien, et que cette supposition continue à être vraie, c'est tout simplement parce qu'il n'y a pas eu d'étude perceptive qui montre le contraire.

Les deux méthodes que nous avons utilisées pour mesurer les distances entre les voyelles longues et brèves dans l'espace vocalique de l'arabe de Mayadin, Schwartz *et al.* (1997b) et Kewley-Port & Zheng (1999), (et qui nous ont permis de savoir si la différence spectrale est suffisante ou non) ne tiennent pas compte de cette non-uniformité de l'espace vocalique, avec unité de mesure le Bark. Si ces deux méthodes sont réellement adaptées à ce type de mesure, il sera alors légitime que nous nous attendions à ce qu'une distance importante (entre une longue et sa correspondante brève) fournie par ces méthodes de mesure se montre plus robuste qu'une distance moins importante. C'est-à-dire que si ces méthodes donnent la distance I/II plus importante que la distance U/UU, alors la distance I/II doit mieux démarquer le /i/ bref du /i:/ long en l'absence de la différence temporelle que la distance U/UU ne démarque le /u/ bref du /u:/ long sous cette même contrainte expérimentale; autrement dit, cette perturbation instrumentale doit affecter l'opposition U/UU plus qu'elle n'affecte l'opposition I/II. Ainsi de suite pour l'opposition I/II vis-à-vis de l'opposition A/AA, et pour l'opposition U/UU vis-à-vis de l'opposition A/AA.

IV.2.2. Méthode

IV.2.2.1. Stimuli

Nos mesures des durées avec un débit normal et un schéma mélodique déclaratif, des voyelles qui ont fait l'objet de notre manipulation expérimentale donnaient les valeurs suivantes (moyennes en ms): /i:/ =196; /a:/ =237 et /u:/ = 209, les correspondantes brèves: /i/ =120, /a/ =143 et /u/ =125.

La neutralisation de l'effet de la durée de la voyelle s'est faite pour les 3 longues et pour les 3 correspondantes brèves. Les 6 voyelles étaient portées par des phrases de la parole naturelle, prononcées par un seul locuteur natif du dialecte. La F0 du locuteur était de 110 Hz. Chaque phrase était formée de deux mots monosyllabiques [CVC CVC], les deux sont

lexicalisés "gal zVr" (il a dit zVr) où V pouvait être l'une des 6 voyelles citées ci-dessus, à chaque voyelle différente correspondait un sens différent du mot "zVr" (cf. paragraphe III.3.3.1).

Il y avait 10 signaux naturels pour chaque voyelle (dix répétitions d'une même phrase [CVC CVC] prononcées par un même locuteur), c'étaient nos signaux de départ. Ces signaux de départ, appelons-les signaux A (A_i , ou l'indice $i=1, 2, \dots, 10$) étaient enregistrés dans une chambre sourde avec un enregistreur PMD 670 et échantillonnés à 44100 Hz.

Nous avons manipulé uniquement la voyelle de la deuxième syllabe de chacune des 10 signaux A. Les signaux A manipulés, appelons-les signaux B (B_i , ou l'indice $i=1, 2, \dots, 10$), étaient alors présentés pour l'identification. Un signal B_i était présenté une seule fois au sujet qui devait l'identifier.

La liste de signaux à identifier contenait 20 signaux, les 10 signaux A et 10 les signaux de B pour chacune des 6 voyelles, pour un total de $(10+10)*6 = 120$ stimuli à identifier par sujet. Finalement, nous avons $20(\text{signaux})*6(\text{voyelles})*10(\text{sujets}) = 1200$ identifications.

IV.2.2.2. Contrôle

Introduire autant de signaux A que de signaux B sur la liste des phrases à identifier avait pour but de savoir si le sujet avait un problème avec la reconnaissance du contraste quantitatif dans la parole naturelle au moment de l'identification des mots. En effet, quand un sujet identifie de manière erronée les signaux A, cela ne nous permet pas de savoir, en ce qui concerne les signaux B, si l'identification est motivée par la perception des indices acoustiques ou si elle est plutôt due à son état physique ou psychologique. Ainsi, les résultats des sujets qui identifieront mal les signaux A ne seront pas pris en compte.

IV.2.2.3. Sujets

Les sujets étaient au nombre de 10, tous volontaires, 4 hommes et 6 femmes âgés de 18 à 42 ans (moyenne d'âge = 30 ans), tous natifs du dialecte et tous connaissent personnellement la personne qui a prononcé les signaux de base et ont bien l'habitude de l'entendre parler, ils exercent tous des métiers intellectuels (7/10 travaillent dans l'enseignement). Ils sont naïfs en ce qui concerne la phonétique. Aucun ne manifestait de troubles psychologiques ou de problèmes de l'audition.

IV.2.2.4. Procédure

Nous avons prévu, lors de la manipulation des signaux, de laisser intacts les indices qui serviraient à l'identification de la qualité de la voyelle (les transitions formantiques, les valeurs des formants, la F0 et l'intensité). En effet, la manipulation des signaux n'a porté que sur la dimension temporelle dans la partie stable de la voyelle. Des méthodes similaires d'allongement et de raccourcissement de voyelles ont été utilisées dans des études comme celle de Nenonen *et al.* (2003) pour les voyelles du finnois, Eerola *et al.* (2003) ont manipulé par le même procédé les durées des voyelles /i/ et /y/ en finnois, Gordon-Salant *et al.* (2006) pour le continuum du /i/ dans l'opposition WHEAT vs. WEED en anglais, Mora & Mott (2003) ont procédé à des allongements et des raccourcissements de consonnes vocaliques [l, r, n] dans le but de déterminer si la syllabation de certaines consonnes était liée ou non à leurs durées.

Notre matière première était formée des 10 signaux A pour chacune des six voyelles, lesquels étaient dupliqués et intégrés ensuite sur la liste des phrases à identifier comme des signaux contrôle. Nous étions face à deux cas de figure:

1) La voyelle est brève. Nous commençons par déterminer la durée de la voyelle en question, parce que l'allongement, comme le raccourcissement, se faisait proportionnellement à la durée initiale. Nos mesures précédentes ont montré que pour les trois qualités vocaliques [i, a, u], la durée de la brève constituait $\approx 0,60$ % de la durée de la longue correspondante. Ainsi, pour savoir de combien il fallait allonger la brève, nous divisons sa durée initiale par 0,6. Le résultat était la valeur de la durée que nous devons atteindre. Garder la proportionnalité permet de produire la même variabilité de longueurs pour les signaux manipulés que celle observée pour les signaux naturels. L'allongement se faisait en copiant un cycle délimité par deux passages par zéro dans le début la partie stable de la voyelle ($T \approx 9$ ms), et en le reproduisant juste après le cycle original (copié). Comme le rapport de l'allongement était quasiment de $2/3$ de la brève, cette opération était répétée sur deux cycles sur trois jusqu'à atteindre la valeur de "durée initiale/0,6", ceci en supprimant ou en rajoutant un cycle, par rapport à ce qui était prévu, pour que la durée atteinte coïncide au mieux avec la durée souhaitée.

2) La voyelle est longue. La procédure est la même que pour la voyelle brève, sauf qu'ici, chaque voyelle s'est vue perdre 40 % de sa durée initiale. Cela s'effectuait en coupant un cycle sur trois, puis un cycle sur deux, puis un cycle sur trois et ainsi de suite jusqu'à atteindre 60 % de la durée initiale de la voyelle objet de la manipulation. Ces signaux étaient

manipulés avec le logiciel Praat, version 4.0.49, et lus, lors de l'identification, par la même version du même logiciel.

Les signaux A et B étaient mélangés et présentés dans un ordre aléatoire. Face à chaque phrase de la liste, le sujet avait deux choix. Soit la phrase était claire auditivement, et dans ce cas, il identifiait et écrivait le deuxième mot de la phrase "zir, zi:r, zar, za:r, zur ou zu:r". Soit la phrase n'est pas claire, parce que le deuxième mot n'était pas distinct, et dans ce cas, il cochait une case prévue pour cette éventualité. Cette possibilité était introduite afin de détecter les signaux qui auraient pu être mal manipulés. Au final, le sujet donnait l'une des trois réponses, soit la voyelle était longue, soit elle était brève soit elle n'était pas claire. Le sujet pouvait écouter la phrase autant de fois qu'il voulait, il prenait le temps dont il avait besoin pour noter le mot qu'il percevait sans souci de vitesse ni de performance.

IV.2.3. Résultats

IV.2.3.1. Signaux A (non-manipulés)

Contrairement à nos attentes, les résultats du test perceptif (tableau 30) montrent que certaines parmi les voyelles qui n'ont subi aucune manipulation ont été soit mal identifiées, soit considérées "pas claires" par les sujets. Le taux d'identification correcte atteint 92 %, brèves et longues confondues. Dans tous les cas, les voyelles longues /i:/, /a:/ et /u:/ sont toutes robustes excepté deux erreurs d'identifications pour le /u:/, mais pas de erreur pour /i:/ ou /a:/. La surprise vient du côté des brèves, surtout pour le /u/ qui a été pris 19 fois sur 100 pour la longue correspondante, totalisant à lui seul plus de 6% des 8% d'erreurs d'identifications des voyelles des signaux A. Le /a/, considéré comme une voyelle qui prêtait à confusion et difficile à reconnaître (pas clair), était cependant rarement confondu avec sa partenaire longue. C'est le /i/ bref qui se maintient le mieux parmi les brèves, il est rarement confondu avec la longue et rarement non reconnaissable, il réalise un taux d'identification correcte qui se rapproche de celui des longues. L'ordre de robustesse des voyelles brèves dans la parole naturelle est le même que celui des longues, c'est-à-dire /i/, /a/ et /u/. Nous notons ici que lorsque nous avons affaire à une voyelle mal identifiée, cela provenait souvent d'un ou de deux sujets, mais ces mêmes sujets identifiaient correctement les autres voyelles des signaux B. Par exemple, toutes les identifications erronées du /i/ des signaux A étaient commises par une même personne. Pour le /u/, 15 des 23 erreurs de reconnaissance provenaient de deux autres sujets, et 8 des 13 erreurs pour le /a/ sont faites par deux sujets seulement. Finalement,

ces résultats, tableau 30, montrent que les sujets identifient massivement correctement la catégorie de la voyelle, Brève/Longue.

Voyelles non manipulées	Identification correcte	Identification erronée	Signal pas clair
/a/	87	3	10
/a:/	98	0	2
/i/	94	4	2
/i:/	100	0	0
/u/	77	19	4
/u:/	95	2	3

Tableau 30. Résultats du test perceptif pour les signaux non-manipulés.

IV.2.3.2. Signaux B (manipulés)

Nous considérons que la manipulation est réussie lorsque le sujet identifie sans ambiguïté le signal B dans le sens de la manipulation. Par exemple, un /i/ bref transformé en /i:/ long et identifié en tant que /i:/ long est une "réussite", alors qu'un /i/ bref transformé en /i:/ long et identifié en tant que /i/ bref est un "échec". Ainsi, les signaux jugés "pas clairs" sont considérés comme faisant partie de la classe "échec". Nous remarquons ici (tableau 31) la même tendance observée auprès des signaux A quant à l'inégalité de la distribution des identifications des voyelles par sujet; par exemple, pour les 10 fois où le signal du /a:/ long était considéré équivoque, c'était un même sujet qui en a fourni 7, et les 9 autres sujets ont fourni les 3 autres.

Quand nous regroupons les deux classes "Échec de la manipulation; Signal pas clair" en une seule classe, qui est "échec", le taux de "réussite" est alors quasiment symétrique pour chacune des catégories (brève et longue), c'est-à-dire que les sujets se faisaient piéger, ou non, presque autant avec la brève manipulée qu'avec la longue manipulée. L'absence de la symétrie parfaite entre l'allongement et le raccourcissement était remarquée aussi par Mora & Mott (2003), dans leur expérience qui a consisté en la manipulation (allongement vs. raccourcissement) de la durée de certaines consonnes vocaliques.

Voyelles manipulées	Réussite de la modification	Échec de la modification	Signal pas clair
/a/ => /a:/	81	4	15
/a:/ => /a/	79	11	10
/i/ => /i:/	21	40	39
/i:/ => /i/	15	68	17
/u/ => /u:/	52	20	28
/u:/ => /u/	42	43	15

Tableau 31. Résultats du test perceptif pour les signaux manipulés.

Nos résultats montrent que c'est la qualité [a] qui est la plus sensible au changement de la durée. Effectivement, 81% des /a/ brefs allongés étaient identifiés en tant que /a:/ longs, et 79 % des /a:/ longs raccourcis étaient identifiés en tant que /a/ brefs. Cela suggère fortement que le contraste entre la voyelle brève et la voyelle longue de qualité [a] est essentiellement basé sur la différence de durée entre brève et longue.

Pour la qualité [i], les résultats enregistrent une nette différence par rapport à ce que nous avons pu observer auprès de la qualité [a]. En effet, seulement 21% des /i/ brefs allongés ont été perçus comme des /i:/ longs, alors que 15% seulement des /i:/ longs raccourcis ont été perçus comme des /i/ brefs. Cela permet d'avancer que le contraste entre les voyelles brève et longue de qualité [i] n'est pas basé sur la différence de durée même s'il est légitime de penser que la différence temporelle puisse jouer un rôle (secondaire). Le taux de "réussite", 18% en moyenne, est très faible comparé à celui de la qualité [a] qui est de 80% en moyenne.

La qualité [u] se trouve dans une position intermédiaire entre la qualité [a] et la qualité [i] en ce qui concerne le taux de "réussite", ce taux est de 47% en moyenne. Effectivement, 52% des /u/ brefs allongés ont été perçus comme des /u:/ longs et 42% des /u:/ longs raccourcis ont été perçus comme des /u/ brefs. Les données concernant le [u] signalent que le contraste entre les voyelles brève et longue de la catégorie [u] nécessite la présence de la différence temporelle.

IV.2.4. Discussion

Nous avons conçu la classe de réponses "pas clair" dans le but d'éliminer l'effet d'éventuels problèmes techniques dus à la synthèse de la durée. Cette conception a réalisé une partie de son objectif puisque nous trouvons que le nombre de signaux qui appartiennent à cette classe a nettement augmenté en passant des signaux A aux signaux B, de 21 à 124 respectivement, ce qui signifie que la synthèse a eu effectivement un effet sur la clarté des

signaux. Nous rappelons ici que les signaux A et B étaient mélangés ensemble et présentés dans un ordre aléatoire. La classe de réponses "pas clair" est une classe problématique, en effet, le fait que cette classe ait reçu des réponses positives avec des signaux A nous laisse penser que l'identification erronée des voyelles naturelles provient du fait que les sujets s'attendaient à ce qu'il y ait des stimuli ambigus, identifier les signaux "pas clairs" a dû être perçu par les sujets comme faisant partie de la tâche. Nous ne savons pas, si cette classe n'avait pas existé quel aurait été le destin des voyelles qui étaient identifiées comme appartenant à cette classe ("pas clair") parmi les signaux A et B. Il est possible cependant que si la consigne était différente et les sujets n'avaient que deux choix forcés "Brève ou Longue", les résultats seraient différents de ce que nous avons trouvé. Un reclassement de ces réponses pourra influencer, dans certaines limites, l'interprétation des résultats de cette expérience. Nous avons envisagé de considérer nos résultats d'un autre point de vue, et cela en éliminant les réalisations qui étaient identifiées comme "pas claire" et ne rapportons ainsi le taux de "réussite" que sur les réalisations qui étaient identifiées à coup sûr. Ce choix s'est avéré périlleux du moment où cette classe n'est pas homogène, les réponses varient considérablement en fonction de la voyelle (entre 0 et 10 pour les signaux A et entre 15 et 38 pour les signaux B), comme elles varient selon le sujet, cela nous a poussé à abandonner ce choix. Un autre phénomène surprenant est l'identification erronée des signaux A, il y en a eu 28 sur les 600 introduits sur la liste, dont 19 pour le /u/ bref seul et 9 pour les 5 autres voyelles de l'expérience, nous rapportons ce résultat alors que nous n'avons pas d'explication pour ces occurrences.

Les résultats de notre traitement statistique et de nos mesures des contrastes perceptifs au niveau quantitatif et qualitatif pour les trois voyelles brèves [i, a, u] et de leurs trois correspondantes longues [i:, a:, u:] de l'arabe de Mayadin sont en accord avec les résultats de cette expérience perceptive.

Concernant l'opposition I/II, nous avons trouvé, chapitre précédent, que F1 et F2 montraient une différence statistique significative sous les 4 conditions expérimentales, quant aux 8 valeurs de la distance spectrale (4 conditions expérimentales et 2 méthodes de mesure) $F1 \times F2$ ou $F1 \times F^2$, elles variaient entre 1 et 1,44 Bark. Cela nous a permis de conclure que la différence qualitative entre les deux termes de la paire /i/ vs. /i:/ est suffisante pour démarquer la brève de la longue. En effet, cette dernière expérience soutient la thèse selon laquelle l'opposition I/II peut être bâtie sur la différence spectrale seule, présente entre la brève et la longue. En effet, en considérant que le refus de la reconnaissance de la catégorie

(en répondant "pas clair") comme une résistance au piège de la manipulation, le taux d'échec a atteint 82% en l'absence du contraste temporel, c'est-à-dire que dans 18% des cas seulement les sujets se faisaient piéger en prenant la brève allongée pour une longue et la longue raccourcie pour une brève. Cela est conforme à nos attentes. Nous ne nous attendions pas cependant à ce que les sujets se fassent piéger autant, autrement dit, nous ne nous attendions pas à ce qu'il y ait une part de l'information sur l'identité de la voyelle qui soit fournie par sa durée à cette hauteur-là. Un résidu de 18% est supérieur à ce que nous prévoyions. Ce résidu sera discuté plus loin dans ce paragraphe dans le but de savoir si la durée pourrait jouer un rôle complémentaire dans la construction de l'opposition I/II.

Concernant l'opposition A/AA, nous avons trouvé que F1 montrait une différence statistique significative sous les 4 conditions expérimentales, alors que F2 n'avait de différence significative que sous la condition expérimentale sans perturbation (débit normal et schéma déclaratif). Quant aux 8 valeurs de la distance spectrale $F1 \times F2$ ou $F1 \times F'2$, elles variaient entre 0,30 et 0,72 Bark suivant la méthode de calcul et la condition expérimentale. Cela nous a persuadé que la différence qualitative entre la brève et la longue, quoique perceptible, n'est pas suffisante pour démarquer catégoriellement les deux voyelles. Notre manipulation instrumentale, chapitre IV, vient étayer notre hypothèse sur l'opposition A/AA. Effectivement, les sujets se faisaient piéger à 80% en prenant la longue pour la brève et brève pour la longue lorsque l'une prend la durée de l'autre, cela signifie que la catégorie de la voyelle (brève vs. longue) ne se maintient pas en l'absence du contraste temporel. Quoique ce résultat soit conforme à nos attentes, nous ne nous attendions pas cependant à ce que les sujets résistent autant à la manipulation, autrement dit, nous n'avons pas prévu que 20% des réponses ne désignent pas la brève allongée comme une longue et la longue raccourcie comme une brève. Ce taux, non négligeable de notre point de vue, signale le rôle que joue la différence spectrale dans l'opposition A/AA. Lorsque nous considérons que la différence au niveau de F2 n'existe pas pour cette paire /a/ vs. /a:/, parce qu'elle n'est pas robuste, les distances spectrales calculées à partir des F1 seuls selon la méthode proposée par Kewley-Port & Zheng (1999), toutes conditions expérimentales confondues, varient entre 0,24 et 0,35 Bark: notre hypothèse selon laquelle la distance au niveau de F1 est une distance perceptible, mais pas suffisante, reste valable.

Concernant l'opposition U/UU, les F1 de la brève et de la longue avaient une différence statistique significative sous les 4 conditions expérimentales, alors que F2 n'avait de différence significative que sous 2 conditions expérimentales sur 4. Les 8 valeurs de la

distance spectrale (4 conditions expérimentales et 2 méthodes de mesure) F1 X F2 ou F1 X F'2, variaient entre 0,43 et 1,03 Bark. Lorsque nous considérons que la différence au niveau de F2 n'existe pas pour cette paire, parce qu'elle n'est pas robuste, les distances spectrales calculées à partir des F1 seuls selon la méthode proposée par Kewley-Port & Zheng (1999), toutes conditions expérimentales confondues, varient entre 0,33 et 0,36 Bark. Cela signifie que la distance au niveau de F1 est une distance perceptible, mais pas suffisante, comme pour A/AA. En effet, notre manipulation instrumentale a montré que l'opposition U/UU est basée et sur la différence temporelle et sur la différence spectrale mais nous signalons la présence d'un avantage de la différence temporelle. Les sujets se faisaient piéger en identifiant la brève allongée en tant qu'une longue et la longue raccourcie en tant qu'une brève à 47%, c'est-à-dire qu'une part considérable de l'information sur l'identité de la voyelle est fournie par sa durée. Cette part reste considérablement inférieure à la part que fournit l'information temporelle pour caractériser les deux termes de l'opposition A/AA, qui est de l'ordre de 80%. Cela prouve aussi par le même biais, que la distance spectrale pour U/UU joue un rôle fondamental dans la construction de l'identité des voyelles brève et longue, puisque nous trouvons que 53% des identifications résistent à la manipulation. Ce taux reste largement inférieur au taux enregistré par le [i]. Donc il est légitime de penser que l'opposition U/UU compte sur les deux informations, temporelle et spectrale. Cette opposition confirme sa position intermédiaire entre les oppositions I/II et A/AA, elle compte sur la durée plus que l'opposition I/II ne fait, mais moins que l'opposition A/AA, et elle se sert de l'information spectrale plus que l'opposition A/AA ne fait mais moins que l'opposition I/II.

Résidu: Nous essayerons ici de répondre aux questions suivantes:

- Nous voudrions montrer qu'il y a un autre indice en dehors des indices temporel et spectral sur la catégorie de la voyelle (Brève vs. Longue): Pourquoi, en ce qui concerne l'opposition I/II, dans 18% des cas, les sujets se faisaient-ils piéger en prenant la brève allongée pour une longue et la longue raccourcie pour une brève? Si la différence spectrale était tout à fait suffisante, alors nous devons nous attendre à ce que ce pourcentage soit proche de 0%.
- Nous voudrions montrer que la distance spectrale sur l'axe de F1 joue effectivement un rôle dans la reconnaissance de la catégorie de la voyelle (Brève vs. Longue): Pourquoi, en ce qui concerne l'opposition A/AA, dans 20% des cas, les sujets ne se faisaient pas piéger, identifiant ainsi la brève allongée en tant que brève et la longue raccourcie en tant que longue? Si l'information n'est fournie que par la durée (thèse abondante dans la littérature de

la quantité vocalique en arabe), alors nous devons nous attendre à ce que les sujets se fassent piéger avec un taux proche de 100%.

Nous posons l'hypothèse que des indices extra segmentaux puissent fournir des informations sur l'identité de la voyelle, notamment les durées des unités phonétiques adjacentes. L'expérience d'Adu Manyah (2003) sur le Twi a montré que la durée de la voyelle n'était pas le seul facteur à l'origine de la distinction Brève/Longue. La complémentarité syllabique jouait, elle aussi, un rôle, puisque la durée de la syllabe était plus ou moins la même quelle que soit la durée de la voyelle, elle était de l'ordre de 230 ms. Adu Manyah (2003) affirme que les voyelles brèves étaient suivies d'une consonne longue et que les voyelles longues étaient suivies d'une consonne brève, il propose que la durée de la consonne post-vocalique, tributaire de la durée de la voyelle, contribue à la distinction entre les deux termes de la paire Brève/Longue. En effet, ses analyses statistiques ont montré que la différence de durées entre les consonnes précédées par des voyelles longues et ces mêmes consonnes lorsqu'elles sont précédées par des voyelles brèves était significative ($p < 0,01$). Concernant nos données, partant de la supposition que la différence spectrale n'a pas d'effet sur l'opposition A/AA, si nous allongeons un /a/ bref, qui est suivi naturellement par une consonne longue, il est possible alors que le sujet reçoive deux informations contradictoires, (i) une durée longue pour la voyelle qui l'oriente vers l'identification d'un /a:/ long, et (ii) un entourage consonantique (une consonne post-vocalique longue essentiellement) qui accompagne normalement une voyelle brève et l'oriente par conséquent vers l'identification d'un /a/ bref. Ce raisonnement pourrait expliquer pourquoi 20% des identifications ont résisté à la manipulation des signaux. (L'hypothèse du rôle contradictoire des deux informations temporelles est valable si nous raccourcissons un /a:/ long).

Nous rappelons que dans cette expérience nous ne nous sommes pas contentés d'éliminer une partie de la durée de la voyelle longue et l'augmenter la durée de la voyelle brève pour arriver à une durée neutre, qui était la même pour la longue et la brève. La manipulation était faite de telle sorte que la plus longue des deux ait pris le statut de la brève et la brève ait occupé celui de la longue, ce qui avait bien évidemment plus de poids pour suggérer une longue à la place de la brève et inversement. Cela a eu pour effet d'accentuer le contraste entre la durée de la voyelle manipulée et la durée des unités adjacentes qui orientent toujours vers l'identification du signal non manipulé.

Dans le cas de l'allongement d'un /i/ bref (le raisonnement est valable pour le raccourcissement d'un /i:/ long), la thèse d'informations contradictoires, comme pour le [a],

est aussi valable. C'est-à-dire que pour un /i/ bref allongé, il y aura (i) une durée plus importante qui dirige vers un /i:/ long, (ii) un entourage consonantique qui dirige vers le /i/ bref, et une troisième information qui n'existe pas pour le [a], (iii) l'information spectrale qui oriente toujours l'identification vers le signal d'avant manipulation, c'est-à-dire qu'elle joue dans le même sens que l'entourage consonantique et dirige vers le /i/ bref. Si c'est le cas, cela voudra dire que l'opposition I/II cherche une partie de l'information sur la catégorie Brève/Longue dans la durée de la voyelle, et que même si la distance spectrale est suffisante, le [i] compte quand même en partie sur l'information temporelle, puisqu'elle sera le seul indice qui expliquera pourquoi 18% ont perçu la voyelle brève allongée en tant que longue et la longue raccourcie en tant que brève.

Afin de vérifier la thèse de l'information fournie par l'entourage consonantique sur la catégorie de la voyelle Brève/Longue, nous nous sommes intéressés à savoir si la durée de la syllabe dépendait de la durée de la voyelle, figure 67. Si la durée de la syllabe ne dépendait pas de la durée de la voyelle, alors les unités de la syllabe devraient se réorganiser afin de maintenir la durée de la syllabe constante: ainsi une voyelle longue qui dure plus longtemps qu'une voyelle brève, comme nous l'avons validé dans le chapitre précédent, devrait contraindre l'entourage consonantique à réduire sa durée. Dans le cas contraire, l'entourage consonantique est insensible à la durée de la voyelle, et par conséquent c'est la durée de la syllabe qui augmenterait en augmentant la durée de la voyelle.

Syllabe: concernant la durée de la syllabe, l'Anova, qui a porté sur 800 observations (406 voyelles brèves et 394 voyelles longues), a montré une dépendance entre la durée de la voyelle et la durée de la syllabe, figure 67. Lorsque nous avons testé cette dépendance pour les 4 conditions expérimentales pour les 3 locuteurs à la fois, toutes donnaient significative la dépendance de la durée de la syllabe par rapport à la durée de la voyelle ($p < 0,01$). Lorsque nous avons testé cette dépendance pour chaque locuteur à part mais pour les 4 conditions expérimentales à la fois, tous donnaient significative cette dépendance ($p < 0,01$). Voir annexe n° X pour les figures représentant les données de chaque locuteur à part.

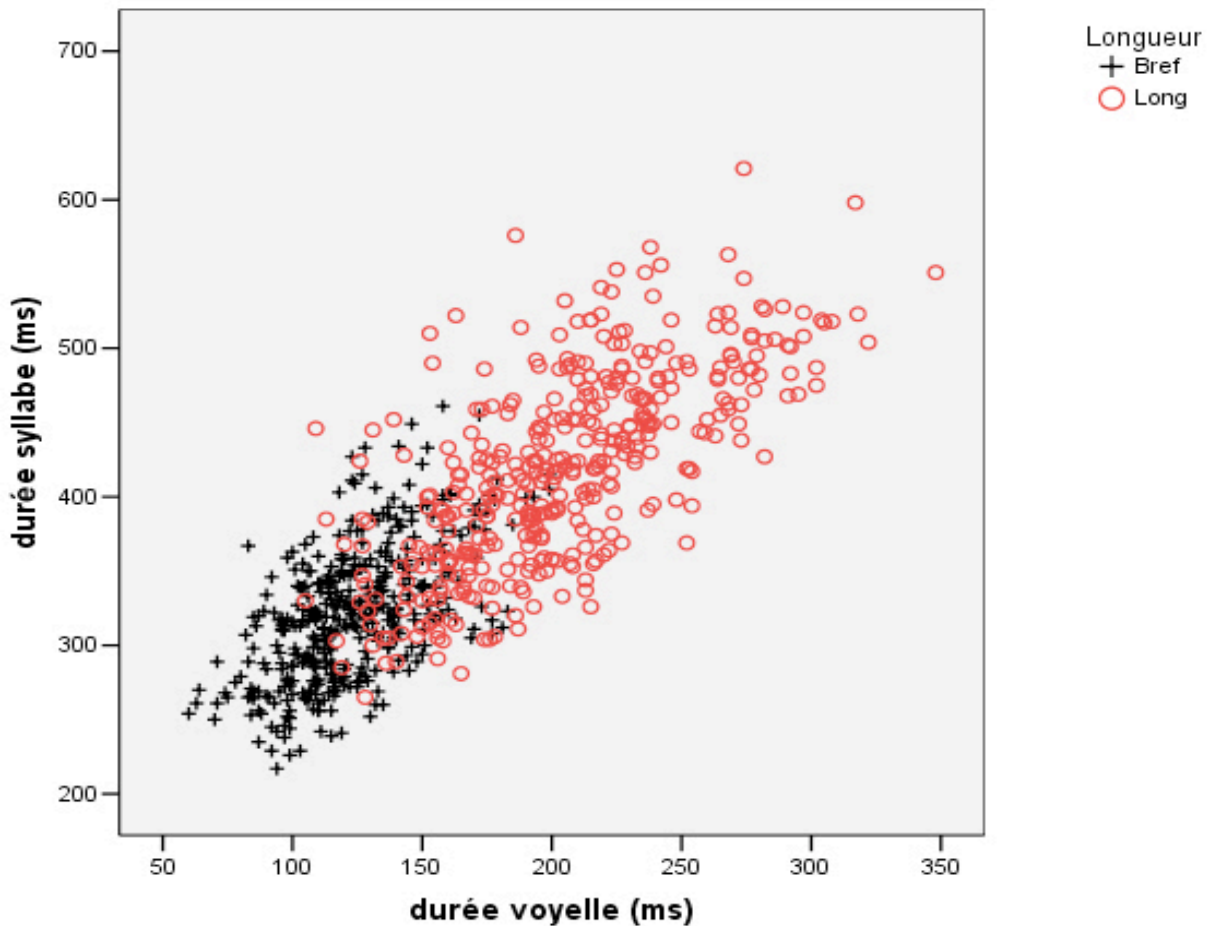


Figure 67. Durée de la syllabe en fonction de la durée de la voyelle en arabe de Mayadin (Syrie). 406 voyelles brèves et 394 voyelles longue. Trois locuteurs masculins. 6 voyelles sont mesurées, trois brèves [i, a, u] et trois longues [i:, a:, u:]. La voyelle est insérée dans la syllabe CVC [zVr]. Les voyelles étaient émises selon deux débits de parole, normal et rapide, et selon deux schémas mélodiques, déclaratif et interrogatif. L'Anova montre un effet significatif ($p < 0,01$) de la durée de la voyelle sur la durée de la syllabe.

Consonne post-voyelle: Les résultats ici sont en harmonie avec les résultats à propos de la corrélation entre la durée de la voyelle et celle de la syllabe. En effet, tous locuteurs et toutes conditions expérimentales confondus, l'Anova ne montre pas de corrélation significative entre la durée de la voyelle et la durée de la consonne ($p = ns$). Nous avons testé aussi la dépendance de la durée de la consonne post-vocalique de la durée de la voyelle pour chacune des 4 conditions expérimentales en mélangeant les données de tous les locuteurs, toutes donnaient non significative la dépendance de la durée de la consonne par rapport à la durée de la voyelle ($p = ns$). Lorsque nous avons testé cette dépendance pour chaque locuteur à part mais pour les 4 conditions expérimentales à la fois, 2 locuteurs sur trois montraient qu'il

y n'avait pas de dépendance ($p = ns$), et un troisième montrait un léger effet significatif ($p < 0,003$). Voir annexe n° XI pour les figures représentant les données de chaque locuteur à part.

Consonne pré-voyelle: tous locuteurs et toutes conditions expérimentales confondus, l'Anova ne montre pas de corrélation significative entre la durée de la voyelle et la durée de la consonne ($p = ns$). Quand nous avons testé la durée de la consonne en position prévoicative en fonction de la durée de la voyelle pour chacune des 4 conditions expérimentales pour tous les locuteurs à la fois, toutes donnaient non significative la dépendance de la durée de la consonne par rapport à la durée de la voyelle ($p = ns$). Lorsque nous avons testé cette dépendance pour chaque locuteur à part mais pour les 4 conditions expérimentales à la fois, 2 locuteurs sur trois montraient qu'il y n'avait pas de dépendance significative ($p = ns$), et un troisième, différent de celui qui a montré un effet significatif pour la consonne post-vocalique, montrait un léger effet significatif ($p < 0,007$). Voir annexe n° XII pour les figures représentant les données de chaque locuteur à part.

Les résultats du traitement statistique réfutent notre thèse selon laquelle les durées des unités phonétiques au sein de la syllabe se réorganisent en fonction de la durée de la voyelle pour maintenir une durée quasi constante de la syllabe, et par conséquent, cette réorganisation constitue un indice acoustique complémentaire sur la catégorie de la voyelle. En effet, la dépendance régulière et robuste et de la durée de la syllabe par rapport à la durée de la voyelle et la corrélation faible et instable de la durée de l'entourage consonantique (consonnes pré et post-vocaliques en syllabe CVC), nous a persuadé qu'il n'y a pas de rapport entre la durée de la voyelle et celle des unités adjacentes, cela nous a ainsi poussé à abandonner la recherche dans le domaine de la durée des unités de la syllabe pour essayer de trouver une explication à la présence du résidu: pourquoi 20% du [a] a résisté à la manipulation et pourquoi 18% du [i] a cédé à la manipulation des durées de nos 6 voyelles, 3 brèves et 3 longues.

Si nous mettons de côté l'hypothèse que la distance spectrale pour l'opposition I/II n'est pas suffisante et que celle de l'opposition A/AA joue un rôle dans la séparation des deux voyelles de la paire /i/ vs. /i:/, notre question concernant le résidu reste jusqu'ici sans réponse. Nous avons réorienté notre recherche vers les traits prosodiques pour entreprendre une réponse à notre question.

En fait, les résultats de l'expérience perceptive d'Abramson (2003) sur la perception de la longueur des consonnes en Pattani Malay, une langue parlée dans le sud de la Thaïlande, ont montré l'existence d'un corrélat évident entre le changement de la perception de la catégorie (brève vs. longue) de la consonne occlusive sourde à l'initiale de la première syllabe

et le changement des deux facteurs prosodiques, Intensité et F0. Les résultats montrent que l'intensité et la F0 sont les seuls facteurs qui permettent la perception de la longueur, et non pas la durée de l'occlusion. Alors, nous nous sommes intéressés à ces deux indices prosodiques pour savoir si la fréquence fondamentale ou intensité en arabe de Mayadin sont bien corrélées à la catégorie de la voyelle (brève vs. longue). Nous rappelons que notre manipulation n'a porté que sur la durée des voyelles, tous les autres indices acoustiques, qualitatifs et prosodiques, sont restés inchangés.

Fréquence fondamentale

Si la F0 est intrinsèquement liée à la catégorie Brève/Longue, elle constituera un indice qui jouera en défaveur de la manipulation, c'est-à-dire qu'elle orientera l'identification vers la catégorie de la voyelle d'avant manipulation de la durée. Nous avons appliqué une analyse de la variance sur nos données afin de savoir si la F0 dépendait de la catégorie de la voyelle Brève/Longue. Tous locuteurs et toutes conditions expérimentales confondus, l'Anova ne montre pas de corrélation significative entre la catégorie de la voyelle et la valeur de la F0 ($p = ns$). Quand nous avons testé la valeur de la F0 en fonction de la catégorie de la voyelle Brève/Longue pour chacune des 4 conditions expérimentales pour tous les locuteurs à la fois, toutes donnaient non significative la dépendance de la F0 par rapport à la catégorie de la voyelle ($p = ns$). Lorsque nous avons testé cette dépendance pour chaque locuteur à part mais pour les 4 conditions expérimentales à la fois, aucun locuteur ne montrait qu'il y avait une dépendance ($p = ns$). Voir annexe n° XIII pour les figures représentant les données de chaque locuteur à part.

Ces résultats, figure 68, nous permettent de penser que la F0 n'est pas différente selon la longueur de la voyelle, et par conséquent, qu'il n'y pas de rôle évident de la F0 sur la reconnaissance de la longueur de la voyelle. Notre conclusion est que la F0 n'est pas un indice acoustique pertinent qui nous permet de comprendre la présence du résidu dans nos données.

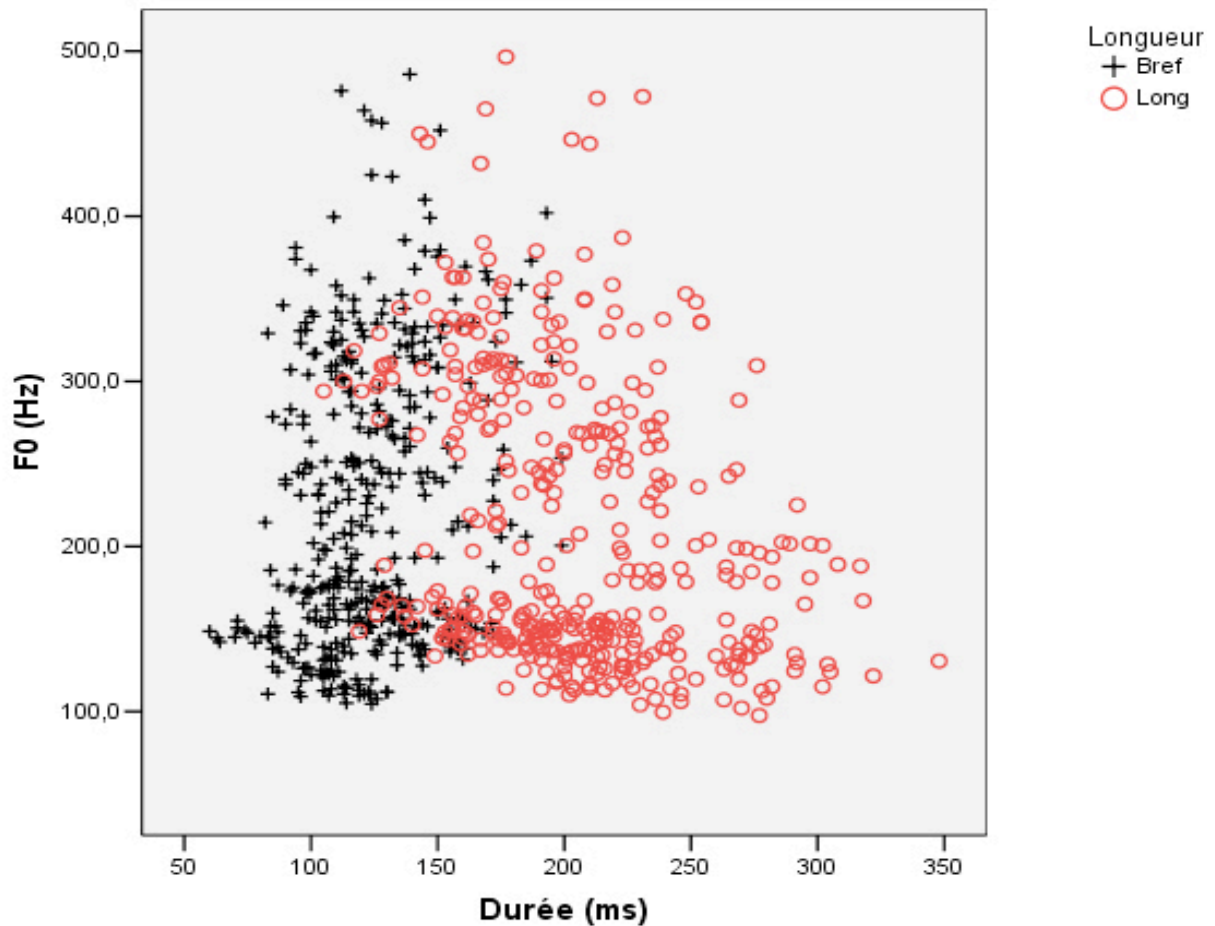


Figure 68. Valeur de la fréquence fondamentale en fonction de la catégorie Brève/Longue des voyelles de l'arabe de Mayadin (Syrie). 406 voyelles brèves et 394 voyelles longues. Trois locuteurs masculins. 6 voyelles sont mesurées, trois brèves [i, a, u] et trois longues [i:, a:, u:]. La voyelle est insérée dans la syllabe CVC [zVr]. Les voyelles étaient émises selon deux débits de parole, normal et rapide, et selon deux schémas mélodiques, déclaratif et interrogatif. L'Anova ne montre pas d'effet significatif de la catégorie de la voyelle sur la F0 ($p = ns$).

Intensité

Nous cherchons à savoir ici si l'intensité est corrélée ou non à la catégorie de la voyelle Brève/Longue. Comme pour la F0, si la valeur de l'intensité de la voyelle dépend substantiellement de sa longueur, alors elle sera un facteur qui nous aidera à mieux comprendre la présence du résidu dans les résultats de notre expérience perceptive. L'analyse de la variance, tous locuteurs et toutes conditions expérimentales confondus, figure 70, donnait significative la dépendance de l'intensité de la voyelle par rapport à sa longueur. Pour chacune des voyelles, nous avons mesuré l'intensité à deux endroits différents de la zone

formantique stable de la voyelle, et puis nous avons obtenu la moyenne à partir de ces deux mesures. Tous locuteurs, tous débits et tous schémas mélodiques confondus, la moyenne pour les voyelles brèves était de 77,98 dB et celle des longues était de 75,86 dB, une différence supérieure à 2 dB.

Lorsque nous avons procédé au test de l'intensité en fonction de la longueur de la voyelle pour chacune des 4 conditions expérimentales à part, en mélangeant les données de tous les locuteurs, sous chacune des 4 conditions la dépendance de l'intensité de la voyelle par rapport à sa longueur était significative ($p < 0,01$). Lorsque nous avons testé cette dépendance pour chaque locuteur à part mais pour les 4 conditions expérimentales à la fois, 2 locuteurs sur trois montraient significative cette dépendance ($p < 0,01$), un troisième avait une p-valeur proche de la signification ($p < 0,019$). Pour les figures représentant les données de chaque locuteur à part voir annexe n° XIV.

En observant les données pour chaque qualité à part, figure 69, la faible différence entre l'intensité du /a/ bref et celle du /a/ long a attiré notre attention sur une éventuelle différence non significative pour certaines de ces trois qualités [i, a, u]. Le tableau des statistiques descriptives relatives aux données de l'intensité est disposé en annexe n° XIII. En effet, l'Anova a montré que la différence qui existe entre le /a/ bref et le /a:/ long n'est pas significative, la tendance demeure la même que pour les deux autres voyelles, c'est-à-dire que la brève des deux a une intensité plus élevée que celle de la longue. L'Anova montre cependant significative la différence entre le /i/ bref et le /i:/ long sous les 4 conditions expérimentales, et la différence entre /u/ bref et /u:/ long n'est significative que dans une condition sur 4 ($p < 0,01$).

Cette différence, brève vs. longue, tous locuteurs, tous schémas mélodiques et tous débit de parole confondus est de (77,67-76,61) 1,06 dB pour la qualité [a], de (78,49-75,11) 3,38 dB pour la qualité [i], enfin elle est de (77,78-75,88) 1,9 dB pour la qualité [u].

Nous trouvons ici un premier indice qui nous permet d'expliquer, en partie, le résidu pour l'opposition A/AA. En effet, l'intensité n'aurait pas joué en faveur de "l'échec" de la manipulation de la durée, elle n'aurait pas orienté l'attention des sujets vers la voyelle, longue ou brève, d'avant manipulation. Autrement dit, concernant la qualité [a], l'intensité du signal ne joue pas de rôle dans la perception de la catégorie (brève ou longue).

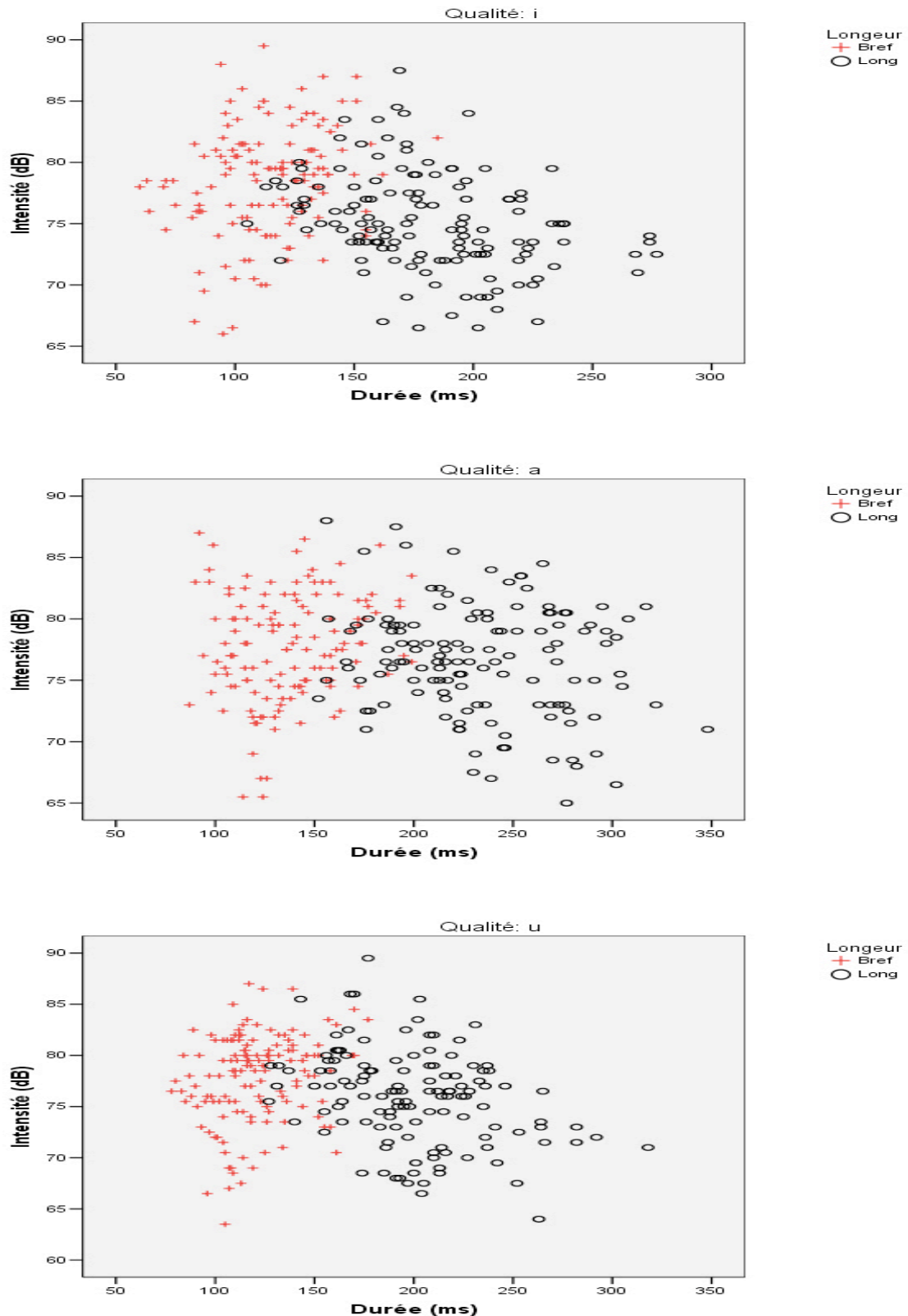


Figure 69. Valeur de l'intensité de la voyelle en fonction de sa catégorie Brève/Longue pour chacune des trois qualités vocaliques de l'arabe de Mayadin (Syrie). Données de trois locuteurs masculins, deux débits de parole, normal et rapide, et selon deux schémas mélodiques, déclaratif et interrogatif. L'Anova montre un effet significatif de la catégorie de la voyelle sur son intensité pour la qualité vocalique [i] mais pas pour les qualités [u], ($p=ns$) et la qualité [a], ($p = ns$).

Cela conforte la thèse selon laquelle la différence spectrale entre le /a/ bref et le /a:/ peut jouer un rôle pour démarquer les catégories de ces deux voyelles. Ainsi, nos soupçons à propos de la différence spectrale d'être à l'origine du résidu (ou d'une partie), peuvent maintenant être justifiés, puisqu'elle est la seule caractéristique qui puisse jouer en défaveur de la "réussite" de la manipulation parmi les caractéristiques que nous avons étudiées: durées des unités contiguës, fréquence fondamentale et intensité acoustique n'ayant pas de rôle significatif. Il est cependant possible que d'autres caractéristiques phonétiques que nous n'avons pas soupçonné jusqu'à maintenant, comme les transitions formantiques, puissent jouer un rôle dans la reconnaissance de la catégorie de la voyelle.

Notre hypothèse concernant l'opposition A/AA reste valable à la lumière de ces dernières révélations à propos des indices acoustiques qui permettent l'identification de la

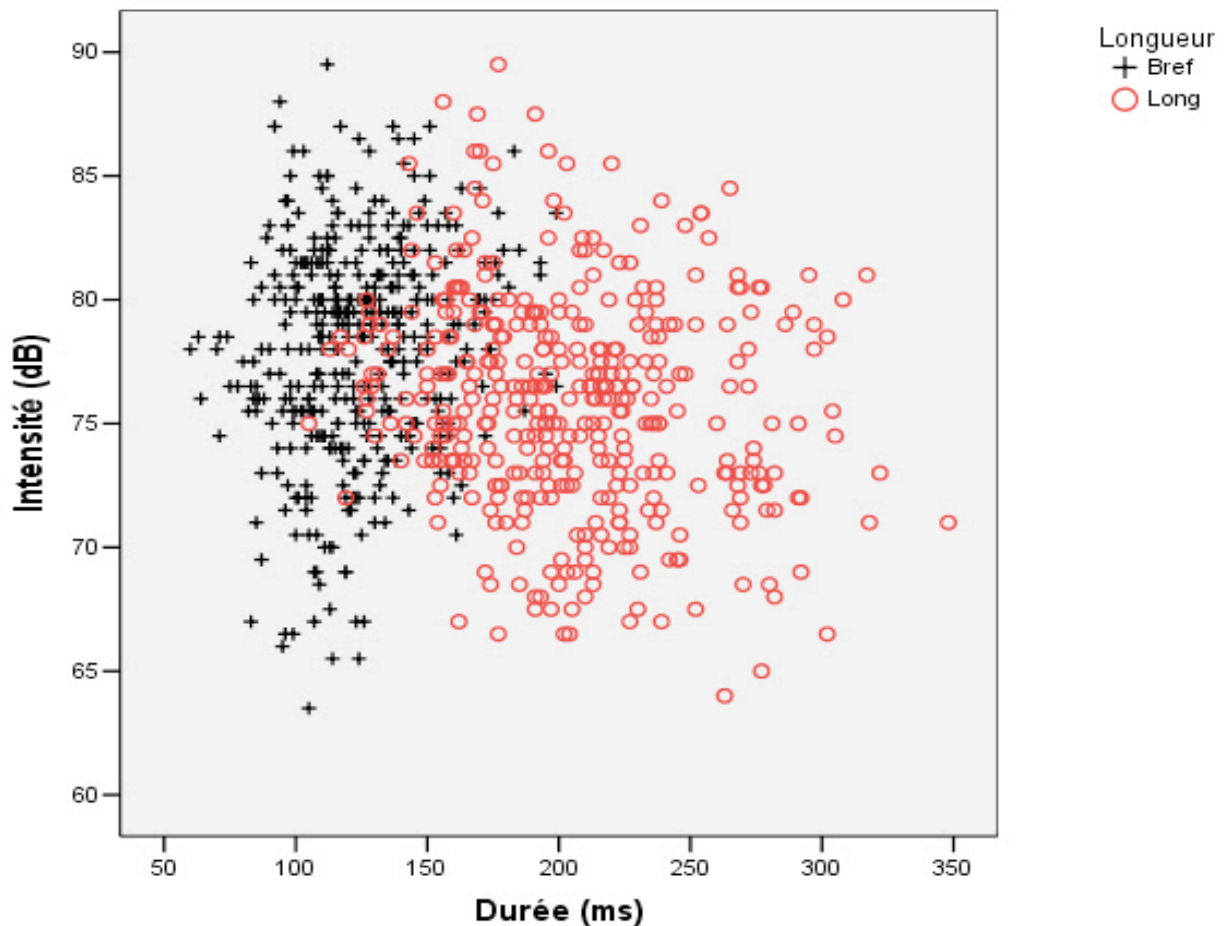


Figure 70. Valeur de l'intensité de la voyelle en fonction de sa catégorie Brève/Longue des voyelles de l'arabe de Mayadin (Syrie). 406 voyelles brèves et 394 voyelles longue. Trois locuteurs masculins. Les voyelles mesurées sont 6, trois brèves [i, a et u] et trois longues [i:, a: et u:]. La voyelle est insérée dans la syllabe CVC [zVr]. Les voyelles étaient émises selon deux débits de parole, normal et rapide, et selon deux schémas mélodiques, déclaratif et interrogatif. L'Anova montre un effet significatif de la catégorie de la voyelle sur l'intensité, ($p < 0,01$).

voyelle. Lorsque nous allongeons un /a/ bref (la démarche est valable si nous raccourcissons un /a:/ long), le sujet reçoit alors deux informations contradictoires, (i) une durée longue de la voyelle qui oriente le sujet vers l'identification d'un /a:/ long (l'inverse dans le cas de raccourcissement) et (ii) une configuration spectrale qui accompagne normalement une voyelle brève et l'oriente par conséquent vers l'identification d'un /a/ bref. Il est cependant certain que le rôle de la durée est massif dans l'identification de la catégorie, puisqu'il n'est pas concurrencé par un autre indice qui dirige vers la même sortie perceptive. La durée est en effet le seul indice que nous ayons trouvé qui dirige les sujets vers l'identification des allongées comme des longues et les raccourcies en tant que brèves. Cela permet de montrer, par le même biais, que le rôle de la différence spectrale reste relativement faible comparé au rôle de l'information temporelle. Pour A/AA, le rôle de la différence spectrale assume totalement les 20% du résidu observé.

Pour l'opposition I/II, cette découverte, concernant le rôle de l'intensité dans la reconnaissance de la catégorie de la voyelle Brève/Longue, conforte l'idée selon laquelle les sujets cherchent toujours une partie de l'information sur la catégorie de la voyelle dans la durée de la voyelle, même si la distance spectrale entre le /i/ bref et le /i:/ long est assez importante pour supporter une différence catégorielle. Parce que les deux indices, intensité et distance spectrale, orientent l'identification dans le même sens. Intensité et distance spectrale sont en concurrence. Le faible rôle de l'intensité, que nous avons pu observer auprès de l'opposition A/AA, permet d'avancer que le grand taux "d'échec" dans l'identification des [i] manipulés, est dû massivement à la distance spectrale.

Nous voudrions reformuler notre hypothèse concernant l'opposition I/II suite aux découvertes concernant les indices acoustiques qui permettent l'identification de la voyelle. Dans le cas de l'allongement d'un /i/ bref (la logique est la même dans le cas du raccourcissement d'un /i:/ long), la thèse d'informations contradictoires, comme pour le [a], est aussi valable. Prenons le cas d'un /i/ bref allongé, il y aura (i) une durée plus importante qui dirige vers un /i:/ long, (ii) une information qui n'existe pas pour le [a], l'intensité, une intensité d'un /i/ bref, elle dirige naturellement vers un /i/ bref, et (iii) une information spectrale qui oriente toujours l'identification vers le signal d'avant manipulation, le /i/ bref en l'occurrence, c'est-à-dire que cette troisième information joue dans le même sens que l'intensité. Donc il est effectif qu'il n'y a qu'un seul indice, parmi ceux que nous avons soupçonnés, qui pousse les sujets à identifier les allongées en tant que longues et les raccourcies en tant que brèves, et qui est par conséquent derrière le résidu de 18%, c'est la

durée de la voyelle, puisque intensité et distance spectrale poussent vers "l'échec" de la manipulation.

Si nous admettons que pour l'opposition U/UU l'intensité ne joue pas de rôle (nous maintenons ici un point d'interrogation), le raisonnement que nous avons développé à propos de l'opposition A/AA est tout à fait valable, sauf que le poids de l'information spectrale est plus important, cela peut expliquer le fort taux de "réussite" de la manipulation, 47%, comparé au 18% de l'opposition I/II. Finalement, le fait que le taux "d'échec" de 53% soit dû uniquement à l'information spectrale, nous permet de confirmer la conclusion que nous avons établie à partir des résultats de la manipulation instrumentale (dans ce chapitre IV), et dans laquelle nous avons conclu qu'il existe une équité de poids: temporel *vs.* spectral, 47% *vs.* 53% respectivement, avec une légère prépondérance de l'information spectrale. Ainsi, l'information la plus importante est sans doute l'information temporelle, suivie par l'information donnée par la distance spectrale qui est suivie finalement, éventuellement, par l'information prosodique donnée par l'intensité.

IV.2.5. Conclusion

Cette manipulation instrumentale a consisté en l'allongement des trois voyelles brèves [i, a, u] et le raccourcissement des trois voyelles longues [i:, a:, u:] afin de neutraliser l'effet de la différence temporelle dans la reconnaissance de la catégorie de la voyelle Brève/Longue. Le but était de valider les résultats statistiques et perceptifs disposés dans le chapitre précédent (chapitre III). Ces résultats-là montrent que la différence spectrale pour I/II est robuste et par conséquent suffisante pour créer le contraste Brève *vs.* Longue toute seule (sans l'information temporelle). La distance spectrale est *semi* robuste pour A/AA et U/UU (étant robuste sur F1 seulement) et joue un rôle dans ces oppositions.

Les résultats de cette manipulation valident les résultats du chapitre précédent, parce qu'ils vont dans le même sens que nos hypothèses ci-dessus. Ils nous révèlent, en outre, les données suivantes: l'information spectrale (quoique suffisante) n'agit pas seule pour démarquer le /i/ bref du /i:/ long, l'information temporelle (quoique suffisante) n'agit pas toute seule pour démarquer le /a/ bref du /a:/ long et pour démarquer le /u/ bref du /u:/ long.

Le poids perceptif de l'information temporelle et celui de l'information spectrale seront effectivement associés à d'autres indices acoustiques. Nos recherches de ces indices nous ont poussé à tester le rôle d'autres caractéristiques phonétiques qui pourraient expliquer cette insuffisance des informations temporelle et spectrale prises séparément. Ainsi nous avons

testé le rôle des durées des unités phonétiques adjacentes à la voyelle étudiée, le rôle de la fréquence fondamentale et finalement le rôle de l'intensité acoustique, cela était dans le but de savoir si elles apportent concrètement de l'information sur la catégorie de la voyelle.

Les durées des consonnes pré et post-vocaliques se sont montrées insensibles à la durée de la voyelle. Encouragés par de précédentes études (Adu Manyah, 2003), nos soupçons vis-à-vis de ces caractéristiques temporelles, *a priori* prometteurs, se sont révélés tout simplement faux. Nous avons conclu que ni la durée de la consonne prévocanique, ni la durée de la consonne post-vocalique n'est porteuse d'information sur la longueur de la voyelle, Brève/Longue.

La fréquence fondamentale n'a pas réussi non plus à prouver d'une corrélation significative avec la longueur de la voyelle. En effet, la F0 de la voyelle brève n'est pas différente de celle de la longue. Notre conclusion, par rapport à la F0, était que cette caractéristique prosodique ne renseigne pas les auditeurs de l'arabe de Mayadin sur la longueur de la voyelle, elle était ainsi exclue.

L'intensité acoustique, contrairement à nos attentes, a montré une corrélation significative à la longueur de la voyelle, les brèves ayant en moyenne, tous locuteurs et toutes conditions expérimentales confondus, une intensité plus élevée que l'intensité des longues, de 3,38 dB pour la qualité [i] et de 1,9 dB pour la qualité [u]. Elle s'est confirmée en tant qu'indice acoustique non négligeable sur la longueur de la voyelle (négligeable, voire inexistant, pour l'opposition U/UU), elle n'est pas un indice pour l'opposition A/AA. La taille de contribution de l'intensité dans la reconnaissance de la longueur de la voyelle n'a pas pu être déterminée avec précision, puisqu'elle pourrait être concurrencée par l'information apportée par la distance spectrale entre la brève et la longue. D'autres expériences peuvent être menées afin de déterminer le poids de cette contribution. Il suffirait de neutraliser le rôle de l'intensité (en donnant à la longue l'intensité de la brève et à la brève celle de la longue) et le rôle de l'information temporelle d'une manière simultanée, et d'observer ensuite la hausse du taux de "l'échec" de la manipulation pour l'opposition I/II. Ces hausses représenteront alors exactement la contribution de l'intensité. La neutralisation de l'intensité pourrait aussi révéler la présence d'autres indices acoustiques sur la longueur de la voyelle, des indices que nous ne suspectons pas à ce stade.

Nous avons réussi à déterminer dans cette étude les sources de l'information sur la catégorie de la voyelle Brève/Longue. Cette information s'est montrée bi factorielle,

temporelle et spectrale, et tri factorielle, temporelle, spectrale et prosodique, selon l'opposition considérée.

Pour l'opposition I/II, l'information sur la catégorie de la voyelle est essentiellement fournie par la distance spectrale entre le /i/ bref et le /i:/ long, puis par la durée de la voyelle et finalement par son intensité. Nous ne pouvons pas classer ces deux dernières selon leur contribution.

Pour l'opposition A/AA, l'information est supportée principalement par la durée de la voyelle, secondairement par la différence spectrale, pas d'évidence pour un éventuel rôle de l'intensité acoustique entre le /a/ bref et /a:/ long.

Pour l'opposition U/UU, l'information sur la catégorie de la voyelle n'est pas fournie massivement par un seul indice. Les indices sont: durée de la voyelle, distance spectrale entre le /u/ bref et le /u:/ long, et (éventuellement) intensité.

Les résultats de cette expérience perceptive (chapitre IV) mettent en évidence une divergence entre la théorie, donnée par les seuils, et la pratique donnée par les perturbations. Ils sont en faveur des perturbations comme stratégie pour la découverte des indices phonologiques, au détriment du jugement théorique fourni par les seuils.

En fait, ils nous enseignent que si nous avons un seul indice et qu'il est jugé suffisant par les seuils et jugé robuste par les perturbations, il va de soi qu'il est indispensable au maintien du contraste. Si maintenant nous avons deux indices et que chacun des deux indices est suffisant (jugement des seuils), si les perturbations les montrent alors robustes tous les deux, c'est que la disparition de l'un n'est pas sans conséquence sur le maintien du contraste (ex. la distance spectrale I/II, quoique suffisante, n'a pas pu supporter l'opposition catégorielle Brève *vs.* Longue toute seule). Autrement dit, si les perturbations jugent un indice robuste, c'est qu'il est indispensable à la communication linguistique. Cela veut dire enfin que le jugement de la spécification phonologique d'un indice acoustique est mieux fourni par les résultats des perturbations que par les résultats des mesures perceptives qui peuvent théoriquement juger un indice capable de supporter un contraste alors que dans la pratique linguistique il en est autrement puisque, associé à un autre indice qui est robuste, il n'est pas capable tout seul de maintenir le contraste.

Nous avons un deuxième et dernier point à examiner dans cette expérience, lequel concerne les deux méthodes que nous avons utilisées pour mesurer les distances entre les

voyelles longues et brèves dans l'espace vocalique de l'arabe de Mayadin, Schwartz *et al.* (1997b) et Kewley-Port & Zheng (1999). Nous avons posé la question suivante: si ces deux méthodes sont adaptées à la mesure des distances spectrales dans les espaces vocaliques, nous devons nous attendre à ce que plus la distance entre deux voyelles, calculée par ces méthodes, est importante et plus elle joue un rôle dans leur discrimination. Les résultats de notre expérience perceptive sont en total accord avec les distances que nous avons obtenues avec les deux méthodes ci-dessus. En effet, les deux méthodes donnent les distances pour les trois oppositions de la plus importante à la moins importante comme suivant: I/II, U/UU et enfin A/AA. Cette expérience perceptive a prouvé que l'opposition I/II est l'opposition qui, parmi les trois oppositions étudiées, compte le plus sur la distance spectrale, suivie de l'opposition U/UU suivie enfin par l'opposition A/AA. Ceci est une démonstration de la cohérence de la démarche de ces méthodes.

En dernier lieu, les données de ce chapitre nous ont incité à faire une distinction entre un indice suffisant et sur lequel les sujets parlants se basent pour distinguer les unités phonétiques. Ainsi, la distance entre /i/ bref et /i:/ long qui est franchement suffisante, ce que nous avons prouvé par divers procédés, n'est pas le seul indice acoustique sur lequel les locuteurs de l'arabe de Mayadin tablent pour discriminer les catégories Brève/Longue, il est cependant le principal indice. La perception des traits distinctifs est alors plus influencée par les habitudes langagières que par un pragmatisme perceptif. Sommes-nous, gens de Mayadin, des traditionalistes de la phonétique? Peut être que oui, du moment où nous préférons faire perdurer un héritage phonétique et nous conformer ainsi à une tradition linguistique plutôt que de minimiser le coût en éliminant un contrôle de surplus tout en restant aussi efficace. En fait, la durée de la voyelle qui joue un rôle tout à fait contingent pour l'opposition I/II, s'est effectivement imposée comme un acteur incontournable dans la reconnaissance de l'identité de la voyelle.

V. Durée et qualité vocaliques en français de Savoie. Étude acoustique intergénérationnelle

V.1. Introduction

Nous effectuons dans ce chapitre un travail expérimental afin de décrire le français de la région de la Savoie. Le but est de comparer le contrôle de la quantité vocalique de l'arabe avec un autre système phonologique qui disposerait *a priori* du procédé phonétique qui consiste à opposer les durées des voyelles à des fins sémantiques, nous avons choisi le français de la région de la Savoie, parce qu'il disposerait, comme l'arabe, de voyelles longues contrastées avec des partenaires brèves.

Ainsi nous parlons de *voyelles longues* et de *voyelles brèves* dans ce parler. Une *longue* et une *brève* forment ainsi une paire, mais elles appartiendront à une même qualité vocalique. Nous étudierons toujours ces voyelles par paire au sein d'une même qualité, c'est-à-dire que nous nous intéresserons aux différences qui existeraient au sein d'une même qualité, et non pas aux différences qui existeraient entre les qualités. Nous avons quatre paires /a/ vs. /ɑ:/, /ɛ/ vs. /ɛ:/, /ɔ/ vs. /o:/ et /u/ vs. /u:/, où la durée serait pertinente phonologiquement.

Le procédé phonétique consistant en l'usage du contraste temporel au sein d'une même qualité vocalique à une fin phonologique, appelé quantité vocalique, est la seconde des stratégies de multiplication des éléments vocaliques à partir des qualités de base, juste après la nasalité. Selon la base UPSID, Maddieson (1986), 11% des systèmes phonologiques des langues du monde y ont eu recours: elle est adoptée par 51 langues sur les 451 langues répertoriées dans la base de données.

La présence de la quantité vocalique en français de la région de la Savoie était attestée par de précédentes études, Martinet (1956) rapporte l'existence de différences temporelles entre voyelles dans le parler franco-provençal de Hauteville. Plus tard, Abry *et al.* (1995), en cherchant à contrôler les descriptions concernant les durées des voyelles et qui étaient données sans mesures expérimentales par Martinet (1956), ont décrit les corrélats de l'opposition de quantité vocalique sous l'accent en franco-provençal de Saint-Pierre-de-Soucy (Savoie) en durée et en qualité vocaliques pour les deux paires de voyelles [a] vs. [æ:] et [ɛ]

vs. [ɛ:] en entourage consonantique non voisé. Ils affirment la présence d'une différence qualitative entre la voyelle brève et la voyelle longue correspondante pour les deux paires qui ont fait l'objet de leur étude, cette différence qualitative est décrite comme robuste puisqu'elle est présente sous les deux débits de parole, normal et rapide. Ils rapportent par ailleurs que les termes de l'opposition [ɛ] vs. [ɛ:] sont mieux distingués (du point de vue de la qualité) que les termes de l'opposition [a] vs. [æ:]. Leurs analyses des durées des brèves et des longues montrent que la durée de la voyelle dépend de sa catégorie (Brève/Longue) et que cette dépendance se maintient sous la perturbation du débit rapide de la parole. Les auteurs de l'étude soutiennent que le /a/ bref constitue de 30 à 40 % la durée de sa partenaire longue, alors que le /ɛ/ bref ne représente que 30 à 40 % la durée de la longue.

Afin de savoir quels sont les facteurs qui sont derrière cette opposition quantitative, Abry *et al.* (1995) testent alors *la résistivité de l'opposition de quantité aux variations de la vitesse d'élocution*. Une vitesse mesurée par l'unité Syllabe/Seconde (soit l'unité syll./s). Le rapport Débit normal/Débit rapide était de 1,5 environ, 4 syll./s pour le débit normal et 6,5 syll./s pour le débit rapide. Les auteurs défendent l'approche de la soumission de la parole au débit rapide par l'absence du bien-fondé de l'attribution des valeurs *arrêtées* à des phénomènes substantiellement instables relevant du domaine de la parole, telles les durées des voyelles ou leurs qualités acoustiques. Les durées des voyelles émergent sous cette dernière contrainte (le débit rapide) étant aussi légitimes que celles qui n'ont pas cette contrainte, puisque cette adaptation (durée moins longue) est une réponse à une exigence de la parole naturelle. Tout en donnant raison aux descriptions auditives de Martinet (1956) du français franco-provençal de Hauteville, Abry *et al.* (1995) se contentent d'une description expérimentale, quantitative et qualitative, pour caractériser l'opposition entre voyelles brèves et longues. Ils constatent ainsi que le système est unique (selon le terme de Schwartz *et al.*, 1997a) mais ils ne parviennent finalement pas à déterminer le trait phonétique pertinent qui est derrière cette opposition de quantité.

Nous émettons donc l'hypothèse que, dans le français de Savoie, il y a des voyelles qui fondent leurs oppositions sur la durée de leurs réalisations et non pas sur des différences provenant de la disposition des zones d'énergies de leurs spectres acoustiques (qualité), parce que la quantité à elle seule est capable de supporter cette opposition catégorielle. Nous étudions ainsi deux aspects de ces voyelles: quantitatif et qualitatif, nous jugeons ensuite les écarts perceptifs entre les deux termes de chaque paire Brève/Longue à la lumière des résultats concernant les seuils différentiels de la perception de la qualité et de la durée

vocaliques que nous avons obtenus en I.1.4.3 et I.2.4. C'est ainsi que nous effectuons des mesures temporelles et spectrales sur les voyelles que nous soupçonnons de disposer du procédé de la quantité vocalique. Comme Abry *et al.* (1995) qui sont allés outre les descriptions de Martinet (1956), ici, nous irons outre les descriptions d'Abry *et al.* (1995) afin de déterminer les corrélats perceptifs responsables de cette opposition Brève/Longue.

Ici, nous supposons que cette opposition est encore vivante dans certains parlars chez des locuteurs d'un certain âge. Certaines de ces voyelles seraient intrinsèquement plus longues d'autres, c'est-à-dire que pour un même contexte phonétique adjacent, ces voyelles s'opposeraient par la durée de leurs réalisations (toutes choses égales par ailleurs). Afin de passer de la supposition à la confirmation, nous étudierons d'une manière détaillée chez chacun de nos locuteurs la durée et la structure acoustique de chacune des 8 voyelles que nous avons citées ci-dessus. Nous cherchons à déterminer au préalable deux points en français de Savoie: (1) la durée, c'est tout simplement la valeur absolue de la différence entre le début et la fin de la voyelle, et (2) la qualité est représentée par les valeurs fréquentielles des formants de la voyelle considérée. Une fois que nous obtenons ces informations, temporelles et spectrales, nous procéderons à des mesures de distances perceptives, comme nous avons fait pour l'arabe en chapitre III, afin de savoir lesquelles des différences temporelles et spectrales participent au processus de démarcation des catégories vocaliques Brève/Longue.

a) Durée

Nous rappelons que nous nous intéressons au français de Savoie parce que son système phonologique de voyelles utilise *a priori* la durée de la réalisation comme un trait distinctif, c'est-à-dire que deux voyelles qui ont la même qualité vocalique n'auront pas le même statut phonologique en fonction de leurs durées respectives. Nous cherchons ainsi à savoir si les mesures expérimentales confirment cette impression de différence de durée, détectée à l'oreille, c'est-à-dire que nous cherchons à savoir si une différence au niveau de la durée entre les voyelles supposées longues d'un côté et leurs homologues supposées brèves de l'autre existe réellement. Nous cherchons la preuve que ces différences existent en traitant un corpus composé d'items lexicalisés en français de Savoie. Nous nous fixons cependant deux objectifs:

- Décrire l'opposition de la durée dans des conditions normales de l'émission de la parole, sans contrainte. Quelques exemples de contraintes: parole forte, rapide, variation de la F0 pour des fins linguistiques (cas de l'interrogation), etc.

• Dans la limite de la présence du contraste temporel, décrire l'opposition de la durée en soumettant l'émission de la parole à des conditions perturbatrices: ici nous appliquons la contrainte de la parole rapide, pour savoir si cette opposition se maintient ou non, autrement dit nous testons la robustesse de l'opposition brève/longue.

b) Qualité

Nous cherchons à savoir, abstraction faite de l'éventuelle existence du contraste temporel, si ces deux voyelles dites *longue* et *brève* présentent ou non une différence spectrale. En fait, nous essayerons de répondre à une question que nous pouvons formuler comme ceci: est-ce que la voyelle brève est une voyelle qui est qualitativement différente de son homologue longue, ou est-ce que ces deux voyelles ont une même qualité?

Le possible contraste de qualité sera soumis à son tour à la perturbation engendrée par le débit rapide afin d'en tester la robustesse, comme l'ont fait Abry *et al.* (1995).

Intergénérationnel, pourquoi? Nous cherchons à savoir si cette différence existait à une certaine époque d'un côté et si cette différence existe encore, si elle est en train de disparaître ou si elle a déjà disparu. C'est dans ce but que nos locuteurs appartiennent à trois générations. Cette opposition qui existait en français, semble complètement avoir disparu du français moderne, et nous voulons dans cette étude apporter une épreuve pour ou contre cette hypothèse-là.

Beaume (1986) plaidait en faveur de cette hypothèse, il écrivait: "*En français moderne, il est impossible de distinguer à l'oral l'ami de l'amie ou bien les amis de les amies hors-texte tandis qu'en ancien français, c'est-à-dire à l'époque où l'opposition de la voyelle longue/brève était vivante et la dernière voyelle d'un mot féminin ou pluriel se prononçait plus longue; ainsi, la prononciation des mots **ami** et **amie**, **dieu** et **dieux** étaient-elles différentes l'une de l'autre. L'omission de plus en plus de ce trait distinctif dans l'oral amène une autre distorsion entre l'oral et l'écrit du français*".

Cette interrogation était bien relevée dans un texte apparu dans le glossaire typographique et linguistique du site Babel (<http://alis.isoc.org/glossaire/breve.fr.htm>):

"*Le français a connu cette opposition de longueur jusqu'au XVIIIe siècle environ. À l'époque, à côté de mots à la voyelle brève comme vit, bout et la, on trouvait vie, boue et las qui se prononçait avec une voyelle longue. Certaines de ces voyelles longues étaient déjà présentes dans la langue depuis le XIIe à la suite de la disparition du s dans des mots tels que beste et teste devenus bête et tête.*

... au Québec on distingue nettement belle et bèle ainsi que tette et tête. Certains auteurs maintiennent cependant qu'il s'agit plus d'une différence de timbre que de durée dans ce cas. En Belgique, on distingue patte et pâte par la durée du a et non par son timbre. De même les formes féminines blessées, amies, crues s'opposent à leur forme masculine respective par une voyelle finale plus longue."

L'enquête de Walter (1982) montrait la présence de différences de durées entre les voyelles de français. Longtemps avant les travaux de Walter sur les variétés régionales du français, Durand (1946) a longuement discuté la question de la présence de la quantité vocalique basée sur l'opposition de la durée entre voyelles longues et voyelles brèves en français à l'époque de l'apparition de son ouvrage.

V.2. Méthode

V.2.1. Stimuli

Les voyelles qui nous donnent l'impression d'avoir des contrastes quantitatifs appartiennent aux 4 qualités suivantes: [a], [ɛ], [o] et [u], lesquelles feront l'objet de notre étude en ce qui concerne la description temporelle et spectrale. Chacune des voyelles concernée, brève et longue, était insérée dans une syllabe de type CVC; ainsi nous avons 8 syllabes, chaque deux syllabes ayant pour noyaux respectifs deux voyelles appartenant à la même qualité, formaient une paire minimale, à savoir [pat] vs. [pa:t], [tɛt] vs. [tɛ:t], [kɔt] vs. [ko:t] et [gut] vs. [gu:t]. Chacune des 8 syllabes était insérée à son tour dans des phrases porteuses qui étaient réalisées avec la même prosodie. Notre corpus était formé des phrases porteuses suivantes:

[a] ~ [ɑ:]

pat = patte. Le jeune compte les pattes.

pa:t = pâtes. Le jeune compte les pâtes.

[ɛ] ~ [ɛ:]

tɛt = tête. Le jeune comte les tête.

tɛ:t = tête. Le jeune compte les têtes.

[ɔ] ~ [o:]

$k\text{ɔ}t = cotte$ Le jeune compte les **cottes**.

$ko:t = côte$ Le jeune compte les **côtes**.

[u] ~ [u:]

$gut = goutte$. Le jeune compte les **gouttes**.

$gu:t = goûte$. Le jeune comte les **goûte**.

Les voyelles entre crochets sont les voyelles dont nous avons mesuré la durée et les deux premiers formants.

V.2.2. Sujets

Nous avons choisi seulement des personnes qui ont des parents et des grands-parents qui sont de la région de la Savoie, qui ont vécu en Savoie toutes leurs vies et qui se considéraient comme Savoyards. Cette dernière condition était exigée pour des raisons sociolinguistiques, parce que nous savons que certaines variantes phonétiques peuvent apparaître, comme elles peuvent disparaître, auprès de certains locuteurs natifs d'un parler régional pour des revendications identitaires et non pas pour des raisons linguistiques (Labov, 1972a, Laks, 1977). Ces contraintes qui, au premier abord, nous paraissaient aisément franchissables, se sont avérées une lourde tâche à remplir, ce qui nous. Nos locuteurs étaient originaires de deux villages du comté de Savoie, la Chavanne et les Molettes, distants de quelque 5 km l'un de l'autre. Puisqu'une partie de notre enquête concerne l'étude intergénérationnelle de l'usage de la durée, nos sujets n'appartenaient pas à la même tranche d'âge. Nos 6 sujets sont répartis sur trois tranches d'âge (générations). Deux hommes et une femme étaient âgés de plus de 76 ans (en 2007, date du recueil des données pour 5 d'entre eux, le sixième était enregistré en 2005), constituant la première génération. La deuxième génération était constituée d'un locuteur féminin âgé de 49 ans. Et la troisième génération était formée d'un garçon et d'une fille âgés de 22 et 25 ans respectivement. Ils étaient tous volontaires, et ils n'étaient pas payés pour leurs contributions.

V.2.3. Procédure et enregistrement

Notre corpus était enregistré avec deux débits de parole. Un débit normal, qui est celui de la parole sans contraintes et sans émotions, et un débit rapide. Ici, nous avons appliqué la méthode des "perturbations prioritaires" utilisée dans Abry *et al.* (1990), c'est-à-dire les

variations de débit d'élocution. Cette méthode de perturbation est en mesure de nous donner des informations sur la robustesse des écarts temporels et spectraux entre voyelles longues et brèves. Sous cette contrainte, le système est amené à se réorganiser, ainsi les oppositions qui disparaîtront ne seront pas des oppositions robustes.

Pour l'émission des phrases en débit rapide, le locuteur était sollicité d'imaginer qu'il était en train de prononcer la phrase à quelqu'un qui était sur le point de partir et qu'il fallait dire la phrase vite, avant que l'autre personne ne parte.

Chacune des 8 phrases du corpus était produite 12 fois, ce qui donnait 96 phrases par locuteur et par débit de parole. Les 96 phrases étaient mélangées ensemble pour donner à la présentation un caractère aléatoire. Finalement nous avons 96 (phrases) * 2 (débits) * 6 (locuteurs) = 1152 signaux à analyser où nous nous intéressons à 3 informations portées par chacun de ces signaux: 1) Durée, 2) F1 et 3) F2.

Les sujets n'étaient pas tous disponibles pour se déplacer jusqu'au laboratoire Gipsa-Lab pour effectuer l'enregistrement dans la chambre sourde du laboratoire. Un seul sujet a pu se déplacer. Pour les autres, nous sommes parti les enregistrer chez eux, et au final nous avons eu trois lieux différents pour nos enregistrements. Les phrases à lire étaient écrites dans un fichier PowerPoint où il n'y avait qu'une seule phrase par diapositive. Le défilement des phrases se faisait manuellement sur l'écran d'un ordinateur portable que nous avons placé le plus loin possible du micro pour éviter la superposition du bruit engendré par l'ordinateur et le signal de parole. Le micro était placé à 25 cm de la bouche du locuteur. Nous avons opté pour le défilement manuel contre le défilement automatique pour éviter que les sujets se trompent par souci de vitesse (comme cela était arrivé avec un premier locuteur, où le locuteur n'avait que 5 secondes pour réfléchir, comprendre et lire la phrase). Avec le défilement manuel, les locuteurs avaient tout le temps qu'ils voulaient pour prononcer les phrases. Les résultats du dépouillement de ces données ont montré que les sujets, notamment ceux de la tranche la plus âgée, se trompaient malgré cette précaution, et cela en prenant la voyelle longue pour une brève et inversement, c'est-à-dire qu'ils confondaient par fois le mot *pattes* avec le mot *pâtes*. Mais il était rare que le sujet se trompe de qualité. Les sujets avaient la possibilité de faire des pauses et de faire des commentaires entre les phrases prononcées, cela afin de donner un cadre naturel à l'enregistrement.

L'enregistrement était effectué avec l'enregistreur PMD-670 du laboratoire Gipsa-Lab, format d'enregistrement PCM, avec une fréquence d'échantillonnage de 22050 Hz, le mode d'enregistrement était MonoMicro. L'analyse était faite avec le logiciel Praat, version 4.0.49.

V.2.4. Dépouillement des données

La numérisation des phrases du corpus était faite automatiquement par l'enregistreur PMD-670. Lors de l'étiquetage, nous avons commencé par la correction des erreurs commises par les locuteurs. Il a fallu ensuite que nous délimitions sur le signal les segments vocaliques qui nous intéressaient afin que nous puissions faire des mesures manuelles et automatiques concernant la durée et la qualité vocalique.

Nos signaux étaient préparés aux mesures automatiques effectuées par Praat de la manière suivante: le premier TIER était destiné à la mesure de la durée de la voyelle, ainsi il avait trois INTERVAL (1) pré-vocalique, (2) la voyelle et (3) post-vocalique. Dans le Script nous avons demandé de mesurer la durée de chaque INTERVAL (2). Sur le deuxième TIER, nous délimitions par deux points la zone que nous jugions stable en ce qui concerne la variation des formants, dans le Script nous avons demandé de mesurer les 5 premiers formants pour la zone délimitée. La valeur du formant donné par Praat est une moyenne calculée sur toute la zone sélectionnée et non pas sur les deux points délimitant la zone sélectionnée. Pour un exemple de segmentation d'un signal, voir annexe n° I.

V.3. Résultats

V.3.1. Durée

Nos résultats montrent la présence d'une différence au niveau de la durée entre les deux termes d'une paire, pour certaines paires brève/longue, mais pas pour toutes. Les durées de certaines voyelles supposées longues sont plus importantes que les durées de certaines partenaires supposées brèves. Ces contrastes temporels sont présents chez les locuteurs de la première et de la deuxième génération seulement. Certains de ces contrastes sont maintenus sous le débit rapide de la parole, mais pas certains autres. Les locuteurs de la troisième génération ne présentent aucune opposition temporelle quelle que soit la qualité, c'est pour cela que nous réservons une étude détaillée de la durée auprès de la première et la deuxième génération, mais pas auprès de la troisième.

V.3.1.1. Première génération

Nous présentons dans le tableau (32) les résultats de la durée pour les trois locuteurs de la première génération. En réalité nous ne pouvons pas mélanger toutes les données provenant des 3 locuteurs parce que leurs réalisations sont très différentes entre elles. Cette contrainte

est fournie par un test statistique qui étudie l'homogénéité des échantillons, ce test permet de savoir si nous pouvons ou non considérer les valeurs comme provenant d'un même locuteur. Nous nous permettons de mettre ensemble toutes les valeurs provenant des 3 locuteurs pour la présentation graphique des résultats, parce qu'une telle présentation n'a pour but que l'illustration des tendances qui restent les mêmes lorsque nous observons chacun de nos locuteurs à part. Nous réservons cependant une présentation détaillée pour les résultats concernant chacun de nos locuteurs avec les études statistiques appropriées. Nous testons pour chacun d'eux, et pour chaque qualité vocalique, les effets des facteurs: Longueur (brève - longue) et Débit (normal - rapide) sur la durée de la voyelle, sur la valeur de F1 et sur la valeur de F2.

Les résultats de nos mesures confirment certaines de nos prédictions, mais pas toutes. En effet, nous nous attendions à ce que les locuteurs de cette tranche d'âge fassent la différence, au niveau de la durée pour toutes les qualités vocaliques qui ont fait l'objet de notre travail de recherche. Une première lecture de nos résultats montre qu'il n'y a pas d'opposition de durée pour les qualités [ɛ] et [u] ni en débit normal ni en débit rapide. En débit normal, la voyelle qui est supposée être la brève pouvant être en effet plus longue que celle qui est censée être la longue.

Plus le rapport (Durée Brève/Durée Longue) est important et moins l'opposition temporelle entre les deux voyelles est importante, s'il est égal à 1, ce qu'il n'y a strictement aucune différence. Ce rapport est le même pour les deux qualités [ɛ] et [u], il est de 1,04. Cela signifie que les voyelles brèves et longues ont sensiblement la même durée. Contrairement à ces deux premières qualités, les deux autres, c'est-à-dire [a] et [o], présentent une opposition temporelle bien marquée entre brèves et longues sous les deux débits. Pour [a], le rapport (Durée Brève/Durée Longue) est de 0,59, cela veut dire que la durée de la brève constitue au peu près 60% de la durée de la longue. Ce rapport est de 0,63 en débit rapide, donc écart moins important. Pour [o] le rapport (Durée Brève/Durée Longue) est plus important que celui que nous avons observé auprès du [a], il est de 0,67 en débit normal et de 0,71 en débit rapide. Le test statistique de Scheffé qui nous permet de comparer les durées des voyelles deux à deux, nous a permis de classer les voyelles en débit normal de la plus longue à la moins longue comme suit¹⁶:

a: > o: > ɛ = ɛ: > u = u: > o > a.

Nous voudrions rappeler ici l'objectif de la soumission de l'émission des voyelles au débit rapide de la parole. Cette méthode, appelée méthode des "perturbations prioritaires" utilisées dans Abry *et al.* (1990), que nous avons appliquée sur l'arabe de Mayadin, permet de tester la robustesse des écarts temporels entre une voyelle brève et sa correspondante longue.

¹⁶ Tous les classements des voyelles en fonction de leurs durées (toujours de la plus longue à la moins longue) dans ce chapitre sont basés sur le même test statistique (Scheffé) qui permet de savoir si les deux durées des voyelles testées sont significativement différentes ou non. Donc le signe '>' signifie différence significative (p < 0,01), le signe '=' dénote l'absence d'une différence significative (p = ns).

Moyenne de la durée de 3 locuteurs âgés de plus de 76 ans

voyelle	durée (ms)	
	normal	rapide
a	100	84
a:	170	134
ɛ	132	105
ɛ:	137	118
ɔ	102	94
o:	153	133
u	126	101
u:	121	105

Tableau 32. Durée des 8 voyelles [a, a:, ɛ, ɛ:, ɔ, o:, u, u:] du français de Savoie pour 2 débits de parole, normal et rapide. Moyenne de 3 locuteurs, deux hommes et une femme âgés de plus de 76 ans en 2007.

Ainsi, le terme "robuste" signifie que l'écart temporel entre la longue et la brève existe sous le débit rapide. Une telle robustesse, si elle existe, permet de savoir, à son tour, si l'écart soumis à la perturbation de l'élocution rapide est contrôlé ou non, parce que son absence marque l'absence du contrôle. Donc la soumission des oppositions au débit rapide sert, en dernier lieu, à savoir quel écart est contrôlé et quel autre ne l'est pas.

Tous locuteurs et toutes qualités vocaliques confondus, nous constatons que le débit rapide abaisse les durées des voyelles longues et brèves, une baisse de 16% en moyenne.

Les voyelles ne se réduisent pas du même pourcentage, elles réagissent différemment. Nous pouvons classer les 8 voyelles en fonction de leur sensibilité (sans test statistique) à l'effet réducteur du débit rapide, de la plus sensible à la moins sensible comme suit:

***a:* > *ɛ:* > *u* > *a* > *ɛ* > *o:* > *u:* > *o*.**

Les représentations graphiques relatives au débit normal et rapide sont disposées dans les figures (71) et (72) respectivement.

Le test post hoc de Scheffé, qui nous permet de comparer les durées des voyelles des locuteurs deux à deux, nous a montré qu'après de la première génération, toutes qualités, toutes longueurs et tous débits confondus, les locuteurs, pris deux à deux, réalisent des voyelles dont les durées sont significativement différentes ($p < 0,01$), ce n'est pas le cas pour la troisième génération.

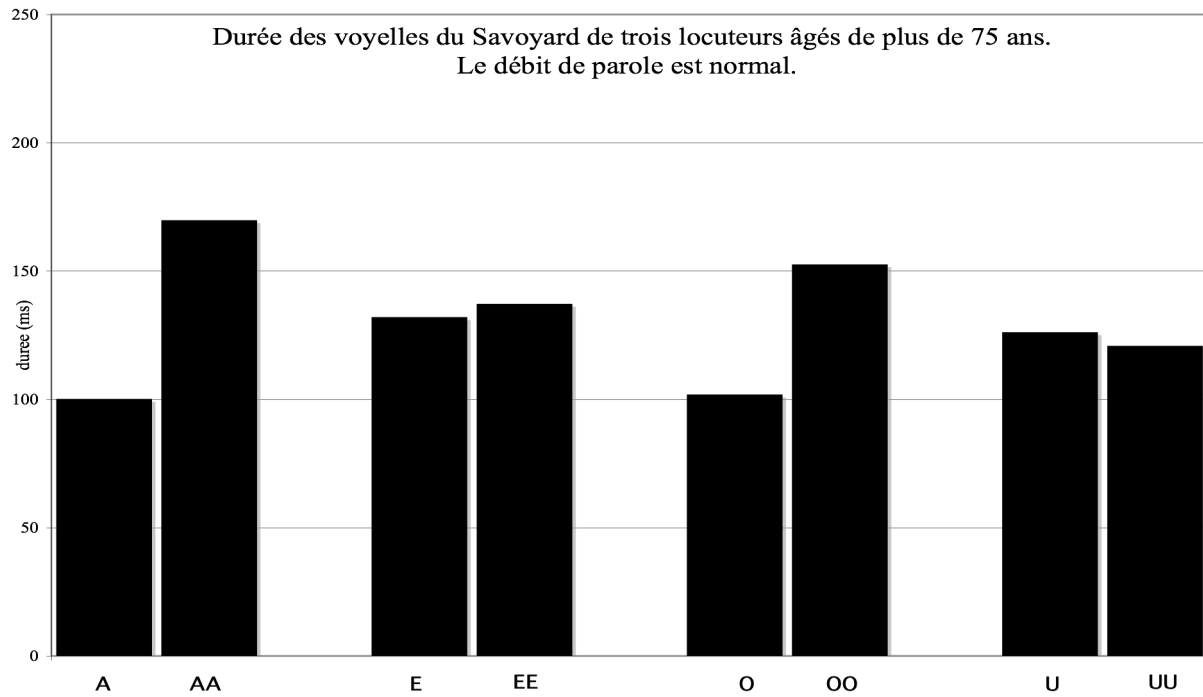


Figure 71. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français de Savoie pour le débit normal. Moyenne de 3 locuteurs, deux hommes et une femme âgés de plus de 76 ans en 2007.

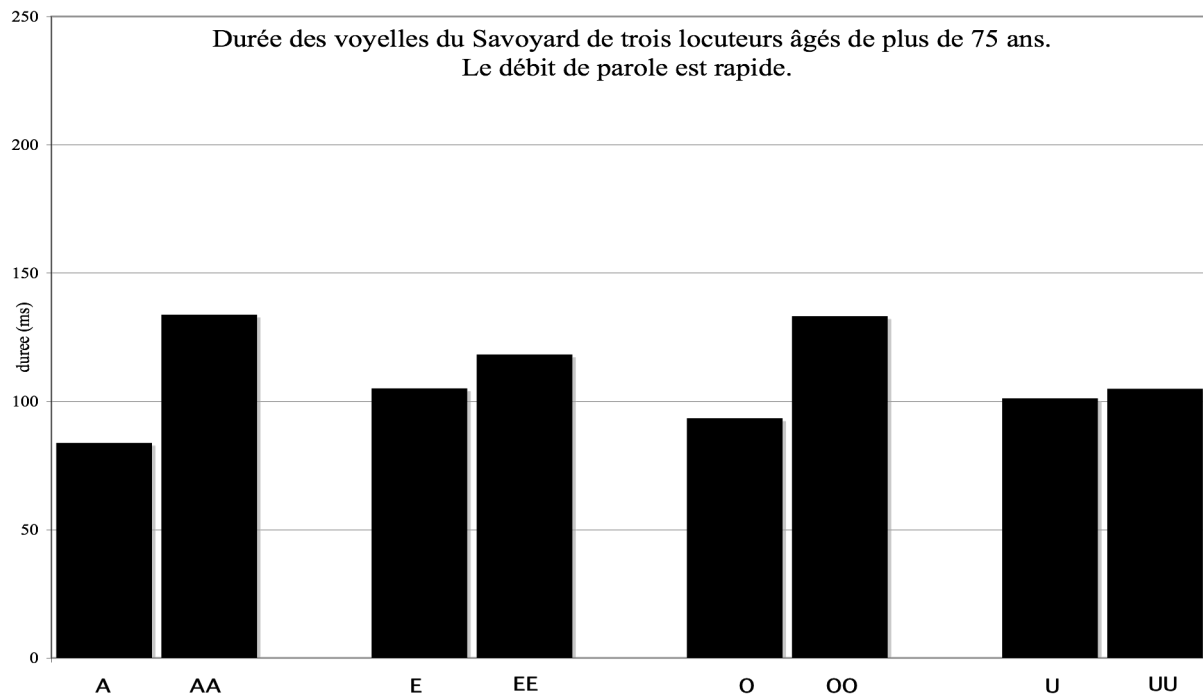


Figure 72. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français de Savoie pour le débit rapide. Moyenne de 3 locuteurs, deux hommes et une femme âgés de plus de 76 ans en 2007.

Pour présenter les résultats de la manière la plus détaillée que nous puissions faire, nous donnons pour chaque locuteur à part un tableau des valeurs numériques de nos mesures sous les deux débits, les données de chaque tableau seront représentées graphiquement sous forme de diagramme en bâtons. Il y aura pour chaque locuteur une conclusion qui est une sorte de résumé de l'essentiel des résultats du locuteur en question. Une représentation sous forme de courbes (plus adaptée au raisonnement statistique) des résultats des mesures de la durée pour les 4 qualités vocaliques étudiées, pour les deux débits de parole et pour chaque locuteur à part est disponible en annexe n° XVI.

Locuteur (Ma).

En débit normal, ce locuteur présente des oppositions très claires pour entre /a/ et /ɑ:/, /ɔ/ et /o:/, alors que ce n'est pas le cas pour les voyelles /ɛ/ et /ɛ:/, /u/ et /u:/, tableau (33). Ces oppositions semblent se maintenir sous la perturbation du débit rapide, une perturbation dont les conséquences sur la durée ne sont pas systématiquement les mêmes, puisque pour certaines voyelles il fait baisser la durée, cas du /o:/, pour d'autres voyelles, il l'augmente, comme pour /ɔ/.

L'analyse de la variance que nous avons effectuée pour tester l'effet de la longueur sur la durée de la voyelle, nous l'avons appliquée pour chacun des sujets de l'expérience, pour chaque débit de parole et pour chaque paire de voyelles, cette analyse a montré qu'après de ce locuteur, les écarts temporels sont significatifs pour /a/ vs. /ɑ:/ ($p < 0,01$), et pour /ɔ/ vs. /o:/ ($p < 0,01$) en débit normal, et que ces 2 écarts demeurent significatifs sous le débit rapide

Locuteur (Ma) - 78 ans

voyelle	durée (ms)	
	normal	rapide
a	73	71
ɑ:	193	160
ɛ	142	128
ɛ:	138	140
ɔ	76	83
o:	160	152
u	99	99
u:	96	102

Tableau 33. Durée des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français de Savoie pour 2 débits de parole, normal et rapide. Locuteur masculin (Ma). Âge: 78 ans en 2007.

($p < 0,01$). Quant aux oppositions / ϵ / vs. / ϵ :/ et /u/ vs. /u:/, elles ne sont pas significatives ni en débit normal ni en débit rapide.

Après du locuteur (Ma), le débit rapide a baissé la durée de 4% en moyenne. Voyelle par voyelle, le débit rapide n'a pas affecté significativement la durée d'aucune des huit voyelles de l'expérience: [a, **a**:, ϵ , ϵ :, ɔ , o:, u et u:]. Nous pouvons classer les voyelles en débit normal de la plus longue à la moins longue comme suit: **a**: > o: > ϵ = ϵ : > u = u: > o > a.

Conclusion pour le locuteur (Ma):

Les oppositions /a/ vs. /**a**:/ et / ɔ / vs. /o:/ existent, elles sont robustes.

Les oppositions / ϵ / vs. / ϵ :/ et /u/ vs. /u:/ n'existent pas.

Le débit rapide n'a pas eu d'effet sur la durée des voyelles.

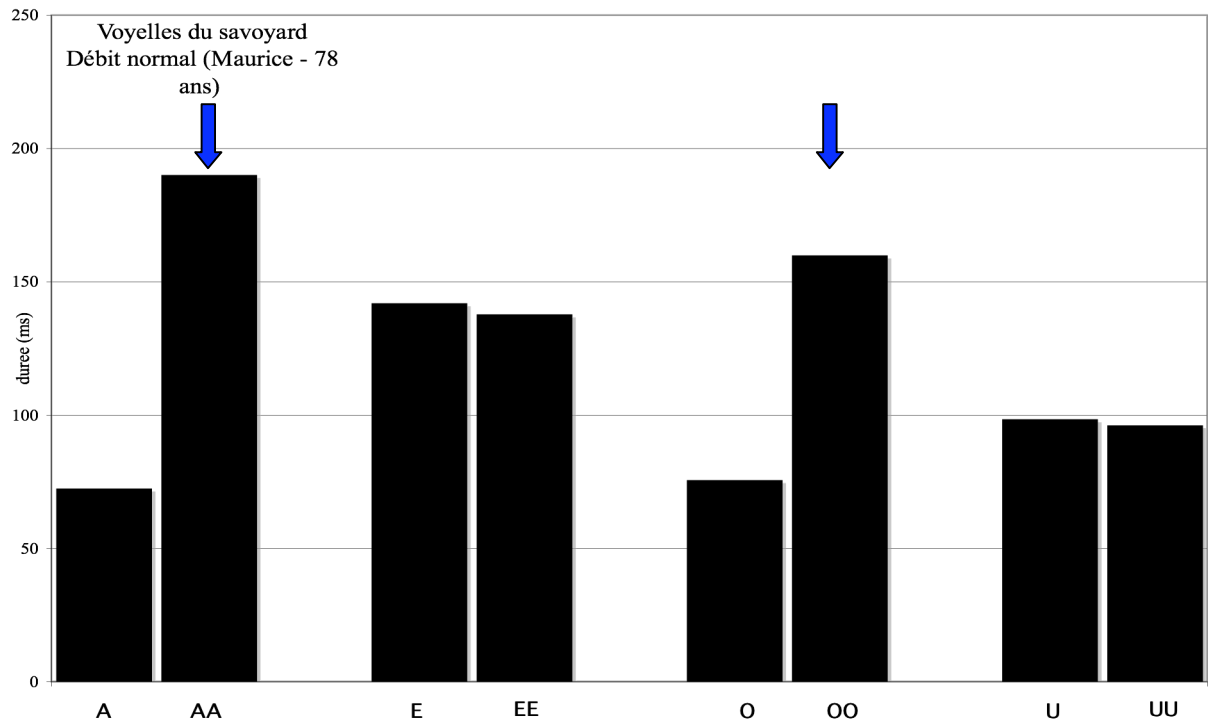


Figure 73. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français de Savoie, le débit est normal. Locuteur masculin (Ma), âge: 78 ans en 2007. Les flèches pointent les qualités pour lesquelles une différence statistiquement significative a été détectée.

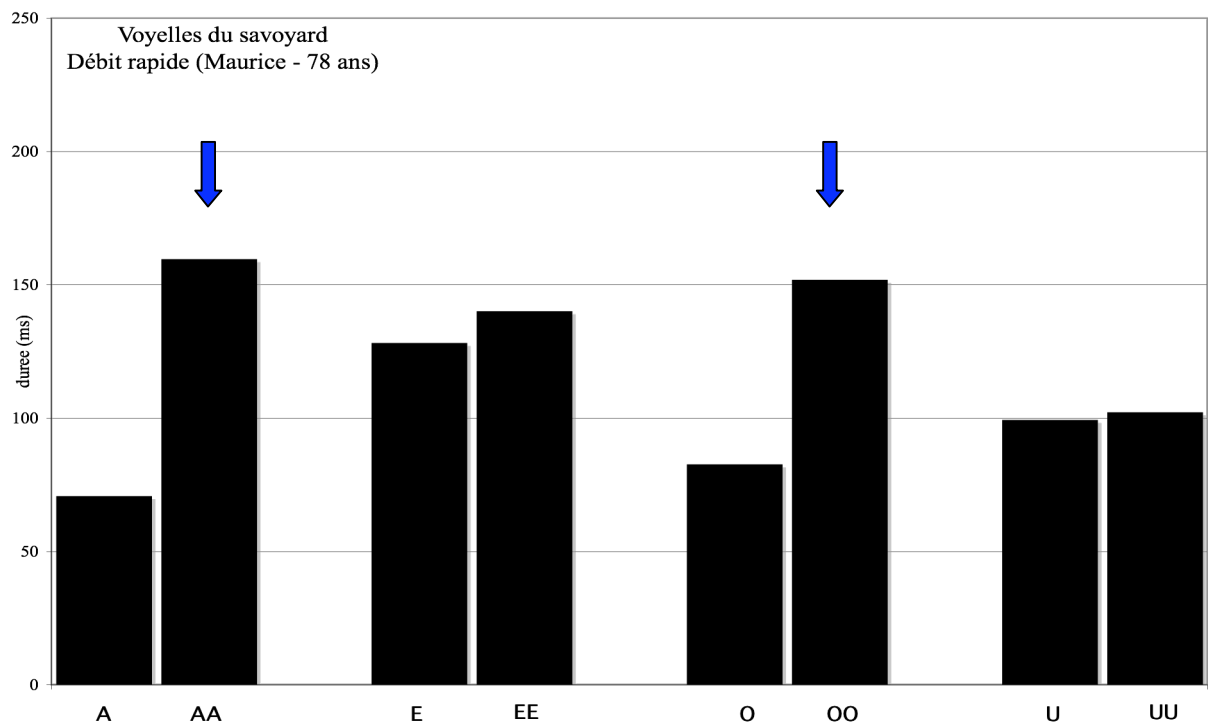


Figure 74. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est rapide. Locuteur masculin (Ma), âge: 78 ans en 2007.

Locuteur (Ge).

En ce qui concerne les oppositions des durées entre voyelles, le locuteur (Ge) montre le même comportement que celui qui était observé auprès du locuteur précédent (Ma), c'est-à-dire qu'en débit normal il présente les oppositions de durée pour /a/ vs. /ɑ:/ et pour /ɔ/ vs. /o:/, mais pas pour /ɛ / vs. /ɛ:/ et pour /u/ vs. /u:/, tableau (34). De même que chez (Ma), ces oppositions se maintiennent sous le débit rapide.

L'analyse de la variance que nous avons effectuée pour tester l'effet de la longueur par locuteur, par débit et par qualité vocalique a montré qu'auprès de ce locuteur, les écarts temporels sont significatifs pour /a/ vs. /ɑ:/ et pour /ɔ/ vs. /o:/ ($p < 0,01$) en débit normal, et que ces écarts restaient significatifs en débit rapide ($p < 0,01$).

Quant aux oppositions /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/, elles ne sont pas significatives ni en débit normal ni en débit rapide.

Contrairement au locuteur (Ma), où la baisse de la durée due au débit rapide n'était pas significative pour aucune des 8 voyelles, pour le locuteur (Ge), où le débit rapide a réduit la durée des voyelles de 12% en moyenne, l'Anova montre que les deux voyelles brèves suivantes: /a/ et /ɔ/ n'étaient pas significativement affectées par le débit rapide, alors que les six autres voyelles [ɑ:, ɛ, ɛ:, o:, u, u:] sont significativement moins longues ($p < 0,01$) en débit rapide.

Nous pouvons classer les voyelles du locuteur (Ge) en débit normal de la plus longue à la moins longue comme suit: *o:* > *ɑ:* > *u = u:* > *ɛ = ɛ:* > *a* > *o*. Nous remarquons que cet ordre est différent de l'ordre que nous avons observé chez le locuteur précédent (Ma), cela explique en partie l'hétérogénéité des données entre les locuteurs de cette génération.

Locuteur (Ge) - 78 ans

voyelle	durée (ms)	
	normal	rapide
a	74	67
ɑ:	113	96
ɛ	86	71
ɛ:	86	77
ɔ	69	66
o:	117	96
u	101	91
u:	98	91

Tableau 34. Durée des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard pour 2 débits de parole, normal et rapide. Locuteur masculin (Ge). Âge: 78 ans en 2007.

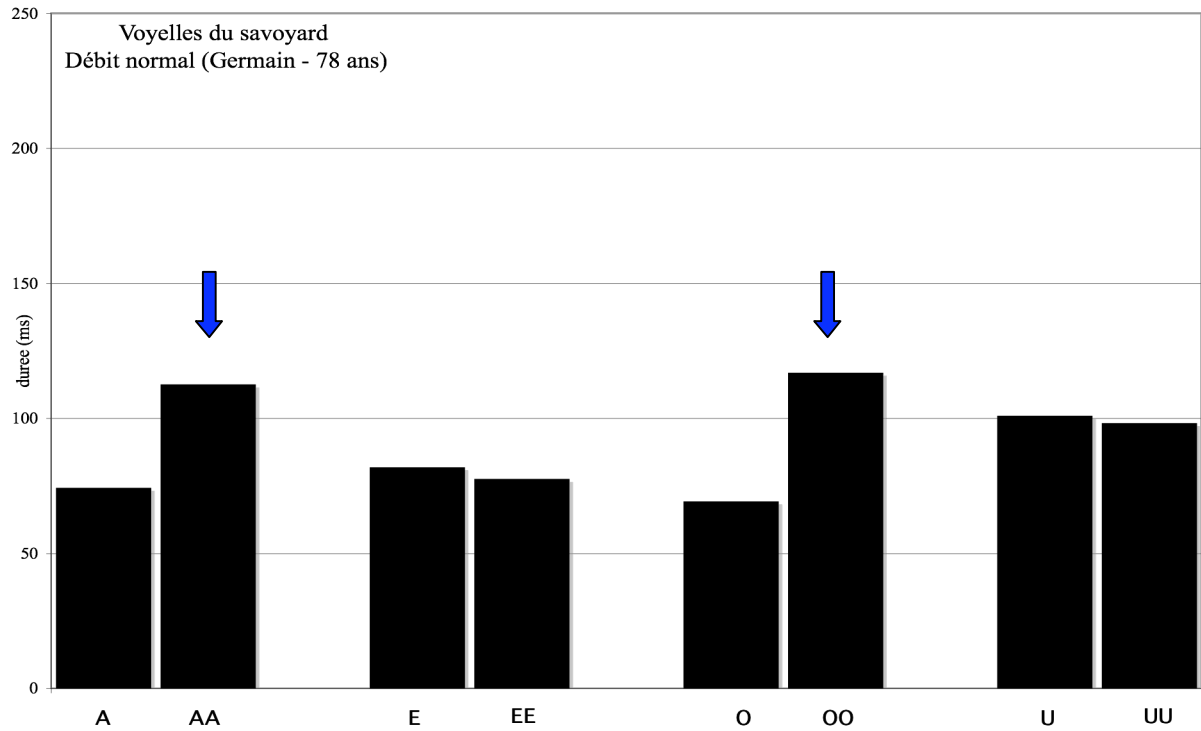


Figure 75. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur masculin (Ge), âge: 78 ans en 2007. Les flèches pointent les qualités pour lesquelles une différence statistiquement significative a été détectée.

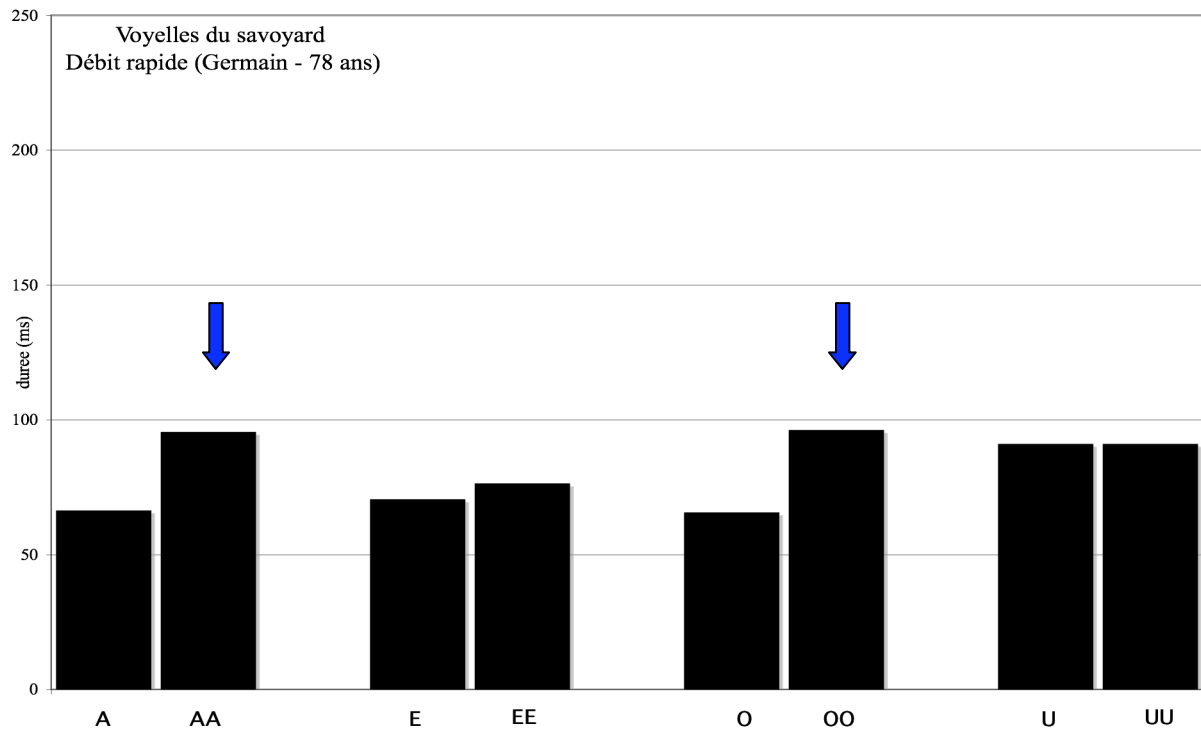


Figure 76. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est rapide. Locuteur masculin (Ge), âge: 78 ans en 2007.

Conclusion pour le locuteur (Ge):

Les oppositions /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ existent, elles sont robustes.

Les oppositions /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ n'existent pas.

Le débit rapide n'a pas eu d'effet sur la durée des voyelles /a/ et /ɔ/. Il a eu de l'effet sur [ɑ:, ɛ, ɛ:, o:, u, u:].

Locutrice (De).

Les durées des voyelles chez la locutrice (De) ne se comportent pas tout à fait de la même manière qu'auprès les locuteurs (Ma) et (Ge), en ce sens que, outre les oppositions trouvées auprès des locuteurs de la même génération, c'est-à-dire /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/, elle oppose /ɛ/ à /ɛ:/ en débit normal et en débit rapide, tableau (35).

L'analyse de la variance que nous avons ciblée pour savoir si la longue des deux voyelles de la paire (brève vs. longue) était réellement plus longue que la brève (autrement dit, l'effet de la longueur sur la durée), par débit et par qualité vocalique, a montré que les écarts temporels étaient significatifs pour /a/ vs. /ɑ:/, /ɛ/ vs. /ɛ:/ et pour /ɔ/ vs. /o:/ en

débit normal. En débit rapide, les écarts /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ restaient significatifs, alors que la différence temporelle entre les deux termes /ɛ/ et /ɛ:/ n'était pas significative (nous

Locutrice (De) - 76 ans

voyelle	durée (ms)	
	normal	rapide
a	148	114
ɑ:	217	147
ɛ	172	117
ɛ:	201	134
ɔ	148	118
o:	189	149
u	176	116
u:	167	120

Tableau 35. Durée des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard pour 2 débits de parole, normal et rapide. Locuteur féminin (De). Âge: 76 ans en 2007.

rappelons que cet écart n'était pas significatif sous aucuns des 2 débits chez les deux locuteurs masculins). Quant à l'opposition /u/ vs. /u:/, elle n'est significative ni en débit normal ni en débit rapide.

La durée des voyelles de la locutrice (De) accuse une baisse importante de 28%. Pour la locutrice (De), l'Anova montre qu'en débit rapide les 8 voyelles sont significativement moins longues qu'en débit normal ($p < 0,01$), elle est ainsi la seule personne de cette génération où la baisse de la durée due au débit rapide était significative pour toutes les voyelles. Nous voudrions mentionner que le locuteur (Ma) qui n'avait aucune qualité de voyelles qui était affectée par le débit rapide, est le frère de la locutrice (De), ils étaient enregistrés dans la même pièce à une demi-heure d'intervalle. Nous pouvons classer les voyelles de locutrice (De) en débit normal de la plus longue à la moins longue comme suit: **a:** > **ɛ:** > **ɛ** > **o:** > **u = u:** > **a** > **o**. Cet ordre est différent de ce que nous avons observé chez les deux autres locuteurs de la même génération.

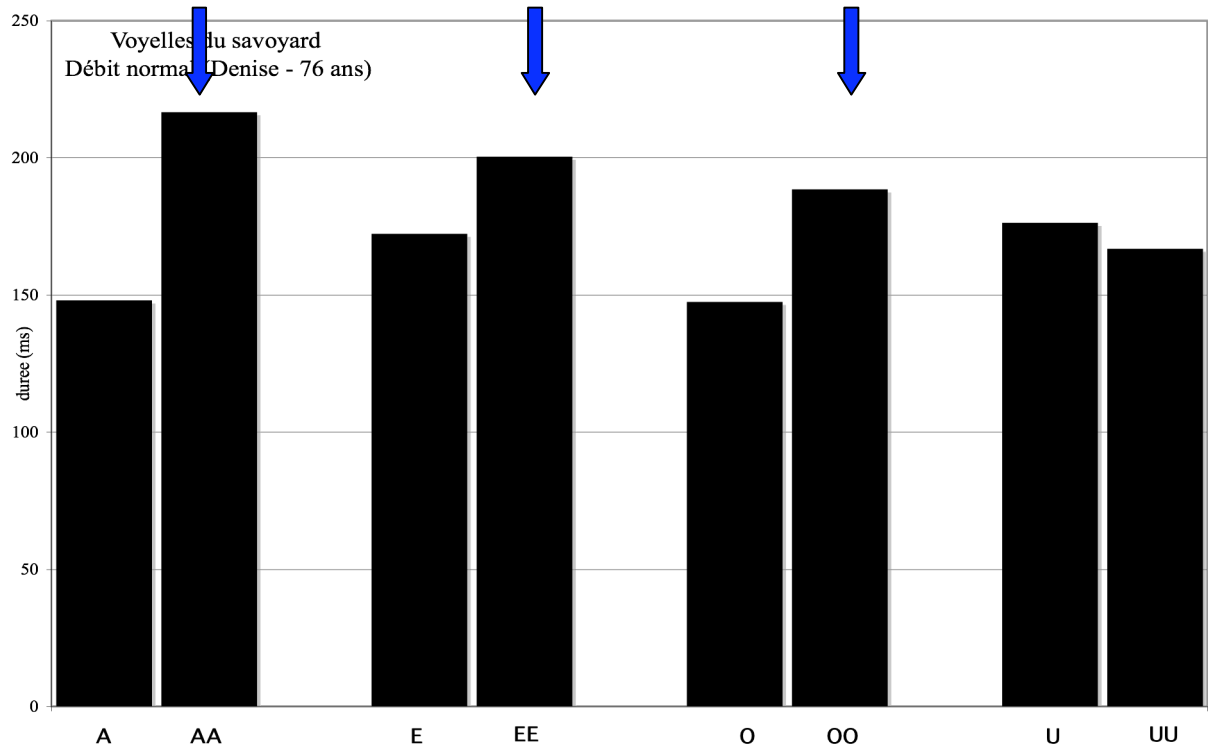


Figure 77. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur féminin (De), âge: 76 ans en 2007. Les flèches pointent les qualités pour lesquelles une différence statistiquement significative a été détectée.

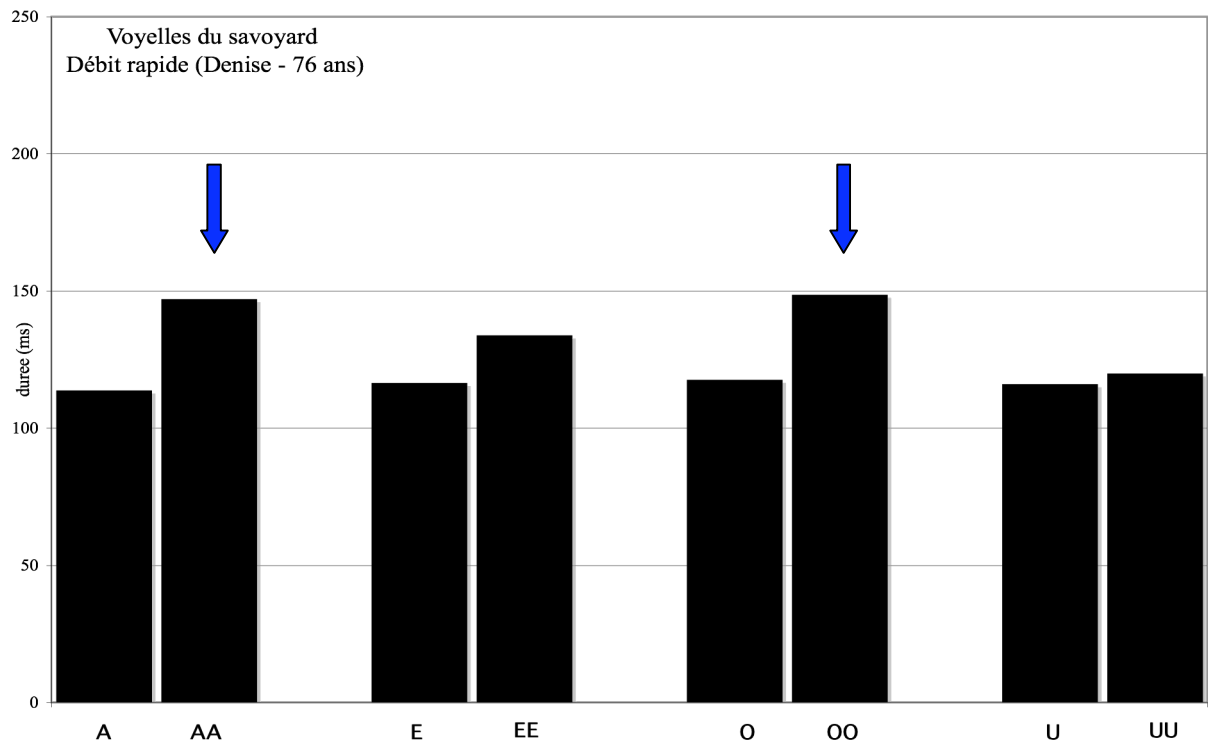


Figure 78. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est rapide. Locuteur féminin (De), âge: 76 ans en 2007.

Conclusion pour la locutrice (De):

Les oppositions /ɛ/ vs. /ɛ:/, /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ existent, seules /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ sont robustes.

L'opposition /u/ vs. /u:/ n'existe pas.

Le débit rapide a réduit la durée de toutes les voyelles de l'expérience.

Résultats de la durée pour la première génération:

Nous avons remarqué que le comportement de la durée chez les trois locuteurs de cette génération sont différents, les durées des voyelles, les étendues des écarts entre brèves et longues et enfin la réaction à l'effet perturbateur du débit rapide (Robuste vs. Non robuste) ne sont pas les mêmes d'un locuteur à l'autre. Nous ne trouvons aucune trace de l'opposition /u/ vs. /u:/, qui aurait disparu à une époque antérieure à celle de la disparition de l'opposition /ɛ/ vs. /ɛ:/ qui est apparu chez la locutrice (De), en débit normal seulement.

Notre lecture des résultats de tous les locuteurs de la première génération dégage cependant quelques points constants, communs à tous les locuteurs:

- 1) Le /a/ bref est significativement moins long que /ɑ:/ long.
- 2) Le /ɔ/ bref est significativement moins long que /o:/ long.
- 3) Il n'y a pas de différence de durée pour les oppositions /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/.

V.3.1.2. Deuxième génération

Nous ne disposons que d'une seule locutrice pour cette génération, parce que notre but était d'observer s'il y avait une transition où certains des contrastes avaient disparu avant d'autres, et non pas l'étude quantitative des durées des voyelles pour cette tranche d'âge intermédiaire où le changement serait en train de s'opérer, contrairement à la troisième génération où l'absence de ce procédé phonétique, si elle est prouvée, serait un phénomène achevé.

Locutrice (S1).

En débit normal, les patrons de la durée chez cette locutrice (S1) ne sont pas tout à fait ceux que nous avons pu observer auprès des locuteurs de la première génération, puisque l'opposition /ɔ/ vs. /o:/, que nous trouvions chez la première génération, nous ne la trouvons pas auprès de cette locutrice. En contrepartie, cette locutrice présente une opposition très faible pour /a/ vs. /ɑ:/, tableau (36).

Le rapport (Durée Brève/Durée Longue) est de 0,91. Pour les trois autres qualités, ce rapport est très proche de 1.

L'analyse de la variance a montré que l'opposition /a/ vs. /ɑ:/ est significative en débit normal ($p < 0,01$), mais pas les trois autres oppositions. En débit rapide, l'opposition /a/ vs. /ɑ:/ demeure significative si nous changeons (seule exception) le seuil de significativité de ($p < 0,01$) à ($p < 0,015$). Par ailleurs, aucune des 3 autres oppositions n'est significative. Cela se situe bien dans nos attentes, puisque nous pensons que la disparition de l'opposition se faisait graduellement. Cette locutrice est peut-être en train de nous renseigner que, du point de vue de la diachronie, l'opposition /ɔ/ vs. /o:/, a disparu à une époque antérieure à celle de la disparition de l'opposition /a/ vs. /ɑ:/.

Nous observons qu'auprès de cette locutrice, le débit rapide n'a pas eu d'effet systématique sur la durée, il a abaissé la durée pour certaines voyelles et il l'a augmenté pour d'autres. L'analyse de la variance révèle que le débit rapide n'a pas eu d'effet sur aucune des durées des 8 voyelles.

Locutrice (S1) - 49 ans		
voyelle	durée (ms)	
	normal	rapide
a	89	96
ɑ:	106	105
ɛ	91	94
ɛ:	90	95
ɔ	84	88
o:	89	92
u	97	91
u:	94	89

Tableau 36. Durée des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard pour 2 débits de parole, normal et rapide. Locuteur féminin (S1). Âge: 49 ans en 2007.

Nous pouvons finalement classer les voyelles de la locutrice (S1) en débit normal de la plus longue à la moins longue comme suit: **a:** > **u = u:** > **ɛ = ɛ:** > **a** > **o = o:**.

Conclusion pour la locutrice (S1):

Seule l'opposition /a/ vs. /a:/ existe, elle est robuste.

Les oppositions /ɔ/ vs. /o:/, /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ n'existent pas.

Le débit rapide n'a pas eu d'effet sur la durée des voyelles.

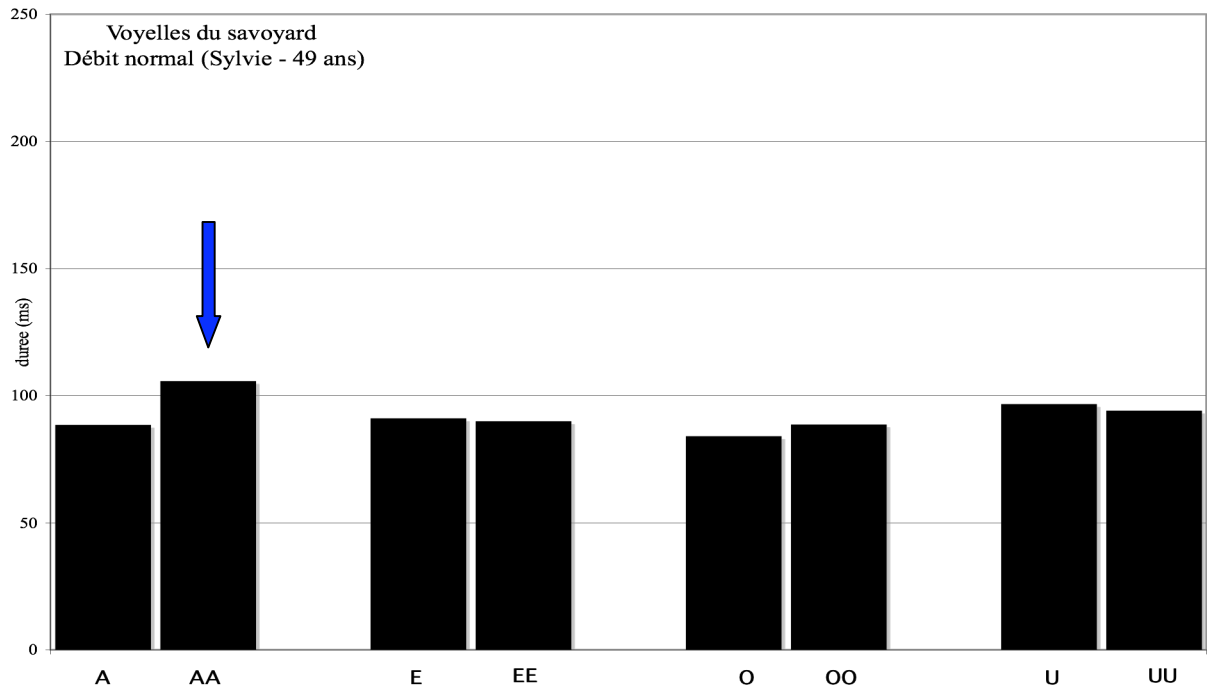


Figure 79. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur féminin (SI), âge: 49 ans en 2007. La flèche pointe la qualité pour laquelle une différence statistiquement significative a été détectée.

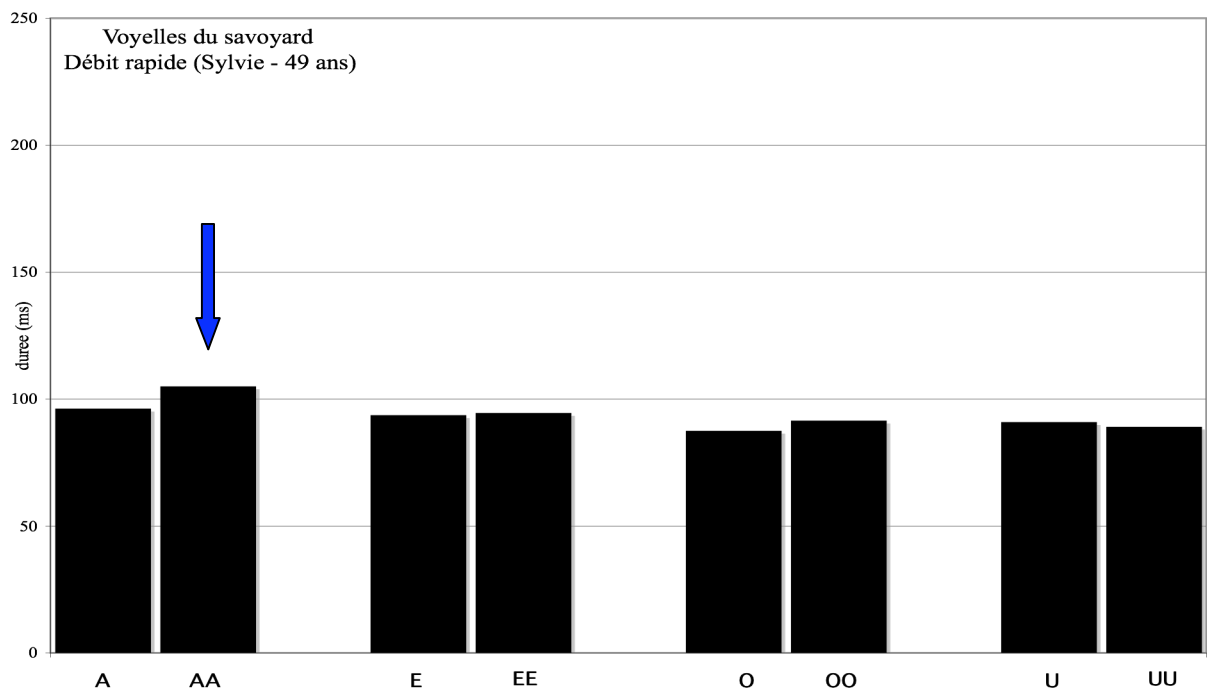


Figure 80. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est rapide. Locuteur féminin (SI), âge: 49 ans en 2007.

V.3.1.3. Troisième génération

Le fait que les valeurs des durées des voyelles chez cette génération soient très proches, ce qui était confirmé par le test d'homogénéité (ce n'était pas le cas pour la première génération), nous a permis de mettre ensemble les données provenant des 2 locuteurs, féminin et masculin, tableau (37).

Nos résultats concernant la troisième génération, constituée de deux locuteurs, montrent qu'il n'y a pas de différence au niveau de la durée des voyelles entre brèves et longues. Celles qui sont supposées brèves sont légèrement plus longues que celles qui sont supposées longues. L'Anova montre cependant qu'il n'y pas de différence significative de durée entre brèves et longues, pour aucune des 4 qualités, chez aucun des locuteurs de cette génération. C'est pour cette raison que nous n'analyserons pas les données sous le débit rapide pour cette génération, notre but étant de tester la robustesse des écarts entre brèves et longues (des écarts qui n'existent pas chez cette génération). Ainsi notre description du français de Savoie pour cette génération s'arrête au débit normal, avec la conclusion que les contrastes temporels qui existaient chez les générations précédentes n'existent plus chez cette génération.

Un regard d'ensemble, sur les résultats obtenus auprès de nos six locuteurs appartenant aux 3 générations, nous laisse penser que nous pouvons émettre l'hypothèse que les 4 oppositions quantitatives /a/ vs. /ɑ:/, /ɛ/ vs. /ɛ:/, /ɔ/ vs. /o:/ et /u/ vs. /u:/ n'ont pas disparu au même moment au cours de l'évolution du français de Savoie vers le français standard. Nous pensons que nous pouvons situer ces disparitions sur une échelle chronologique: l'opposition /u/ vs. /u:/ serait la première à disparaître, en effet, aucun locuteur sous aucun débit n'a manifesté une différence de durée pour cette paire: nous n'avons plus aucune trace de sa

2 locuteurs de la 3e génération.

Moyenne d'âge 23 ans

Voyelle	Durée (ms)
Débit normal	
a	116
ɑ:	115
ɛ	90
ɛ:	99
ɔ	97
o:	96
u	121
u:	118

Tableau 37. Durée des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard avec un débit de parole normal. 2 locuteurs, 1 féminin, âgé de 25 ans et masculin âgé de 22 ans en 2007.

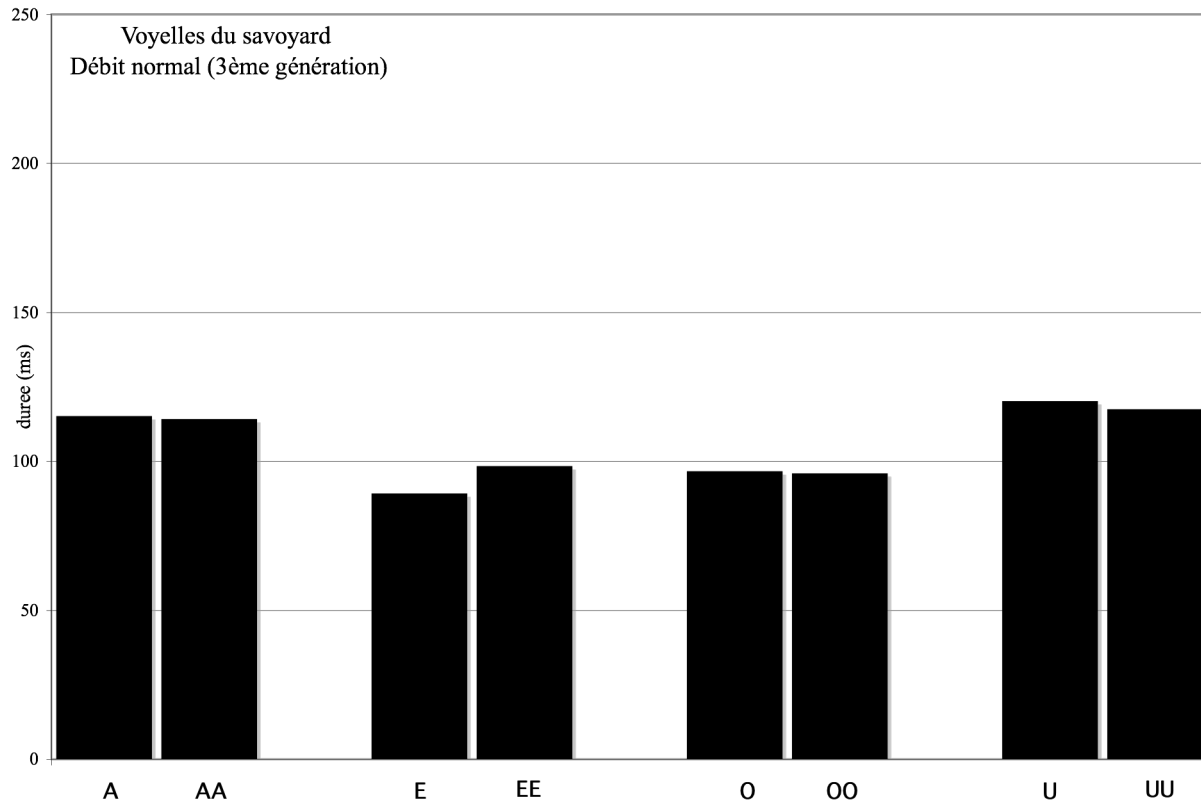


Figure 81. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard pour le débit normal. Moyenne de 2 locuteurs, un homme et une femme âgés de moins de 25 ans en 2007.

présence. Après /u/ vs. /u:/, ce serait l'opposition /ε/ vs. /ε:/ qui a disparu, nous avons remarqué sa présence auprès de la locutrice (De) en débit normal, mais pas en débit rapide. Les deux autres locuteurs de la même génération, bien qu'ils opposent /a/ à /a:/ et /ɔ/ à /o:/ comme la locutrice (De), n'opposent pas /ε/ à /ε:/. La troisième opposition qui a disparu serait l'opposition /ɔ/ vs. /o:/, puisque cette opposition (existante avec la première génération) n'existe pas avec la deuxième génération.

Le test Post hoc de Scheffé, qui nous permet de comparer les durées des voyelles des locuteurs deux à deux, nous a montré qu'auprès de la troisième génération, toutes qualités, toutes longueurs et tous débits confondus, les 2 locuteurs réalisent des voyelles dont les durées ne sont pas significativement différentes (rappel: ce n'était pas le cas pour les locuteurs de la première génération).

Une conclusion générale pour ce paragraphe peut être donnée comme suit:

1) Le /a/ bref est significativement moins long que le /a:/ long chez la première et la deuxième générations (pas chez la troisième génération), sous les deux débits de parole.

2) Le /ɔ/ bref est significativement moins long que le /o:/ long chez la première génération sous les deux débits de parole.

3) À une exception près¹⁷, il n'y a pas de différence de durée pour les 2 oppositions /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ chez aucune génération sous aucun débit

¹⁷ La locutrice (De) oppose /e/ à /e:/ en débit normal, pas en débit rapide.

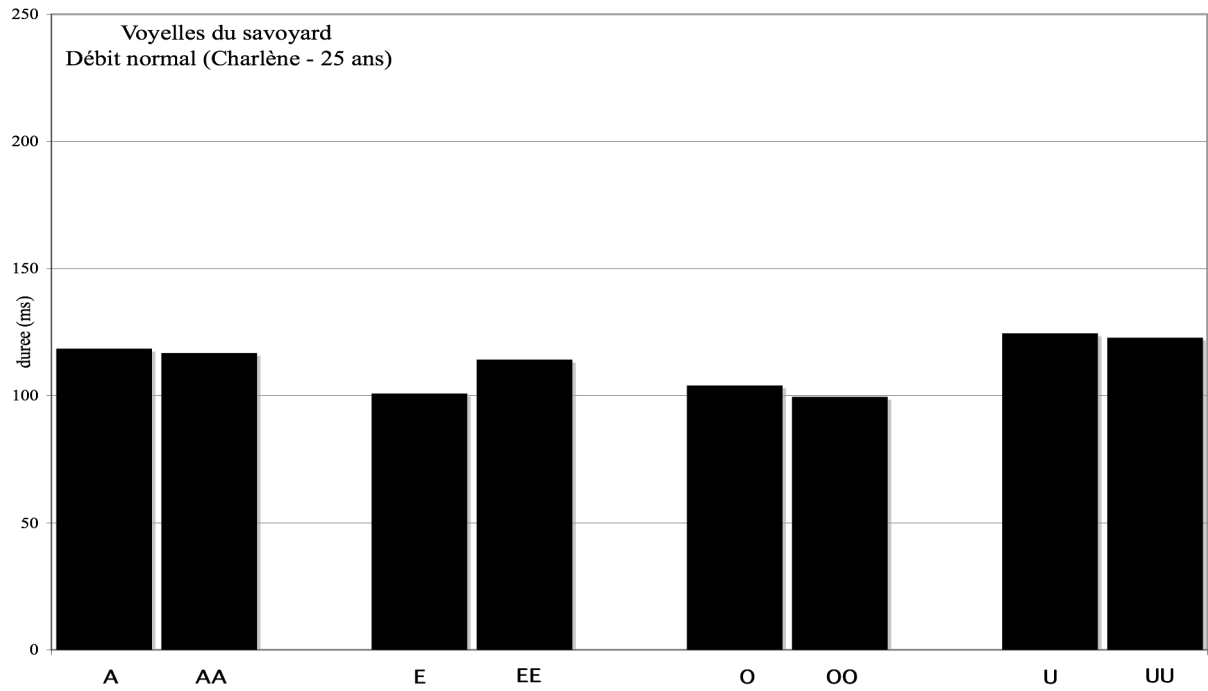


Figure 82. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur féminin (Ch), âge: 25 ans en 2007.

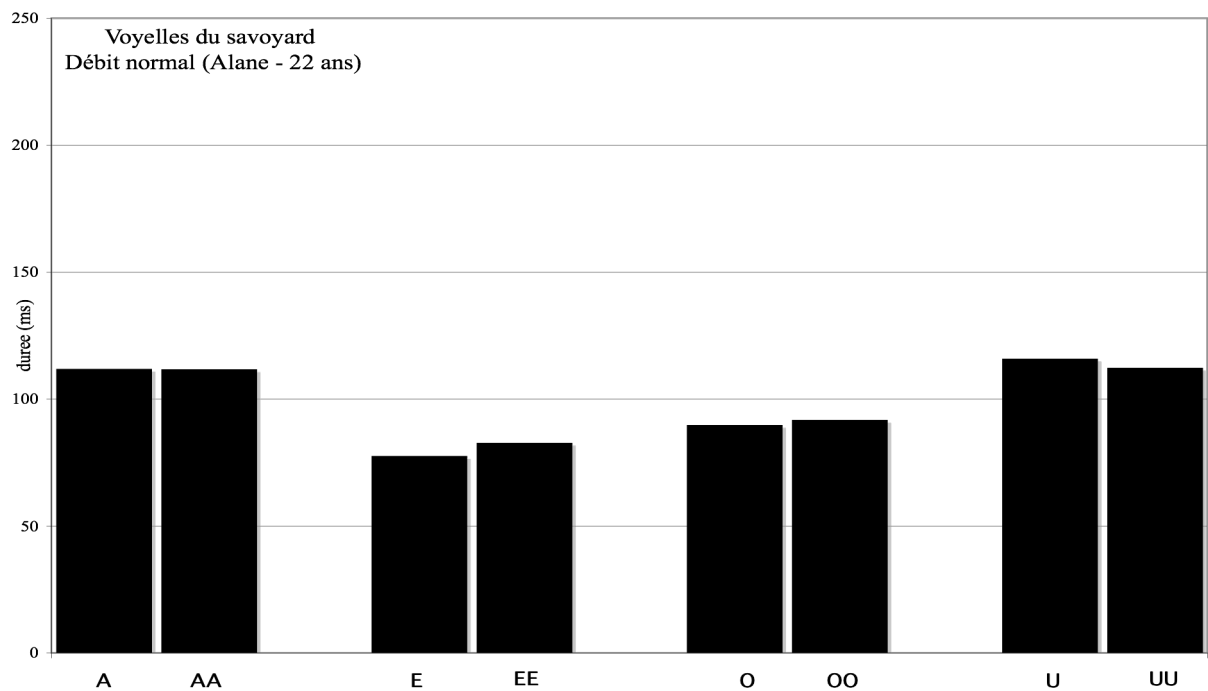


Figure 83. Durée des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur masculin (Al), âge: 22 ans en 2007.

V.3.2. Qualité

V.3.2.1. Première génération

Locuteur (Ma).

L'observation de la figure (84) et de la figure (85), qui représentent l'espace vocalique du locuteur (Ma), valeurs des deux premiers formants en tableau (38), sous le débit normal et sous le débit rapide respectivement (nous avons entouré les qualités pour lesquelles une différence statistiquement significative a été détectée sur toutes les figures qui représentent les valeurs des deux premiers formants), nous laisse penser que ce locuteur n'oppose pas /ɛ/ à /ɛ:/ du point de vue qualitatif, comme il n'oppose pas /u/ à /u:/. En effet, nous remarquons que, pour chaque qualité, les points de dispersion des valeurs formantiques des deux réalisations sont superposés. Les deux réalisations /a/ et /ɑ:/ sont distantes sur l'axe de F2, mais pas sur l'axe de F1, alors que les deux réalisations /ɔ/ et /o:/ sont distantes sur les deux axes, celui de F1 et celui de F2. Vu l'étendue de cette distance, nous nous attendons à ce que ces deux dernières réalisations soient deux qualités distinctes.

Les analyses de la variance qui testent l'effet de la longueur sur le F1 d'un côté et sur le F2 de l'autre pour chaque paire, chaque locuteur et chaque débit, montre que pour ce locuteur, en débit normal, les paires /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ ne sont pas différents ni sur F1 ni sur F2, et que cela demeure vrai sous le débit rapide. Concernant la paire /a/ vs. /ɑ:/, l'analyse de la variance donne significative la différence au niveau de F2 ($p < 0,01$); mais pas au niveau de F1. Le débit rapide ne change rien à ce portrait. Conformément à nos attentes, les deux réalisations de la paire /ɔ/ vs. /o:/ sont différentes et au niveau de F1 et au niveau de F2 ($p <$

	Débit normal	
	F1	F2
a	618	1311
ɑ:	625	1065
ɛ	446	1681
ɛ:	451	1665
ɔ	497	1187
o:	362	906
u	292	890
u:	295	910
	Débit rapide	
	F1	F2
a	646	1402
ɑ:	627	1047
ɛ	515	1698
ɛ:	493	1692
ɔ	516	1076
o:	431	914
u	371	794
u:	361	794

Tableau 38. Moyennes des valeurs formantiques (F1 et F2) des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard pour 2 débits de parole, normal et rapide. Locuteur masculin (Ma). Âge: 78 ans en 2007.

0,01), ces différences demeurent significatives sous le débit rapide. Auprès du locuteur (Ma), le débit rapide a affecté significativement ($p < 0,01$) les valeurs des F1 des voyelles suivantes: [ɛ, ɛ:, o:, u, u:] pas les F1 des autres, c'est-à-dire, [a, a:, o]. Il a affecté significativement aussi les F2 des voyelles suivantes: [a, ɔ, u, u:] seulement.

Conclusion pour le locuteur (Ma):

L'opposition /a/ vs. /a:/ existe seulement au niveau de F2, elle est robuste. Elle n'existe pas au niveau de F1.

L'opposition /ɔ/ vs. /o:/ existe et au niveau de F1 et au niveau de F2, elle est robuste pour les deux formants.

Les oppositions /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ n'existent pas, ni sur F1 ni sur F2.

Le débit rapide a eu de l'effet, non systématique, sur certaines valeurs de F1 et de F2 sans changer pour autant les distances obtenues avec le débit normal.

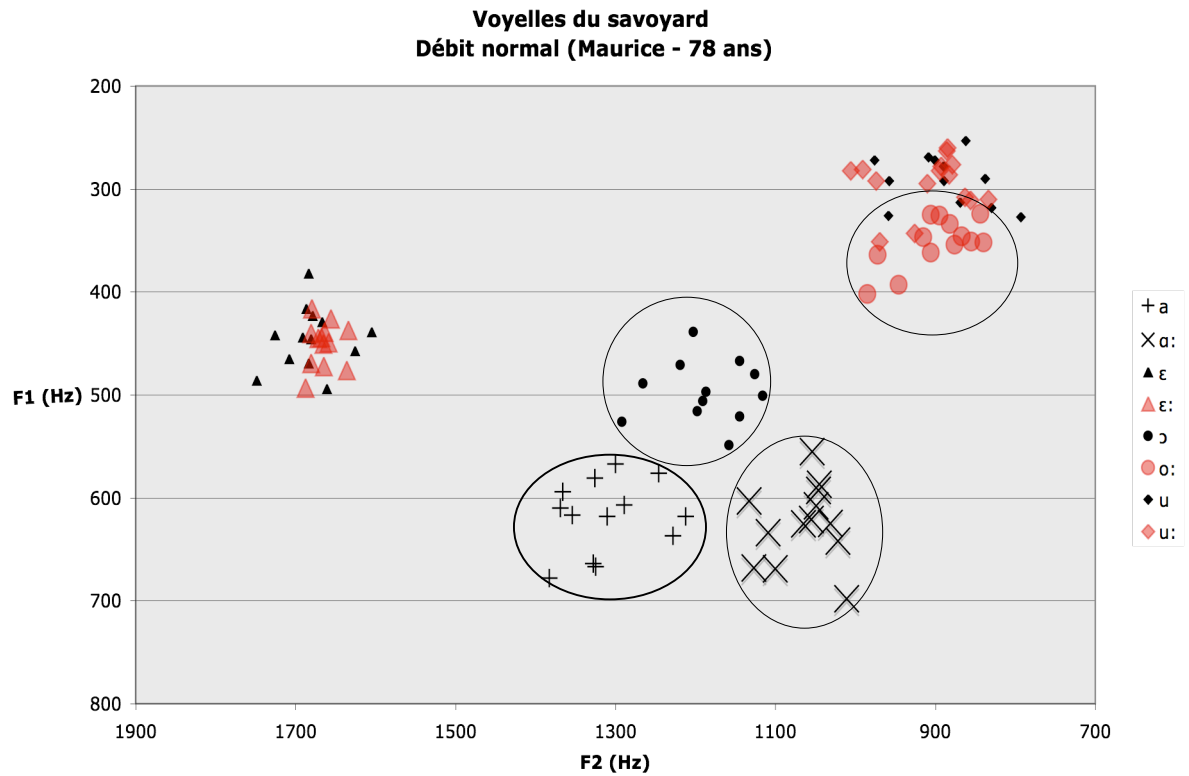


Figure 84. Espace vocalique des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur masculin (Ma), âge: 78 ans en 2007.

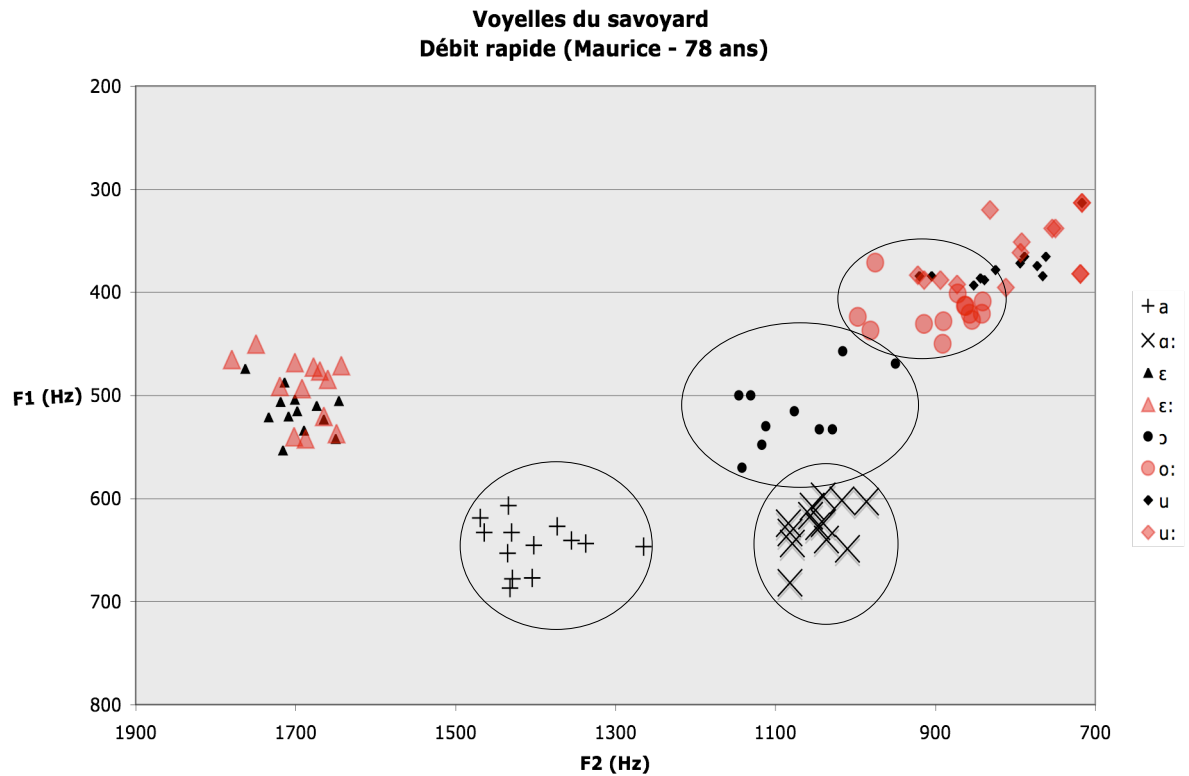


Figure 85. Espace vocalique des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard, le débit est rapide. Locuteur masculin (Ma), âge: 78 ans en 2007.

<u>Locuteur (Ge).</u>	Débit normal		
	F1	F2	
Les patrons des formants chez le locuteur (Ge) tableau (39), sont similaires à ce que nous avons pu observer auprès du locuteur (Ma). En ce sens qu'il n'oppose pas /ɛ/ à /ɛ:/, comme il n'oppose pas /u/ à /u:/, et que les deux réalisations /a/ et /ɑ:/ montrent une différence pour F2, mais pas pour F1, alors que les deux réalisations /ɔ/ et /o:/ sont différentes au niveau des deux formants, figures (86) et (87). L'analyse statistique confirme la lecture graphique de nos résultats, puisque l'Anova montre une différence significative au niveau de F2 seulement ($p < 0,01$) pour /a/ vs. /ɑ:/ et une différence significative pour F1 et pour F2 ($p < 0,01$) pour /ɔ/ vs. /o:/. Les deux réalisations /u/ et /u:/ont leurs F2 proches de la signification. Finalement, la réalisation /u/ n'est pas différente de la réalisation /u:/, ni sur F1 ni sur F2, idem pour les deux réalisations /ɛ/ et /ɛ:/. Bien que le débit rapide ait eu un effet significatif sur les valeurs des F1 des réalisations [a, ɑ:, ɛ:, o:, u:] et sur les valeurs des F2 des réalisations [a, ɑ:, o:, u], l'image de l'espace vocalique observée en débit normal reste la même en débit rapide, c'est-à-dire que les différences significatives sont restées significatives, et celles qui ne l'étaient pas sont restées non significatives.	a	650	1308
	ɑ:	582	1033
	ɛ	466	1548
	ɛ:	474	1531
	ɔ	499	1298
	o:	386	850
	u	327	968
	u:	327	1009
		Débit rapide	
		F1	F2
	a	715	1400
	ɑ:	715	1128
	ɛ	501	1532
	ɛ:	501	1543
	ɔ	478	1347
	o:	412	983
u	344	1051	
u:	350	1040	

Tableau 39. Moyennes des valeurs formantiques (F1 et F2) des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard pour 2 débits de parole, normal et rapide. Locuteur masculin (Ge). Âge: 78 ans en 2007.

Conclusion pour le locuteur (Ge):

L'opposition /a/ vs. /ɑ:/ existe seulement au niveau de F2, elle est robuste. Elle n'existe pas au niveau de F1.

L'opposition /ɔ/ vs. /o:/ existe et au niveau de F1 et au niveau de F2, elle est robuste pour les deux formants.

Les oppositions /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ n'existent pas, ni sur F1 ni sur F2.

Le débit rapide a eu de l'effet, non systématique, sur certaines valeurs du F1 et du F2 sans changer pour autant les distances obtenues avec le débit normal.

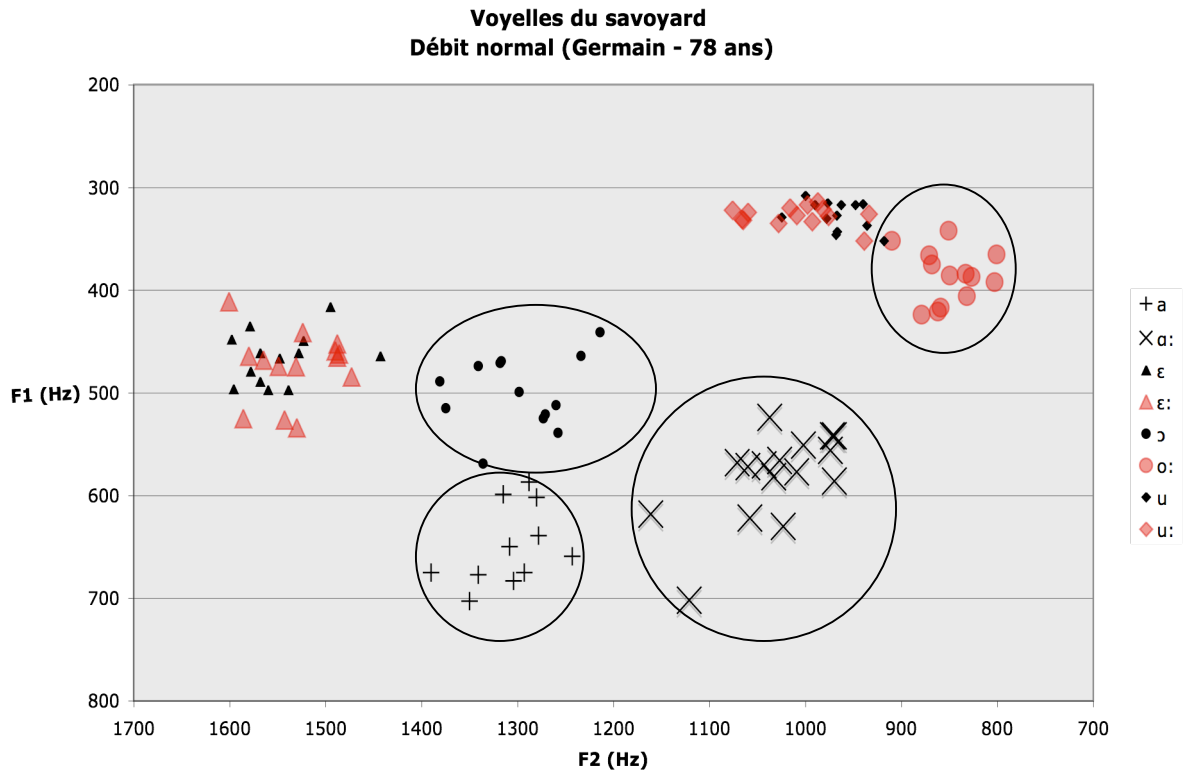


Figure 86. Espace vocalique des 8 voyelles [a, ɑ:, ε, ɛ:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur masculin (Ge), âge: 78 ans en 2007.

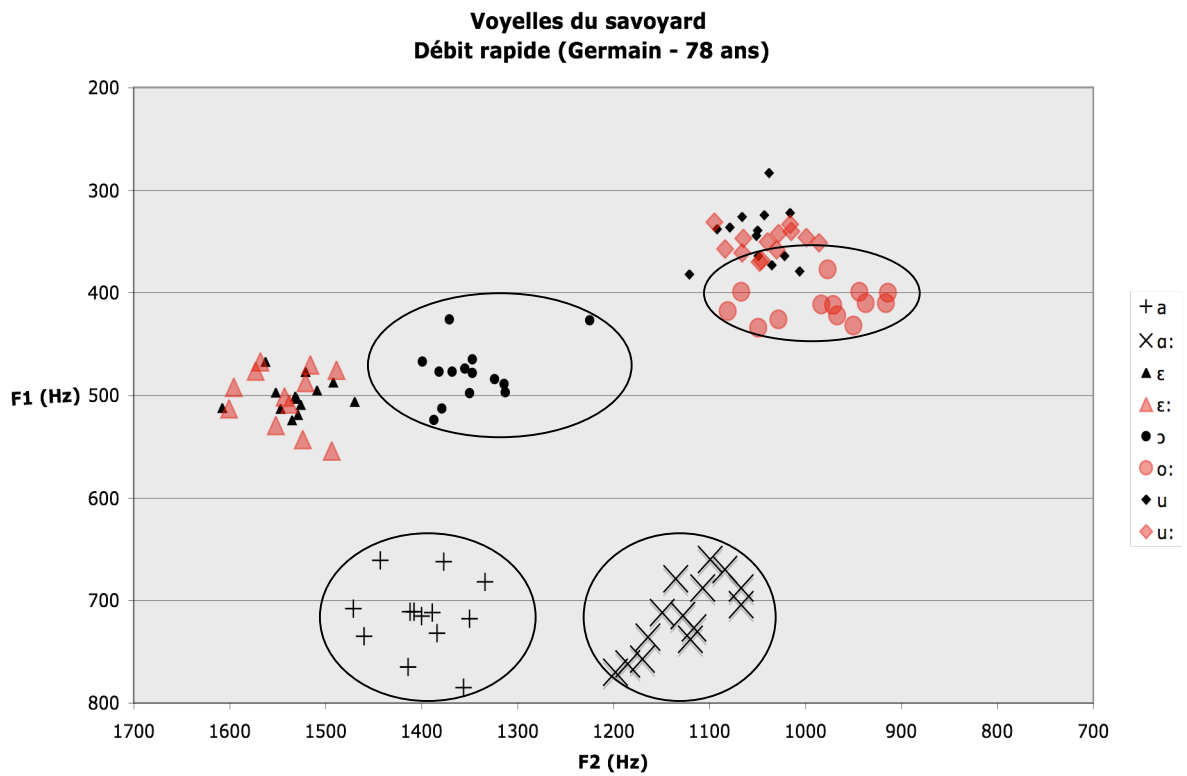


Figure 87. Espace vocalique des 8 voyelles [a, ɑ:, ε, ɛ:, ɔ, o:, u, u:] du français savoyard, le débit est rapide. Locuteur masculin (Ge), âge: 78 ans en 2007.

Locutrice (De).

Notre troisième, et dernière, locutrice de la première génération montre le même comportement, tableau (40), en ce qui concerne les formants, que nous avons trouvé auprès des locuteurs (Ma) et (Ge).

Elle ne montre pas de différence entre le F1 et le F2 de la réalisation /ɛ/ et le F1 et le F2 de la réalisation /ɛ:/, autrement dit, /ɛ/ et /ɛ:/ sont une seule qualité vocalique et non pas deux.

Elle n'oppose pas /u/ à /u:/ non plus, ni sur l'axe de F1 ni sur l'axe de F2, ce qui nous permet de confirmer, maintenant que nous avons les résultats des 3 locuteurs, que, comme pour /ɛ/ et /ɛ:/, /u/ et /u:/ sont une seule qualité vocalique et non pas deux. Les deux réalisations /a/ et /ɑ:/ montrent une différence au niveau de F2, mais pas au niveau de F1, alors que les deux réalisations /ɔ/ et /o:/ sont différentes au niveau des deux formants, figures (88) et (89).

L'analyse statistique va de pair avec ce que nous avons pu remarquer, en effet, l'Anova montre, malgré le fait que le /ɛ/ a une zone de dispersion plus importante que le /ɛ:/, en débit normal et en débit rapide, que ces deux réalisations ne sont pas différentes sur aucun axe et avec aucun débit. Les deux réalisations /u/ et /u:/ ne montrent pas non plus, sous les 2 débits de parole, aucune différence statistique. Nous notons une différence significative, en débit normal, au niveau de F2 ($p < 0,01$), pour /a/ vs. /ɑ:/, le F1 ne réalise pas une différence significative. Le débit rapide ne change pas cette configuration ni pour F1 ni pour F2. Une autre différence significative que nous pouvons noter est celle que nous trouvons pour F1 et pour F2 ($p < 0,01$) pour l'opposition /ɔ/ vs. /o:/, les deux réalisations sont effectivement distinctes du point de vue statistique au niveau des deux formants. Le débit rapide n'a quasiment pas eu d'effet significatif sur les valeurs des formants, aucun pour F1, seulement sur F2 des deux réalisations [ɔ, u].

	Débit normal	
	F1	F2
a	635	1597
ɑ:	650	1300
ɛ	456	2041
ɛ:	444	2076
ɔ	456	1244
o:	378	936
u	267	972
u:	264	1004
	Débit rapide	
	F1	F2
a	626	1596
ɑ:	613	1359
ɛ	490	2044
ɛ:	478	2095
ɔ	478	1122
o:	408	970
u	268	1036
u:	263	987

Tableau 40. Moyennes des valeurs formantiques (F1 et F2) des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard pour 2 débits de parole, normal et rapide. Locuteur féminin (De). Âge: 76 ans en 2007.

Conclusion pour la locutrice (De):

L'opposition /a/ vs. /ɑ:/ existe seulement au niveau de F2, elle est robuste. Elle n'existe pas au niveau de F1.

L'opposition /ɔ/ vs. /o:/ existe et au niveau de F1 et au niveau de F2, elle est robuste pour les deux formants.

Les oppositions /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ n'existent pas, ni sur F1 ni sur F2.

Le débit rapide n'a pas eu d'effet sur les F1, a eu très peu d'effet sur les F2, il n'a pas changé les distances obtenues avec le débit normal.

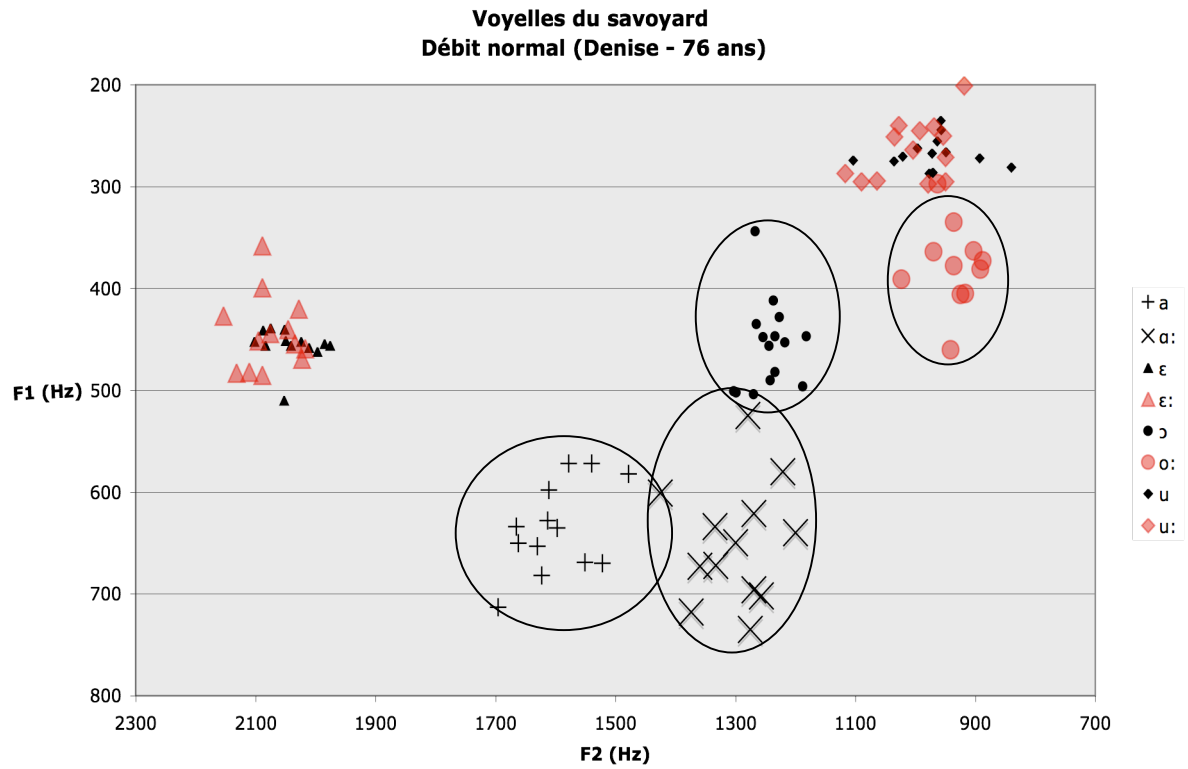


Figure 88. Espace vocalique des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur féminin (De), âge: 76 ans en 2007.

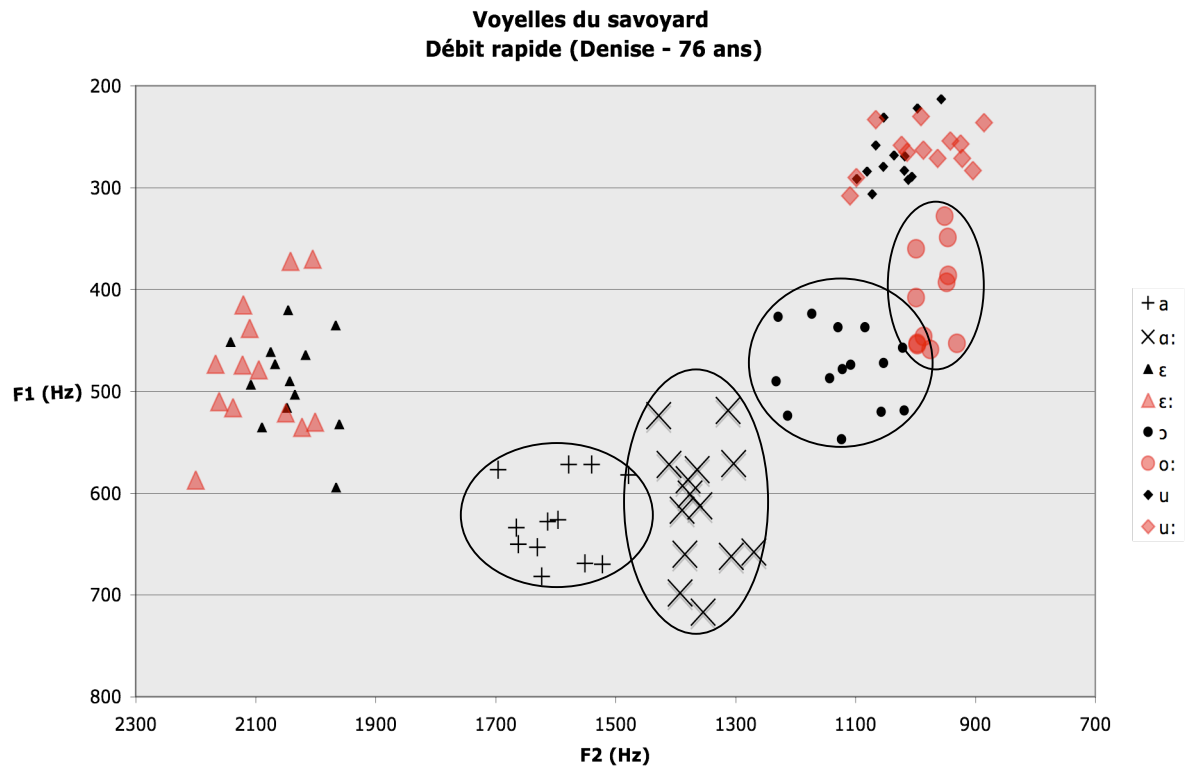


Figure 89. Espace vocalique des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur féminin (De), âge: 76 ans en 2007.

V.3.2.2. Deuxième générationLocutrice (S1).

Notre seule locutrice de la deuxième génération qui, lors de l'opposition quantitative, s'est démarquée de la première génération en ce qui concerne l'opposition /ɔ/ vs. /o:/ pour rejoindre la troisième génération (pas d'opposition), alors qu'elle rejoignait la première génération pour l'opposition /a/ vs. /ɑ:/ (opposition robuste). (S1) montre exactement le même comportement que les locuteurs de la première génération en ce qui concerne l'opposition qualitative, tableau (41). Nous pouvons en effet observer sur la représentation graphique de l'espace vocalique de cette locutrice deux nuages de points assez bien séparés pour les deux réalisations /ɔ/ et /o:/ sur l'axe de F1 et sur l'axe de F2. Les 2 nuages de points ne sont séparés que sur l'axe de F2 pour les deux réalisations /a/ et /ɑ:/. Alors que pour les réalisations /u/ et /u:/, et pour les réalisations /ɛ/ et /ɛ:/, nous observons dans l'ensemble une superposition des points, figures (90) et (91).

Nous notons que le débit rapide a eu pour effet d'élargir la dispersion. L'analyse statistique montre que le débit rapide a affecté significativement ($p < 0,01$) les F1 des réalisations [a, ɑ:, ɛ, ɛ:] et les F2 des réalisations [ɛ, ɔ, u:], sans modifier la significativité des oppositions.

L'Anova montre que les deux réalisations /u/ et /u:/ ne sont pas différentes entre elles sur aucun axe et avec aucun débit. Le constat est le même pour les réalisations /ɛ/ et /ɛ:/.

	Débit normal	
	F1	F2
a	659	1668
ɑ:	648	1510
ɛ	562	1901
ɛ:	547	1933
ɔ	534	1413
o:	420	1033
u	361	1146
u:	336	1151
	Débit rapide	
	F1	F2
a	769	1664
ɑ:	770	1551
ɛ	489	1882
ɛ:	490	1881
ɔ	526	1523
o:	426	1127
u	381	1054
u:	350	903

Tableau 41. Moyennes des valeurs formantiques (F1 et F2) des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard pour 2 débits de parole, normal et rapide. Locuteur féminin (S1). Âge: 49 ans en 2007.

Cette locutrice (S1) présente une différence significative, en débit normal, au niveau de F2 ($p < 0,01$), pour /a/ vs. /ɑ:/. Le F1 ne réalise pas une différence significative. Le débit rapide ne change pas cette configuration ni pour F1 ni pour F2. Malgré le rapprochement significatif entre les dispersions des deux réalisations /ɔ/ et /o:/ en débit rapide, elles restent significativement différentes ($p < 0,01$) sur F1 et sur F2.

Conclusion pour la locutrice (S1):

L'opposition /a/ vs. /ɑ:/ existe seulement au niveau de F2, elle est robuste. Elle n'existe pas au niveau de F1.

L'opposition /ɔ/ vs. /o:/ existe et au niveau de F1 et au niveau de F2, elle est robuste pour les deux formants.

Les oppositions /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ n'existent pas, ni sur F1 ni sur F2.

Le débit rapide a eu très peu d'effet, sur les F2 seulement, sans changer les distances obtenues avec le débit normal.

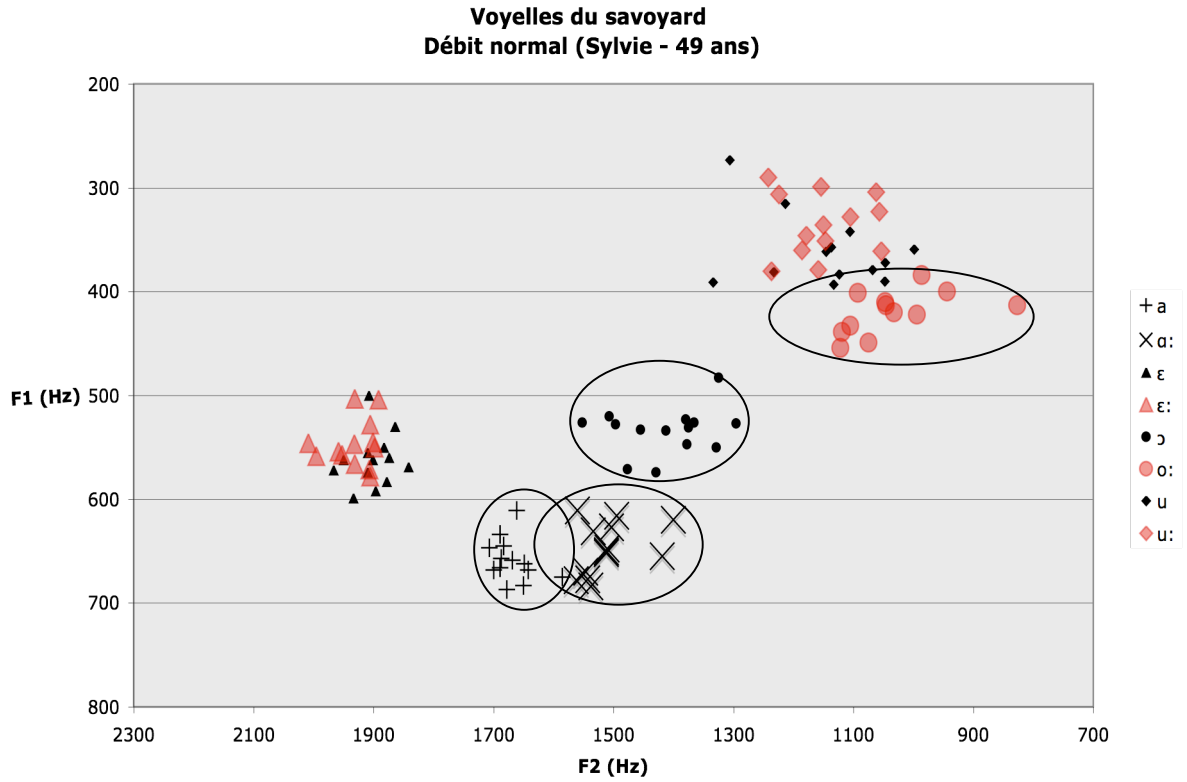


Figure 90. Espace vocalique des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur féminin (S1), âge: 49 ans en 2007.

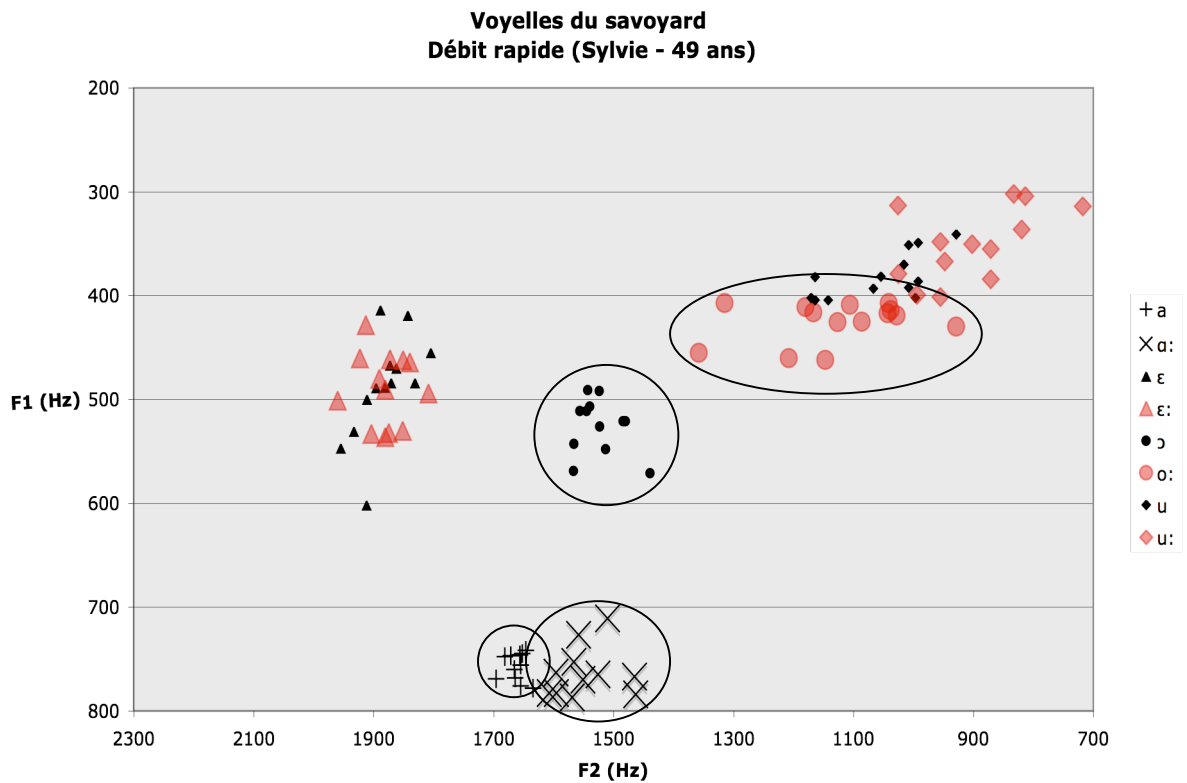


Figure 91. Espace vocalique des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur féminin (S1), âge: 49 ans en 2007.

V.3.2.3. Troisième génération

Les 2 locuteurs de cette génération, comme nous avons pu le constater plus haut, n'opposent pas de voyelles par leurs durées. C'est pour cette raison que nous ne nous sommes pas intéressés à l'émission des 8 voyelles de l'expérience sous le débit rapide par cette génération, parce que, rappelons-le, la soumission des voyelles à cette condition expérimentale a pour but de tester la robustesse des écarts quantitatifs et qualitatifs pour une paire de voyelles présentant un contraste temporel. Ainsi notre exposition des résultats des valeurs des formants, tableau (42), ne portera que sur les valeurs obtenues en débit normal.

L'analyse de la variance qui a testé les différences de qualité au sein de chacune des 4 paires (brève vs. longue) montre que les deux réalisations /ɔ/ et /o:/ sont différentes ($p < 0,01$)

entre elles et au niveau de F1 et au niveau de F2, chez les deux locuteurs. Ce n'est pas le cas pour les réalisations des 3 autres paires.

L'espace vocalique du locuteur féminin, figure (92), est nettement plus étendu que celui du locuteur masculin, figure (93). La proximité de certaines voyelles selon F1-F2 chez ce dernier a attiré notre curiosité, et nous nous sommes bien demandés si le /a/ est bien différent du /ɔ/. Nous avons comparé alors les voyelles "supposées" longues deux à deux et les voyelles "supposées" brèves deux à deux chez ce locuteur. Le F1 du /ɛ/, le F1 du /a/ et celui du /ɔ/ ne sont pas différents, alors que le F2 du /a/ et celui du /ɔ/, seulement, ne sont pas différents, ce qui signifie, comme nous avons pu le remarquer visuellement sur le graphique, que les deux réalisations /a/ et /ɔ/ ne sont pas différentes chez ce locuteur. Tous les F1 et les F2 des longues sont deux à deux différents.

Locutrice (Ch). Débit normal.

	F1	F2
a	811	1793
ɑ:	803	1786
ɛ	487	2225
ɛ:	462	2239
ɔ	482	1623
o:	370	1188
u	291	1123
u:	276	1098

Locuteur (Al). Débit normal.

	F1	F2
a	451	1439
ɑ:	474	1453
ɛ	428	1743
ɛ:	436	1761
ɔ	415	1384
o:	345	1155
u	296	1018
u:	287	1026

Tableau 42. Moyennes des valeurs formantiques (F1 et F2) des 8 voyelles [a, ɑ:, ɛ, ɛ:, ɔ, o:, u, u:] du français savoyard pour 2 locuteurs avec un débit de parole normal. Locuteur féminin (Ch). Âge: 25 ans en 2007. Locuteur masculin (Al). Âge: 22 ans en 2007.

Conclusion pour la troisième génération:

L'opposition /ɔ/ vs. /o:/ existe et au niveau de F1 et au niveau de F2 chez les deux locuteurs, (Ch) et (Al), de cette génération. Aucune autre opposition qualitative n'existe.

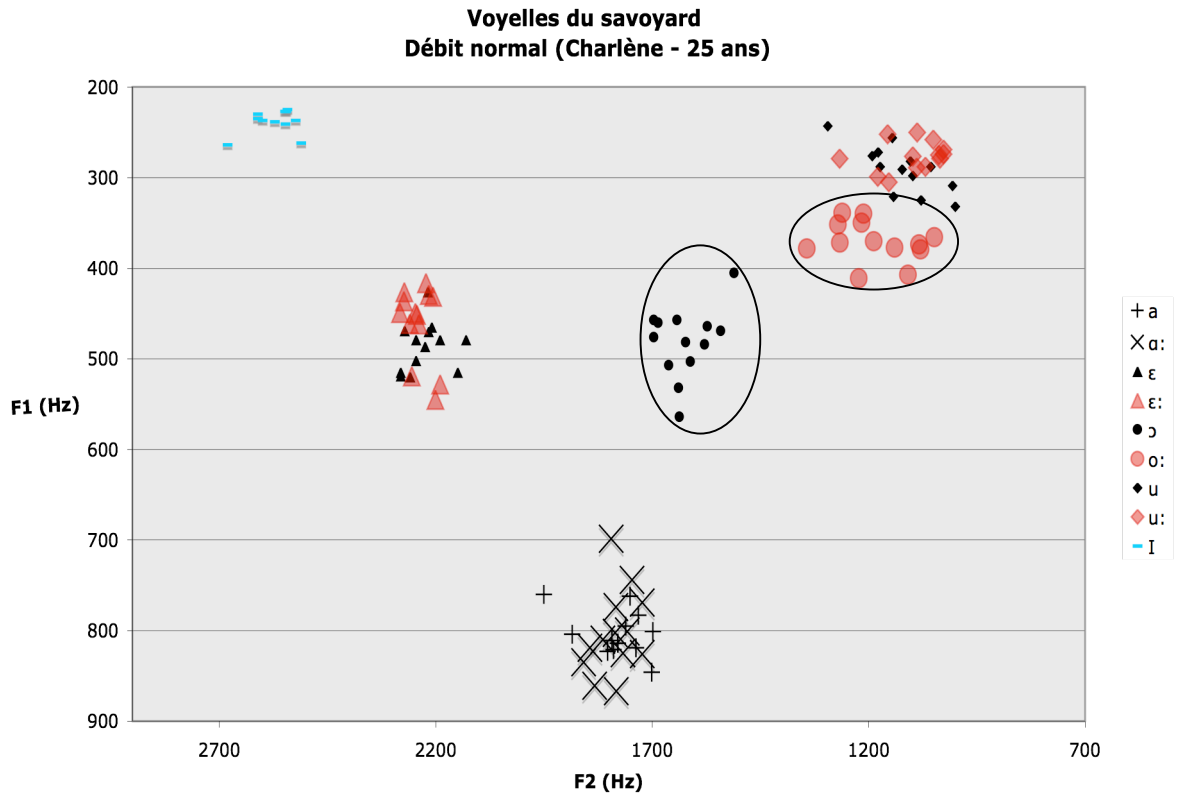


Figure 92. Espace vocalique des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur féminin (Ch), âge:25 ans en 2007.

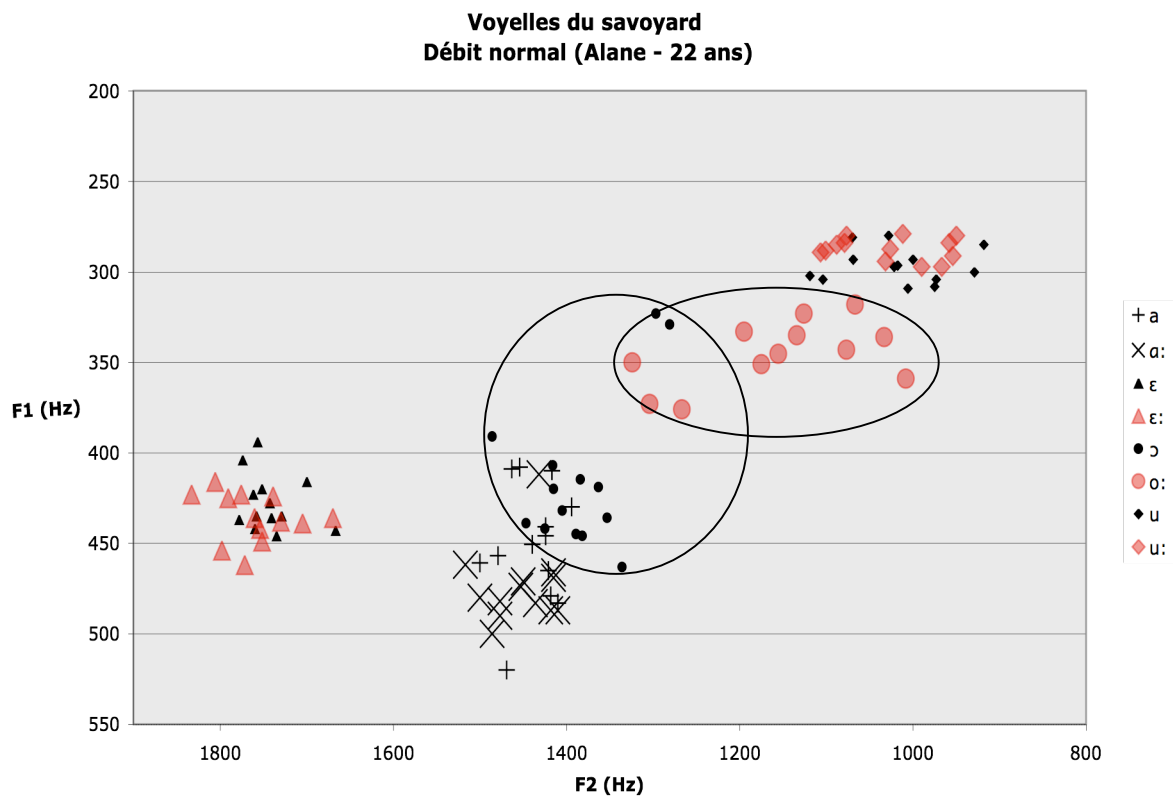


Figure 93. Espace vocalique des 8 voyelles [a, a:, ε, ε:, ɔ, o:, u, u:] du français savoyard, le débit est normal. Locuteur masculin (Al), âge: 22 ans en 2007.

V.3.3. Écarts perceptifs

Nous avons pu conclure en V.3.1. que, chez la première génération, seuls le /a/ et le /ɔ/ montraient des durées inférieures à celles de leurs partenaires longs /ɑ:/ et /o:/ sous les deux débit de parole. Chez la deuxième génération, seul le /a/ montrait une durée inférieure à celle de son partenaire long /ɑ:/ sous les deux débit de parole. Pour la troisième génération, aucune des 4 paires ne montrait une différence temporelle entre ses deux termes. Ainsi notre étude des écarts perceptifs ne portera que sur ces deux paires de voyelles, c'est-à-dire /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/, pour les deux autres paires, c'est-à-dire /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ l'étude s'arrête ici, parce qu'aucune opposition, quantitative comme qualitative, n'a été observée.

Toutes les paires qui présentent des oppositions temporelles présentent en même temps des oppositions spectrales, à l'exception de la paire /ɔ/ vs. /o:/ chez la deuxième et la troisième générations, où nous observons effectivement une opposition spectrale avec une durée égale pour les deux voyelles de la paire.

Nous essayerons de déterminer en fonction des résultats que nous avons obtenus dans le premier chapitre de ce manuscrit, laquelle des deux caractéristiques est à l'origine de la perception du contraste entre les deux réalisations.

V.3.3.1. Écart temporel

Nous voudrions rappeler les résultats de nos recherches donnés en paragraphe I.2.4. concernant l'importance de l'étendue de l'écart temporel entre deux voyelles pour qu'elles puissent se démarquer au niveau perceptif uniquement par leur durée: Pour des durées situées entre 50 et 250 ms, pour qu'une variation de la durée puisse être exploitée linguistiquement, il faudrait que la durée initiale accuse une hausse ou une baisse supérieure à 10%, soit un rapport de Weber supérieur à 0,1. Une hausse ou une baisse, de la durée initiale, entre 10% et 20%, peut être, mais pas nécessairement, suffisante pour créer une opposition temporelle linguistiquement exploitable au-delà de 20% il est fort probable que la hausse soit suffisante pour permettre aux deux durées, initiale et initiale+hausse (ou +baisse) d'opérer indépendamment l'une par rapport à l'autre dans un système phonologique.

Comme les données de nos locuteurs sont relativement hétérogènes, nous donnerons le rapport de Weber pour chaque locuteur à part en tableau 43: [(longue-brève)/longue], c'est-à-dire ce que représente la différence entre la longue et la brève par rapport à la durée de la

longue. Donc nous rapportons la différence de durée entre les deux voyelles à la durée d'une voyelle, et non pas sur la durée d'une syllabe, parce que l'unité syllabe est très variable (elle peut avoir une structure plus ou moins importante, ce qui modifie sensiblement le rapport de Weber).

Locuteur (Ma).

Les deux paires de voyelles sous les deux débits semblent présenter, auprès de ce locuteur, des écarts temporels qui sont largement supérieurs au seuil différentiel de la perception de la durée vocalique. Parmi les locuteurs de la première génération, c'est lui qui présente les écarts les plus importants. Le rapport de Weber est plus élevé pour la paire /a/ vs. /ɑ:/ que pour la paire /ɔ/ vs. /o:/, comme il est plus élevé en débit normal qu'en débit rapide.

Locuteur (Ge).

Il présente des écarts significatifs en termes de perception, cela pour les deux paires de voyelles testées et sous les deux débits de parole. Contrairement au locuteur (Ma), auprès du locuteur (Ge) le rapport de Weber est plus élevé pour la paire /ɔ/ vs. /o:/ que pour la paire /a/ vs. /ɑ:/, et ce rapport est plus élevé en débit normal qu'en débit rapide.

Locutrice (De).

Même si les écarts temporels auprès de la locutrice (De) sont plus importants que ceux du locuteur (Ge) en valeurs absolues, elle présente les rapports de Weber les moins importants parmi tous les locuteurs de la première génération. Comme les autres locuteurs, le débit rapide réduit le rapport de Weber. Comme le locuteur (Ma), le rapport de Weber est plus élevé pour la paire /a/ vs. /ɑ:/ que pour la paire /ɔ/ vs. /o:/.

Locuteur (Ma) - 78 ans

	normal	rapide
[(a:-a)/a:]	0,62	0,56
[(o:-o)/o:]	0,53	0,45

Locuteur (Ge) - 78 ans

	normal	rapide
[(a:-a)/a:]	0,35	0,3
[(o:-o)/o:]	0,41	0,31

Locuteur (De) - 76 ans

	normal	rapide
[(a:-a)/a:]	0,32	0,22
[(o:-o)/o:]	0,14	0,21

Locuteur (Sl) - 49 ans

	normal	rapide
[(a:-a)/a:]	0,16	0,086

Tableau 43. Rapport de Weber pour deux paires de voyelles: /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ avec 2 débits de parole, normal et rapide pour 4 locuteurs du français savoyard. Âge en 2007.

Conclusion: tous les locuteurs de la première génération, à une exception près¹⁸, pour les deux paires de voyelles mesurées, c'est-à-dire pour /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ et sous les deux débits de parole, normal et rapide, présentent un rapport de Weber supérieur à 0,20.

Selon Rossi (1972) tous ces écarts sont suffisants pour démarquer les voyelles brèves des voyelles longues. Que nous considérons les brèves se trouvant entre 60 et 140 ms, et leur seuil constant en valeur absolue de 30 ms, ou que nous considérons les longues se trouvant entre 130 et 290 ms, et leur rapport de Weber qui doit être au moins de 0,225 pour qu'elles se démarquent des brèves, les deux conditions sont vérifiées auprès des 3 locuteurs de cette génération pour les deux oppositions /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ et sous les deux débits de parole.

Donc, nous pouvons avancer l'idée que, pour cette génération, les différences de durée entre voyelles longues et voyelles brèves sont suffisamment importantes pour permettre à ces deux types de voyelles de se distinguer uniquement par leurs durées. Ainsi, nos résultats du point de vue statistique et du point de vue perceptif convergent vers une conclusion selon laquelle le français de Savoie de cette génération, et très vraisemblablement celui des générations précédentes, est une langue qui fait usage de la quantité vocalique pour les deux paires de voyelles /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/, comme l'arabe, pour multiplier le nombre de voyelles de son système phonologique.

Locutrice (S1).

Pour la deuxième génération, avec une seule locutrice, l'interprétation des résultats est moins évidente, puisque pour la seule paire /a/ vs. /ɑ:/, qui a montré une différence temporelle statistiquement significative, le rapport de Weber est inférieur à 0,20 avec les deux débits de parole, il est inférieur à 0,10 en débit rapide. Les résultats de cette locutrice ne sont pas concluants, parce que le rapport de Weber tombe dans un intervalle critique où nous ne savons pas si cette taille de l'écart temporel est plutôt suffisante ou plutôt insuffisante pour que les deux voyelles puissent se démarquer par leurs durées seulement dans un système phonologique d'une langue.

Cet écart est insuffisant selon les critères de Rossi (1972). Selon lui, pour les voyelles entre 60 ms et 140 ms, il faut une différence de 30 ms, alors que les différences ici sont de 17 et de 9 ms en débit normal et en débit rapide respectivement. S'il faut toutefois pencher pour une hypothèse, nous dirons que la taille de cet écart est plutôt suffisante lorsque le débit de parole est normal, et qu'elle devient non suffisante lors de la parole rapide. Cette hypothèse

¹⁸ Locutrice (De), le rapport relatif à la qualité [o] en débit normal est de 0,14 (< 0,20).

est étayée par les statistiques en quelque sorte, parce que la seule fois où nous avons changé le seuil de signification de ($p < 0,010$) à ($p < 0,015$) c'était pour l'opposition de durée entre les deux réalisations /a/ et /ɑ:/ en débit rapide pour la locutrice de la deuxième génération, ce qui montre aussi que cet écart est relativement faible du point de vue statistique.

V.3.3.2. Écart spectral

Nous cherchons dans ce paragraphe à établir les distances perceptives entre voyelles longues et brèves en termes de F1 et de F2. Cela est pour les deux paires de voyelles /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/, pour lesquelles nous avons pu constater une différence qualitative, auprès de la première et de la deuxième génération. Nous rappelons qu'auprès de la troisième génération seule l'opposition /ɔ/ vs. /o:/ existe, au niveau spectral uniquement. Nous déterminerons ces distances pour chaque locuteur à part et pour chaque débit de parole à part.

Pour calculer la distance au sein de chaque paire présentant une opposition qualitative, nous procéderons comme suit: (1) Nous convertissons nos valeurs en Hz à l'échelle des Bark, nous utilisons pour cette conversion la formule proposée dans Traunmüller (1990), (l'équation n° 6) et utilisée dans Kewley-Port & Zheng (1999):

$$Z \text{ (en Bark)} = ((26,81 * f \text{ (en Hz)}) / (1960 + f \text{ (en Hz)})) - 0,53$$

Comme nous l'avons mentionné en I.1.3.2.3., l'oreille humaine regroupe des fréquences voisines dans des bandes, appelées "bandes critiques", où la valeur de la largeur de chacune des bandes vaut 1 Bark, cette réalité perceptive fait qu'en convertissant des valeurs des formants de l'échelle des Hz à l'échelle des Bark, les valeurs du seuil du F2 ont des valeurs proches de celles du seuil du F1, figure (18) pour les seuils en Hz et figure (19) pour les seuils en Bark.

En Hz, l'espace vocalique des locuteurs féminins est significativement plus étendu que l'espace vocalique des locuteurs masculins, Hillenbrand *et al.* (1995), figure (24). Cette conversion nous permet par ailleurs de ne pas tenir compte du fait que le locuteur soit de sexe féminin ou de sexe masculin, comme le montre les travaux de Ménard *et al.* (2002) sur la normalisation perceptive pour les 10 voyelles du français /i y u e ø o ε œ ɔ a/.

La proximité entre les seuils des formants permet par conséquent d'obtenir une moyenne qui sera relative à la totalité de la voyelle. (2) Nous mesurons la distance avec la méthode de mesure utilisée par Kewley-Port & Zheng (1999), laquelle consiste tout simplement à calculer une moyenne arithmétique qui base la distance uniquement sur F1-F2. Cette méthode est résumée comme suit:

$$d_{ViVj} = [(|F1V2-F1V1| + |F2V2-F2V1|)/2]$$

Pour la paire /a/ vs. /ɑ:/, la distance ne sera calculée qu'au niveau de F2, puisque nous considérons que la distance au niveau de F1 est due au hasard, par conséquent, elle ne sera pas intégrée dans l'équation ci-dessus.

Les résultats de la conversion en Bark et la mesure des distances pour toutes les paires concernées sont données dans le tableau (44) pour la première génération et dans le tableau (45) pour la 2^e et la 3^e générations.

	Valeurs (Hz)		Valeurs (Bark)		Distance (Bark)	
1e génération						
Locuteur (Ma)						
Débit normal						
Voyelle	F1	F2	F1	F2		
a	618	1311	5,90	10,21	/a/ vs. /ɑ:/	0,65
ɑ:	625	1065	5,95	8,91		
ɔ	497	1187	4,89	9,58	/ɔ/ vs. /o:/	1,44
o:	362	906	3,65	7,95		
Débit rapide						
Voyelle	F1	F2	F1	F2		
a	646	1402	6,11	10,65	/a/ vs. /ɑ:/	0,92
ɑ:	627	1047	5,97	8,81		
ɔ	516	1076	5,05	8,97	/ɔ/ vs. /o:/	0,86
o:	431	914	4,30	8,00		
Locuteur (Ge)						
Débit normal						
Voyelle	F1	F2	F1	F2		
a	650	1308	6,15	10,20	/a/ vs. /ɑ:/	0,74
ɑ:	582	1033	5,61	8,72		
ɔ	499	1298	4,91	10,15	/ɔ/ vs. /o:/	1,80
o:	386	850	3,88	7,58		
Débit rapide						
Voyelle	F1	F2	F1	F2		
a	715	1400	6,64	10,64	/a/ vs. /ɑ:/	0,69
ɑ:	715	1128	6,63	9,26		
ɔ	478	1347	4,73	10,39	/ɔ/ vs. /o:/	1,29
o:	412	983	4,12	8,43		
Locutrice (De)						
Débit normal						
Voyelle	F1	F2	F1	F2		
a	635	1597	6,03	11,51	/a/ vs. /ɑ:/	0,68
ɑ:	650	1300	6,14	10,16		
ɔ	456	1244	4,53	9,88	/ɔ/ vs. /o:/	1,24
o:	378	936	3,80	8,13		
Débit rapide						
Voyelle	F1	F2	F1	F2		
a	626	1596	5,96	11,50	/a/ vs. /ɑ:/	0,53
ɑ:	613	1359	5,85	10,45		
ɔ	478	1122	4,73	9,23	/ɔ/ vs. /o:/	0,76
o:	408	970	4,09	8,35		

Tableau 44. Conversion en Bark des deux premiers formants pour les 4 voyelles [a, ɑ:, ɔ, o:] du français savoyard, et mesure de la distance perceptive en termes de F1 et de F2 pour les deux paires de voyelles: /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ avec 2 débits de parole, normal et rapide pour les 3 locuteurs de la 1e génération.

	Valeurs (Hz)		Valeurs (Bark)		Distance (Bark)
2e génération					
Locutrice (SI)					
Débit normal					
Voyelle	F1	F2	F1	F2	
a	659	1668	6,21	11,80	/a/ vs. /ɑ:/ 0,33
ɑ:	648	1510	6,13	11,14	
ɔ	534	1413	5,21	10,70	/ɔ/ vs. /o:/ 1,49
o:	420	1033	4,20	8,72	
Débit rapide					
Voyelle	F1	F2	F1	F2	
a	769	1664	7,02	11,78	/a/ vs. /ɑ:/ 0,24
ɑ:	770	1551	7,03	11,31	
ɔ	526	1523	5,14	11,19	/ɔ/ vs. /o:/ 1,41
o:	426	1127	4,25	9,26	
3e génération					
Locutrice (Ch)					
Débit normal					
Voyelle	F1	F2	F1	F2	
a	811	1793	7,32	12,28	/a/ vs. /ɑ:/ 0,02
ɑ:	803	1786	7,26	12,25	
ɔ	482	1623	4,76	11,61	/ɔ/ vs. /o:/ 1,53
o:	370	1188	3,73	9,59	
Locuteur (Al)					
Débit normal					
Voyelle	F1	F2	F1	F2	
a	451	1439	4,48	10,82	/a/ vs. /ɑ:/ 0,03
ɑ:	474	1453	4,69	10,88	
ɔ	415	1384	4,15	10,57	/ɔ/ vs. /o:/ 0,91
o:	345	1155	3,48	9,41	

Tableau 45. Conversion en Bark des deux premiers formants pour les 4 voyelles [a, ɑ:, ɔ, o:] du français savoyard, et mesure de la distance perceptive en termes de F1 et de F2 pour les deux paires de voyelles: /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ avec 2 débits de parole, normal et rapide pour la locutrice (SI) de la 2e génération, avec un débit normal pour les 2 locuteurs de la 3e génération.

Première génération: En nous appuyant, en particulier, sur les travaux de Kewley-Port & Watson (1994) et Kewley-Port & Zheng (1999) sur le seuil de la qualité vocalique pour l'anglais américain, nous avons pu conclure en I.1.4.3. que, pour un système vocalique relativement dense, tel le système du français, le seuil catégoriel de la qualité vocalique est absolu, il est approximativement de 0,6 Bark.

Les locuteurs de cette génération semblent tous produire des écarts qualitatifs supérieurs à 0,6 Bark (seule exception: locutrice (De) pour la paire /a/ vs. /ɑ:/ en débit normal, la distance = 0,53 Bark), pour les deux paires de voyelles et sous les deux débits de parole (2

paires * 2 débits * 3 locuteurs = 12 écarts). À une exception près (l'opposition /ɔ/ vs. /o:/ chez le locuteur (Ma) en débit rapide), les deux termes de la paire /ɔ/ vs. /o:/ présentent un écart spectral nettement supérieur à l'écart entre les deux termes de la paire /a/ vs. /ɑ:/, pouvant être le double dans certaines conditions (cas du locuteur (Ma) en débit normal).

Le débit rapide affecte négativement les écarts qualitatifs (sans test statistique), ceci en réduisant la taille de l'écart entre la voyelle brève et la voyelle longue correspondante (exception: locuteur (Ma) pour l'écart entre /a/ et /ɑ:/). Cependant, le débit rapide ne réussit à effacer aucun des écarts.

Première génération.

En débit normal de la parole, les locuteurs de cette génération présentent des écarts spectraux perceptibles pour /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/, cela demeure vrai sous le débit rapide.

Deuxième génération.

La seule locutrice de la deuxième génération semble présenter un écart à la limite du perceptible pour la paire /a/ vs. /ɑ:/ en débit normal, et probablement non perceptible en débit rapide. Nous avons pensé que cela était dû au fait qu'elle réalisait cette opposition qualitative d'une manière non systématique, c'est-à-dire que pour certaines réalisations l'écart est perceptible, alors qu'il est absent pour d'autres, et en obtenant la moyenne de ses réalisations, nous avons eu un écart faible, qui, selon nos critères, ne serait pas perceptible. Afin de vérifier cette hypothèse nous avons procédé au calcul des 12 distances qui existent entre les 12 brèves et les 12 longues produites par cette locutrice avec un débit normal. Nous avons trouvé que pour 2 des 12 écarts, la distance était supérieure à 0,60 Bark et pour 5 autres l'écart était inférieur à 0,35 Bark. Pour l'une des paires, l'écart était quasiment nul, mais même en enlevant la valeur de ce dernier écart, la valeur moyenne de l'écart reste inférieure à 0,40 Bark. Cela signifie que cette locutrice réalise de manière systématique un écart faible qui serait tout de même perceptible.

Troisième génération.

Pour les deux locuteurs de cette génération les résultats sont clairs. Conformément à nos attentes, il n'y a pas d'écart spectral entre /a/ et /ɑ:/, puisque ces deux réalisations n'ont présenté aucune différence à aucun niveau. Alors que pour la paire /ɔ/ vs. /o:/, ces deux termes présentent un écart largement supérieur au seuil de 0,60 Bark: ils sont par conséquent qualitativement distincts.

V.4. Conclusions

Dans cette étude, nous avons étudié le rôle de deux indices acoustiques, à savoir la durée de la voyelle et son spectre acoustique, dans la création des identités de certaines voyelles du français du comté de Savoie. Nous ne pouvons pas restreindre l'origine de l'information acoustique à ces deux paramètres, nous sommes conscients que d'autres éléments peuvent intervenir pour jouer un rôle important dans l'opposition entre deux voyelles. Nous avons démontré en chapitre IV de ce manuscrit que des paramètres prosodiques peuvent intervenir pour participer à la démarcation des catégories Brève/Longue. En effet, l'arabe, qui oppose de manière systématique et robuste des voyelles brèves à des voyelles longues a dévoilé la présence non seulement de différences temporelles et spectrales mais aussi de différence d'intensité, laquelle est plus élevée pour la voyelle brève.

Indépendamment du positionnement des zones d'énergie sur l'échelle des fréquences, la variation de la structure acoustique de la voyelle elle-même au cours de son évolution (Nearey 1989) et la variation des unités phonétiques contiguës notamment (Strange 1989) peuvent porter une part non négligeable de l'information acoustique sur l'identité de la voyelle. En effet, Adu Manyah (2003) a mesuré les durées et les qualités vocaliques des voyelles longues et brèves du Twi (parlé au Ghana), il a observé que dans une syllabe VC la durée des consonnes est dynamique et change ainsi en fonction de la durée de la voyelle pour maintenir une même durée au niveau de la syllabe. Effectivement, alors que le rapport Longue/Brève = 2,4, la durée de la syllabe est plus ou moins la même quelle que soit la durée de la voyelle, elle est de l'ordre de 230 ms. Ainsi, une voyelle brève était suivie d'une consonne longue et inversement. Cela était vrai quelque soit le timbre de la voyelle. Cette observation est conforme au principe de « patrons de complémentarité », qui implique que la séquence VC: a la même durée que la séquence V:C. Cette variation de la durée de la consonne en fonction de la durée de la voyelle qui la précède peut constituer un indice acoustique sur la durée de la voyelle. Denes (1955), en manipulant la durée de voyelles synthétisées en 4 durées différentes, montre que la durée de la voyelle peut servir d'indice pour la perception de l'état du voisement de la consonne en position post-vocalique, il affirme ainsi que l'information ne provient pas seulement du spectre de l'unité considérée, mais d'autres éléments extra-spectraux peuvent participer à sa reconnaissance, l'auteur affirme que c'est un exemple qui montre qu'il y a effectivement un effet significatif du contexte phonétique environnant sur l'unité perçue.

Maddieson (2006) argumente pour une co-variation phonétique plus importante dans la parole claire par rapport à la parole réduite, c'est parce que la valeur d'un paramètre peut-être donné, en partie, par la variation des autres. Pour donner ainsi plus d'informations sur les identités des unités phonétiques, nous magnifions la co-variation des unités environnantes. Selon lui, beaucoup de patrons de la co-variation ont une contribution significative à l'identification perceptive des unités phonétiques. En ce qui nous concerne, Maddieson (2006) affirme que dans des langues à quantité vocalique, la durée de la consonne qui suit une voyelle longue est plus courte de la durée de cette même consonne quand elle est précédée par une voyelle brève, la durée de la consonne post-vocalique est un effet de la co-variation, lequel constitue un indice sur la durée de la voyelle qui la précède, Maddieson rapporte que certains (Diehl & Kluender, 1989) vont jusqu'à proposer que la co-variation est intentionnellement produite dans le but de magnifier le contraste entre les unités que nous cherchons à distinguer.

L'étude d'Abramson (2003) sur la durée des consonnes occlusives en Pattani Malay, montre que la sensation de la durée peut être fournie par d'autres indices que le temps de la réalisation, comme la hauteur de la fréquence fondamentale et l'intensité.

Nous croyons, cependant, après avoir uniformisé l'effet de la co-variation phonétique en introduisant les voyelles qui ont fait l'objet de notre étude dans des paires minimales, que ces deux informations, c'est-à-dire l'information temporelle et l'information spectrale, sont deux informations déterminantes dans la création des identités des voyelles.

Nous discutons brièvement en annexe n° XVII du rôle des transitions formantiques (contexte), de la variation de la qualité de la voyelle au cours de son émission et du rapport entre les valeurs des formants dans la reconnaissance de l'identité de la voyelle.

Nos résultats concernant l'écart temporel pour les 4 paires /a/ vs. /ɑ:/, /ɛ/ vs. /ɛ:/, /ɔ/ vs. /o:/ et /u/ vs. /u:/ montrent que seules les paires de voyelles /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ s'opposent par la durée de leurs réalisations chez la génération la plus âgée, des oppositions confirmées par les calculs statistiques, lesquelles vont au-delà du seuil différentiel de la perception de la durée vocalique. Seul /a/ s'opposant à /ɑ:/ par la durée chez la génération intermédiaire, avec un écart moins important comparé à celui présent chez la génération la plus âgée, il est statistiquement significatif sous les deux débits ($p < 0,02$), mais il ne serait perceptible qu'en débit normal, le rapport de Weber descendant sous la barre critique de 0,10 en débit rapide; et enfin aucune de ces 4 paires ne présente de contraste temporel entre ses deux termes chez la

jeune génération. Cela met en évidence la tendance selon laquelle l'opposition quantitative dans le système phonologique de voyelles du français de Savoie est en train de disparaître.

Nos résultats pour l'écart spectral pour les 4 paires /a/ vs. /ɑ:/, /ɛ/ vs. /ɛ:/, /ɔ/ vs. /o:/ et /u/ vs. /u:/ montrent que, chez la génération la plus âgée, seules les paires /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ présentent des différences spectrales, des différences confirmées par les calculs statistiques et qui dépassent le seuil catégoriel de la qualité vocalique. Les deux informations, temporelle et spectrale, sont perceptibles chez cette génération, ainsi l'une d'elles suffirait pour démarquer les deux identités vocaliques d'une paire.

Ces mêmes différences sont présentes chez la locutrice (S1) de la génération intermédiaire, nous pensons que la distance spectrale entre /a/ et /ɑ:/ est insuffisante. Ainsi, pour démarquer ces deux réalisations /a/ et /ɑ:/ l'une de l'autre, (S1) compte sur les deux informations temporelle et spectrale, parce que ni l'une ni l'autre n'est suffisante toute seule. La différence spectrale entre /ɔ/ et /o:/ est suffisante pour démarquer qualitativement les deux voyelles.

Nos résultats concernant la jeune génération, constituée de deux locuteurs, moyenne d'âge 23 ans en 2007, montrent qu'il n'y a pas de différence au niveau de la durée des voyelles entre brèves et longues. Certaines de celles qui sont supposées brèves pouvant se montrer légèrement plus longues que celles qui sont supposées longues. L'Anova ne montre cependant pas de différence significative de durée entre brèves et longues, pour aucune des 4 qualités, chez aucun des locuteurs de cette génération.

La différence spectrale, seule en jeu, entre /ɔ/ et /o:/ est suffisante pour démarquer qualitativement les deux voyelles, ces deux voyelles présentent effectivement une distance supérieure au seuil de 0,60 Bark. Aucun écart spectral entre les deux termes de chacune des trois autres paires n'a pu être observé.

VI. Conclusion

VI.1. Contrôle de la quantité vocalique

La première partie du premier objectif de notre travail était la caractérisation des deux corrélats acoustiques de la quantité vocalique (oppositions temporelles et spectrales entre voyelles brèves et longues) en arabe dialectal de la ville de Mayadin pour les trois paires de voyelles: /i/ vs. /i:/, /a/ vs. /a:/ et /u/ vs. /u:/ et sous 4 conditions expérimentales: 1) débit normal et schéma déclaratif; 2) débit rapide et schéma déclaratif; 3) débit normal et schéma interrogatif et 4) débit rapide et schéma interrogatif. Et du français du comté de Savoie pour les quatre paires de voyelles: /a/ vs. /ɑ:/, /ɛ/ vs. /ɛ:/, /ɔ/ vs. /o:/ et /u/ vs. /u:/ sous deux conditions expérimentales: 1) débit normal et schéma déclaratif et 2) débit rapide et schéma déclaratif. La deuxième partie consiste à déterminer quels sont les corrélats acoustiques de la quantité vocalique, parmi ceux dont nous tenons alors la preuve de l'existence, ceux qui sont contrôlés et ceux qui ne le sont pas.

VI.1.1. Arabe dialectal

Écart temporel

Nous avons enregistré trois locuteurs masculins (âgés de 27 à 39 ans). Nos résultats concernant la durée indiquent que, tous locuteurs, toutes qualités et toutes conditions expérimentales confondus, la brève constitue 57% de la durée de la longue.

Au plan statistique

Nous distinguons deux points sur ce plan: l'écart entre deux valeurs de deux indices et la valeur d'un indice. Le fait que les valeurs des indices testés séparément ne soient pas significativement affectées par la perturbation n'implique pas que l'écart ne soit pas affecté, puisque l'écart (la différence ou la distance) concerne deux valeurs simultanément.

Les résultats des tests statistiques indiquent que les écarts temporels entre voyelles brèves et longues pour les paires: /i/ vs. /i:/, /a/ vs. /a:/ et /u/ vs. /u:/ sont significatifs ($p < 0,01$), et ceci sous les 4 conditions expérimentales. Cela signifie que l'opposition temporelle a résisté à la perturbation de la parole rapide et à la perturbation engendrée par l'interrogation.

Comme ces deux procédés de perturbation, rythmique et intonatif, ont des effets *a priori* antagonistes, et pour savoir si l'un des deux procédés de perturbation a eu de l'effet sur

la durée ou la qualité de la voyelle, nous ne considérons qu'un seul procédé à la fois. C'est-à-dire que pour savoir si le débit rapide a de l'effet, nous considérons les deux débits sous un schéma déclaratif, ou pour savoir si le schéma mélodique interrogatif a de l'effet, nous considérons les deux schémas sous le même débit de parole. Parce que si nous trouvons un effet sous l'interrogation rapide, nous ne saurons pas à quel procédé est dû cet effet.

La perturbation de la durée des voyelles par le débit rapide est effective. Avec un schéma déclaratif, la perturbation de la parole rapide consiste en l'abaissement significatif de la durée de la voyelle ($p < 0,01$) quelle que soit sa catégorie (brève ou longue) et quelle que soit sa qualité [i, a, u]. En augmentant significativement les valeurs de la F0 et les valeurs de l'intensité des voyelles brèves et longues, le débit rapide s'est montré un élément perturbateur des indices prosodiques également.

La perturbation de la durée par l'interrogation, elle aussi s'est montrée effective. Avec un débit normal, la perturbation de l'interrogation consiste en l'augmentation de la durée, de certaines voyelles et non pas toutes, son effet est moins régulier que celui du débit rapide en ce qui concerne la durée. Elle augmente significativement la F0 et l'intensité de toutes les voyelles (toujours en débit normal). Nous notons alors que le débit rapide reste le procédé le plus efficace que nous ayons observé pour perturber les durées des voyelles.

Donc nous pouvons conclure que l'opposition temporelle en arabe dialectal de Mayadin, ayant traversé les épreuves de la perturbation tout en se préservant, est une opposition robuste, et par conséquent c'est une opposition linguistiquement spécifiée. Ces résultats nous permettent finalement d'avancer que le geste articulatoire qui permet la production de l'écart temporel, le premier corrélat acoustique de la quantité vocalique, entre voyelles longues et brèves est un geste **contrôlé**.

Au plan perceptif

Les résultats sur le plan perceptif n'entrent pas en contradiction avec la conclusion qui juge contrôlé le geste articulatoire qui permet la production de l'écart temporel. En effet, nos mesures montrent que quel que soit le débit de la parole et quel que soit le schéma mélodique, l'écart temporel entre chaque voyelle brève et sa correspondante longue est perceptible, le rapport de Weber [(longue-brève)/longue] étant supérieur à 35% quelle que soit la condition expérimentale, alors que le seuil différentiel de la perception de la durée vocalique signale qu'un rapport supérieur à 20% est suffisamment important pour démarquer perceptivement la brève de la longue. Donc ces résultats indiquent finalement que l'écart temporel entre la brève

et la longue en arabe dialectal de Mayadin est un écart **perceptible**. Les résultats sur le plan perceptif sont enfin en totale harmonie avec les données statistiques.

Distance spectrale

La distance spectrale s'exprime en fonction des deux principales composantes de la qualité vocalique, le F1 et le F2. Nous avons testé l'écart entre voyelles brèves et longues pour les paires: /i/ vs. /i:/, /a/ vs. /a:/ et /u/ vs. /u:/ au niveau de F1 et au niveau de F2.

Au plan statistique

Les résultats montrent que les écarts au niveau de F1 entre la brève et la longue sont significatifs ($p < 0,01$) et ceci sous les 4 conditions expérimentales et pour les trois paires, ce qui signifie que ces écarts sont insensibles aux perturbations rythmique et intonative, et sont par conséquent robustes.

Concernant les effets perturbateurs du débit et de l'interrogation sur les deux premiers formants, les résultats montrent que ces deux composantes de la qualité vocalique, contrairement à la durée qui est une composante quantitative, restent pratiquement insensibles aux effets de ces deux procédés de perturbation. En effet, quand nous considérons chaque valeur à part, en schéma déclaratif, le débit rapide n'a pas d'effet sur les F1 (sauf sur le F1 du /i:/ long qui accuse une hausse de 15 Hz). L'interrogation, aussi faible que le débit rapide, n'a pas d'effet sur les F1 (sauf pour le F1 du /i:/ long avec une hausse de 8 Hz). Les écarts sur l'axe de F1 sont significativement présents quel que soit le débit et quel que soit le schéma. Selon le principe des perturbations prioritaires (Abry *et al.*, 1990), si cette opposition se maintient c'est parce qu'elle joue un rôle distinctif dans notre système vocalique.

La significativité des écarts au niveau de F2 dépend de la paire de voyelles, du débit et du schéma mélodique considérés. Contrairement à ce que nous avons observé pour les écarts sur l'axe de F1, le débit rapide et l'interrogation ont significativement affecté les écarts sur l'axe de F2 pour les paires /a/ vs. /a:/ et /u/ vs. /u:/ comme le montre le tableau ci-dessous.

F2

	Déclaratif		Interrogatif	
	Normal	Rapide	Normal	Rapide
A/AA	(+)	(-)	(-)	(-)
I/II	(+)	(+)	(+)	(+)
U/UU	(-)	(-)	(+)	(+)

Le signe (+) signifie que l'écart est significatif, le signe (-) signifie que l'écart n'est pas significatif.

Quand nous considérons chacune des valeurs de F2 à part, nous constatons qu'elles ne sont pas affectées par le débit rapide (à l'exception du F2 du /i:/ long, un abaissement de 29 Hz). L'interrogation n'a pas d'effet non plus sur les valeurs des F2 (elle n'a affecté que celui du /a:/ long, occasionnant une hausse de 49 Hz). L'écart sur l'axe de F2 n'est significativement présent que pour la paire /i/ vs. /i:/.

Ces résultats nous permettent de conclure que les écarts sur l'axe de F1 sont des écarts contrôlés, puisqu'ils sont présents sous les 4 conditions expérimentales que nous avons appliquées à notre corpus. Les écarts sur l'axe de F2 ne sont contrôlés que pour la paire /i/ vs. /i:/. Ces résultats nous permettent finalement d'avancer que le geste articulatoire qui permet la production de l'écart spectral, le deuxième corrélât acoustique de la quantité vocalique, entre voyelles longues et brèves, est un geste en partie **contrôlé** pour les paires /a/ vs. /a:/ et /u/ vs. /u:/ et tout à fait **contrôlé** pour la paire /i/ vs. /i:/

Au plan perceptif

Nous voudrions rappeler ici que nous distinguons deux seuils concernant les qualités vocaliques: un seuil pour la perception d'un changement vocalique (Just Noticeable Difference, JND), et un seuil catégoriel de la qualité vocalique. En fait, le premier nous permet de savoir si une modification dans la qualité de la voyelle est perceptible, le deuxième indique si la distance entre deux entités vocaliques est suffisante pour démarquer perceptivement les deux entités en question. En nous basant sur les propositions de Kewley-Port & Watson (1994), Kewley-Port & Zheng (1996, 1999) et Kewley-Port (2001), concernant cette question, nous pouvons admettre qu'une distance spectrale combinée (F1 X

F2) d'une valeur avoisinant les 0,30 Bark peut franchir le premier seuil (JND), alors que pour franchir le deuxième seuil cette valeur doit être double, soit 0,60 Bark.

Partant, la distance entre le /i/ bref et le /i:/, calculée en termes de F1 et de F2, est une distance qui permet évidemment de démarquer les deux qualités vocaliques, elle varie entre 1,22 et 1,44 Bark en fonction de la condition expérimentale. Ce résultat est en totale conformité avec les données statistiques qui montrent que la brève est différente de la longue sur F1 et sur F2 quelle que soit la condition expérimentale.

Nous considérons que les différences au niveau de F2 n'existent pas pour les 2 paires /a/ vs. /a:/ et /u/ vs. /u:/, parce que, selon nos données statistiques, elles ne se sont pas montrées robustes puisqu'elles n'ont pas pu résister aux effets perturbateurs du débit rapide de la parole et du schéma interrogatif. Ainsi, les distances spectrales calculées à partir des F1 seuls, toutes conditions expérimentales confondues, varient entre 0,30 et 0,35 Bark (excepté la distance entre le /a/ bref et le /a:/ long qui est égale à 0,24 Bark en débit rapide avec un schéma déclaratif). Cela signifie que les distances au niveau de F1 sont des distances perceptibles, mais pas suffisantes pour franchir le deuxième seuil, c'est-à-dire que ces distances contribuent à l'élaboration du contraste Brève/Longue pour les deux paires /a/ vs. /a:/ et /u/ vs. /u:/, mais elles ne sont pas suffisantes pour supporter seules ce contraste. En ce lieu également, nous observons une cohérence entre les données statistiques et les calculs perceptifs, parce que si les seuils donnaient non perceptible un écart qui est donné significatif par les calculs statistiques, cela aurait créé une inadéquation qui nous aurait obligé à reconsidérer l'interprétation de nos résultats.

Nous admettons que les perturbations que nous avons appliquées sur les voyelles de notre corpus sont des perturbations suffisantes pour révéler les gestes obligatoires. Nos études détaillées nous permettent d'affirmer que le geste qui place le triangle vocalique des voyelles brèves dans le centre de l'espace vocalique sur l'axe de F1 est un geste contrôlé pour toutes les paires, puisque nous avons prouvé qu'il est impératif et perceptible. Contrairement aux thèses qui abondent dans la littérature de la quantité vocalique, cette réduction vocalique n'est finalement pas la conséquence d'une durée moins importante, cette centralité n'est pas produite accidentellement, c'est un geste prévu, les articulateurs sont ainsi intentionnellement programmés pour distancer qualitativement la brève de la longue au niveau F1.

La centralité (réduction) sur l'axe de F2, pour la paire /i/ vs. /i:/, est produite par un geste contrôlé qui abaisse délibérément la valeur de F2. En ce lieu également, il ne s'agit pas d'un geste occasionnellement moins ample, mais d'un geste volontairement et décidément moins

ample. Pour les paires /a/ vs. /a:/ et /u/ vs. /u:/, la centralité sur l'axe de F2 n'est pas produite par un geste obligatoire, ce geste est produit de manière fortuite pour répondre aux contraintes articulatoires, conséquences des perturbations naturelles de l'émission de la parole, puisque ce geste ne se maintient pas sous la perturbation.

Résultats de la perturbation par la manipulation instrumentale

Nous avons annoncé dans l'introduction de ce manuscrit que nous procéderons à une perturbation non naturelle dans le but de retrouver les gestes contrôlés. En effet, nous avons procédé à la manipulation instrumentale des durées des voyelles brèves et longues. Nous ne nous sommes pas contentés de donner à la brève et la longue une même durée (une moyenne des deux durées), mais nous avons donné à la voyelle brève la durée de la longue et à la voyelle longue la durée de la brève, c'est-à-dire que cette manipulation ne se limitait pas à éliminer la différence temporelle entre les deux catégories, mais elle dirigeait les sujets vers la perception de la brève en tant que longue et la perception de la longue en tant que brève. Notre hypothèse est que les termes des paires qui ne comptent que sur la dimension temporelle pour s'opposer devraient alors se confondre, alors que ceux des paires qui basent leurs oppositions sur d'autres corrélats acoustiques de la quantité vocalique devraient préserver leurs oppositions sous cette perturbation. Ainsi les oppositions qui survivent à la disparition de la différence temporelle, en se maintenant, sont des oppositions qui disposent d'autres corrélats acoustiques, lesquels sont alors contrôlés puisqu'ils sont à valeurs phonologiques dans le système vocalique de notre dialecte.

Les résultats de la perturbation par la manipulation instrumentale de la durée montrent que l'information spectrale n'est pas seule derrière la séparation catégorielle du /i/ bref du /i:/ long, contrairement à nos attentes. L'information temporelle n'agit pas seule non plus pour démarquer le /a/ bref du /a:/ long et le /u/ bref du /u:/ long, cela est conforme au rôle de la distance spectrale détectée sur l'axe de F1 et que nous avons jugée perceptible.

Les résultats concernant l'opposition /i/ vs. /i:/ nous ont appris que les indices acoustiques, celui porteur de l'information temporelle et celui porteur de l'information spectrale, sont associés à d'autres indices acoustiques. Nous avons testé alors le rôle d'autres caractéristiques phonétiques. Ainsi nous avons testé le rôle des durées des unités phonétiques adjacentes à la voyelle étudiée, le rôle de la fréquence fondamentale et finalement le rôle de l'intensité acoustique, cela était dans le but de savoir si elles apportaient concrètement de l'information sur la catégorie de la voyelle.

Les durées des consonnes pré et post-vocaliques se sont montrées insensibles à la durée intrinsèque de la voyelle (Brève/Longue). Nous avons conclu que ni la durée de la consonne pré-vocalique, ni la durée de la consonne post-vocalique n'est porteuse d'information sur la catégorie de la voyelle, Brève/Longue.

La fréquence fondamentale, elle non plus, n'est pas porteuse d'information sur la catégorie de la voyelle, Brève/Longue. En effet, la F0 de la voyelle brève n'est pas différente (au sens statistique du terme) de la F0 de la voyelle longue. Notre conclusion par rapport à la F0 était que cette caractéristique prosodique ne renseigne pas les auditeurs de l'arabe de Mayadin sur la catégorie de la voyelle.

L'intensité acoustique, contrairement à nos attentes, est significativement corrélée à la catégorie de la voyelle pour la qualité [i] où brève et longue n'ont pas la même intensité. La brève ayant en moyenne, tous locuteurs et toutes conditions expérimentales confondus, une intensité plus élevée, de 3,38 dB pour la qualité [i]. La taille de la contribution de l'intensité dans la reconnaissance de la catégorie de la voyelle n'a pas pu être déterminée avec précision, puisqu'elle pourrait être concurrencée par l'information apportée par la distance spectrale entre la brève et la longue.

Donc, nous avons effectivement réussi à déterminer dans cette étude les sources de l'information sur la catégorie de la voyelle Brève/Longue. Cette information s'est montrée temporelle et spectrale, et temporelle, spectrale et prosodique selon l'opposition considérée.

Pour l'opposition I/II, l'information sur la catégorie de la voyelle est essentiellement fournie par la distance spectrale, puis par la durée et l'intensité. Nous ne pouvons pas classer ces deux dernières selon leur contribution. Pour l'opposition A/AA, l'information est supportée principalement par la durée de la voyelle, ensuite par la différence spectrale, pas d'évidence pour un éventuel rôle de l'intensité acoustique entre le /a/ bref et /a:/ long. Pour l'opposition U/UU, l'information sur la catégorie de la voyelle n'est pas fournie massivement par un seul indice, les indices sont: durée de la voyelle et distance spectrale.

Nous avons un deuxième et dernier point à examiner en effectuant cette manipulation, lequel concerne les deux méthodes que nous avons utilisées pour mesurer les distances entre les voyelles longues et brèves dans l'espace vocalique de l'arabe de Mayadin, Schwartz *et al.* (1997b) et Kewley-Port & Zheng (1999). Nous avons posé la question suivante: si ces deux méthodes sont adaptées à la mesure des distances spectrales dans les espaces vocaliques, nous devons nous attendre à ce que plus la distance entre deux voyelles, calculée par ces méthodes,

est importante et plus elle joue un rôle dans leur discrimination. Les résultats de notre étude dans ce chapitre sont en total accord avec les distances que nous avons obtenues avec les deux méthodes ci-dessus. En effet, les deux méthodes donnent les distances pour les trois oppositions de la plus importante à la moins importante comme suit: I/II, U/UU et enfin A/AA. Les résultats de cette manipulation prouvent en effet que l'opposition I/II est l'opposition qui, parmi les trois oppositions étudiées, compte le plus sur la distance spectrale, suivie par l'opposition U/UU suivie enfin par l'opposition A/AA. Ceci est une démonstration de la cohérence de la démarche de ces méthodes.

Nous sommes maintenant en mesure de répondre aux questions que nous nous sommes posées dans l'introduction de ce manuscrit, nous les rappelons:

(1) Différence temporelle: est-ce que la différence temporelle entre voyelles brèves et longues est contrôlée? Autrement dit, est-ce que l'appareil phonatoire produit délibérément cette différence de durée?

Tous les résultats convergent vers une conclusion selon laquelle la différence temporelle entre la voyelle brève et la voyelle longue, pour les trois paires de voyelles /i/ vs. /i:/, /a/ vs. /a:/ et /u/ vs. /u:/, en arabe dialectal de Mayadin est une différence linguistiquement spécifiée. Elle est ainsi intentionnellement, et non pas accidentellement, produite, donc le geste responsable de cette différence est un geste **contrôlé**.

(2) Distance spectrale: est-ce que le fait que les voyelles brèves se trouvent déplacées vers le centre de l'espace vocalique (par rapport aux longues) est une réduction vocalique (undershoot, pour rendre le terme à Lindblom, 1963) due à un geste moins ample disposant de moins de temps pour se réaliser ou est-ce que cette distance est la production d'un geste délibérément moins ample qui prévoit l'emplacement des brèves dans le centre de l'espace vocalique?

Là aussi, toutes nos données indiquent que la distance au niveau de F1 pour les trois paires n'est pas une distance due au fait que le geste articulatoire dispose de moins de temps pour atteindre sa cible vocalique et qu'il déplace de manière occasionnelle la voyelle brève vers le centre de l'espace vocalique. Elles indiquent que cette distance est délibérément produite et qu'elle a une valeur phonologique dans notre système, donc le geste responsable de cette distance est un geste **contrôlé**. Nos données indiquent aussi que le déplacement sur l'axe de F2 n'est contrôlé que pour la voyelle brève /i/, donc il est effectivement possible

d'affirmer que le déplacement occasionnel des /a/ et /u/ brefs sur l'axe de F2 vers le centre de l'espace vocalique est une réduction vocalique (undershoot), cela ne change pas notre interprétation selon laquelle la séparation du triangle des brèves de celui des longues en dialecte arabe de Mayadin s'effectue d'une manière **contrôlée**.

VI.1.2. Français de la région de la Savoie

Nous rappelons que nous avons étudié le comportement de la quantité vocalique du français du comté de Savoie pour les quatre paires de voyelles suivantes: /a/ vs. /ɑ:/, /ɛ/ vs. /ɛ:/, /ɔ/ vs. /o:/ et /u/ vs. /u:/. Nous avons appliqué un seul type de perturbation, c'est l'élocution rapide de la parole. En fait, notre étude de ce système s'inscrit dans 2 cadres: (i) comparatif, où nous avons pour but de savoir si d'autres systèmes opposant des voyelles longues et des voyelles brèves observent des réactions similaires à ceux que nous observerons avec le système de l'arabe dialectal de Mayadin lorsqu'ils sont soumis à la contrainte de la parole rapide. (ii) descriptif: dans ce cadre nous réalisons une étude acoustique intergénérationnelle afin de confirmer, ou de réfuter, l'hypothèse selon laquelle le trait de la quantité vocalique dans le français de cette région a réellement existé dans le passé et a réellement disparu aujourd'hui chez la jeune génération. Ainsi nos locuteurs appartenaient à 3 générations: 3 locuteurs (deux hommes et une femme) septuagénaires pour la première génération, une femme quadragénaire pour la deuxième et un homme et une femme âgés de 22 et 25 respectivement pour la troisième génération.

Nous développons ci-dessous nos résultats à propos de cette étude.

VI.1.2.1. Première génération

Écart temporel

Au plan statistique

Nous avons constaté que les différences temporelles sont plus marquées chez la première génération que chez les deux autres générations. Le rapport (Durée Brève/Durée Longue) est le même pour les deux qualités [ɛ] et [u], il est de 1,04. Cela signifie que les voyelles brèves et longues ont sensiblement la même durée. Contrairement à ces deux premières qualités, les deux autres, c'est-à-dire [a] et [o], présentent une opposition temporelle bien marquée entre brèves et longues sous les deux débits de parole, normal et rapide. Pour [a], le rapport (Durée Brève/Durée Longue) est de 0,59 (cela signifie que la durée de la brève constitue 59% de la durée de la longue). Ce rapport est de 0,63 en débit rapide, donc écart moins important. Pour la qualité [o], ce rapport est plus important que celui que nous avons

observé auprès de la qualité [a], il est de 0,67 en débit normal et de 0,71 en débit rapide. Nous notons que plus ce rapport est proche de 1 et moins l'écart temporel est important.

En réalité nous ne pouvons pas mélanger toutes les données provenant des 3 locuteurs parce que leurs réalisations sont très différentes entre elles du point de vue statistique, c'est ainsi que nous présentons l'essentiel des résultats pour chaque locuteur à part. Chez le premier locuteur de cette génération (Ma), les résultats des tests statistiques indiquent que les écarts temporels entre voyelles brèves et longues pour les paires: /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ sont significatifs ($p < 0,01$), et ceci sous les 2 conditions expérimentales. Cela signifie que l'opposition temporelle a résisté à la perturbation de la parole rapide. Ces résultats indiquent aussi que les écarts temporels pour les paires /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ sont nuls ($p = ns$). Chez ce locuteur, et contrairement aux 2 autres locuteurs de la même génération, la perturbation de la durée des voyelles par le débit rapide ne s'est pas montrée effective. Pour le deuxième locuteur masculin de cette génération (Ge), les résultats montrent que les écarts temporels pour les paires /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ sont significatifs, et comme pour le premier locuteur, les écarts temporels pour les paires /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ n'existent pas. Le débit rapide a significativement abaissé les durées des voyelles /ɑ:, ɛ, ɛ:, o:, u, u:/, alors qu'il n'a pas affecté les durées des voyelles /a/ et /ɔ/. Malgré cette perturbation, les écarts pour /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ demeurent significatifs sous le débit rapide. Là aussi l'opposition temporelle a résisté à la perturbation de la parole rapide.

Pour la locutrice de cette génération (De), les résultats montrent que les écarts temporels entre voyelles brèves et longues pour les paires /ɛ/ vs. /ɛ:/, /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ sont significatifs, elle est la seule locutrice de cette génération qui présente une différence temporelle significative entre le /ɛ/ bref et le /ɛ:/ long. Cependant, seuls les écarts pour /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ ont résisté à la perturbation de la parole rapide, donc l'écart temporel entre le /ɛ/ bref et le /ɛ:/ long n'est pas robuste. L'opposition /u/ vs. /u:/ n'existe pas. Le débit rapide s'est montré effectivement perturbateur et ceci en réduisant les durées de toutes les voyelles, et en éliminant l'écart temporel entre le /ɛ/ bref et le /ɛ:/ long.

Au plan perceptif

Nous ne nous intéressons qu'aux paires de voyelles où l'écart temporel s'est montré significatif sous les deux débits de parole lors des tests statistiques, c'est-à-dire les paires /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/. Nos calculs des rapports de Weber [(longue-brève)/longue], chez tous les locuteurs de la première génération, pour les deux paires de voyelles mesurées, /a/ vs. /ɑ:/

et /ɔ/ vs. /o:/ et sous les deux débits de parole, normal et rapide, montrent, à une exception près¹⁹, que ces rapports sont supérieurs à 0,20. Ces résultats nous permettent d'affirmer que l'écart temporel entre la brève et la longue en français de la région de la Savoie est un écart **perceptible**, ce qui n'est pas en contradiction avec nos données statistiques. L'étendue de l'écart est assez importante pour que cet écart soit capable, seul, de supporter l'opposition catégorielle (Brève/Longue).

En considérant les résultats des tests statistiques et perceptifs pour les trois locuteurs de cette génération, nous pouvons dresser le bilan suivant: seuls les écarts pour les paires /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ existent, ces écarts sont robustes. Nous pouvons conclure que l'opposition temporelle en français de la région de la Savoie pour les paires /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ chez la première génération, se maintenant sous la perturbation du débit rapide, n'est pas une opposition qui se produit de manière accidentelle, et par conséquent c'est une opposition linguistiquement spécifiée. Ces résultats nous permettent finalement d'affirmer que le geste articulatoire qui permet la production de l'écart temporel entre ces voyelles longues et brèves est un geste **contrôlé**.

Distance spectrale

Au plan statistique

Les distances spectrales significatives sont observées seulement auprès des voyelles qui présentent une opposition temporelle robuste, autrement dit, les deux termes d'une paire qui ont la même durée ont aussi la même qualité vocalique. Le débit rapide a changé significativement certaines valeurs de F1 et de F2 du débit normal chez les trois locuteurs (différemment selon le locuteur), cela prouve que cette condition expérimentale est effectivement perturbatrice de la qualité vocalique. Cette perturbation n'a pas affecté cependant les distances spectrales entre voyelles brèves et longues observées en débit normal. La configuration des distances spectrales entre voyelles dans l'espace vocalique du français de la région de la Savoie est commune à tous les locuteurs de cette génération. En fait, le /a/ bref et /ɑ:/ long sont distants seulement sur l'axe de F2, cette distance est robuste, ils ne sont pas différents au niveau de F1. Le /ɔ/ bref et le /o:/ long sont distants et sur l'axe de F1 et sur l'axe de F2, ces distances sont robustes pour les deux formants.

¹⁹ Locutrice (De), le rapport relatif à la qualité [o] en débit normal est de 0,14 (< 0,20).

Les oppositions qualitatives pour les paires /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ n'existent pas, ni sur F1 ni sur F2.

Au plan perceptif

Pour savoir si les distances spectrales qui se sont montrées statistiquement significatives, c'est-à-dire la distance entre le F2 du /a/ bref et le F2 du /ɑ:/ (le F1 n'est pas intégré dans cette distance parce que nous avons constaté que la distance sur cet axe est le produit du hasard), et les distances entre le F1 et le F2 du /ɔ/ bref et le F1 et F2 du /o:/ long. Nous utilisons la même méthode déjà utilisée pour l'arabe dialectal de Mayadin, c'est-à-dire la méthode basée sur les travaux de Kewley-Port & Watson (1994) et Kewley-Port & Zheng (1999) sur le seuil catégoriel de la qualité vocalique qu'ils évaluent approximativement à 0,60 Bark.

Les résultats de nos calculs montrent que les locuteurs de cette génération produisent tous des distances spectrales supérieures à 0,60 Bark pour les deux paires de voyelles étudiées et sous les deux débits de parole, ce qui est conforme à nos données statistiques (seule exception: locutrice (De) pour la paire /a/ vs. /ɑ:/ en débit normal, la distance = 0,53 Bark).

Les deux termes de la paire /ɔ/ vs. /o:/ présentent une distance spectrale nettement supérieure à la distance entre les deux termes de la paire /a/ vs. /ɑ:/ (une exception: la distance /ɔ/ vs. /o:/ chez le locuteur (Ma) en débit rapide est moins importante que la distance /a/ vs. /ɑ:/). La distance /ɔ/ vs. /o:/ pouvant être le double de la distance /a/ vs. /ɑ:/ dans certaines conditions.

Le débit rapide affecte négativement les écarts qualitatifs (sans test statistique), ceci en réduisant la taille de l'écart entre la voyelle brève et la voyelle longue correspondante (exception: locuteur (Ma) pour l'écart entre /a/ et /ɑ:/). Cependant, le débit rapide ne réussit à effacer aucun des écarts. En nous basant sur ces derniers résultats, nous pouvons finalement affirmer que le /ɔ/ et le /o:/ sont deux qualités vocaliques bien distinctes, et que le /a/ et le /ɑ:/ sont aussi deux qualités distinctes. Cela signifie que ces distances sont intentionnellement, et non pas accidentellement, produites. Les gestes articulatoires qui distancent le /ɔ/ du /o:/ et le /a/ du /ɑ:/ sont finalement des gestes obligatoires, et par conséquent **contrôlés**.

VI.1.2.2. Deuxième génération

Écart temporel

Au plan statistique

Nous rappelons que pour cette génération intermédiaire nous n'avions qu'une seule locutrice (S1) 49 ans en 2007. Les résultats des tests statistiques indiquent que seul l'écart temporel entre le /a/ bref et le long /ɑ:/ existe, comme ils indiquent que cet écart se maintient sous le débit rapide de la parole. Ainsi, les écarts pour les paires: /ɔ/ vs. /o:/, /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/ n'existent pas ($p = ns$), donc nous précisons que l'écart temporel pour la paire /ɔ/ vs. /o:/ qui existe chez la première génération disparaît chez la deuxième. Nous notons ici que le débit rapide n'a pas eu d'effet sur les durées des voyelles.

Au plan perceptif

Nous avons calculé le rapport de Weber [(longue-brève)/longue] pour la seule paire de voyelles /a/ vs. /ɑ:/ et sous les deux débits de parole, normal et rapide. Ce rapport est égal à 0,16 en débit normal et 0,086 en débit rapide. Donc ce rapport est inférieur aux 0,20 requis. Notre conclusion vis-à-vis de ces résultats est que cet écart temporel est perceptible en débit normal, il est cependant peu probable que cet écart puisse seul démarquer perceptivement la brève de la longue. En débit rapide, il est certain que cet écart, à la limite du perceptible, ne peut pas fonder l'opposition catégorielle Brève/Longue. Nous rappelons que nous avons changé le seuil de signification de ($p < 0,010$) à ($p < 0,015$) pour l'opposition de durée entre les deux réalisations /a/ et /ɑ:/ en débit rapide pour la locutrice de la deuxième génération, cela prouve que cet écart est relativement faible du point de vue statistique, ce qui témoigne par ailleurs d'une cohérence entre données statistiques et perceptives.

Les résultats de cette locutrice en débit normal ne sont pas concluants, parce que le rapport de Weber tombe dans un intervalle critique où nous ne saurons pas trancher si cette taille de l'écart temporel est plutôt suffisante ou plutôt insuffisante pour que les deux voyelles puissent se démarquer par leurs durées seulement dans le système phonologique du français de la région de la Savoie. Nous admettons finalement, et sur la base de notre conclusion en I.2.4., que l'écart temporel entre /a/ et /ɑ:/ est **perceptible** sous les deux débits de parole. Selon le même raisonnement que nous avons suivi pour l'arabe dialectal et pour le français de la région de la Savoie pour la première génération, nous pouvons finalement affirmer que le geste articulatoire qui permet la production de l'écart temporel entre le /a/ bref et le long /ɑ:/ est ici **contrôlé**.

Distance spectrale

Au plan statistique

Les distances spectrales significatives ne sont pas observées seulement auprès des voyelles qui présentent des oppositions temporelles robustes, comme nous avons remarqué auprès de la première génération. En effet, nous remarquons que chez la locutrice (S1) les deux réalisations /ɔ/ et /o:/ sont distantes sur l'axe de F1 et sur l'axe de F2 alors qu'elles ne présentaient pas d'opposition temporelle. Nous observons aussi que les deux réalisations /a/ et /ɑ:/ sont distantes sur l'axe de F2 seulement (pour ces deux réalisations il y a eu une opposition temporelle). Les réalisations /u/ et /u:/, comme les réalisations /ɛ/ et /ɛ:/ ne montrent aucune distance spectrale sur aucun axe. Après avoir soumis ces oppositions au débit rapide nous avons constaté que le débit rapide a eu très peu d'effet, sur les F2 seulement, sans changer les distances obtenues avec le débit normal.

L'opposition /a/ vs. /ɑ:/ qui n'existe qu'au niveau de F2, s'est montrée robuste. Elle n'existe pas au niveau de F1. L'opposition /ɔ/ vs. /o:/ qui existe au niveau de F1 et au niveau de F2, s'est montrée robuste pour les deux formants.

Pas de distances spectrales observées pour les paires /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/, ni sur F1 ni sur F2.

Au plan perceptif

Les données de cette locutrice à propos de l'opposition /a/ vs. /ɑ:/ sont décidément problématiques, puisque les différences ici non plus ne permettent pas de trancher, se situant autour du seuil de la perception de la qualité vocalique. En effet, nous avons évalué la distance spectrale à 0,33 Bark en débit normal et à 0,24 Bark en débit rapide alors que le seuil est estimé à 0,30 Bark. En nous faisant aider par les résultats des tests statistiques qui montrent robuste cette opposition, nous admettons plus facilement la thèse selon laquelle cette distance est perceptible en débits normal et rapide.

Les résultats de nos calculs montrent que notre locutrice produit des distances spectrales supérieures à 0,60 Bark pour la paire de voyelles /ɔ/ vs. /o:/, en débit normal, 1,49 Bark, et en débit rapide, 1,41 Bark. Ces distances spectrales, nettement supérieures aux distances entre les deux termes de la paire /a/ vs. /ɑ:/, sont suffisantes pour démarquer les deux qualités vocaliques perceptivement.

Nous pouvons soutenir ici que chez notre locutrice (S1) l'opposition /a/ vs. /ɑ:/ est basée à la fois sur la différence temporelle et sur la distance spectrale, et que l'une des deux composantes, seule, ne saurait pas supporter l'opposition catégorielle Brève/Longue. Nous pouvons soutenir aussi que l'opposition /ɔ/ vs. /o:/ n'est basée que sur la distance spectrale.

En nous appuyant sur ces derniers résultats, nous pouvons finalement affirmer que le /ɔ/ et le /o:/ sont deux qualités vocaliques bien distinctes, et que le /a/ et le /ɑ:/ ne sont pas deux qualités distinctes. Cependant, toutes les distances spectrales pour les deux paires /ɔ/ vs. /o:/ et /a/ vs. /ɑ:/ sont délibérément, et non pas fortuitement, produites. Les gestes articulatoires qui distancient le /ɔ/ du /o:/ et le /a/ du /ɑ:/ sont finalement des gestes obligatoires, et par conséquent **contrôlés**.

VI.1.2.3. Troisième génération

Écart temporel

Au plan statistique

Nos tests statistiques concernant la troisième génération montrent qu'il n'y a pas de différence au niveau de la durée entre voyelles brèves et longues pour aucune des 4 qualités étudiées et chez aucun des locuteurs de cette génération. Celles qui sont supposées brèves pouvant être légèrement plus longues que celles qui sont supposées être longues. C'est pour cette raison que nous n'analyserons pas les données sous le débit rapide pour cette génération, notre but étant de tester la robustesse des écarts entre brèves et longues (des écarts qui n'existent pas chez cette génération).

Distance spectrale

Au plan statistique

L'analyse de la variance qui a testé les différences de qualité au sein de chacune des 4 paires (brève vs. longue) montre que les deux réalisations /ɔ/ et /o:/ sont différentes ($p < 0,01$) entre elles et au niveau de F1 et au niveau de F2, chez les deux locuteurs. Ce n'est pas le cas pour les réalisations des 3 autres paires.

Au plan perceptif

Les deux termes de la paire /ɔ/ vs. /o:/, présentent un écart largement supérieur au seuil de 0,60 Bark auparavant détaillé, ils sont par conséquent deux entités vocaliques qualitativement distinctes.

En effet, la locutrice de cette génération présente ainsi une distance égale à 1,53 Bark et le locuteur une distance égale à 0,91 Bark, ces distances sont suffisantes, à elles seules, pour supporter une opposition qualitative. Nos résultats statistiques et perceptifs convergent enfin vers une conclusion selon laquelle il n'y pas de différence temporelle entre voyelles chez cette génération. Et que la distance séparant le /ɔ/ du /o:/ sur l'axe de F1 et sur l'axe de F2 est ainsi volontairement, et non pas accidentellement, produite, donc le geste responsable de cette différence qualitative est un geste **contrôlé**.

Donc si nous prenons en compte les résultats des trois générations, au plan statistique et au plan perceptif, nous constatons que le procédé phonétique de la quantité vocalique a réellement existé dans le français de la région de la Savoie moyennant essentiellement deux corrélats acoustiques, temporel et spectral. Le corrélat temporel a entièrement disparu chez la jeune génération. Un seul corrélat spectral a perduré, celui de l'opposition /ɔ/ vs. /o:/.

Notre observation intergénérationnelle de la mutation de la quantité vocalique nous a permis de savoir dans quel ordre les oppositions temporelles ont disparu. En effet, nous avons suspecté les quatre paires de voyelles: /a/ vs. /ɑ:/, /ɛ/ vs. /ɛ:/, /ɔ/ vs. /o:/ et /u/ vs. /u:/ d'avoir toutes des oppositions quantitatives (oppositions temporelles associées éventuellement à des différences qualitatives), nous n'avons détecté aucune trace pour l'opposition /u/ vs. /u:/ (contrairement aux trois autres oppositions), aucun résultat aux niveaux statistique et perceptif auprès des trois générations ne peut nous inciter à penser que cette opposition ait existé. Donc si une opposition quantitative a existé pour /u/ vs. /u:/, elle était certainement la première, parmi les quatre étudiées, à disparaître du système phonologique du français de la région de la Savoie. La deuxième opposition à disparaître est sans doute l'opposition /ɛ/ vs. /ɛ:/. En effet, cette opposition s'est manifestée auprès de la locutrice (De) de la première génération avec un débit normal, mais elle n'a pas résisté à la perturbation du débit rapide, contrairement aux deux autres oppositions /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/ qui se sont montrées significatives et significativement robustes chez les trois locuteurs de la première génération.

La troisième opposition à perdre le trait quantitatif, est l'opposition /a/ vs. /ɑ:/. Cette opposition, qui existe encore au niveau temporel et spectral chez la deuxième génération, disparaît intégralement chez la troisième génération.

Nos données indiquent que la quatrième, et la dernière, opposition à perdre le trait quantitatif, est l'opposition /ɔ/ vs. /o:/, qui est basée sur la différence temporelle et sur la

différence spectrale chez la première génération, devient une opposition qui ne base son contraste que sur la distance spectrale, cette opposition qualitative est le seul héritage qu'a reçu la jeune génération de la quantité vocalique des générations précédentes. Nous pouvons dire alors que l'opposition /ɔ/ vs. /o:/, qui est devenue purement qualitative, est une opposition qui n'a perdu que le corrélat temporel de son contraste mais qui continue à vivre.

VI.1.3. Comparaison relative au comportement de la quantité vocalique en arabe dialectal de Mayadin et en français de la région de la Savoie

Nous voulons comparer dans ce paragraphe quelques aspects du comportement de la quantité vocalique dans les deux systèmes qui ont fait l'objet de notre étude. Nous nous intéresserons au phénomène de la redondance en contrôle des gestes articulatoires responsables des corrélats acoustiques de la quantité vocalique, aux rapports de durée entre brèves et longues, aux axes et aux étendues des distances spectrales et enfin à l'évolution de nos deux systèmes.

Nous observons un comportement commun à l'arabe de Mayadin et au français de la Savoie chez la première et la deuxième générations en ce qui concerne la quantité vocalique. En effet, toutes les paires qui présentent des différences temporelles entre brèves et longues associent régulièrement à ces différences temporelles des distances spectrales significatives. Effectivement, nous n'avons observé, dans les deux systèmes, aucune paire qui base son opposition sur le corrélat temporel uniquement, aucun locuteur faisant usage de la quantité vocalique ne fait exception à cette règle.

Pour les trois paires de voyelles /i/ vs. /i:/, /a/ vs. /a:/ et /u/ vs. /u:/ de arabe dialectal de la ville de Mayadin, brèves et longues s'opposent temporellement et spectralement, les deux corrélats sont alors contrôlés. Les écarts temporels sont suffisants pour démarquer perceptivement la catégorie de la brève de celle de la longue. La distance spectrale pour /i/ vs. /i:/ est suffisante pour une démarcation catégorielle, pas celles que nous observons pour les paires /a/ vs. /a:/ et /u/ vs. /u:/. Les distances spectrales sont cependant redondantes pour les trois paires.

Cette redondance est un trait commun à nos deux systèmes. En effet, chez la première génération du français de la Savoie, pour les deux paires de voyelles /a/ vs. /ɑ:/ et /ɔ/ vs. /o:/, brèves et longues s'opposent temporellement et spectralement, comme pour l'arabe dialectal de Mayadin, les deux corrélats sont contrôlés. Les écarts temporels sont suffisants pour démarquer perceptivement la catégorie de la brève de celle de la longue. Les distances spectrales pour les deux paires, pourtant contingentes, sont aussi suffisantes pour une démarcation catégorielle. Ainsi nous pouvons admettre que le système du français de la région de la Savoie comporte plus de redondance, puisque l'un des deux corrélats est suffisant pour préserver la mission informationnelle de ce système vocalique pour les deux paires en question, alors que pour le système de Mayadin le corrélat spectral n'est pas suffisant pour les

oppositions /a/ vs. /a:/ et /u/ vs. /u:/, mais dans tous les cas les deux systèmes comportent de la redondance.

En observant les durées des voyelles dans les deux systèmes vocaliques qui ont fait l'objet de notre étude, nous trouvons des rapports de durée similaires entre voyelles brèves et longues. En fait, nous avons deux manières de considérer les rapports entre les durées, soit nous considérons ce que la durée de la brève représente de celle de la longue: (Durée Brève/Durée Longue), soit nous considérons le rapport de Weber: [(longue-brève)/longue], nous trouvons qu'avec les deux visions, les résultats sont comparables.

(Durée Brève/Durée Longue): en arabe dialectal (tous locuteurs et toutes conditions expérimentales confondues), concernant l'opposition /i/ vs. /i:/, la brève constitue 57% de la durée de la longue, pour l'opposition /a/ vs. /a:/ la brève constitue 59% et pour l'opposition /u/ vs. /u:/ la brève constitue 60% de la durée de la longue. Chez la première génération de locuteurs du français de la Savoie (tous locuteurs et tous débits confondus), concernant l'opposition /a/ vs. /ɑ:/, la brève constitue 61% de la durée de la longue et pour l'opposition /ɔ/ vs. /o:/ la brève constitue 69%.

Rapport de Weber [(longue-brève)/longue]: en arabe dialectal de Mayadin (tous locuteurs et toutes conditions expérimentales confondues), le rapport de Weber pour l'opposition /i/ vs. /i:/ est de 0,43, il est de 0,41 pour l'opposition /a/ vs. /a:/ et il est de 0,40 pour l'opposition /u/ vs. /u:/. Le rapport de Weber pour l'opposition /a/ vs. /ɑ:/ chez la première génération du français de la Savoie (tous locuteurs et tous débits de parole confondus) est de 0,40, alors qu'il est de 0,34 pour l'opposition /ɔ/ vs. /o:/.

Ces rapports montrent que les oppositions de durées, malgré une légère avance de l'arabe de Mayadin (le contraste temporel en arabe de Mayadin entre brèves et longues est plus important), sont aussi importantes en français de la région de la Savoie qu'en arabe dialectal de Mayadin et sont tout à fait comparables, donc si elles constituent la base du contraste quantitatif en arabe, elles peuvent très bien l'être en français aussi. Ainsi, la comparaison de l'évolution des deux systèmes reste permise puisque les proportions ont le même ordre de grandeur. Les durées ont eu les mêmes patrons de réactions au débit rapide de la parole, puisque dans les deux systèmes le débit rapide a abaissé les durées des voyelles sans affecter les écarts temporels entre les brèves et les longues. Seule exception est l'opposition /ɛ/ vs. /ɛ:/ qui a perdu son écart temporel sous l'effet du débit rapide chez la locutrice de la première génération.

L'arabe de Mayadin, comme le français de la Savoie, montre que les oppositions qualitatives peuvent se situer sur les deux axes, F1 et F2, c'est le cas du /i/ vs. /i:/ en arabe de Mayadin comme c'est le cas du /ɔ/ vs. /o:/ en français de la Savoie, ou sur un seul axe, en effet, les oppositions qualitatives /a/ vs. /a:/ et /u/ vs. /u:/ en arabe de Mayadin se trouvent uniquement sur l'axe de F1, l'opposition qualitative de /a/ vs. /ɑ:/ en français de la Savoie se trouve uniquement sur l'axe de F2. Cette dernière comparaison nous révèle que le corrélat spectral de la quantité vocalique ne suit pas une configuration déterminée, c'est-à-dire que l'allongement de la voyelle dû à la quantité ne contraint pas la brève à se démarquer spectralement de la longue toujours avec le même comportement, puisque la distance spectrale peut se situer sur l'axe de F1 seul, sur l'axe de F2 seul ou sur les deux axes en même temps. Autre révélation, c'est qu'une distance, même si elle ne se situe que sur un seul axe, peut être suffisante pour fonder l'opposition catégorielle entre deux voyelles (cas du /a/ vs. /ɑ:/ en français de la région de la Savoie).

Notre observation de l'évolution du français de la région de la Savoie nous a incité à nous poser la question suivante: est-ce que la disparition du corrélat temporel de l'opposition quantitative est un signe qui nous permet de prédire l'avenir du corrélat temporel de l'opposition quantitative en arabe dialectal de Mayadin?

En fait, il est théoriquement possible que l'opposition /i/ vs. /i:/ de l'arabe dialectal connaisse la même évolution qu'a connue l'opposition /ɔ/ vs. /o:/ du français de la région de la Savoie, qui, chez la première génération, se basait sur les deux corrélats temporel et spectral, et qui, chez la jeune génération, est devenue uniquement qualitative. La distance spectrale pour cette paire chez la jeune génération n'a pourtant pas augmenté par rapport à la première génération, c'est la preuve qui atteste que l'opposition qualitative est capable, toute seule, de supporter l'opposition catégorielle (Brève vs. Longue), et que l'opposition temporelle est un corrélat redondant du point de vue de l'économie des gestes articulatoires, comme nous l'avons mentionné plus haut. Ainsi, l'opposition /i/ vs. /i:/ où chacun des écarts, temporel et spectral est suffisant, a réellement le potentiel de devenir une opposition uniquement qualitative.

En effet, si nous calculions les distances spectrales selon la proposition de Kewley-Port & Zheng (1999), la distance entre le /ɔ/ et /o:/, variant entre 0,91 et 1,53 Bark chez la deuxième et la troisième génération (où la différence est uniquement spectrale), et entre 0,76

et 1,80 Bark pour la première génération (où la différence est à la fois temporelle et spectrale) est du même ordre de grandeur que la distance entre /i/ vs. /i:/ qui varie entre 1,22 et 1,44 Bark selon la condition expérimentale. Donc si l'évolution des langues suit des parcours communs, cette comparaison est en train de nous apprendre que la différence temporelle entre le /i/ bref et le /i:/ long en arabe dialectal est vouée à disparaître au profit de la distance spectrale perceptivement suffisante et robuste.

Un autre renseignement peut être tiré de l'observation de l'évolution de l'opposition /ɔ/ vs. /o:/ en français de la Savoie, c'est que le corrélat spectral est préféré au corrélat temporel dans l'évolution des oppositions quantitatives. Ainsi, les distances sur l'axe de F1 pour les paires /a/ vs. /a:/ et de /u/ vs. /u:/ seraient les prémisses d'un changement qui signalent la mutation du système de Mayadin vers un système où les oppositions ne se situent qu'au niveau qualitatif. En fait, ces distances, comme nous l'avons affirmé à plusieurs reprises, ne sont pas produites par hasard, il est donc légitime de penser qu'il existe un objectif derrière leur présence, autre que le rôle démarcatif au niveau perceptif dont nous avons démontré la redondance. Autrement, elles sont vouées disparaître conformément au principe de l'économie des gestes articulatoires.

Quel avenir pour le système de Mayadin?

Les différences quantitatives peuvent reposer seulement sur le corrélat spectral et devenir ainsi uniquement qualitatives. Nous ne prévoyons pas cependant une disparition d'aucune des oppositions Brève vs. Longue en arabe dialectal.

Nous pensons que si l'opposition /a/ vs. /ɑ:/ a définitivement disparu (et peut-être avant elle les oppositions /ɛ/ vs. /ɛ:/ et /u/ vs. /u:/), c'est parce que le lexique du français de la Savoie ne comportait pas assez d'items qui fondent leurs contrastes uniquement sur la différence acoustique entre /a/ et /ɑ:/, donc garder dans la mémoire de cette langue un procédé coûteux (la quantité vocalique), pour peu d'usage, semble être une tâche linguistique non rentable. En outre, contrairement à l'arabe, la quantité vocalique en français de la Savoie n'est pas un produit d'un procédé linguistique systématique, un procédé enraciné dans la morphologie de la langue. En arabe, où pour certains items lexicaux qui suivent une structure syllabique précise (un schème), l'allongement systématique de la durée d'une voyelle précise du schème, entraîne un changement du sens de l'item résultant, les deux items entretiennent alors un lien sémantique relativement fort. En effet, la morphologie dérivationnelle de l'arabe produit

fréquemment, par le procédé de la quantité vocalique, à partir d'une racine consonantique, des unités lexicales appartenant au même champ sémantique, sans ou avec changement de catégorie syntaxique. Voir annexe n° XVIII pour quelques exemples sur la proximité sémantique entre deux unités lexicales qui ne sont différenciées que par le trait de la quantité vocalique.

En outre, l'observation de la fréquence des oppositions lexicales en arabe qui se basent sur les corrélats de la quantité vocalique nous indique qu'il est tout simplement impossible de renoncer (en fusionnant brève et longue en une seule voyelle) à tant d'unités lexicales dans le lexique dialectal et sûrement standard. Un autre point qui argumente en faveur du maintien de ces oppositions, ce sont les propriétés distributionnelles des items concernés, en fait, deux items différenciés par la quantité vocalique peuvent se rencontrer dans les mêmes contextes lexicaux puisqu'ils peuvent appartenir au même champ sémantique, ainsi l'entourage lexical ne peut pas être un indice qui oriente le locuteur vers la perception du mot émis, l'information sur l'identité du mot n'est véhiculée que par le trait phonétique. Il nous semble que ce n'est pas le cas pour les mots qui se contrastent par la quantité vocalique en français de la Savoie. Nous rappelons que 5 des 6 items du corpus de l'arabe de Mayadin appartiennent à 2 champs sémantiques, alors que les 8 items du corpus du français de la Savoie appartiennent à 8 champs sémantiques différents, aucun lien de sens ne peut être établi entre 2 de ces items. Finalement, comme les contextes, lexical et pragmatique, contribuent à la construction de l'information sur l'identité sémantique de l'item considéré, et comme en français de la Savoie les items contrastés par la quantité vocalique n'avaient pas, en règle générale, les mêmes occurrences contextuels, l'entourage lexical restait un indice fort sur l'identité de l'item, d'où une importance moindre du trait quantitatif pour distinguer les deux items objets du contraste.

Donc une fréquence très élevée d'oppositions basées sur la quantité vocalique et des contiguïtés sémantiques et distributionnelles font que le trait phonétique de la quantité vocalique en arabe dialectal de Mayadin joue un rôle déterminant dans la perception d'un nombre considérable d'unités du lexique de ce dialecte (et de l'arabe standard sûrement), par conséquent, une suppression des oppositions Brève *vs.* Longue, comme c'était le cas pour l'opposition /a/ *vs.* /ɑ:/ en français de la région de la Savoie, n'est certainement pas concevable.

Enfin, la thèse selon laquelle le système de Mayadin serait en phase de mutation vers un système où les oppositions ne se situent qu'au niveau qualitatif, reste une thèse plausible tant que cette mutation ne compromet pas la distinction phonétique entre les brèves et les longues.

VI.2. *Plasticité des systèmes vocaliques*

VI.2.1. Langues des aborigènes d'Australie

Les résultats du travail de recherche effectué par Fletcher & Butcher (2003) qui avait pour but de sonder les effets de la structure prosodique dans la mise en avant des contrastes paradigmatiques pour les voyelles de trois langues des aborigènes du nord de l'Australie (Kayardild, Dalabon et Mayali), tendent à montrer que les systèmes vocaliques se construisent selon le principe du contraste suffisant, contraire au principe de la dispersion maximale. Voir figure 94 pour la situation géographique de ces langues.

La Théorie de Dispersion - Focalisation, la TDF (Schwartz *et al.*, 1997b), avance que les voyelles occupent l'espace vocalique de telle manière que les distances qui les séparent sont maximales, cela explique pourquoi les trois premières voyelles à apparaître dans un système vocalique sont /i/, /a/ et /u/, la première voyelle intérieure (non périphérique) à apparaître est la voyelle /i/ barré et les cinq premières voyelles sont /i/, /a/, /u/, /e/ et /o/. Cette hypothèse a été ébranlée par les données de Butcher (1994) au profit du contraste suffisant.

En effet, Butcher (1994) a précédemment procédé à l'analyse des formants de certaines des langues des aborigènes d'Australie, telles Arrernte, Warlpiri et Burarra. Ces langues disposent de systèmes à deux, à trois et à cinq voyelles respectivement. Ses résultats montrent que la taille de l'espace vocalique est relative à la taille de l'inventaire de voyelles du système considéré, c'est-à-dire que l'espace vocalique est d'autant plus petit que le nombre de voyelles du système en question est faible. Cela va effectivement à l'encontre du principe de la dispersion maximale. Butcher affirme que les espaces vocaliques de ces langues tendent à être compactes, comparés aux espaces vocaliques d'autres langues à un inventaire de voyelles plus important, tels l'anglais ou le suédois. Ces données plaident finalement en faveur du principe de la dispersion suffisante au détriment du principe de la dispersion maximale.

Les langues étudiées dans Fletcher & Butcher (2003) sont le Kayardild, le Dalabon et le Mayali. Les trois langues sont décrites comme des langues à accent. Les deux dernières langues appartiennent à la même famille de langues alors que le Kayardild appartient à une autre famille de langues australiennes. Le système du Kayardild dispose de trois voyelles avec un contraste quantitatif, le Dalabon et le Mayali ont des systèmes à six et cinq voyelles respectivement. L'objectif de l'étude de Fletcher & Butcher (2003) est d'explorer les effets de la position dans le groupe intonatif et les effets de l'accent phrastique sur les valeurs des formants dans ces trois langues. Il y a un autre objectif relatif au contraste quantitatif en

Kayardild, les auteurs veulent en effet vérifier si les voyelles brèves se situent dans une zone centrale de l'espace vocalique par rapport aux voyelles longues, comme c'est le cas pour l'arabe.

Pour le Dalabon et le Mayali, le corpus était constitué de 6 minutes de discours spontané de deux locuteurs mâles. Pour le Kayardild, il y a eu deux narrations séparées d'un

homme et d'une femme, de 9 et de 11 minutes respectivement. Le Dalabon et le Kayardild sont parlés par moins que 10 locuteurs alors que le Mayali dispose de plus de 1000 locuteurs (Données vérifiées en 2003).

Sur la base de l'analyse de l'intonation, les auteurs répartissent les syllabes en trois catégories, soit la syllabe est identifiée comme accentuée (la syllabe est associée à un accent de F0), soit comme non accentuée, soit la syllabe est identifiée comme finale (c'est lorsqu'elle se rencontre en fin de groupe intonatif). Les auteurs rapportent que l'accentuation finale est souvent absente en Mayali et en Kayardild n'avaient pas, l'accent frappe la syllabe pénultième du constituant intonatif. Le Dalabon, par contre, pouvait avoir une accentuation finale.

En ce qui concerne les deux langues le Mayali et le Dalabon, l'espace acoustique est réduit (compact) en quelque sorte, avec un effet faible de l'accent.

Alors que les voyelles accentuées en Mayali occupent un espace légèrement plus grand que l'espace qu'occupent leurs homologues non accentuées, cette différence, due au passage de la non accentuation à l'accentuation, n'est pas significative statistiquement ($p = ns$).

Fletcher & Butcher (2003) rapportent qu'en le Dalabon, les voyelles accentuées ont aussi tendance à être différentes de celles qui ne sont pas accentuées, à part les /i/ qui ont à peu près les mêmes formants dans les deux conditions d'expérimentation. La seule voyelle qui a présenté une différence claire et statistiquement significative dans les deux langues était



Figure 94. Carte du continent australien et origines géographiques des trois langues des aborigènes du nord de l'Australie étudiées par Fletcher & Butcher (2003).

la voyelle /a/, le Dalabon ($p < 0,05$); le Mayali ($p < 0,05$). La valeur de F1 est légèrement augmentée (voyelle plus basse) dans les deux langues, donc l'accentuation a un effet de décentralisation. Cette tendance à la décentralisation, nous l'avons aussi observée en arabe dialectal d'Amman, Jordanie (cf. de Jong & Zawaydeh, 2002), lorsque l'accent frappe les voyelles basses. En arabe dialectal de Mayadin, l'interrogation a augmenté la valeur de F2 du /a:/ long et a augmenté très peu le F1 du /i:/ long.

Selon Fletcher & Butcher (2003), les voyelles antérieures fermées dans les deux langues (le Dalabon et le Mayali) sont plus décentralisées quand elles sont frappées par l'intonation de la fin de phrase que lorsqu'elles sont sous la condition de l'accentuation. Le Dalabon ($p < 0,02$); le Mayali ($p < 0,01$). Cela s'est surtout manifesté sur F1 en Dalabon, et sur F1 et F2 du Mayali. En Dalabon, les voyelles mi-antérieures sont devenues légèrement plus antérieures ($p < 0,05$) en fin de groupe intonatif, mais les différences entre les 3 conditions²⁰ d'expérimentation ne sont pas significatives pour la voyelle /e/, parce qu'il y a eu seulement 3 occurrences pour cette voyelle en finale intonative, ce qui n'était pas suffisant pour l'étude statistique. Le même problème s'était produit pour les voyelles mi-postérieures arrondies dans les deux langues. Les voyelles ouvertes étaient sensibles à la présence de l'accent, puisqu'il y a eu une différence significative en ce qui concerne le F2 de /a/ entre les conditions accentuée et fin de groupe intonatif en Dalabon. Par contre, les voyelles postérieures arrondies en Dalabon étaient significativement centralisées en fin de groupe intonatif, avec une valeur de F2 significativement plus élevée ($p < 0,05$). En Mayali, F2 était plus bas ($p < 0,02$), aboutissant ainsi à un /u/ plus périphérique.

Aucun effet n'a été détecté pour le contexte prosodique sur la sixième voyelle du Dalabon. Cette voyelle centrale fermée arrondie reste située vers le centre de l'espace vocalique à proximité de la voyelle postérieure fermée. Le contraste entre le /i/ barré et le /u/ n'est pas produit par tous les locuteurs du Dalabon. Il est à noter qu'il n'y a pas de différence significative F1/F2 entre les versions: fin de groupe intonatif et non accentuation pour ces deux voyelles.

Pour la langue Kayardild, alors que le contexte prosodique n'a pas d'effet significatif sur F1/F2 des voyelles fermées chez aucun des locuteurs, la voyelle brève était significativement différente de la longue chez le locuteur homme (F1, $p < 0,05$; F2, $p < 0,05$), comme il y avait une interaction significative entre la longueur de la voyelle et le contexte

²⁰ Nous rappelons les trois catégories de syllabes: 1- non accentuation. 2- accentuation. 3- intonation de fin de phrase.

prosodique pour le locuteur femme ($F1$, $p < 0,001$; $F2$, $p < 0,05$). En d'autres termes, les voyelles brèves fermées sont plus centralisées dans des contextes non accentués, et les voyelles longues fermées sont plus décentralisées dans le contexte de fin de groupe intonatif. Concernant les voyelles postérieures arrondies, le contexte prosodique ($p < 0,01$) et la longueur de la voyelle ($p < 0,05$) ont eu tous deux un effet significatif pour le locuteur homme du Kayardild. Les voyelles longues en condition de fin de groupe intonatif sont plus postérieures que toutes les autres voyelles postérieures arrondies fermées, avec des voyelles longues postérieures accentuées légèrement plus décentralisées que les postérieures non accentuées. Bien que les voyelles longues tendent à être plus périphériques que les voyelles brèves chez le locuteur femme, il n'y a pas eu d'effet significatif du contexte prosodique ou de la longueur de la voyelle sur les voyelles postérieures.

Finalement, concernant les voyelles basses centrales, il y a eu un effet significatif du contexte prosodique et de la longueur sur $F1$ pour les 2 locuteurs homme et femme, ces effets sont cependant plus marqués chez le locuteur homme que chez le locuteur femme.

Les voyelles brèves montrent une centralisation significative par rapport aux voyelles longues. Une différence, cependant non significative, était détectée entre les effets de la condition de fin de groupe intonatif et ceux de la condition de non accentuation sur les voyelles brèves.

En guise de conclusion de leur étude Fletcher & Butcher (2003) rappellent que l'espace acoustique des voyelles des trois langues australiennes n'est pas augmenté sous les conditions de l'accentuation, à part pour le /a/. La tendance à la décentralisation sous l'effet de l'accentuation se manifeste le plus auprès de la langue Mayali, en comparaison avec les deux autres langues. En condition de fin de groupe intonatif, les syllabes sont plus longues par rapport à la condition où elles ne sont pas en position finale. Cependant, les cibles des voyelles sont mieux établies en condition de fin de groupe intonatif, ce qui veut dire que cette zone est très importante pour mettre en évidence des effets prosodiques.

En Kayardild, la longueur est en rapport direct avec la périphéricité de la voyelle, en effet, en condition de fin de groupe intonatif les voyelles longues occupent des points plus extrêmes dans l'espace vocalique. Dans les deux autres langues, les voyelles en position finale sont plus périphériques aussi, même si ce n'est pas toujours le cas pour toutes les voyelles. Dans les 3 langues, la syllabe finale est plus longue que la syllabe non finale. En Kayardild, les voyelles brèves ont tendance à être centralisées, ce qui laisse penser à une expansion de l'espace vocalique reliée à la durée des voyelles. L'intonation de fin de phrase

se montre un endroit prosodiquement proéminent dans les 3 langues. Finalement, les auteurs affirment qu'un regard global sur les espaces vocaliques des trois langues étudiées permet de soutenir le principe du contraste suffisant plutôt que la dispersion maximale.

Fletcher & Butcher (2003) et Butcher (1994) soutiennent le principe du contraste suffisant parce que les trois langues que nous venons d'examiner sont essentiellement défectives en voyelles hautes. En effet, malgré le fait que ces langues possèdent un système à plus de trois voyelles, leurs voyelles ne remplissent pas la périphérie de l'espace vocalique comme le prédit la TDF. Fletcher & Butcher (2003) utilisent le terme "compact" pour désigner la réduction de ces espaces. Nous voudrions montrer ici qu'il est possible que les langues ne se conforment pas à la TDF, parce que la configuration de leurs espaces vocaliques n'est pas achevée alors, elle est en évolution.

Nous avons mentionné dans l'introduction de ce manuscrit que les perturbations sont des contraintes qui poussent un système vocalique vers son évolution, puisqu'elles permettent aux systèmes d'explorer leurs potentiels organisationnels. Ainsi l'adaptation momentanée à une contrainte donnée indique au système considéré qu'il est capable de recourir à une telle stratégie adaptative pour que l'information linguistique survive à la perturbation, laquelle peut finalement être adoptée définitivement comme un élément faisant partie intégrante du mécanisme du système en question.

Ainsi les perturbations, en faisant apparaître ou disparaître un ou des traits distinctifs, qui, dans des conditions normales de l'émission de la parole, se manifestent autrement, contribuent à l'évolution en révélant les propriétés organisationnelles. En arabe dialectal de Mayadin, la perturbation par la prosodie (l'interrogation) a fait en effet disparaître l'écart sur l'axe de F2 entre le /a/ bref et le /a:/ long qui devient non significatif sous l'interrogation. Cette même perturbation a fait apparaître un écart sur l'axe de F2 entre le /u/ bref et le /u:/ long qui n'est pas significatif sous le schéma mélodique déclaratif. Cette dernière révélation nous a permis de prédire une opposition uniquement qualitative entre le /u/ bref et le /u:/ long puisque cette potentialité existe.

Nous cherchons maintenant dans ces langues australiennes des indices qui nous indiquent que sous la perturbation par la prosodie, les espaces vocaliques ne sont pas hermétiques aux voyelles hautes, ce qui constitue la preuve que même si ces langues sont

défectives en voyelles hautes, il ne leur est pas impossible d'en produire lorsque la stratégie adaptative l'exige.

Si nous observons les espaces vocaliques de ces langues, une par une, nous trouvons que, comme pour l'arabe dialectal de Mayadin ou d'Amman (de Jong & Zawaydeh, 2002), les voyelles sont en général sensibles aux variations intonatives d'accent ou de durée intrinsèque (Bref vs. Long) et co-intrinsèque (fin de groupe intonatif vs. autres positions).

Pour la langue Dalabon (figure 95), l'espace vocalique est effectivement réduit dans des conditions normales de l'émission de la parole, le F1 pour [i] est de 350 Hz, il est de 400 Hz pour [u], la thèse selon laquelle cette langue est déficiente en voyelles haute est plausible. Cette thèse devient moins facilement défendable lorsque nous considérons ce même espace vocalique sous la condition de fin de groupe intonatif (où les syllabes sont globalement plus longues par rapport à la condition où elles sont en d'autres positions), notamment pour [i] où un F1 autour de 325 Hz pour un locuteur masculin n'est pas vraiment curieux.

Moyennes des valeurs formantiques (Hz)-Dalabon.

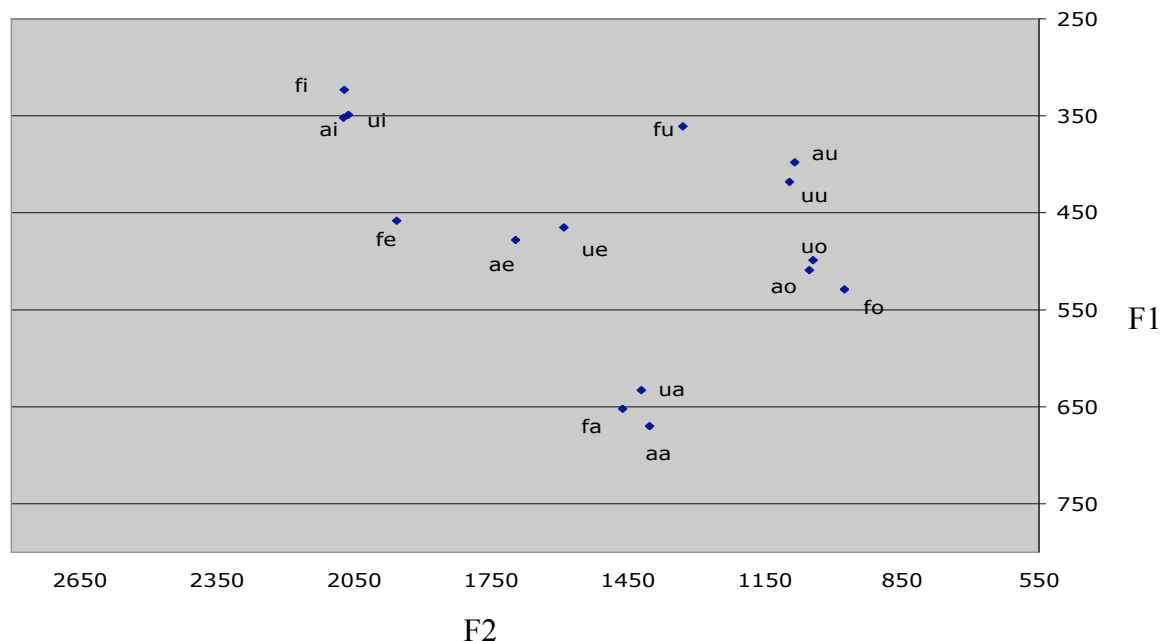


Figure 95. Espace vocalique de la langue Dalabon, langue du nord de l'Australie parlée par moins de 10 locuteurs en 2003. Moyenne de deux locuteurs masculins. La 1^{re} lettre désigne la condition de mesure: accentuée (a), non accentuée (u) et en fin de groupe intonatif (f), et la 2^e désigne la qualité vocalique mesurée.

Lorsque nous considérons la langue Mayali, nous trouvons que la configuration de l'espace vocalique est différente de celle de la langue Dalabon, puisque cette langue est clairement déficiente en voyelles hautes (figure 96). En effet, aucune des conditions de

l'émission de la parole n'a réussi à restituer la partie haute de l'espace, le F1 du [i] demeurant supérieur à 400 Hz, et celui du [u] autour de 450 Hz. Donc cette langue est clairement déficiente en voyelles hautes.

Moyennes des valeurs formantiques (Hz)-Mayali.

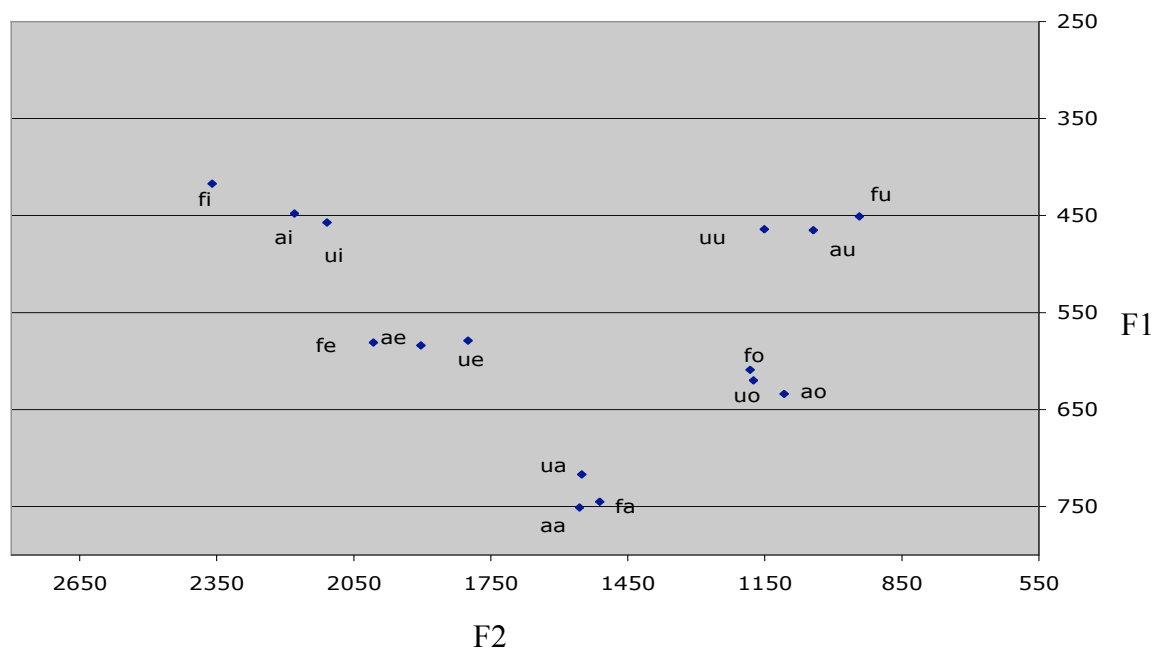


Figure 96. Espace vocalique de la langue Mayali, langue du nord de l'Australie parlée par plus de 1000 locuteurs en 2003. Moyenne de deux locuteurs masculins. La 1^e lettre désigne la condition de mesure: accentuée (a), non accentuée (u) et en fin de groupe intonatif (f), et la 2^e désigne la qualité vocalique mesurée.

Concernant la langue Kayardild, comme en arabe dialectal de Mayadin, les voyelles brèves forment un triangle à l'intérieur du triangle des longues correspondantes, donc les longues sont sans surprise plus périphériques que les brèves. Même lorsque nous considérons le triangle des voyelles longues, l'espace vocalique du locuteur masculin de la langue Kayardild (figure 97) reste plutôt réduit, il est d'une taille effectivement intermédiaire, moins étendu que celui de la langue Dalabon mais plus étendu que celui de la langue Mayali. En effet, les F1 pour le [i:] et le [u:] longs sont autour de 300 Hz, cette valeur de F1 est relativement élevée pour des voyelles hautes, donc nous admettons que cette langue est plutôt déficiente en voyelles hautes.

L'espace vocalique du locuteur féminin de la langue Kayardild (figure 98) est similaire à celui du locuteur masculin, elle est aussi plutôt déficiente en voyelles hautes. Nous avons remarqué auprès de la langue Dalabon une tendance à faire apparaître une voyelle de qualité

[i] quand la syllabe se produit en fin de groupe intonatif, cette tendance est ici plus marquée. En effet, cette locutrice produit un [i:] long caractéristique, où le F1 est inférieur à 300 Hz. Cette tendance à faire apparaître des voyelles hautes là où elles se produisent normalement est ici très claire. Ce phénomène est la preuve que cette langue n'a pas un espace vocalique imperméable aux voyelles hautes, mais qu'elle dispose d'une stratégie adaptative qui a besoin de recourir à cette zone de l'espace pour répondre aux exigences de la communication linguistique.

VI.2.2. Vers un changement réversible vs. irréversible

Nous voudrions finalement montrer que les langues ont des exemplaires extrêmes [i a u], intercalant d'autres voyelles sur la périphérie [i e a o u], pas au centre, conformément à notre théorie, la TDF. La variation des trois extrêmes [i a u] a pour effet, soit de remplir la périphérie par un nouveau prototype (le [e] en arabe dialectal); soit de retrouver un extrême disparu (le [i] du kayardild, une des langues d'Australie défectives en voyelles hautes [i u]). Ce sont deux changements —réversible vs. irréversible— qui contribuent pour finir au maintien des voyelles extrêmes

Un regard d'ensemble sur les données de notre système vocalique de Mayadin et des systèmes vocaliques des langues australiennes que nous venons de développer, nous signale que nous sommes face à deux remarques qui s'imposent par rapport à la théorie des systèmes vocaliques (en nous référant aux héritages de Lindblom et de Stevens, cumulés dans la TDF de Schwartz *et al.*, 1997b).

1. Le système de Mayadin, qui n'aurait pas besoin de différencier plus de trois qualités ([i a u] dans le cas de l'arabe classique) réalise des exemplaires extrêmes pour les voyelles hautes ou basses, qui occupent ainsi les zones des voyelles dites focales. Le triangle vocalique formé par les voyelles brèves est en effet robuste, et les distances entre voyelles brèves, deux à deux, sont supérieures à 1 Bark quelle que soient la condition expérimentale, le système de Mayadin ne présente pourtant pas de réduction générale vers une distinctivité "suffisante". En effet, il est théoriquement possible que le système de Mayadin se restreigne au triangle vocalique réduit formé par les voyelles brèves (cf. figure 64), qu'il peut dédoubler par la différence temporelle et ainsi retrouver les 6 entités vocaliques qui le constituent. Ce n'est en fait pas le cas, nous avons pu démontrer que les deux triangles sont qualitativement séparés (sur l'axe de F1 pour les paires /a/ vs. /a:/ et /u/ vs. /u:/ et sur les deux axes, F1 et F2, pour la paire /i/ vs. /i:/) non pas pour répondre aux contraintes biomécaniques de la parole, mais ils

sont intentionnellement séparés quelles que soient les conditions de l'émission de la parole. Donc le système de Mayadin est construit selon le principe de la dispersion maximale (défendu par la TDF, Schwartz *et al.*, 1997b) et non pas selon le principe de la dispersion suffisante (défendu par Butcher, 1994 et Fletcher & Butcher, 2003).

2. La voyelle brève, en opposition avec [i] long, a fini par occuper la zone de [e] (phénomène courant en arabe dialectal). Et elle s'y maintient définitivement, ne se confondant jamais en qualité, quelles que soient les perturbations, avec sa partenaire longue, en effet nous n'avons pas pu détecter chez aucun locuteur à aucun des axes (F1 et F2) et sous aucune condition expérimentale une fusion entre la brève et la longue. Cette opposition qualitative est réellement massive, puisqu'elle permet à elle seule la création de deux entités vocaliques perceptivement distinctes et phonologiquement indépendantes. Nous avons pu prouver de facto cette assertion lorsque nous avons mesuré la distance spectrale pour la paire /ɔ/ vs. /o:/ en français de la région de la Savoie chez la troisième génération. Cette distance, qui n'est pas associée à une différence temporelle, réussit, donc seule, à démarquer les identités des deux termes de cette paire, le /ɔ/ et /o:/ se comportent effectivement l'un indépendamment de l'autre dans le système phonologique de ce parler du français. Cette distance est cependant du même ordre de grandeur que la distance spectrale qui sépare les deux termes de la paire /i/ vs. /i:/. Donc nous assistons ici à la stabilité d'un nouveau prototype avec la persévération de la variante extrême de cette qualité vocalique, ce dernier phénomène est prédit par la théorie "Perception-for-Action-Control Theory" (cf. Schwartz *et al.*, 2002).

Ces deux observations sont à considérer, dans le cadre de la théorie des systèmes vocaliques, par rapport aux systèmes réduits des langues d'Australie, dont les formants ont été mesurés depuis quelques années par Andrew Butcher et son équipe (Butcher, 1994 pour les langues Arrernte, Warlpiri et Burarra et Fletcher & Butcher, 2003 pour les langues Kayardild, Dalabon et Mayali). Les systèmes du Kayardild, du Dalabon et du Mayali à trois, six et cinq qualités vocaliques respectivement, ne présentent pas, comme l'arabe ou l'inuit, de voyelles hautes [i u] (ceci reste vrai en ajoutant d'autres langues en cours de description à celles qui sont déjà disponibles, Butcher, com. pers.). Cependant récemment Fletcher & Butcher (2003), ont testé l'influence de conditions prosodiques perturbatrices d'accentuation, réductrices ou magnifiantes, et en finale de groupe prosodique par exemple, Nous avons pu observer la réapparition, sous cette dernière condition, des exemplaires extrêmes, comme [i], nous avons vérifié cela auprès de la langue Dalabon, et surtout auprès de la locutrice de la langue Kayardild où nous avons noté la présence d'un [i] caractéristique, annonçant un retour à

l'univers perdu de [i a u]. Cette réapparition fut la preuve que les espaces vocaliques de ces langues ne sont que localement défectifs en voyelles hautes, et qu'elles sont capables de loger des qualités vocaliques dans cette zone pour répondre aux contraintes de la parole.

Quelle leçon peut-on en tirer pour notre propos? Nous pouvons certainement distinguer entre deux types de perturbations adaptatives: (i) les perturbations réversibles comme celles que nous trouvons à Mayadin pour [i:] réduit ou magnifié selon les variations rythmique et intonative mais restant essentiellement [i:] et en langue Kayardild pour [e] redevenant [i] en fin de groupe intonatif, réversible parce que les qualités vocaliques regagnent leurs positions initiales, occupées dans des conditions idéales de l'émission de la parole, lorsque la perturbation disparaît; (ii) les perturbations non réversibles comme la brève [e] de Mayadin opposée à [i:], qui ne se rapproche plus jamais de cette zone de voyelles extrêmes, une réalité que nous avons démontrée par plusieurs procédés de perturbation et de mesure. Cette non réversibilité est le signe d'un changement phonétique accompli. Et c'est ainsi que la violation apparente d'un universel comme [i a u] (ou "les langues tendent toutes à réaliser des exemplaires extrêmes" selon la TDF) dans plusieurs langues d'Australie ne serait qu'un accomplissement local récent (de Sprachbund pour reprendre le terme de Troubetzkoy), qui se révélerait comme un évolutif universel encore à venir... si les conditions ethno- sociologiques prêtent localement vie à ces langues, comme à notre dialecte.

Moyennes des valeurs formantiques (Hz)-Kayardild (M)

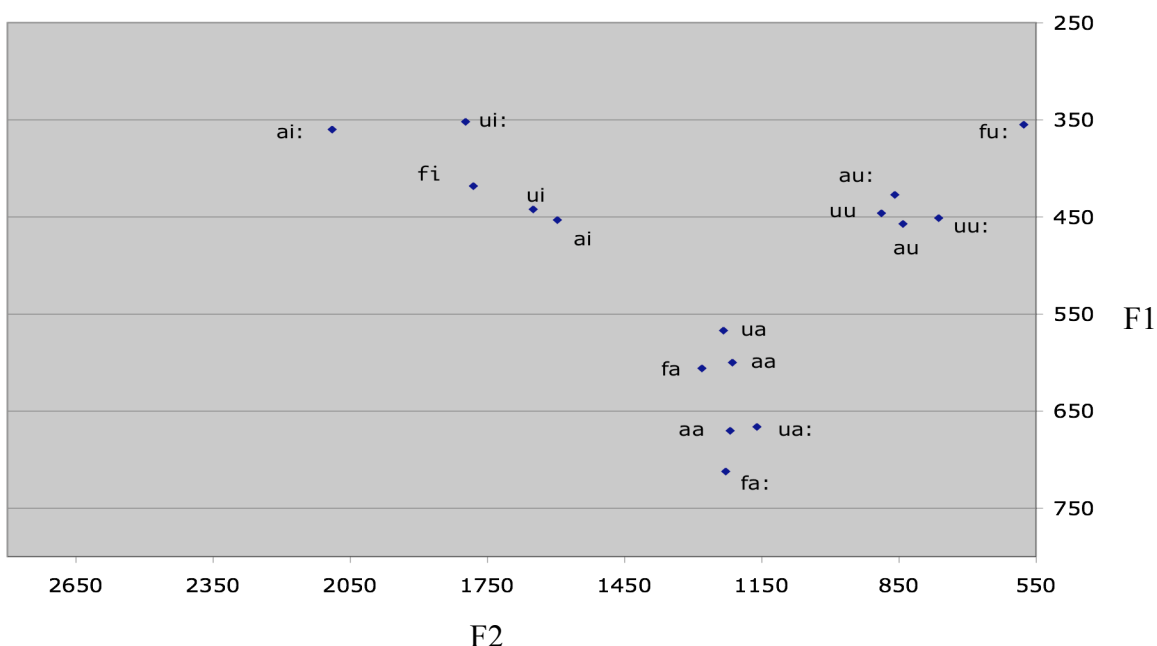


Figure 97. Espace vocalique de la langue Kayardild, langue du nord de l'Australie parlée par moins de 10 locuteurs en 2003. Moyenne d'un locuteur masculin. La 1^e lettre désigne la condition de mesure: accentuée (a), non accentuée (u) et en fin de groupe intonatif (f), et la 2^e désigne la qualité vocalique mesurée.

Moyennes des valeurs formantiques (Hz)-Kayardild (F)

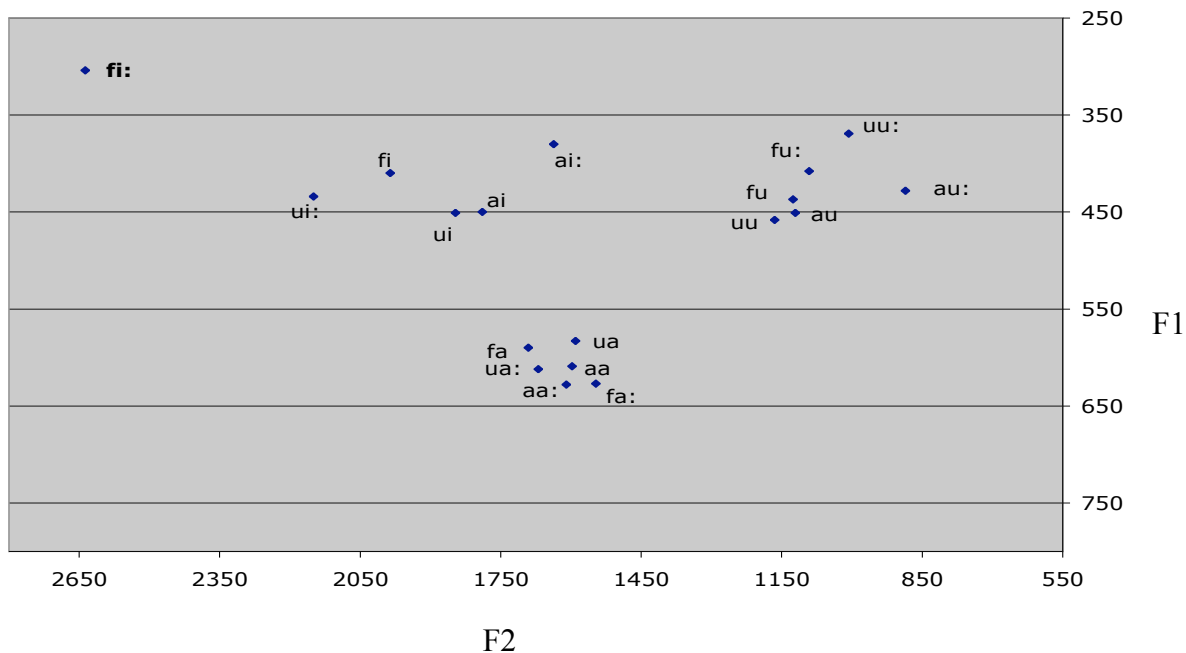


Figure 98. Espace vocalique de la langue Kayardild, langue du nord de l'Australie parlée par moins de 10 locuteurs en 2003. Moyenne d'un locuteur féminin. La 1^e lettre désigne la condition de mesure: accentuée (a), non accentuée (u) et en fin de groupe intonatif (f), et la 2^e désigne la qualité vocalique mesurée.

Bibliographie

- Abel, S. (1972a). Temporal Discrimination of Noise Bursts. *Journal of the Acoustical Society of America*, 51, pp. 519-524.
- Abel, S. (1972b). Duration Discrimination of Noise and Tone Bursts. *Journal of the Acoustical Society of America*, 51, pp. 1219-1223.
- Abou-Haidar, L. (1994). Norme linguistique et variabilité dialectale: analyse formantique du système vocalique de la langue arabe. *Revue de Phonétique Appliquée*, 110, pp. 1-15.
- Abramson, A. S. (1986). The perception of word-initial consonant length: Pattani Malay. *Journal of the International Phonetic Association*, 16, pp. 8-16.
- Abramson, A. S. (1987). Word-initial consonant length in Pattani Malay. *11th International Congress of Phonetic Sciences*, Tallinn, pp. 68-70.
- Abramson, A. S. (1999). Fundamental frequency as a cue to word-initial consonant length: Pattani Malay. *14th International Congress of Phonetic Sciences*, Berkeley, pp. 591-594.
- Abramson, A. S. (2003). Acoustic Cues to Word-Initial Stop Length in Pattani Malay. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 387-390.
- Abry, C., Boë, L. J., Beutemps, D., Laboissière, R., & Schwartz, J. L. (2000). Les sosies vocaliques, Inversion et focalisation. *23e Journées d'Étude sur la Parole*, Aussois, pp. 257-260.

- Abry, C., Boë, L. J., & Cheikh-Rouhou, M. (1995). Facteurs de variation et stabilité de la quantité vocalique en franco-provençal de Saint-Pierre-de-Soucy (Savoie): Situation typologique. In J.-F. Bonnot (Ed.), *Parole régionale: Normes, variétés linguistiques et contexte social*. Strasbourg, Presses Universitaires de Strasbourg, pp. 289-316.
- Abry, C., Orliaguet, J. P., & Sock, R. (1990). Patterns of Speech Phasing. Their robustness in the production of a timed linguistic task: Single vs. Double (abutted) consonants in French. *Current Psychology of Cognition*, 10, pp. 269-288.
- Abry, C., & Schwartz, J. L. (1987). Quelques éléments pour une théorie des objets phonétiques du langage. Autour d'une voyelle inconnue. *Bulletin du laboratoire de la communication parlée*, 1, pp. 191-210.
- Adu Manyah, K. (2003). Vowel Quantity Contrasts in Twi. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 3185-3158.
- Adu Manyah, K., & Sock, R. (2001). Durée et qualité acoustique en Twi. *Travaux de l'Institut de Phonétique de Strasbourg*, 31, pp. 1-10.
- Ahn, M. J. (2000). *Phonetic and functional bases of syllable weight for stress assignment*. Unpublished Ph. D. Dissertation, University of Illinois, Urbana-Champaign.
- Ainsworth, W. A. (1972). Duration as a cue in the recognition of synthetic vowels. *Journal of the Acoustical Society of America*, 51, pp. 648-651.
- Al-Ani, S. (1970). *Arabic Phonology: An Acoustical and Physiological Investigation*. The Hague/Paris: Mouton
- Alioua, A. (2003). Variations co-intrinsèques des durées des voyelles longues en arabe littéral moderne. *Travaux de l'Institut de Phonétique de Strasbourg*, 33, pp. 25-54.
- Allatif, O. (2003). *Caractérisation de la différencel/ressemblance des voyelles cardinales longues et brèves de l'Euphrate arabe irako-syrienne*. Unpublished Mémoire de DEA, Sciences du langage, Université Stendhal (Grenoble III), Grenoble.

-
- Allatif, O. (2005). Perturbation, adaptabilité et réversibilité dans les systèmes vocaliques. *Un siècle de Phonétique expérimentale: histoire et développement. De Théodore Rosset à John Ohala*, Grenoble, p. 28 [Absrtacts].
- Allatif, O. (2007). Le poids perceptif de la durée de la voyelle face à sa qualité acoustique dans la discrimination des voyelles longues et brèves. *7e Rencontres Jeunes Chercheurs en Parole*, Paris, pp. 8-11.
- Allatif, O., & Abry, C. (2004). Adaptabilité des paramètres temporels et spectraux dans l'opposition de quantité vocalique de l'arabe de Mayadin (Syrie). *25e Journées d'Étude sur la Parole*, Fès, pp. 13-16.
- Allatif, O., & Abry, C. (2004). Perturbation, adaptabilité et réversibilité dans les systèmes vocaliques. *Perturbations et Réajustements, Journées Fédératrices COGNIEST*, Haguenau.
- Allatif, O., & Abry, C. (2006). Reversibility in vowel systems: When prosody succeeds and when it fails. *10th Conference on Laboratory Phonology*, Paris, pp. 151-152 [Absrtacts].
- Barkat, M. (2001). Vers l'identification automatique des parlers arabes. *Langue et Linguistique*, 7, pp. 47-95.
- Barkat-Defradas, M., Al-Tamimi, J. E., & Benkirane, T. (2003). Phonetic variation in production and perception of speech: a comparative study of two Arabic dialects. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 857-860.
- Barkat-Defradas, M., Vasilescu, I., & Pellegrino, F. (2003). Stratégies perceptuelles et identification automatique des langues: application au continuum dialectal arabe. *Revue PArole*, 25, pp. 1-44.
- Barry, M. C. (2000). A phonetic and phonological investigation of English clear and dark syllabic /l/. *Les Cahiers de l'I.C.P., Bulletin de la Communication Parlée*, 5, pp. 77-87.

-
- Beaume, E. (1986). *La lecture: Préalable à sa pédagogie*. Paris: Association française pour la Lecture.
- Belkaid, Y. (1984). Les voyelles de l'arabe littéraire moderne. Analyse spectrographique. *Travaux de l'Institut de Phonétique de Strasbourg*, 16, pp. 217-240.
- Best, C. T., & Faber, A. (2000). Developmental increase in infants' discrimination of nonnative vowels that adults assimilate to a single native vowel. *International Conference on Infant Studies*, Brighton.
- Boersma, P., & Weenink, D. (2006). *Praat. Doing Phonetics by Computer. version 4.4.49*
- Bonnot, J. F. P. (1979). Etude expérimentale de certains aspects de la gémination et de l'emphase en arabe. *Travaux de l'Institut de Phonétique de Strasbourg*, 11, pp. 109-118.
- Bradlow, A. R. (1993). A cross-language comparison of vowel production and perception: Language-specific and universal aspects. *Working Papers of the Cornell Phonetics Laboratory*, 8, pp. 29-86.
- Butcher, A. (1994). On the phonetics of small vowel systems: evidence from Australian languages. *5th Australian International Conference on Speech Science and Technology*, Canberra, pp. 28-33.
- Carré, A., Ainsworth, W. A., Jospa, P., Maeda, S., & Padeloup, V. (2001). Perception of vowel-to-vowel transitions with different formant trajectories. *Phonetica*, 58, pp. 163-178.
- Cathiard, M. A. (1994). *La perception visuelle de l'anticipation des gestes vocaliques: cohérence des événements audibles et visibles dans le flux de la parole*. Unpublished Thèse de doctorat, Psychologie Cognitive, Université Pierre-Mendès-France (Grenoble II).
- Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. *Phonetica*, 22, pp. 125-159.

-
- Chistovich, L. A., & Lublinskaya, V. V. (1979). The center of gravity effect in vowel spectra and critical distance between the formants. *Hearing Research, 1*, pp. 185-195.
- Creelman, C. D. (1962). Human Discrimination of Auditory Duration. *Journal of the Acoustical Society of America, 34*, pp. 582-593.
- Cruttenden, A. (2001). *Gimson's Pronunciation of English*. London: Edward Arnold
- Cutler, A., & Otake, T. (1994). Mora or Phoneme? Further evidence for Inanguage-specific listening. *Journal of Memory and Language and Speech, 33*, pp. 824-844.
- de Jong, K. J. (1991). An articulatory study of consonant-induced vowel duration changes in English. *Phonetica, 48*, pp. 1-18.
- de Jong, K. J. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America, 97*, pp. 491-504.
- de Jong, K. J. (2004). Stress, lexical focus, and segmental focus in English: patterns of variation in vowel duration. *Journal of Phonetics, 32*, pp. 493-516.
- de Jong, K. J., & Zawaydeh, B. A. (1999). Stress, Duration, and Intonation in Arabic Word-level Prosody. *Journal of Phonetics, 27*, pp. 3-22.
- de Jong, K. J., & Zawaydeh, B. A. (2002). Comparing stress, lexical focus, and segmental focus: patterns of variation in Arabic vowel duration. *Journal of Phonetics, 30*, pp. 53-75.
- Dejardins, R. N., & Trainor, L. J. (1998). Fundamental frequency influences vowel discrimination in 6-month-old infants. *Can. Acoust. Proc., 26*, pp. 96-97.
- Denes, P. (1955). Effect of Duration on the Perception of Voicing. *Journal of the Acoustical Society of America, 27*, pp. 761-764.

-
- DiCristo, A. (1985). *De la microprosodie à l'intonosyntaxe*. Aix-en-Provence: Publications de l'Université de Provence
- Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1*, pp. 121-144.
- Dooley, G. J., & Moore, B. C. J. (1988). Duration discrimination of steady and gliding tones: A new method for estimating sensitivity to rate of change. *Journal of the Acoustical Society of America, 84*, pp. 1332-1337.
- Dorman, M. F., Marton, K., Hannley, M. T., & Lindholm, J. M. (1985). Phonetic identification by elderly normal and hearing-impaired listeners. *Journal of the Acoustical Society of America, 77*, pp. 664-670.
- Durand, M. (1946). *Voyelles longues et voyelles brèves. Essai sur la nature de la quantité vocalique*. Paris: Klincksieck
- Edwards, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America, 89*, pp. 369-382.
- Eerola, O., Laaksonen, J. P., Savela, J., & Aaltonen, O. (2003). Perception and production of the short and long Finnish [i] vowels: individuals seem to have different perceptual and articulatory templates. *15th International Congress of Phonetics Sciences, Barcelona*, pp. 989-992.
- Eilers, R. E., Bull, D. H., Oller, D. K., & Lewis, D. C. (1984). The discrimination of vowel duration by infants. *Journal of the Acoustical Society of America, 75*, pp. 1213-1218.
- Embarki, M. (2002). L'acquisition de l'allongement vocalique en arabe marocain: productions de jeunes enfants marocains en âge préscolaire. *24e Journées d'Étude sur la Parole*, pp. 173-176.
- Embarki, M. (2004). Quantité vocalique et nature du lexique en arabe marocain. *25e Journées d'Étude sur la Parole, Fès*, pp. 201-204.

-
- Embarki, M., & Guilleminot, C. (2003). The Moving boundaries of the first-acquired variety's phonological features: evidence from production/perception of Moroccan Arabic's vowels. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 639-642.
- Fennan, M. (1986). *Phonologie et Morphologie d'un Parler Arabe Marocain de Rabat*. Unpublished Thèse de doctorat, Université de Paris V.
- Fintoft, K. (1961). The duration of some Norwegian speech sounds. *Phonetica*, 7, pp. 19-39.
- Fitzgibbons, P. J., & Gordon-Salant, S. (1994). Age Effects on Measures of Auditory Duration Discrimination. *Journal of Speech and Hearing Research*, 37, pp. 662-670.
- Fitzgibbons, P. J., & Gordon-Salant, S. (1995). Age effects on duration discrimination with simple and complex stimuli. *Journal of the Acoustical Society of America*, 98, pp. 3140-3145.
- Flanagan, J. L. (1955). A difference limen for vowel formant frequency. *Journal of the Acoustical Society of America*, 30, pp. 613-617.
- Flanagan, J. L., & Saslow, M. G. (1958). Pitch Discrimination for Synthetic Vowels. *Journal of the Acoustical Society of America*, 30, pp. 435-442.
- Flege, J. E., & Port, R. (1981). Cross-Language Phonetic Interference: Arabic to English. *Language and Speech*, 24, pp. 125-146.
- Fletcher, J., & Butcher, A. (2003). Local and global influences on vowel formants in three Australian languages. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 905-908.
- Freyman, R. L., & Nelson, D. A. (1986). Frequency discrimination as a function of tonal duration and excitation-pattern slopes in normal and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 79, pp. 1034-1044.

-
- Freyman, R. L., & Nelson, D. A. (1987). Frequency Discrimination of Short- versus Long-Duration Tones by Normal and Hearing-Impaired Listeners. *Journal of Speech and Hearing Research, 30*, pp. 28-36.
- Gagne, J. P., & Zurek, P. M. (1988). Resonance-frequency discrimination. *Journal of the Acoustical Society of America, 83*, pp. 2293-2299.
- Ghazeli, S. (1979). Du statut des voyelles en arabe. *Analyse & Théorie, N° spécial: Etudes Arabes, 2/3*, pp. 199-219.
- Gordon-Salant, S., Yeni-Komshian, G. H., Fitzgibbons, P. J., & Barrett, J. (2006). Age-related differences in identification and discrimination of temporal cues in speech segments. *Journal of the Acoustical Society of America, 119*, pp. 2455-2466.
- Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics, 12*, pp. 91-114.
- Gottfried, T. L., & Beddor, P. S. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech, 31*, pp. 57-75.
- Hall, J. W., & Wood, E. J. (1984). Stimulus Duration and Frequency Discrimination for Normal-Hearing and Hearing-Impaired Subjects. *Journal of Speech and Hearing Research, 27*, pp. 252-256.
- Harrington, J., Fletcher, J., & Beckman, M. E. (2000). Manner and place conflicts in the articulation of accent in Australian English. *5th Laboratory Phonology, Cambridge*, pp. 40-51.
- Harrington, J., Fletcher, J., & Roberts, C. (1995). Coarticulation and the accented/unaccented distinction. *Journal of Phonetics, 23*, pp. 305-322.
- Harris, J. D. (1952). Pitch Discrimination. *Journal of the Acoustical Society of America, 24*, pp. 750-755.

-
- Hawks, J. W. (1994). Difference limens for formant patterns of vowel sounds. *Journal of the Acoustical Society of America*, 95, pp. 1074-1084.
- Henry, F. M. (1948). Discrimination of the duration of a sound. *Journal of Experimental Psychology: Human Perception and Performance*, 38, pp. 734-742.
- Hillenbrand, J., Laura, A., Getty, M., Clark, J., & Wheeler, K. (1995). Acoustic Characteristics of American English Vowels. *Journal of the Acoustical Society of America*, 97, pp. 3099-3111.
- Jacewicz, E. (2005). Listener sensitivity to variations in the relative amplitude of vowel formants. *Journal of the Acoustical Society of America*, 6, pp. 118-124.
- Jacobson, H. (1951). Information and human ear. *Journal of the Acoustical Society of America*, 23, p. 464.
- Jomaa, M. (1991). *Organisation temporelle, acoustique et articulatoire de la quantité vocalique en arabe tunisien: Comparaison avec l'arabe standard et l'arabe koweïtien*. Unpublished Thèse de doctorat, Sciences du Langage, Université Stendhal (Grenoble III).
- Kewley-Port, D. (2001). Vowel formant discrimination II: Effects of stimulus uncertainty, consonantal context, and training. *Journal of the Acoustical Society of America*, 110, pp. 2141-2155.
- Kewley-Port, D., & Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America*, 95, pp. 485-496.
- Kewley-Port, D., & Zheng, Y. (1996). Vowel formant discrimination in ordinary listening conditions I. *Journal of the Acoustical Society of America*, 100, p. 2689.
- Kewley-Port, D., & Zheng, Y. (1999). Vowel formant discrimination: Towards more ordinary listening conditions. *Journal of the Acoustical Society of America*, 106, pp. 2945-2958.

-
- Klatt, D. H. (1982c). Prediction of Perceived Phonetic Distance from Critical-Band Spectra: A First Step. *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP*, pp. 1278-1281.
- Kondo, M., & Shinohara, S. (2003). Durational evidence of the psychological reality of the mora in Japanese speakers' French. *15th International Congress of Phonetic Sciences, Barcelona*, pp. 463-466.
- Kuhl, P. K. (1991). Human adults and human infants exhibit a prototype effect for phoneme categories: Monkeys do not. *Perception & Psychophysics*, 50, pp. 93-107.
- Laabi, T. (1975). *Phonétique et Phonologie du Parler Arabe de Fès*. Unpublished Thèse de doctorat, Université de Toulouse II.
- Labov, W. (1972a). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press
- Ladefoged, P., & Maddieson, I. (1996). *Sounds of the World's Languages*. Oxford: Blackwell
- Laks, B. (1977). Contribution empirique à l'analyse socio-différentielle de la chute de /r/ dans les groupes consonantiques finals. *Langue Française*, 34, pp. 109-125.
- Lehiste, I. (1970). *Suprasegmentals*. Cambridge: M.I.T. Press
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, pp. 467-477.
- Liberman, A., & Mattingly, I. (1985). The motor theory of speech perception revised. *Cognition*, 21, pp. 1-36.
- Liberman, A. M., Delattre, P. C., Gerstman, L. C., & Cooper, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 52, pp. 127-137.

-
- Lindblom, B. E. F. (1987). Adaptative variability and absolute constancy in speech signals: Two themes in the quest for phonetic invariance. *11th International Congress of Phonetic Sciences*, Tallin, pp. 2-20.
- Lindblom, B. E. F. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, 35, pp. 1773-1781.
- Lindblom, B. E. F. (1967). On the role of formant transitions in vowel recognition. *Journal of the Acoustical Society of America*, 42, pp. 830-842.
- Maddieson, I. (1986). The Size and Structure of Phonological Inventories: Analysis of UPSID. In J. J. Ohala & J. J. Jaeger (Eds.), *Experimental phonology*. New York: Academic Press, pp. 105-123.
- Maddieson, I. (2006). Variation in co-variation: the search for explanatory principles. *10th Conference on Laboratory Phonology*, Paris, pp. 21-22 [Abstracts].
- Malmberg, B. (1971). *La phonétique*. Paris: Presses Universitaires de France
- Mantakas, M., Schwartz, J. L., & Escudier, P. (1986). Modèle de prédiction du deuxième formant effectif F²- application à l'étude de la labialité des voyelles avant du français. *15e Journées d'Etudes sur la Parole*, Aix-en-Provence, pp. 157-161.
- Martin, P. (1998). Les voyelles d'aperture moyenne en français du Québec. *Cahiers de l'ILSL*, 11, pp. 215-242.
- Martin, P. (2001). Les voyelles nasales en français du Québec. *La Linguistique*, 37, pp. 49-70.
- Martinet, A. (1956). *La description phonologique avec application au parler franco-provençal d'Hauteville (Savoie)*. Genève: Droz, Paris: Minard
- Medin, D. L., & Barsalou, L. W. (1987). Categorization processes and categorical perception. In S. Harnad (Ed.), *Categorical Perception*. Cambridge: Cambridge University Press, pp. 455-490.

-
- Ménard, L., Schwartz, J. L., Boë, L. J., Kandel, S., & Vallée, N. (2002). Auditory normalization of French vowels synthesized by an articulatory model simulating growth from birth to adulthood. *Journal of the Acoustical Society of America*, *111*, pp. 1892-1905.
- Mermelstein, P. (1978). Difference limens for formant frequencies of steady-state and constant bound vowels. *Journal of the Acoustical Society of America*, *54*, pp. 610-619.
- Messaoud-Galusi, S., & Sperenger-Charolles, L. (2003). The effect of age and reading level on the perception of temporal speech information. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 2011-2014.
- Meunier, C., Frenck-Mestre, C., Lelekov-Boissard, T., & Besnerais, M. L. (2003). Production and perception of vowels: does a density of the system play a role? *15th International Congress of Phonetic Sciences*, Barcelona, pp. 723-726.
- Miller, J. L., & Grosjean, F. (1997). Dialect effects in vowel perception: The role of temporal information in French. *Language and Speech*, *40*, pp. 277-288.
- Minagawa, Y., Kagomiya, T., & Maekawa, K. (2003). Durational variations of Japanese long/short vowels in different speaking rates: analysis of a spontaneous speech corpus. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 2127-2130.
- Mitleb, F. (1984). Vowel length contrast in Arabic and English: A spectrographic test. *Journal of Phonetics*, *12*, pp. 229-236.
- Mitleb, F. M. (1984a). Voicing effect on vowel duration is not an absolute universal. *Journal of Phonetics*, *12*, pp. 23-27.
- Mochizuki-Sudo, M., & Kiritani, S. (1991). Production and perception of stress-related durational patterns in Japanese learners of English. *Journal of Phonetics*, *19*, pp. 231-248.
- Moore, B. C. J. (1971). *Temporal parameters in pitch perception*. Unpublished Ph. D. thesis, University of Cambridge.

-
- Moore, B. C. J. (1973). Frequency difference limens for short-duration tones. *Journal of the Acoustical Society of America*, 54, pp. 610-619.
- Mora, J. C., & Mott, B. (2003, 3-9 August). Length as an Acoustic Cue in the Perception of Consonant Syllabicity. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 411-414.
- Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic: Acoustic measurements and accentedness ratings. *Language and Speech*, 36, pp. 39-66.
- Munson, W. A. (1947). The growth of auditory sensation. *Journal of the Acoustical Society of America*, 19, pp. 584 - 591.
- Nearey, T. M. (1989). Static, dynamic and relational factors in vowel perception. *Journal of the Acoustical Society of America*, 85, pp. 2088-2113.
- Nelson, D. A., Stanton, M. E., & Freyman, R. L. (1983). A general equation describing frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, 73, pp. 2117-2123.
- Nenonen, S., Shestakova, A., Alku, P., & Huotilainen, M. (2003). Perception of vowel length in native speakers and second-language users of a quantity language. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 531-534.
- Newman, D., & Verhoeven, J. (2002). Frequency analysis of Arabic vowels in connected speech. *Antwerp Papers in Linguistics*, 100, pp. 77-86.
- Nishinuma, Y. (1977). *Contribution à l'étude prosodique du japonais*. Unpublished Thèse de doctorat, Université de Provence, Aix-en-Provence.
- Nord, L., & Sventelius, E. (1979). *Analysis and prediction of difference limen data for formant frequencies* (Quarterly Progress Status Report No. 3-4). Stockholm, Sweden: Speech Transmission Laboratory, Royal Institute of Technology.

-
- Norlin, K. (1987). A phonetic study of emphasis and vowels in Egyptian Arabic. *Working Papers of the Lund University Department of Linguistics*, 30, pp. 1-119.
- Öhman, S. E. G. (1967). Numerical Model of Coarticulation. *Journal of the Acoustical Society of America*, 41, pp. 310-320.
- Otake, T., Hatano, G., Cultler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32, pp. 258-278.
- Peterson, G. E., & Lehiste, I. (1960). Duration of Syllable nuclei in English. *Journal of the Acoustical Society of America*, 32, pp. 693-703.
- Plomp, R. (1964). Rate of Decay of Auditory Sensation. *Journal of the Acoustical Society of America*, 36, pp. 277-282.
- Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, 100, pp. 577-592.
- Polka, L., & Bohn, O. S. (2003). Asymmetries in vowel perception. *Speech Communication*, 41, pp. 221-231.
- Polka, L., & Werker, J. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, 20, pp. 421-435.
- Port, R. F., Al-Ani, S., & Maeda, S. (1980). Temporal compensation and universal phonetics. *Phonetica*, 37, pp. 235-252.
- Rebollo, L., Barbosa, F., Costa, R., Lannes, R., Nicolau, J., Rosa, C., & Resende, F. (2003). Length and identification of Brazilian Portuguese segments in continuous speech signal. *15th International Congress of Phonetic Sciences, Barcelona*, pp. 3073-3076.

- Repp, B. H., Healy, A. F., & Crowder, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 5, pp. 129-145.
- Rhardisse, N. (1989). *La résistivité de la quantité vocalique et consonantique de l'arabe marocain aux variations de la vitesse de l'élocution*. Unpublished Thèse de doctorat, Sciences du langage, Université Stendhal (Grenoble III).
- Rossi, M. (1972). Le seuil différentiel de durée. In *Mélange à la mémoire de P. Delattre*. La Haye: Mouton, pp. 435-449.
- Rossi, M. (1981). *L'intonation: de l'acoustique à la sémantique*. Paris: Klincksieck
- Rostolland, D., Parant, A., Takahashi, H., & Pandales, E. (1985). Durée vocalique intrinsèque et co-intrinsèque en français: contraintes physiologiques et variations temporelles dans les syllabes CVC. *14e Journées d'Étude sur la Parole*, Paris E.N.S.T, pp. 179-182.
- Ruhm, H. B., Mencke, E. O., Milbrun, R., Cooper, W. A., & Rose, D. (1966). Differential sensitivity to duration of acoustic signals. *Journal of Speech and Hearing Research*, 9, pp. 371-384.
- Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997a). Major trends in vowel system inventories. *Journal of Phonetics*, 25, pp. 233-253.
- Schwartz, J. L., Boë, L. J., Vallée, N., & Abry, C. (1997b). The Dispersion - Focalisation Theory of vowel systems. *Journal of Phonetics*, 25, pp. 255-286.
- Schwartz, J. L., Abry, C., Boë, L. J., & Cathiard, M. A. (2002). Phonology in a Theory of Perception-for-Action-Control. In J. Durand & B. Laks (Eds.), *Phonetics, Phonology, and Cognition*. Oxford: Oxford University Press, pp. 254-280.
- Simpson, A. P. (1998). Accounting for the phonetics of German r without processes. *ZAS Working Papers in Linguistics*, 11, pp. 91-104.

-
- Simpson, A. P. (2002). Gender-specific articulatory-acoustic relations in vowel sequences. *Journal of Phonetics*, 30, pp. 417-435.
- Simpson, A. P., & Ericsson, C. (2003). Sex-specific durational differences in English and Swedish. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 1113-1116.
- Sinnott, J. M., & Adams, F. S. (1987b). Differences in human and monkey sensitivity to acoustic cues underlying voicing contrasts. *Journal of the Acoustical Society of America*, 82, pp. 1539-1547.
- Sinnott, J. M., Beecher, M. D., Moody, D. B., & Stebbins, W. C. (1976). Speech sound discrimination by monkeys and humans. *Journal of the Acoustical Society of America*, 60, pp. 687-695.
- Sinnott, J. M., & Kreiter, N. A. (1991). Differential sensitivity to vowel continua in Old World monkeys (*Macaca*) and humans. *Journal of the Acoustical Society of America*, 89, pp. 2421-2429.
- Sinnott, J. M., Owren, M. J., & Petersen, M. R. (1987a). Auditory duration discrimination in Old World monkeys (*Macaca*, *Cercopithecus*) and humans. *Journal of the Acoustical Society of America*, 82, pp. 465-470.
- Small, A. M., & Campbell, R. A. (1962). Temporal differential sensitivity for auditory stimuli. *American Journal of Psychology*, 75, pp. 401-410.
- Sock, R. (1983). *L'organisation temporelle de l'opposition de quantité vocalique en Wolof de Gambie, sa résistivité aux conditions de durée segmentale et suprasegmentale*. Unpublished Thèse de doctorat, Sciences du Langage, Université Stendhal (Grenoble III).
- Stevens, K. N., & House, A. S. (1963). Perturbation of vowel articulation by consonantal context: An acoustical study. *Journal of speech, language, and hearing research*, 6, pp. 111-128.

-
- Strange, W. (1989). Evolving theories of vowel perception. *Journal of the Acoustical Society of America*, 85, pp. 2081-2087.
- Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America*, 82, pp. 847-863.
- Swoboda, P. J., Kass, J., Morse, P. A., & Leavitt, L. A. (1978.). Memory factors in infant vowel discrimination of normal and at-risk infants. *Child Development*, 49, pp. 332-339.
- Tajima, K., Kato, H., Rothwell, A., & G.Munhall, K. (2003). Perception of phonemic length contrasts in Japanese by native and non-native listeners. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 1585-1588.
- Tajima, K., Zawaydeh, B. A., & Kitahara, M. (1999). A comparative study of speech rhythm in Arabic, English and Japanese. *14th international congress of phonetic sciences*, San Francisco, pp. 285-288.
- Tallal, P., & Piercy, M. (1975). Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia*, 13, pp. 69-74.
- Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, 88, pp. 97-100.
- Tubach, J. P. (1989). *La parole et son traitement automatique*. Paris: Masson
- Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch & B. Lloyd (Eds.), *Cognition and Categorization*. Hillsdale, N.J./New York: Lawrence Erlbaum Associates, pp. 81-99.
- Tyler, R. S., Summerfield, Q., Wood, E. J., & Fernandes, M. A. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 72, pp. 740-752.

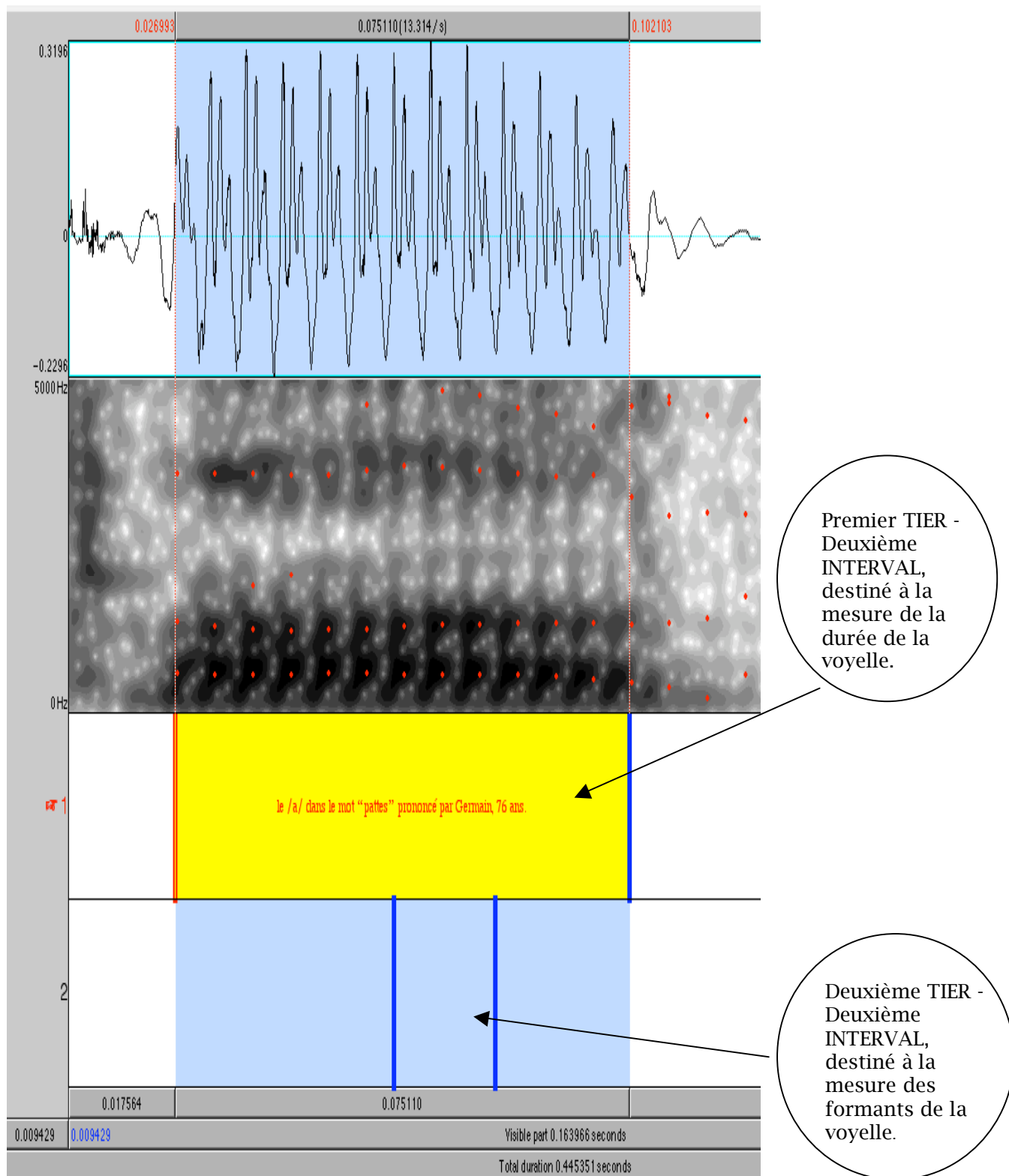
-
- Uguzzoni, A., Azzaro, G., & Schmid, S. (2003). Short vs. long and/or abruptly vs. smoothly cut vowels. New The Journal of the Acoustic Society of America perspectives on a debated question. *15th International Congress of Phonetic Sciences*, Barcelona, pp. 2717-2720.
- Uguzzoni, A., & Busà, M. G. (1995). Acoustic correlates of vowel quantity contrasts in an Italian dialect. *13th International Congress of Phonetic Sciences*, Stockholm, pp. 390-393.
- Vallée, N. (1994). *Systèmes vocaliques: de la typologie aux prédictions. Modèles et traitements en communication parlée, langue naturelle et géolinguistique*. Unpublished Thèse de Doctorat, Sciences du Langage, Université Stendhal (Grenoble III).
- Vallée, N., & Kandel, S. (2003). Can we recover vowel gestures from speech sounds? An experimental study based on an original psychophysical paradigm. *15th international congress of phonetic sciences*, Barcelona, pp. 817-820.
- Van-Heuven, V. J. (1994). What is the smallest prosodic domain? *Phonological Structure and Phonetic Form: 3rd Laboratory Phonology*, Cambridge, pp. 76-98.
- Van-Summuers, W., & Leek, M. R. (1991). Use of spectral and temporal cues in vowel identification by normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 89, p. 1959.
- Walter, H. (1982). *Enquête phonologique et variétés régionales du français*. Paris: PUF
- Watson, C. S., & Foyle, D. C. (1985). Central factors in the discrimination and identification of complex sounds. *Journal of the Acoustical Society of America*, 78, pp. 375-380.
- Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, 10, pp. 193-216.
- Werker, J. F., Gilbert, J. H. V., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 52, pp. 349-355.

- Wier, C. C., Jesteadt, W., & Green, D. M. (1977). Frequency discrimination as a function of frequency and sensation level. *Journal of the Acoustical Society of America*, 61, pp. 178-184.
- Wieringen, A. V., & Pols, L. C. W. (1994). Frequency and duration discrimination of short first-formant speechlike transitions. *Journal of the Acoustical Society of America*, 95, pp. 502-511.
- Yamasaki, H. (1996). *Unité minimale de perception de la parole chez le nouveau-né: Capacité de discrimination moraique*. Unpublished Mémoire de DEA, Phonétique, ILPGA Sorbonne Nouvelle (Paris 3).
- Zawaydeh, B. A. (1999). *The phonetics and phonology of gutturals in Arabic*. Unpublished Ph. D. dissertation, Indiana University.

Annexes

Annexe I: exemple de segmentation des signaux.

Dans cette annexe, nous montrons comment nous avons procédé à la segmentation des signaux (du corpus de l'arabe et du français de la Savoie) afin d'obtenir la durée de la voyelle et les valeurs de ses formants.



Annexe II: Algorithme du calcul du deuxième formant effectif avec le logiciel Matlab.

Nous devons disposer des valeurs des 4 premiers formants pour pouvoir déterminer le F prime 2. Nous aurons une matrice de n lignes (n étant le nombre de voyelles pour lesquelles nous voulons effectuer ce calcul) de p colonnes (p est toujours égal 4, puisque le calcul est basé sur les 4 premiers formants), le calcul est basé sur la distance de 3,5 Bark de Chistovich, L. A., & Lublinskaya, V. V. (1979) à propos du centre de gravité de la voyelle et la distance critique entre les formants.

```

for i = 1: size(Z,1)
    if Z(i,3) - Z(i,2) > 3.5
        c = 0;
    else
        if Z(i,3) - Z(i,2) > 2.5
            c = 3.5 -(Z(i,3) - Z(i,2));
        else
            if Z(i,4) - Z(i,2) > 3.5
                c = 1;
            else
                if Z(i,4) - Z(i,3) > Z(i,3) - Z(i,2)
                    c = 1;
                else
                    c = 2;
                end
            end
        end
    end
end
end
end
end
end
if c <= 1
    c2 = 1;
    c3 = 0.5*c;
    c4 = 0;
else

```

```
c2 = 0;  
c3 = 1;  
c4 = 0.5;  
end  
z2P =(c2*Z(i,2) + c3*Z(i,3) + c4*Z(i,4))/(c2 + c3 + c4);
```

Annexe III: déplacements des voyelles dans l'espace vocalique en fonction du pourcentage affecté à leurs formants.

Dans cette annexe, nous affectons des pourcentages aux formants de nos voyelles (de l'arabe de Mayadin) pour observer leurs déplacements dans l'espace vocalique, le but étant de savoir si à un pourcentage égal qui affecte les 4 premiers formants des trois voyelles cardinales [i, a et u] les déplacements dans l'espace vocalique sont égaux ou non.

Les valeurs, en Hz et en Bark, des 4 premiers formants des trois voyelles cardinales [i, a, u] sont disposées dans les tableaux ci-dessous.

[i]

Hz				Bark				
F1	F2	F3	F4	F1	F2	F3	F4	F'2
277	2208	3079	3560	2.899	13.559	15.816	16.813	16.1488 [i] valeurs de référence
291	1987	3387	3560	3.038	12.854	16.471	16.813	12.8541 F1*5% F2F3*10% F4*100%.
291	1987	3387	3916	3.038	12.854	16.471	17.471	12.8541 F1*5% F2F3F4*10%.
305	1987	3387	3916	3.175	12.854	16.471	17.471	12.8541 F1F2F3F4*10%.
332	1766	3695	4272	3.436	12.074	17.070	18.072	12.0744 F1F2F3F4*20%.
360	1545	4003	4628	3.702	11.204	17.623	18.627	11.2037 F1F2F3F4*30%.

[a]

Hz				Bark				
F1	F2	F3	F4	F1	F2	F3	F4	F'2
800	1228	2500	3560	7.249	9.751	14.397	16.813	9.7508 [a] valeurs de référence
760	1351	2750	3560	6.973	10.347	15.044	16.813	10.3473 F1*5% F2F3*10% F4
760	1351	2750	3916	6.973	10.347	15.044	17.471	10.3473 F1*5% F2F3F4*10%
720	1351	2750	3916	6.689	10.347	15.044	17.471	10.3473 F1F2F3F4*10%
640	1474	3000	4272	6.093	10.901	15.639	18.072	10.9012 F1F2F3F4*20%
560	1596	3250	4628	5.460	11.414	16.187	18.627	11.4137 F1F2F3F4*30%

[u]									
Hz				Bark					
F1	F2	F3	F4	F1	F2	F3	F4	F'2	
277	553	2420	3560	2.899	5.403	14.177	16.813	5.4028	[u] valeurs de référence
291	608	2662	3560	3.038	5.845	14.823	16.813	5.8446	F1*5% F2F3*10% F4
291	608	2662	3916	3.038	5.845	14.823	17.471	5.8446	F1*5% F2F3F4*10%
305	608	2662	3916	3.175	5.845	14.823	17.471	5.8446	F1F2F3F4*10%
332	664	2904	4272	3.436	6.276	15.416	18.072	6.2757	F1F2F3F4*20%
360	719	3146	4628	3.702	6.681	15.964	18.627	6.6814	F1F2F3F4*30%
388	774	3388	4984	3.9633	7.0704	16.4727	19.1407	7.0704	F1F2F3F4*40%

Mesure de la distance euclidienne pondérée selon la proposition de Schwartz *et al.* (1997b): $dViVj = [(F1i-F1j)^2 + 0,3^2 (F'2i- F'2j)^2]^{1/2}$

Toute distance en Bark: pour le [i] et le [a] la hausse de 10% de F4 n'a pas été apportée.

[i]

$$dii F1*5\% F2F3*10\% F4 \text{ fixe} = [((3.0376-2.8995)^2)+0.3^2*(12.8541-16.1488)^2]^{1/2} = 0.9980$$

$$dii F1*5\% F2F3F4*10\% = [((3.0376-2.8995)^2)+0.3^2*(12.8541-16.1488)^2]^{1/2} = 0.9980$$

$$dii F1F2F3F4*10\% = [((3.1747-2.8995)^2)+0.3^2*(12.8541-16.1488)^2]^{1/2} = 1.0260$$

$$dii F1F2F3F4*20\% = [((3.4358-2.8995)^2)+0.3^2*(12.0744-16.1488)^2]^{1/2} = 1.3348$$

[a]

$$daa F1*5\% F2F3*10\% F4 \text{ fixe} = [((6.9730-7.2487)^2)+0.3^2*(10.3473-9.7508)^2]^{1/2} = 0.3288$$

$$daa F1*5\% F2F3F4*10\% = [((6.9730-7.2487)^2)+0.3^2*(10.3473-9.7508)^2]^{1/2} = 0.3288$$

$$daa F1F2F3F4*10\% = [((6.6890-7.2487)^2)+0.3^2*(10.3473-9.7508)^2]^{1/2} = 0.5880$$

$$daa F1F2F3F4*20\% = [((6.0930-7.2487)^2)+0.3^2*(10.9012-9.7508)^2]^{1/2} = 1.2059$$

[u]

$$\text{duu F1*5\% F2F3*10\% F4 fixe} = [((3.0376-2.8995)^2)+0.3^2*(5.8446-5.4028)^2]^{(1/2)} = 0.1914$$

$$\text{duu F1*5\% F2F3F4*10\%} = [((3.0376-2.8995)^2)+0.3^2*(5.8446-5.4028)^2]^{(1/2)} = 0.1914$$

$$\text{duu F1F2F3F4*10\%} = [((3.1747-2.8995)^2)+0.3^2*(5.8446-5.4028)^2]^{(1/2)} = 0.3055$$

$$\text{duu F1F2F3F4*20\%} = [((3.4358-2.8995)^2)+0.3^2*(6.2757-5.4028)^2]^{(1/2)} = 0.5968$$

$$\text{duu F1F2F3F4*30\%} = [((3.7020-2.8995)^2)+0.3^2*(6.6814-5.4028)^2]^{(1/2)} = 0.8895$$

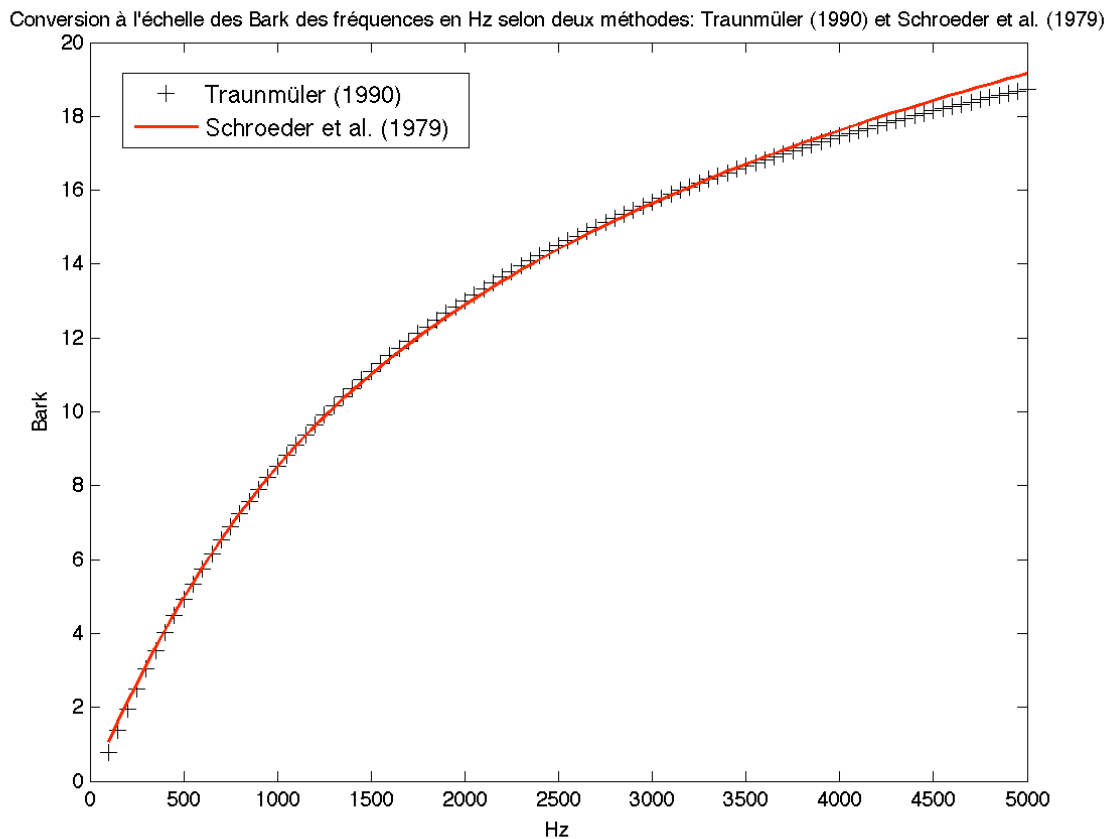
$$\text{duu F1F2F3F4*40\%} = [((3.9633-2.8995)^2)+0.3^2*(7.0704-5.4028)^2]^{(1/2)} = 1.1756$$

Annexe IV: bande critique.

Le champ auditif humain [20-20 kHz] est divisé traditionnellement en 24 bandes critiques dont la largeur de chacune est (approximativement) 1 Bark. Ci-dessous la table des largeurs des bandes critiques selon deux unités de mesures, le Hz et le Bark.

Tonie z (Bark)	f (Hz)	Δf (Hz)
0	0	0
1	100	100
2	200	100
3	300	100
4	400	100
5	510	110
6	630	120
7	770	140
8	920	150
9	1080	160
10	1270	190
11	1480	210
12	1720	240
13	2000	280
14	2320	320
15	2700	380
16	3150	450
17	3700	550
18	4400	700
19	5300	900
20	6400	1100
21	7700	1300
22	9500	1800
23	12000	2500
24	15500	3500

Annexe V: comparaison de deux méthodes de conversion des Hz vers l'échelle des Bark.



Nous comparons dans cette figure deux méthodes de conversion des Hz vers l'unité de mesure le Bark. Schwartz *et al.* (1997b) utilisent la formule proposée par Schroeder *et al.* (1979): $1\text{Bark} = 7 * \text{asinh}(f \text{ en Hz}/650)$; alors que Kewley-Port & Zheng (1999) utilisent la formule proposée par Traunmüller (1990): $1\text{Bark} = ((26,81 * f \text{ (en Hz)})/(1960 + f \text{ (en Hz)})) - 0,53$.

Ci-dessous nous appliquons les deux formules de conversion de nos données sur l'arabe. Le tableau du haut est obtenu selon la formule de Schroeder *et al.* (1979), et celui du bas par la formule de Traunmüller (1990).

Voyelle	F1		F2		F3	
	Hz	Bark	Hz	Bark	Hz	Bark
AND	548	5,33	1295	10,14	2567	14,67
AAND	636	6,04	1196	9,63	2486	14,46
IND	431	4,30	1549	11,30	2554	14,64
IIND	320	3,23	2030	13,11	2642	14,86
UND	436	4,35	1040	8,76	2509	14,52
UUND	360	3,63	921	8,04	2691	14,98
ANI	593	5,70	1313	10,23	2543	14,61
AANI	671	6,31	1245	9,88	2606	14,77
INI	430	4,29	1592	11,49	2502	14,50
IINI	328	3,31	2055	13,19	2619	14,80
UNI	438	4,37	1066	8,91	2484	14,46
UUNI	367	3,70	856	7,62	2485	14,46
ARD	565	5,47	1257	9,95	2469	14,42
AARD	626	5,96	1237	9,84	2467	14,41
IRD	422	4,22	1576	11,42	2560	14,65
IIRD	335	3,38	2001	13,01	2633	14,84
URD	430	4,29	1026	8,68	2521	14,55
UURD	356	3,59	999	8,52	2545	14,62
ARI	575	5,55	1356	10,43	2520	14,55
AARI	657	6,20	1330	10,31	2456	14,38
IRI	453	4,50	1603	11,53	2504	14,51
IIRI	354	3,57	2016	13,06	2588	14,73
URI	465	4,61	1136	9,31	2477	14,44
UURI	392	3,94	903	7,93	2566	14,67

La conversion de Hs en Bark était effectuée comme suivant: $F \text{ en Bark} = ((26,81 * F \text{ en Hz}) / (1960 + F \text{ en Hz})) - 0,53$.

Voyelle	F1		F2		F3	
	Hz	Bark	Hz	Bark	Hz	Bark
AND	548	5,36	1295	10,08	2567	14,58
AAND	636	6,06	1196	9,59	2486	14,36
IND	431	4,36	1549	11,22	2554	14,54
IIND	320	3,32	2030	13,00	2642	14,77
UND	436	4,40	1040	8,74	2509	14,42
UUND	360	3,70	921	8,03	2691	14,90
ANI	593	5,73	1313	10,17	2543	14,51
AANI	671	6,33	1245	9,84	2606	14,68
INI	430	4,35	1592	11,40	2502	14,40
IINI	328	3,40	2055	13,08	2619	14,71
UNI	438	4,42	1066	8,89	2484	14,35
UUNI	367	3,77	856	7,62	2485	14,36
ARD	565	5,50	1257	9,90	2469	14,31
AARD	626	5,99	1237	9,80	2467	14,31
IRD	422	4,27	1576	11,33	2560	14,56
IIRD	335	3,46	2001	12,90	2633	14,75
URD	430	4,35	1026	8,66	2521	14,45
UURD	356	3,66	999	8,51	2545	14,52
ARI	575	5,58	1356	10,37	2520	14,45
AARI	657	6,22	1330	10,25	2456	14,28
IRI	453	4,55	1603	11,44	2504	14,41
IIRI	354	3,65	2016	12,95	2588	14,63
URI	465	4,66	1136	9,27	2477	14,33
UURI	392	4,00	903	7,92	2566	14,57

La conversion de Hs en Bark était effectuée comme suivant: $F \text{ en Bark} = 7 \text{ASINH}(F \text{ en Hz}/650)$.

Afin d'avoir le seuil en Bark, Kewley-Port & Zheng (1999) procèdent comme suit: après avoir obtenu le seuil en Hz, colonne (2), pour les formants en colonne (1), ils additionnent (1) et (2), ce qui donnent la colonne (3). Ils convertissent alors les colonnes (1) et (3) en Bark ce qui donnent les colonnes (4) et (5) respectivement, ils soustraient la colonne (4) à la colonne (5), ce qui donne la colonne (6) qui est donc le seuil en Bark.

(1)	(2)	(3)	(4)	(5)	(6)
formant Hz	seuil Hz	sum Hz	formant Bark	sum Bark	seuil Bark
450	12,6	462,6	4,476	4,589	0,113
600	17,5	617,5	5,754	5,893	0,139
700	12,4	712,4	6,525	6,617	0,092
1000	22,0	1022,0	8,527	8,658	0,131
1400	27,3	1427,3	10,641	10,767	0,126
1950	23,4	1973,4	12,841	12,921	0,080
2200	36,7	2236,7	13,648	13,759	0,110
2300	28,5	2328,5	13,945	14,027	0,082
					0,109
formant Hz	seuil Hz	sum Hz	formant Bark	sum Bark	seuil Bark
450	32,3	482,3	4,476	4,764	0,288
600	27,5	627,5	5,754	5,972	0,218
700	27,9	727,9	6,525	6,730	0,205
1000	54,8	1054,8	8,527	8,850	0,323
1400	83,4	1483,4	10,641	11,020	0,379
1950	80,0	2030,0	12,841	13,110	0,269
2200	38,9	2238,9	13,648	13,765	0,117
2300	46,5	2346,5	13,945	14,078	0,133
					0,242

Exemples du calcul du seuil en Bark. Ci-dessus 2 exemples d'application de la méthode de détermination du seuil en Bark à partir des valeurs obtenues en Hz. Le tableau du haut est un exemple pour des formants isolés, le tableau du bas est un exemple pour des formants insérés dans des voyelles.

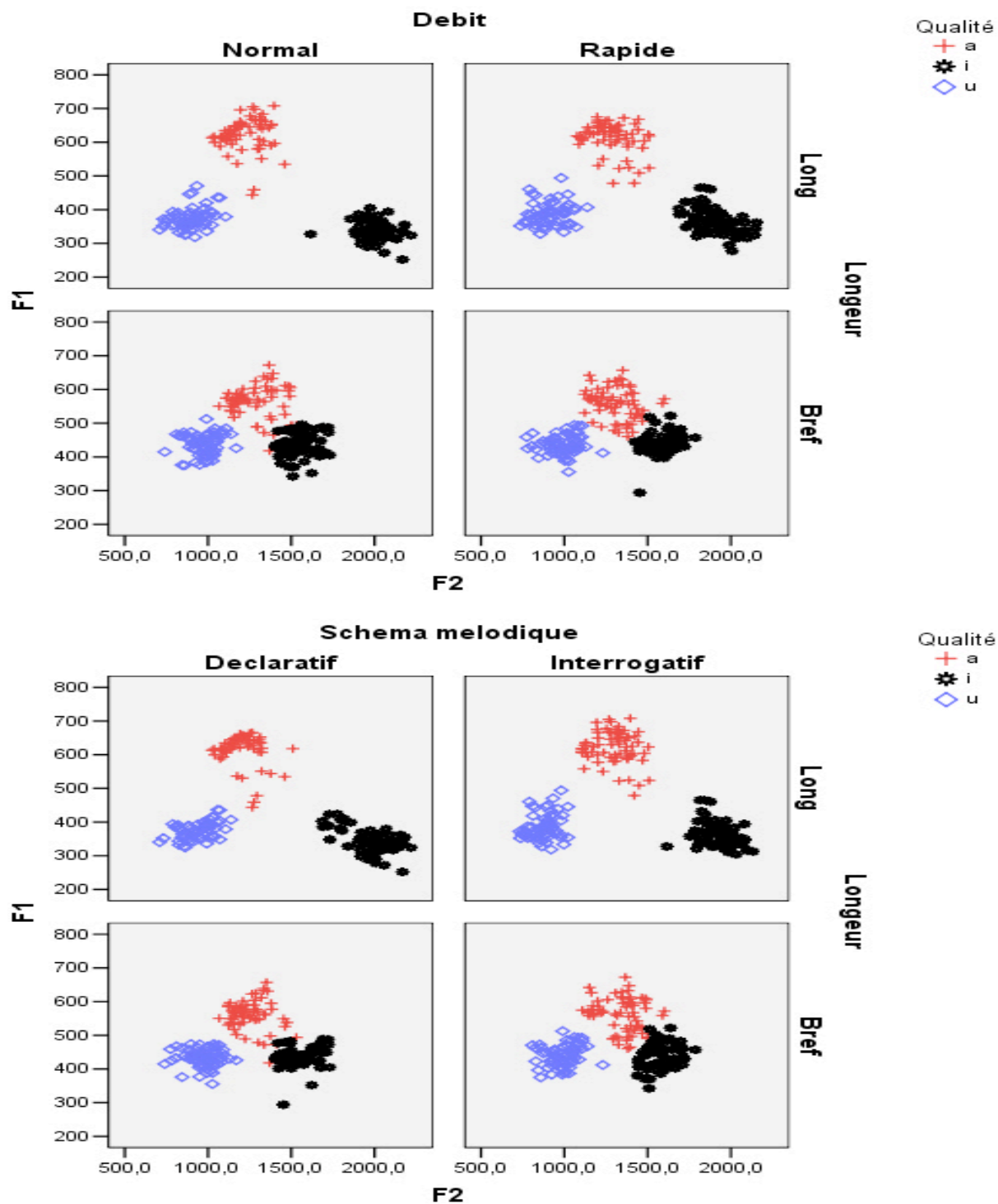
Annexe VI: résumé des résultats des études qui ont porté sur le seuil différentiel de la perception de la durée.

Nous exposons dans cette annexe les résultats des études qui ont porté sur le seuil différentiel de la perception de la durée vocalique, consonantique, de bruit et de silence que nous avons examinées pour établir nos résultats par rapport à ce seuil et qui sont disposés en chapitre I de ce manuscrit.

	Étude	Résultat ($\Delta T/T$)	Nature stimulus
1	Rossi (1972)	(de 60 à 140 ms) constant= 30 ms.	segment vocalique
		(de 13 à 290) $\Delta T/T = 22,5\%$.	segment vocalique
2	Abel (1972a)	(fonction) = 20 %	Silence.
3	Tyler <i>et al.</i> (1982)	(19/30) = 63 %	Bruit
		(51/30)	silence
		(73/100)	silence
		(11/15)	VOT
		(30/30)	VOT
4	Eliers <i>et al.</i> (1984)	(100/540) = 18,5 %	Voyelle/Syllabe
5	Sinnott <i>et al.</i> (1987a)	(5/25) = 20 %	
		(30/200) = 15 %	Bruit
6	Sinnott <i>et al.</i> (1987)	(de 25 à 100 ms) $\Delta T/T = 20\%$	Bruit
		(de 150 à 400 ms) $\Delta T/T = 10\%$	Bruit
7	Dooley <i>et al.</i> (1988)	(47/750) = 6%	Bruit
8	Wieringen <i>et al.</i> (1994)	(3/50) = 6%	segment vocalique
		(2,7/20) = 13%	transition vocalique
		(4,5/30) = 15%	transition vocalique

		(4,9/50) = 9,8 %	transition vocalique
9	Fitzgibbon <i>et al.</i> (1994)	(48/250) = 19%	Bruit = silence
10	Fitzgibbons <i>et al.</i> (1995)	(46,8/250) = 19%	bruit
		(48,5/250) = 19%	silence
11	Gordon-Salant <i>et al.</i> (2006)	(19/483) = 3,9 %	Silence/syllabe
		(29 /310) = 9,4 %	VOT/syllabe
		(38/249) = 15,3%	Voyelle/syllabe
		(28/247) = 11,3 %	Transition formantique/syllabe

Annexe VIII: illustration graphique de la réduction/magnification de l'espace vocalique de l'arabe dialectal de Mayadin selon la condition expérimentale.



Annexe IX: manipulation non-formelle des durées des voyelles de l'arabe de Mayadin.

Nous présentons dans cette annexe une expérience perceptive qui consiste en la manipulation des voyelles pour effacer la différence au niveau de la quantité et ne laisser que la qualité en jeu. Les résultats de cette expérience ne sont pas en mesure de nous renseigner sur la significativité des écarts, mais elle nous fournit plutôt des informations sur la prédominance de l'une ou de l'autre composante (quantité/qualité) dans la discrimination Brève/Longue.

Une expérience perceptive PRÉLIMINAIRE: Le poids de la quantité dans la reconnaissance de chacune des trois voyelles cardinales.

I. Introduction

Nous effectuons ici une expérience qui nous permettra de contrôler les deux composantes de notre dichotomie Quantité - Qualité l'une indépendamment de l'autre. Une expérience où nous raccourcirons des voyelles longues et nous rallongerons des voyelles brèves, puis nous demanderons à des locuteurs d'identifier ces voyelles. Si une longue abrégée est identifiée en tant que brève c'est parce que c'est la quantité qui fait la différence; si la différence est maintenue malgré la modification de la quantité c'est parce qu'il s'agit d'une opposition assurée par la qualité.

Cette expérience a pour but de sonder le poids de la quantité vocalique vis-à-vis de la qualité acoustique dans la discrimination entre une voyelle longue et sa correspondante brève.

Pour maintenir cet écart perceptif entre deux voyelles d'une paire, les voyelles ne procèdent pas toutes de la même manière; l'opposition [a:] ~ [a] sera due à une différence dans la quantité vocalique, alors que les oppositions [u:] ~ [u] et [i:] ~ [i] seront dues à une différence dans la qualité acoustique.

Nous ne prétendons pas pouvoir donner le taux de contribution de chacune de ces deux composantes perceptives dans la création de cette opposition, mais nous donnons la prédominance de l'une sur l'autre pour chaque voyelle, parce que, en effet, une différence et

au niveau de la quantité et au niveau de la qualité est constatée pour toutes les 3 paires de voyelles [a:] ~ [a], [i:] ~ [i] et [u:] ~ [u] sous différentes conditions expérimentales.

L'expérience a consisté en l'écoute de 120 signaux auditifs, dont 60 signaux modifiés (10 signaux par voyelle), lus par le logiciel Praat 4.0.4, avec un casque audio ATH-610. Les signaux ont été échantillonnés à 44100 Hz. L'expérimentateur faisait défiler les mots, 2 répétitions par mot à 1 seconde d'intervalle, puis le locuteur avait 3 secondes pour écrire le mot qu'il entendait, la confrontation des mots qui ont été présentés avec les mots qui ont été identifiés par les locuteurs a donné les tableaux 22 et 23.

II. méthode de modification

Les six voyelles de l'expérience que nous avons modifiées sont toutes émises avec un débit normal et avec un schéma déclaratif. La durée des voyelles brèves a été rallongée (augmentée) de telle sorte qu'elle atteigne la durée moyenne de celle des voyelles longues, et la durée des voyelles longues a été réduite de telle sorte qu'elle atteigne la durée moyenne de celle des voyelles brèves. Le processus de rallongement et d'abréviation a suivi les démarches suivantes:

- Nous calculons la durée moyenne des longues d'une part et celle des brèves de l'autre.
- Nous calculons pour chaque voyelle de combien, en millièmes de seconde, il faudra augmenter ou réduire sa durée pour qu'elle soit, à quelques ms près, égale à la durée moyenne de son homologue.
- On regarde ensuite le signal de la voyelle à modifier. Nous savons que le cycle vocalique change de forme au cours de l'émission de la voyelle; nous mesurons la durée de chaque forme et calculons la proportion de chaque forme. Nous augmentons ou diminuons la durée de chaque forme proportionnellement à la durée totale de la voyelle à modifier. Ainsi l'augmentation ou réduction, que la voyelle subit, est répartie sur chaque forme en fonction du taux de sa présence dans la voyelle.

Exemple: supposons que la durée moyenne des voyelles brèves est de 110 ms et que celle des longues est de 215 ms. Soit une voyelle brève dont la durée est de 123 ms, nous devons augmenter sa durée pour qu'elle atteigne une valeur proche de 215 ms, donc nous allons rajouter à cette voyelle 92 ms.

Nous mesurons la durée de chaque forme du cycle vocalique. Supposons que nous en avons 3, le 1er est de 65 ms donc il constitue 52.8% de la durée de la brève, le 2è est de 34 ms donc il constitue 27.6% de la durée voyelle, et le 3è est de 24 ms donc 19.5% de la durée totale. Nous divisons les 92 ms suivant ces pourcentages, c'est-à-dire que nous calculons 52.8% de 92 soit 48,6 ms; 27.6% de 92 soit 25,4 ms; et 19.5% de 92 soit 18 ms, et nous les accordons respectivement à 65, à 34 et à 24. En effet toute la voyelle était concernée par le rallongement, y compris les événements VOT et VTT. Pour éviter toute dissonance, les cycles rajoutés sont des cycles de la même voyelle qui subit le rallongement. Si nous voulions rallonger une forme cyclique quelconque, appelons-le Ω , qui compte, admettons, 10 cycles d'une durée qui équivaut à 3 de cycle (la durée d'un cycle $T= 1/f$), nous copions le cycle n°3 et nous le collions juste après le cycle n° 3, puis nous copions le cycle n°6 et nous le collions après le cycle n°6, ensuite nous copions le cycle n°9 et nous le collions après le cycle n°9, tout cela pour dire que nous veillions à ce que les nouveaux cycles que nous introduisions soient répartis dans le signal à modifier à équidistance.

Pour les longues, leurs raccourcissements ont suivi exactement les mêmes démarches que les rallongements des voyelles brèves, sauf que pour les longues on coupait, alors que pour les brèves on rajoutait.

II. Résultats

II.1. Résultat de l'identification des voyelles modifiées

Après nous être assuré que la liste des signaux qui passeront à l'écoute est exhaustive (signaux modifiés et signaux non modifiés) et que l'ordre de la présentation de ces signaux est bien aléatoire, dans un milieu calme, les locuteurs nous ont donné leurs réponses que nous avons converti en un langage plus concis et plus expressif, le langage +/- . Nous avons trié ensuite les réponses qui ont porté sur les voyelles modifiées et celles qui ont porté sur les voyelles non modifiées, tableau ci-dessous.

Résultats du test perceptif des signaux modifiés											
Locuteur 1(AB)						locuteur 2(AH)					
A	AA	I	II	U	UU	A	AA	I	II	U	UU
(+)	(+)	(-)	(-)	(-)	(+)	(+)	(+)	(-)	(-)	(-)	(-)
(+)	(-)	(-)	(-)	(-)	(-)	(+)	(+)	(-)	(-)	(-)	(-)
(+)	(-)	(-)	(-)	(-)	(-)	(+)	(+)	(-)	(-)	(-)	(-)
(-)	(+)	(-)	(-)	(-)	(-)	(+)	(+)	(-)	(-)	(-)	(-)
(+)	(+)	(-)	(-)	(-)	(+)	(+)	(+)	(-)	(-)	(-)	(-)
(-)	(+)	(-)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(-)
(+)	(-)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(-)	(-)
(+)	(-)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(-)	(-)
(+)	(+)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(-)	(-)
(+)	(+)	(-)	(-)	(-)	(+)	(+)	(+)	(-)	(-)	(-)	(-)

La consonne simple = brève, et double = longue. Toutes les brèves ont été rallongées, et toutes les longues ont été abrégées. Le signe (+) veut dire que la voyelle était identifiée conformément à la modification, c'est-à-dire qu'une brève modifiée en longue était identifiée en tant que longue, et qu'une longue modifiée en brève était identifiée en tant que brève (le locuteur s'est fait piégé). Le signe (-) veut dire que la voyelle était identifiée comme si elle n'était pas modifiée (le locuteur ne s'est pas fait piégé).

Aucune des erreurs d'identification, représentées ici par le signe (+), n'a porté sur la qualité de la voyelle, c'est-à-dire qu'une voyelle pouvait être identifiée en tant que la longue ou la brève correspondante, mais pas en tant qu'une autre voyelle.

Voici tous locuteurs confondus, les taux de réponses correctes [pour nous, la réponse est correcte quand le locuteur n'entend pas la modification; quand la réponse est (-)], voyelle par voyelle:

A: 3/20, soit 15%; AA: 7/20 soit 35%; I: 20/20 soit 100%; II: 20/20 soit 100%; U: 20/20 soit 100%; UU: 17/20 soit 85%.

III.2. Résultat de l'identification des voyelles non modifiées

Les abréviations et les significations des signes du tableau des signaux modifiés valent aussi pour ce tableau ci-dessous.

Résultats du test perceptif des **signaux non modifiés**

Locuteur 1(AB)						locuteur 2(AH)					
A	AA	I	II	U	UU	A	AA	I	II	U	UU
(-)	(-)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(-)	(-)
(-)	(-)	(-)	(-)	(-)	(-)	(?)	(-)	(-)	(-)	(-)	(-)
(+)	(-)	(-)	(-)	(-)	(-)	(+)	(-)	(-)	(-)	(-)	(-)
(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)
(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)	(-)

À part trois mauvaises identifications pour la voyelle [a] bref, à trois reprises il était identifié en tant que [a] long, et une 4^e fois elle n'a pas pu être identifiée. Le signe "?", nous le comptons comme 1/2 bonne réponse, cette interprétation des réponses de doute était inspirée des travaux de M. Rossi (1971), il procédait comme suit: "Dans l'interprétation des résultats nous comptons les réponses de doute comme des réponses intermédiaires: 50% + et 50% -. Ce qui fait que cette voyelle a marqué 97,1% de bonnes réponses, toutes les autres voyelles ont marqué un taux d'identification correcte de 100%. Les résultats de ce deuxième tableau mettent en évidence deux choses: - Les voyelles non modifiées, à part le /a/ bref, sont bien des voyelles caractéristiques, puisqu'elles capitalisent 100% d'identification correcte. - Le /a/ est une voyelle dont la qualité passe pour celle de la longue pour la moindre perturbation, les qualités A/AA sont proches, mais ce n'est pas longue qui passe pour la brève.

Résultat: notre lecture des résultats de cette expérience montre que les voyelles n'utilisent pas les mêmes paramètres pour affirmer leurs identités vocaliques par rapport à leurs brèves correspondantes, nous observons aussi que l'effet de la qualité est présent, puisqu'à aucun niveau nous n'avons pas trouvé que les voyelles se fusionnent à cent pour cent. Peut-être nous pouvons raffiner notre lecture en disant: toutes les voyelles se servent des mêmes paramètres, mais pas au même degré.

En effet, l'opposition A/AA est la plus affectée par la neutralisation de la différence qui se situe au niveau des durées des voyelles, les deux voyelles sont loin d'arriver à 50% de bonnes réponses, un seuil critique que nous adoptons pour juger de la prédominance d'un paramètre sur l'autre, donc [a] et [a:] même qualité réalisée avec deux durées différentes.

L'opposition I/II est absolument insensible à la modification de la quantité, les deux voyelles réalisent 100% de bonnes réponses en l'absence de l'opposition au niveau de leurs

durées. La qualité de la brève et celle de la longue sont tellement opposées que de par cette différence elles ne se confondent pas, donc [i] et [i:] sont deux qualités vocaliques bien distinctes.

Quant à l'opposition U/UU, elle est sensible au changement de la durée, c'est la longue des deux qui passe pour la brève, de toutes les manières, le taux de son identification reste largement au-dessus de 50%, donc une prédominance de la qualité sur la quantité, mais une prédominance moins marquée par rapport au [i]. Ainsi [u] et [u:] seront deux qualités vocaliques bien différentes.

Conclusion

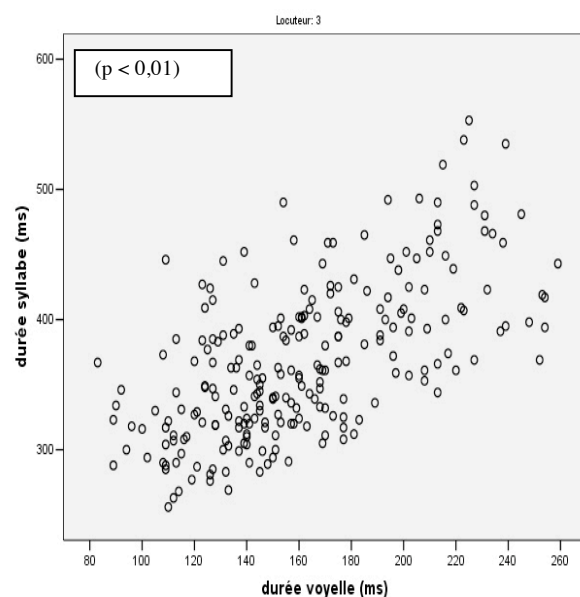
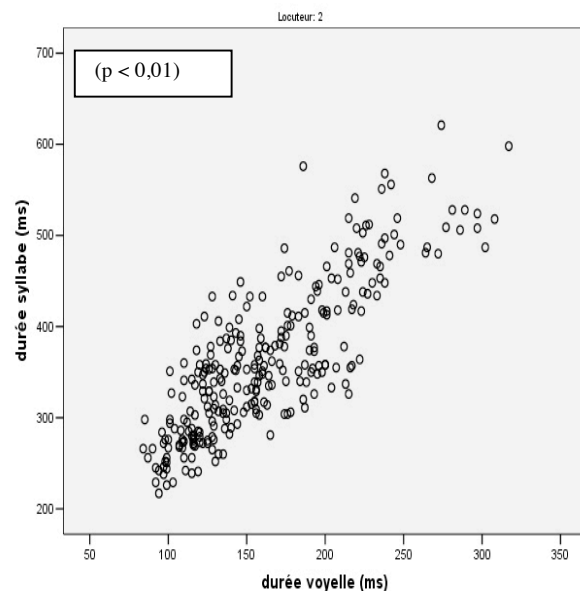
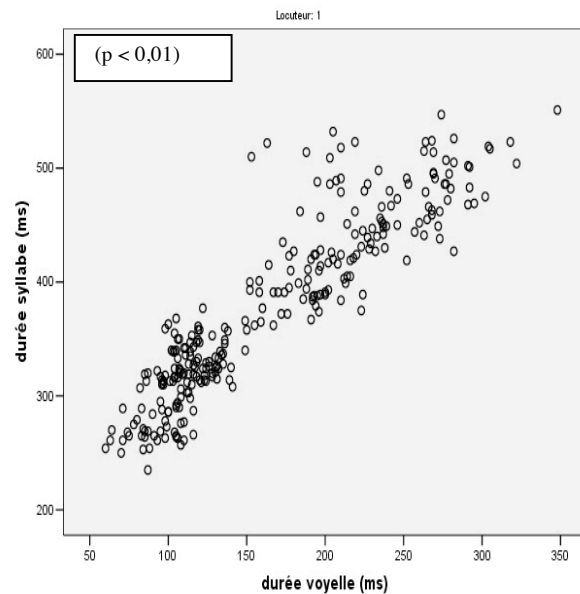
Finalement, ce que nous venons de constater peut constituer une réponse à une question que nous nous sommes posés dès le début de ce travail: Savoir si l'on arrive à manipuler, en ce qui concerne les voyelles, des qualités acoustiques et des squelettes temporels les uns indépendamment des autres. D'après tout ce que nous avons vu jusqu'à là, notre réponse est: non, un changement de la quantité vocalique était toujours accompagné d'un changement au niveau de la qualité, même pour le /a/ qui présente un petit changement au niveau de la qualité, nous ne pouvons pas dire que ce changement n'existe pas.

Annexe X: la durée de la syllabe en fonction de la durée de la voyelle en arabe de Mayadin (Syrie) pour les trois locuteurs.

Les figures ci-contre représentent la durée de la syllabe en fonction de la durée de la voyelle en arabe de Mayadin (Syrie) pour les trois locuteurs masculins de l'étude, 1, 2 et 3 classés selon leurs âges 40, 38 et 28 ans respectivement. Les voyelles mesurées sont 6, trois brèves [i, a et u] et trois longues [i:, a: et u:]. Les voyelles étaient émises selon deux débits de parole, normal et rapide, et selon deux schémas mélodiques, déclaratif et interrogatif. La voyelle est insérée dans la syllabe CVC [zVr]. L'Anova montre un effet significatif ($p < 0,01$) de la durée de la voyelle sur la durée de la syllabe pour chaque locuteur à part.

Effectifs

Locuteur 1: 144 brèves et 137 longues.
 Locuteur 2: 143 brèves et 138 longues.
 Locuteur 3: 119 brèves et 120 longues.

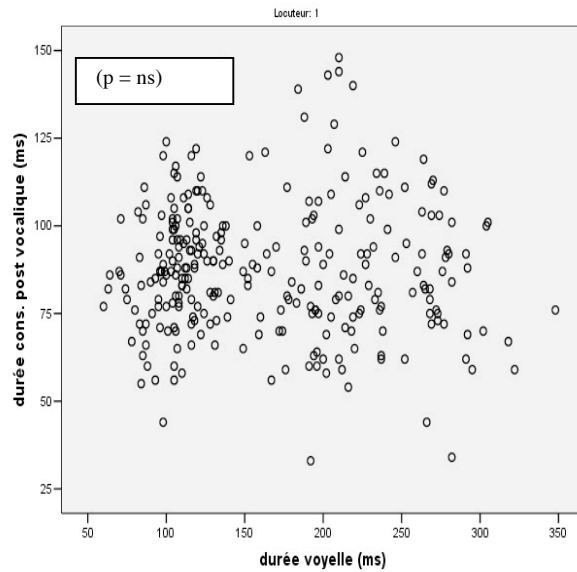


Annexe XI: la durée de la consonne post-vocalique en fonction de la durée de la voyelle en arabe de Mayadin pour les trois locuteurs.

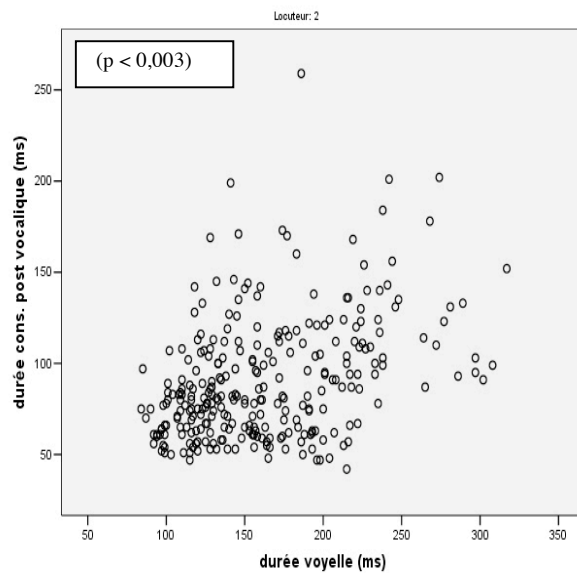
Les figures ci-contre représentent la durée de la consonne post-vocalique en fonction de la durée de la voyelle en arabe de Mayadin (Syrie) pour les trois locuteurs masculins de l'étude, 1, 2 et 3 classés selon leurs âges 40, 38 et 28 ans respectivement. Les voyelles mesurées sont 6, trois brèves et u] et trois longues [i:, a: et u:]. Les voyelles étaient émises selon deux débits de parole, normal et rapide, et selon deux schémas mélodiques, déclaratif et interrogatif. La voyelle est insérée dans la syllabe CVC [zVr]. L'Anova ne montre d'effet significatif pour le locuteur n°2 ($p < 0,003$) de la durée de la voyelle sur la durée de la consonne post-vocalique.

Effectifs

Locuteur 1: 144 brèves et 137 longues. Locuteur 2: 143 brèves et 138 longues. Locuteur 3: 119 brèves et 120 longues.

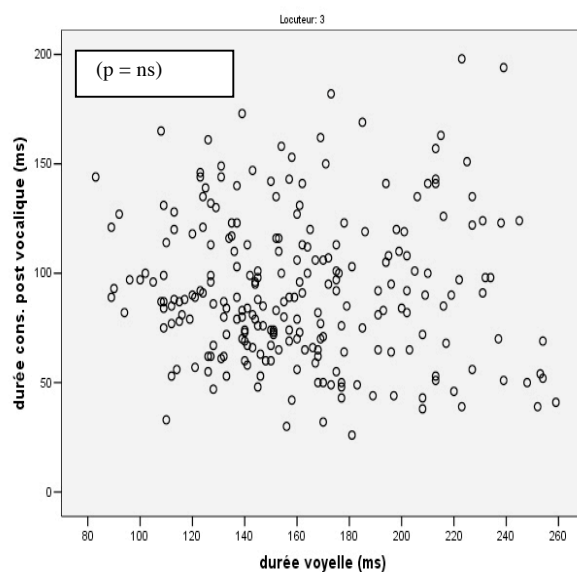


de



Les
[i, a

que

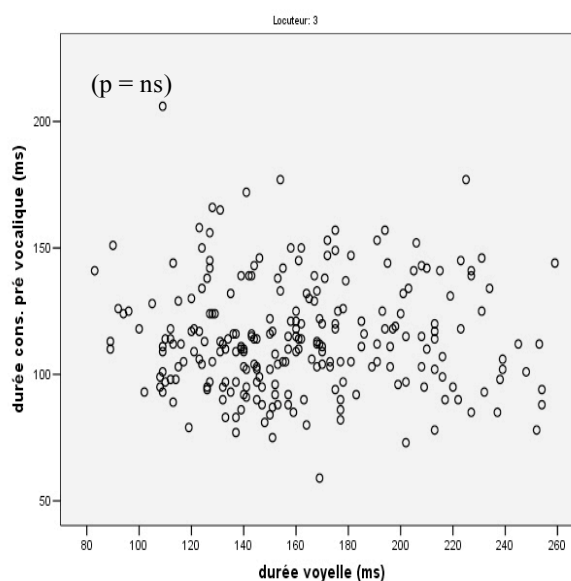
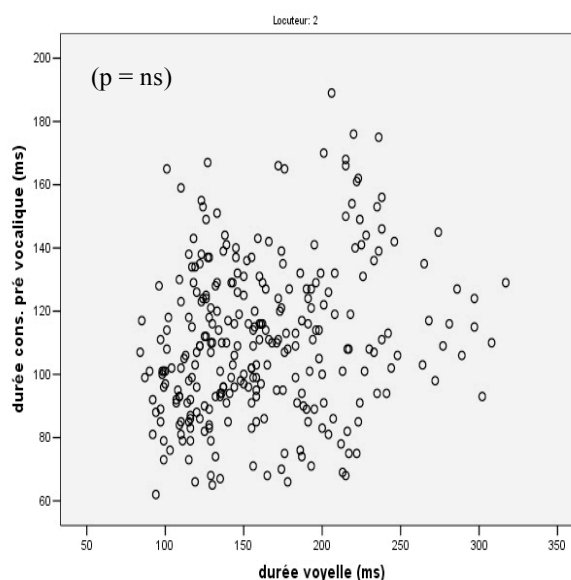
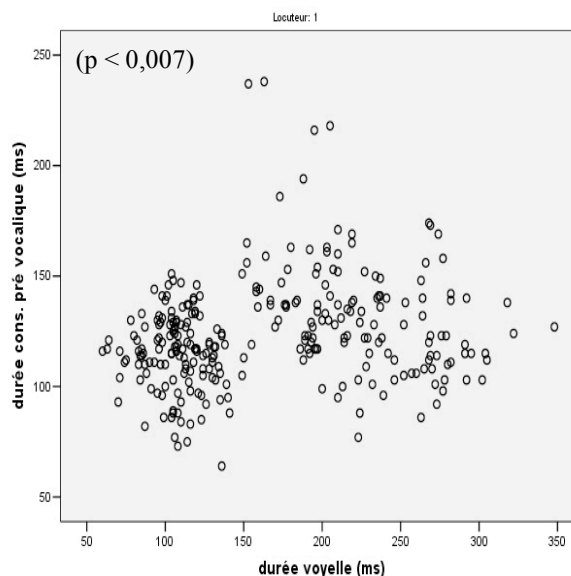


Annexe XII: la durée de la consonne prévocale en fonction de la durée de la voyelle en arabe de Mayadin pour les trois locuteurs.

Les figures ci-contre représentent la durée de la consonne prévocale en fonction de la durée de la voyelle en arabe de Mayadin (Syrie) pour les trois locuteurs masculins de l'étude, 1, 2 et 3 classés selon leurs âges 40, 38 et 28 ans respectivement. Les voyelles mesurées sont 6, trois brèves [i, a et u] et trois longues [i:, a: et u:]. Les voyelles étaient émises selon deux débits de parole, normal et rapide, et selon deux schémas mélodiques, déclaratif et interrogatif. La voyelle est insérée dans la syllabe CVC [zVr]. L'Anova ne montre pas d'effet significatif que pour le locuteur n°1 ($p < 0,007$) de la durée de la voyelle sur la durée de la consonne prévocale.

Effectifs

Locuteur 1: 144 brèves et 137 longues.
 Locuteur 2: 143 brèves et 138 longues.
 Locuteur 3: 119 brèves et 120 longues.

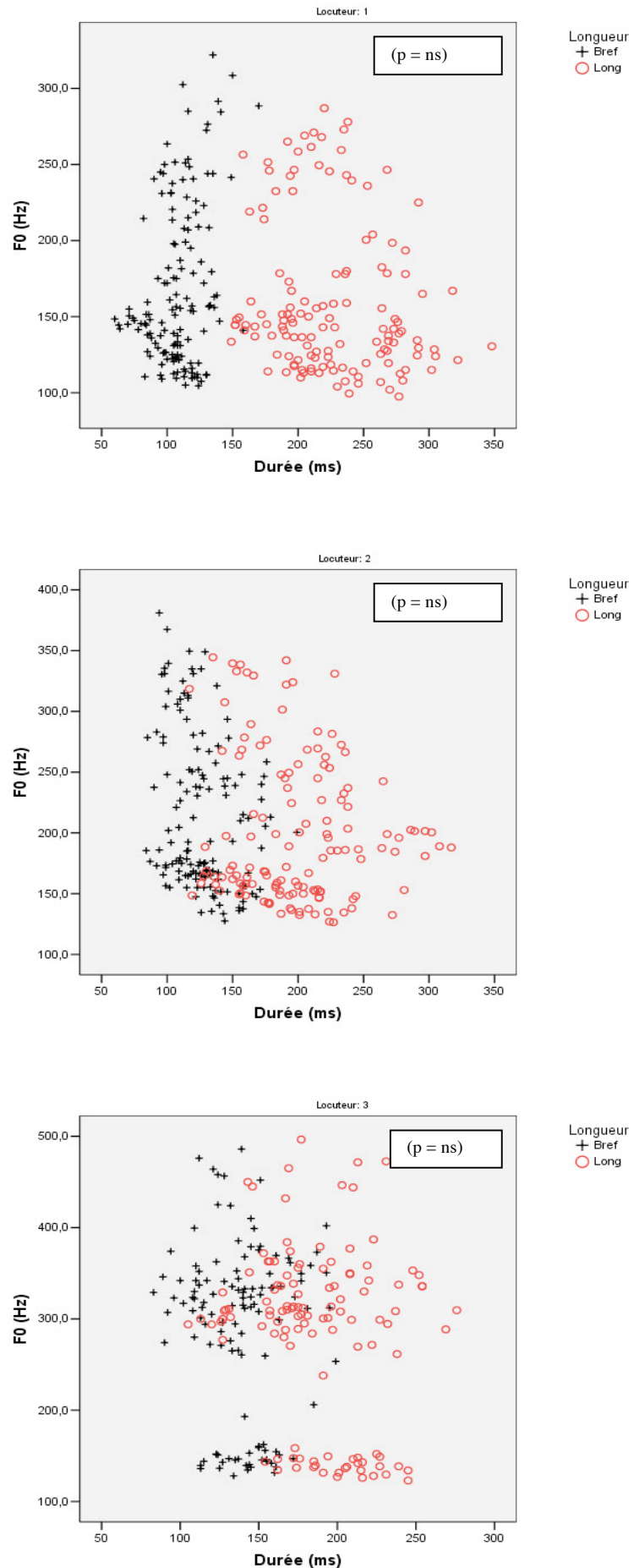


Annexe XIII: la fréquence fondamentale en fonction de la catégorie de la voyelle Brève/Longue en arabe de Mayadin pour les trois locuteurs

Les figures ci-contre représentent la fréquence fondamentale en fonction de la catégorie de la voyelle Brève/Longue en arabe de Mayadin (Syrie) pour les trois locuteurs masculins de l'étude, 1, 2 et 3 classés selon leurs âges 40, 38 et 28 ans respectivement. Les voyelles mesurées sont 6, trois brèves [i, a et u] et trois longues [i:, a: et u:]. Les voyelles étaient émises selon deux débits de parole, normal et rapide, et selon deux schémas mélodiques, déclaratif et interrogatif. La voyelle est insérée dans la syllabe CVC [zVr]. L'Anova ne montre pas d'effet significatif pour aucun des trois locuteurs de la catégorie de la voyelle sur la F0.

Effectifs

Locuteur 1: 144 brèves et 137 longues. Locuteur 2: 143 brèves et 138 longues. Locuteur 3: 119 brèves et 120 longues.

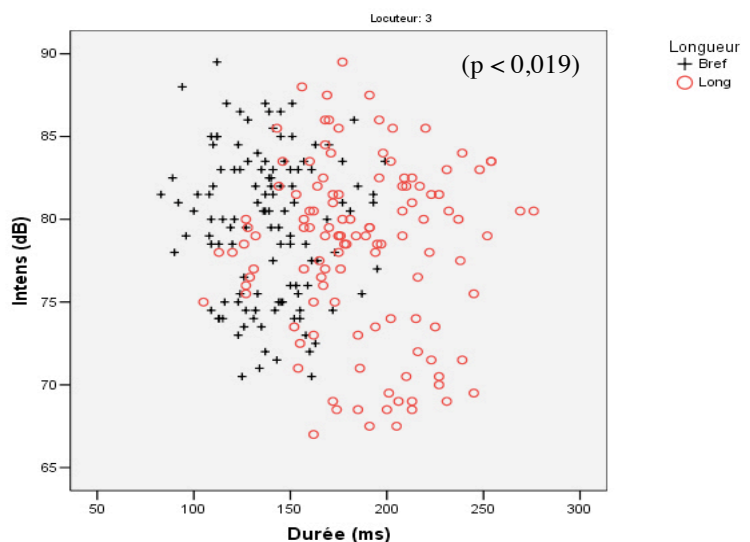
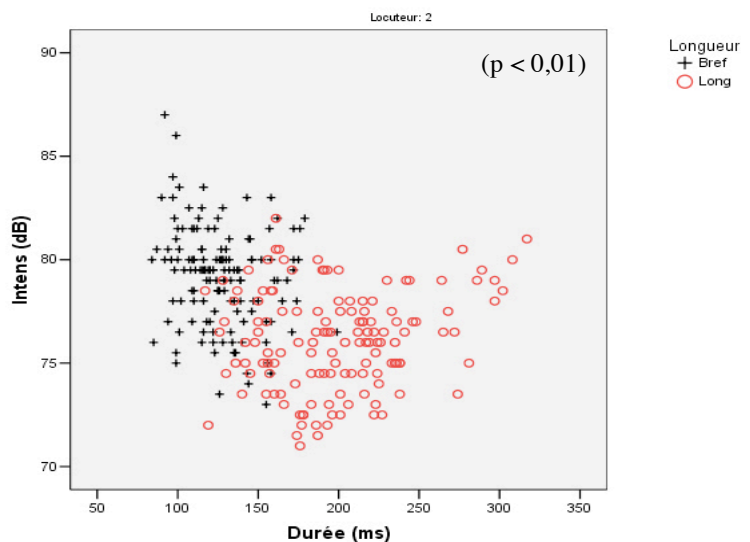
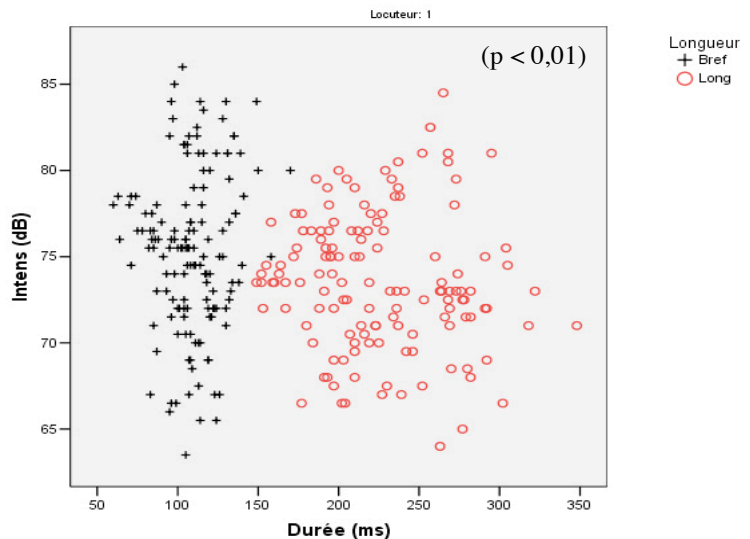


Annexe XIV: Intensité de la voyelle en fonction de sa catégorie Brève/Longue en arabe de Mayadin pour les 3 locuteurs.

Les figures ci-contre représentent l'intensité de la voyelle en fonction de sa catégorie Brève/Longue en arabe de Mayadin (Syrie). Trois locuteurs masculins, 1, 2 et 3 classés selon leurs âges 40, 38 et 28 ans respectivement. Les voyelles mesurées sont 6, trois brèves [i, a et u] et trois longues [i:, a: et u:]. Les voyelles étaient émises selon deux débits de parole, normal et rapide, et selon deux schémas mélodiques, déclaratif et interrogatif. La voyelle est insérée dans la syllabe CVC [zVr]. L'Anova montre un effet significatif, pour chacun des trois locuteurs, de la catégorie de la voyelle sur son intensité.

Effectifs

Locuteur 1: 144 brèves et 137 longues. Locuteur 2: 143 brèves et 138 longues. Locuteur 3: 119 brèves et 120 longues.



Annexe XV: intensité pour chaque voyelles de l'étude de l'arabe de Mayadin, et effet du débit.

Les tableaux dans cette annexe représentent l'intensité pour chacune des 6 voyelles de l'étude de l'arabe de Mayadin, tous locuteurs, tous schémas mélodiques et tous débits de parole confondus.

Qualité = a, Longueur = Bref

Statistiques descriptives					
	N	Minimum	Maximum	Moyenne	Ecart type
Intensité (dB)	134	66	87	77,67	4,353

Qualité = a, Longueur = Long

Statistiques descriptives					
	N	Minimum	Maximum	Moyenne	Ecart type
Intensité (dB)	133	65	88	76,61	4,504

Qualité = i, Longueur = Bref

Statistiques descriptives					
	N	Minimum	Maximum	Moyenne	Ecart type
Intensité (dB)	136	66	90	78,49	4,470

Qualité = i, Longueur = Long

Statistiques descriptives					
	N	Minimum	Maximum	Moyenne	Ecart type
Intensité (dB)	135	67	88	75,11	3,982

--

Qualité = u, Longueur = Bref

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
Intensité (dB)	136	64	87	77,78	4,381

Qualité = u, Longueur = Long

Statistiques descriptives

	N	Minimum	Maximum	Moyenne	Ecart type
Intensité (dB)	126	64	90	75,88	4,604

Intensité en fonction du débit de la parole.

En passant du débit normal au débit rapide, nous avons observé une hausse de l'intensité de l'ordre de 5,5 dB, tous locuteurs, toutes longueurs et toutes voyelles confondues.

LOCUTEUR	VOY.	Normal	Rapide
1	A		
	AA		
	I		
	II	70,29 (dB)	76,62 (dB)
	U		
2	UU		
	A		
	AA		
	I		
	II	77,38 (dB)	79,5 (dB)
3	U		
	UU		
	A		
	AA		
	II	75,7 (dB)	83,7 (dB)

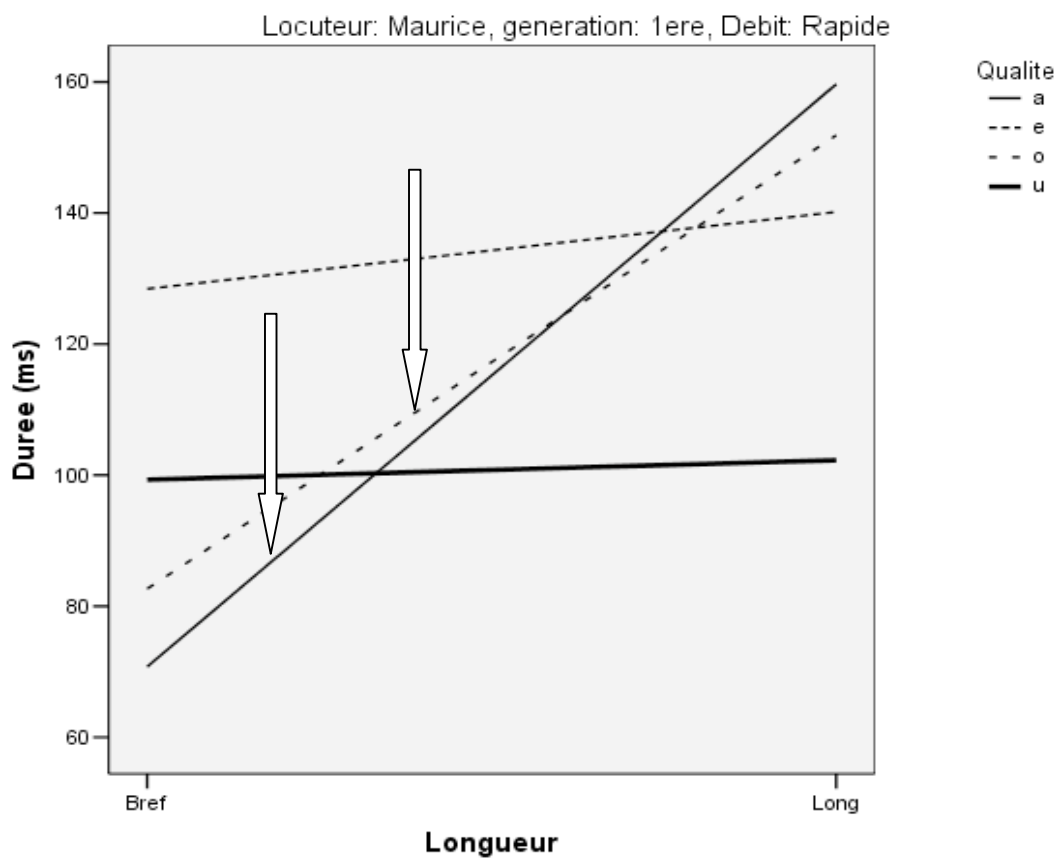
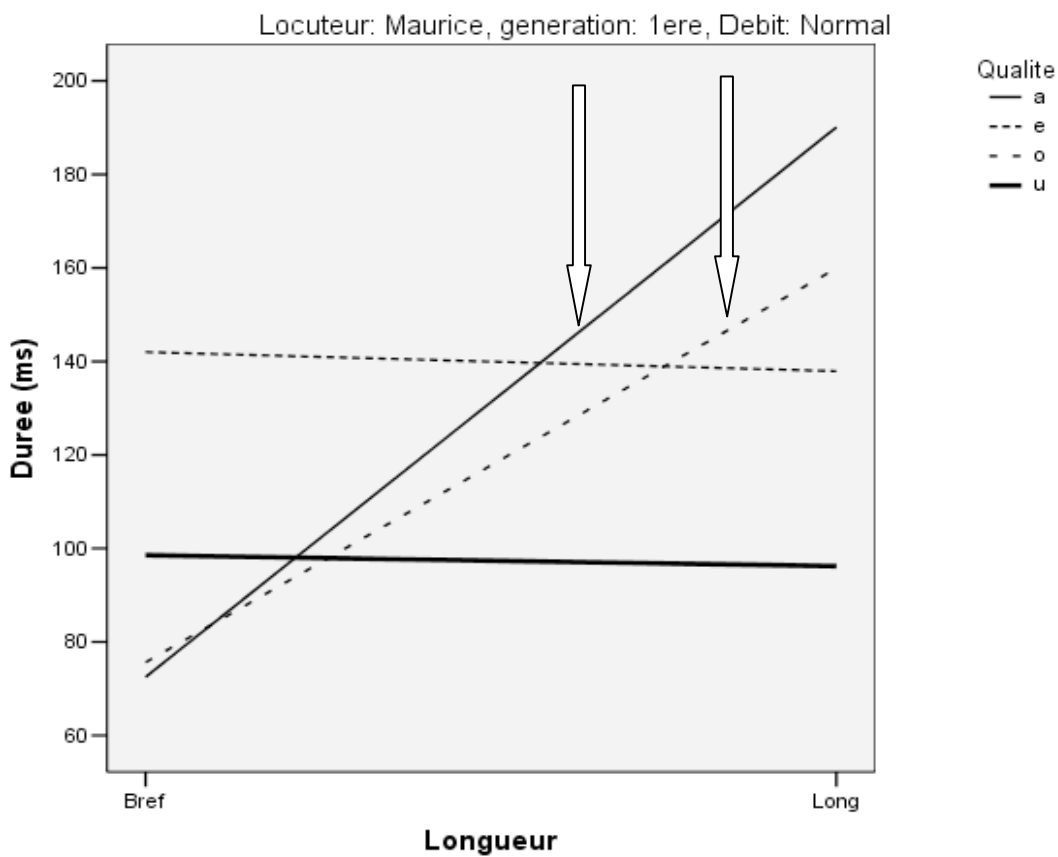
L'effet du débit rapide sur l'intensité des voyelles d'après notre expérience sur l'arabe de Mayadin.

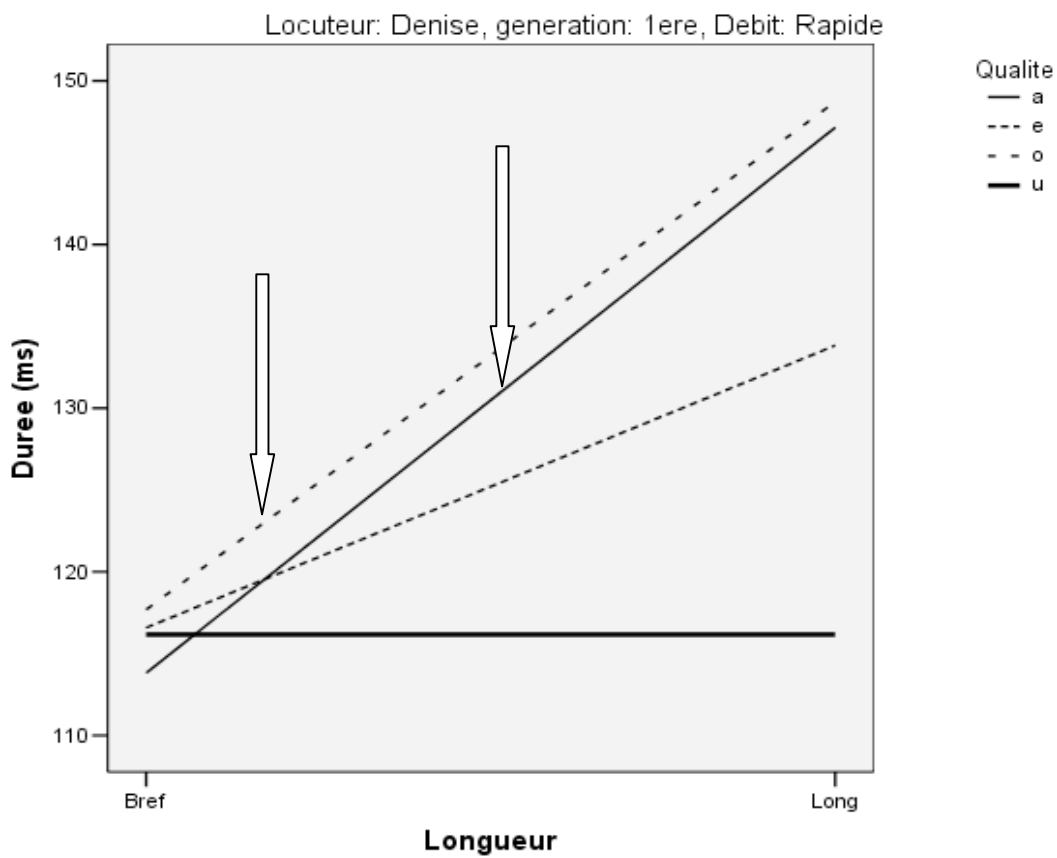
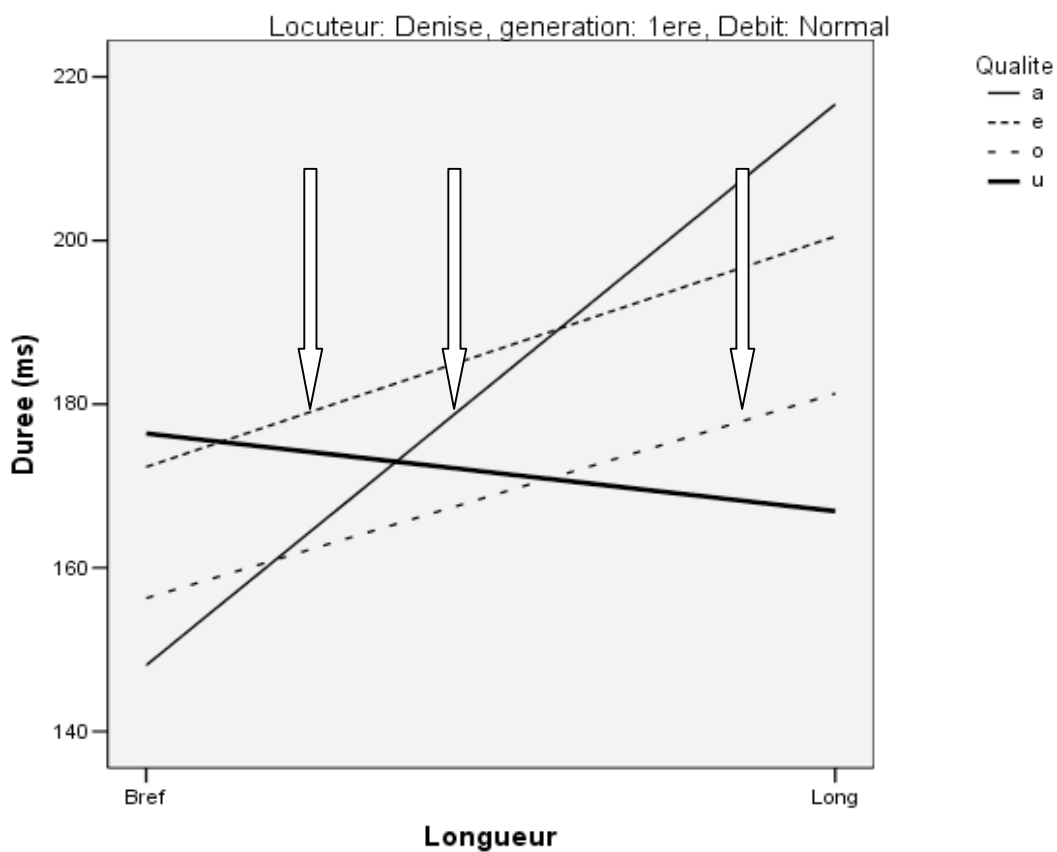
Ce tableau montre que le débit rapide dans la parole est marqué par une hausse systématique²¹ de l'intensité, toutes voyelles confondues. Ce tableau montre aussi que les locuteurs réalisent cette hausse à différents degrés.

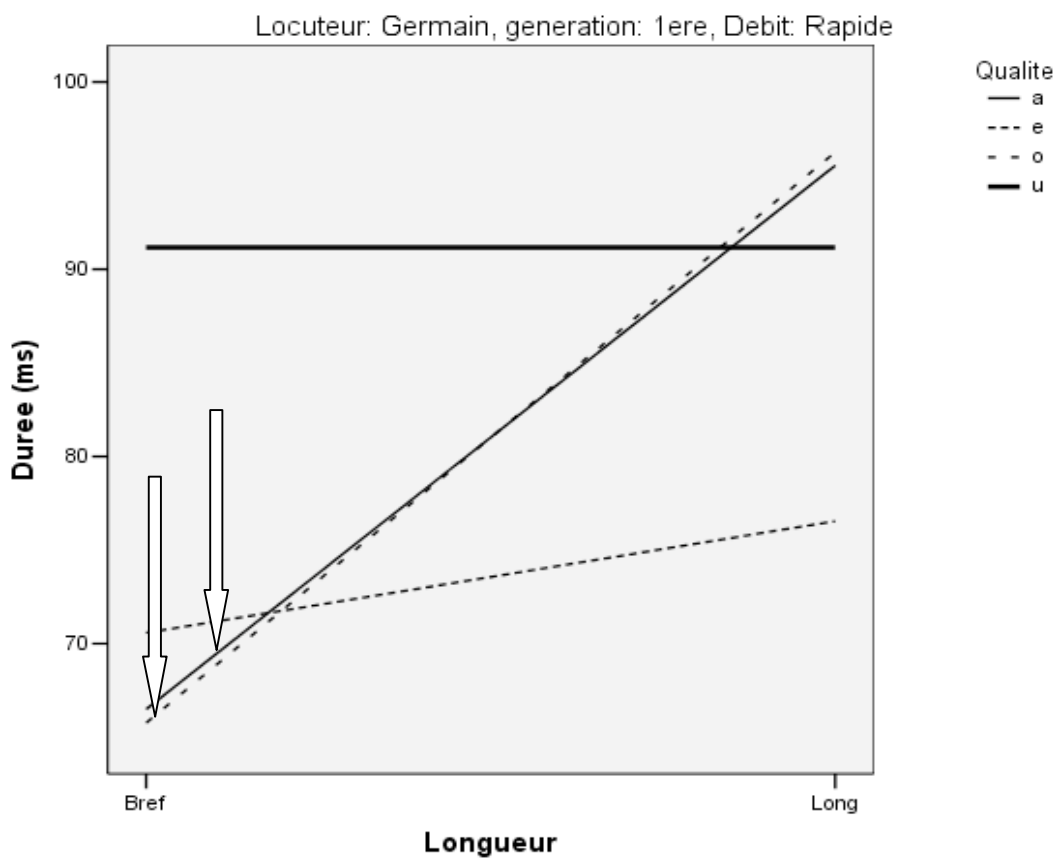
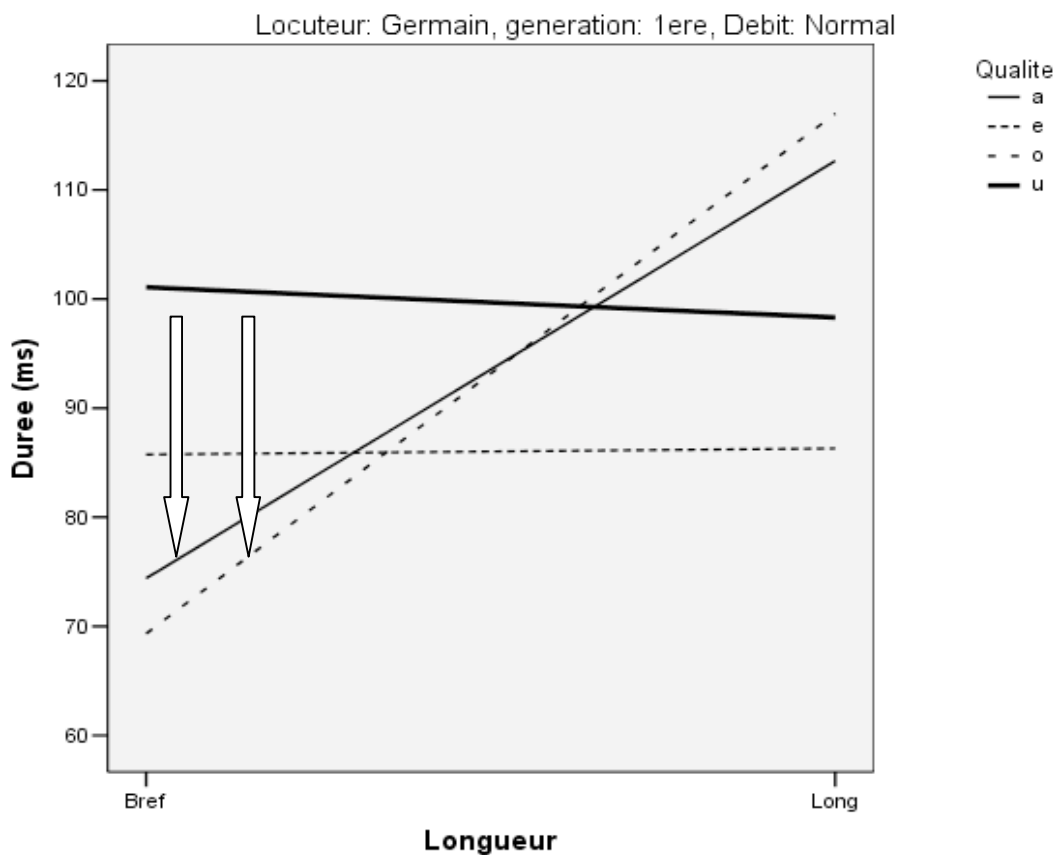
²¹ Nous nous sommes permis de dire *systématique* parce que la hausse était effectuée par tous nos locuteurs, donc cette vérité se limiterait à notre étude à nous.

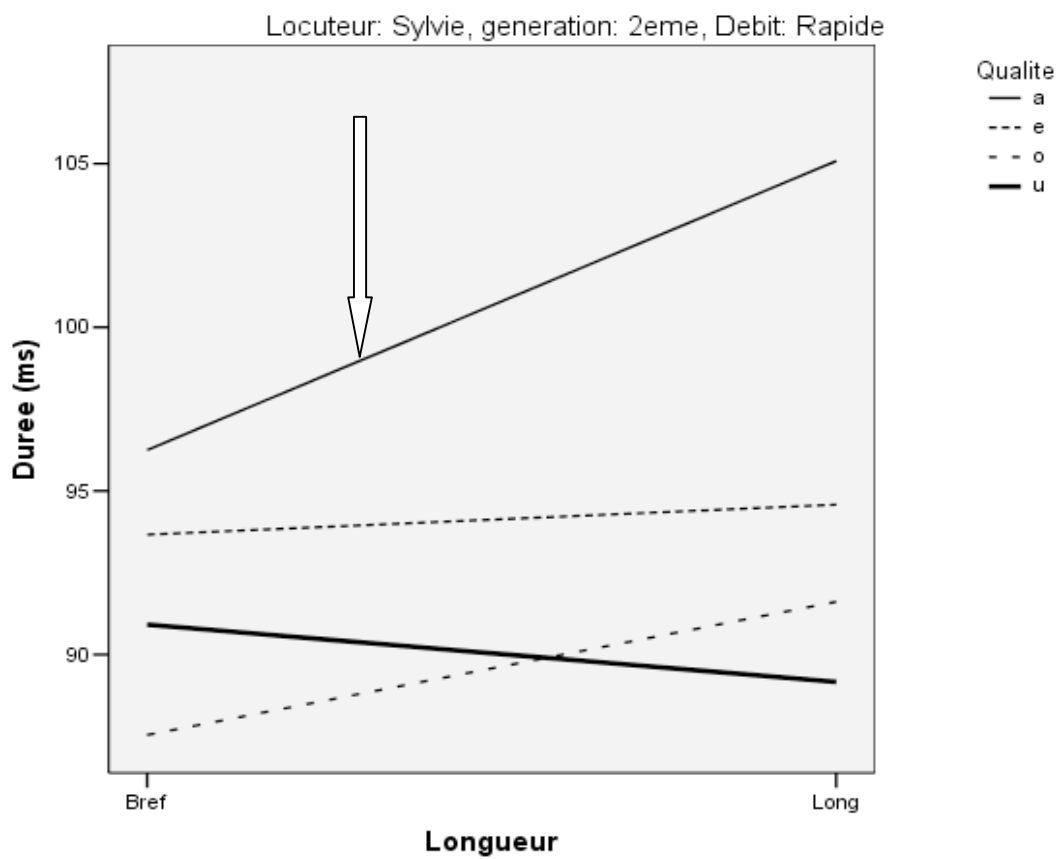
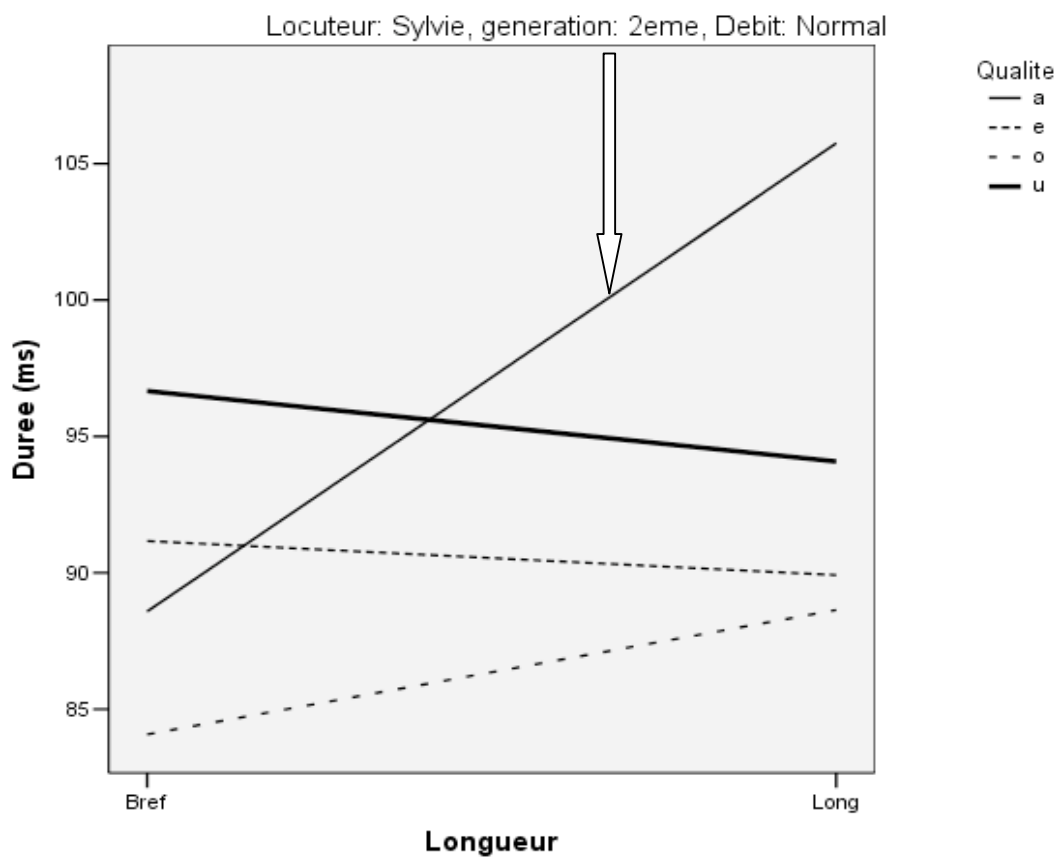
Annexe XVI: résultats concernant la partie qui a porté sur les mesures de la durée des voyelles en français de la Savoie pour chaque locuteur et chaque débit de parole à part.

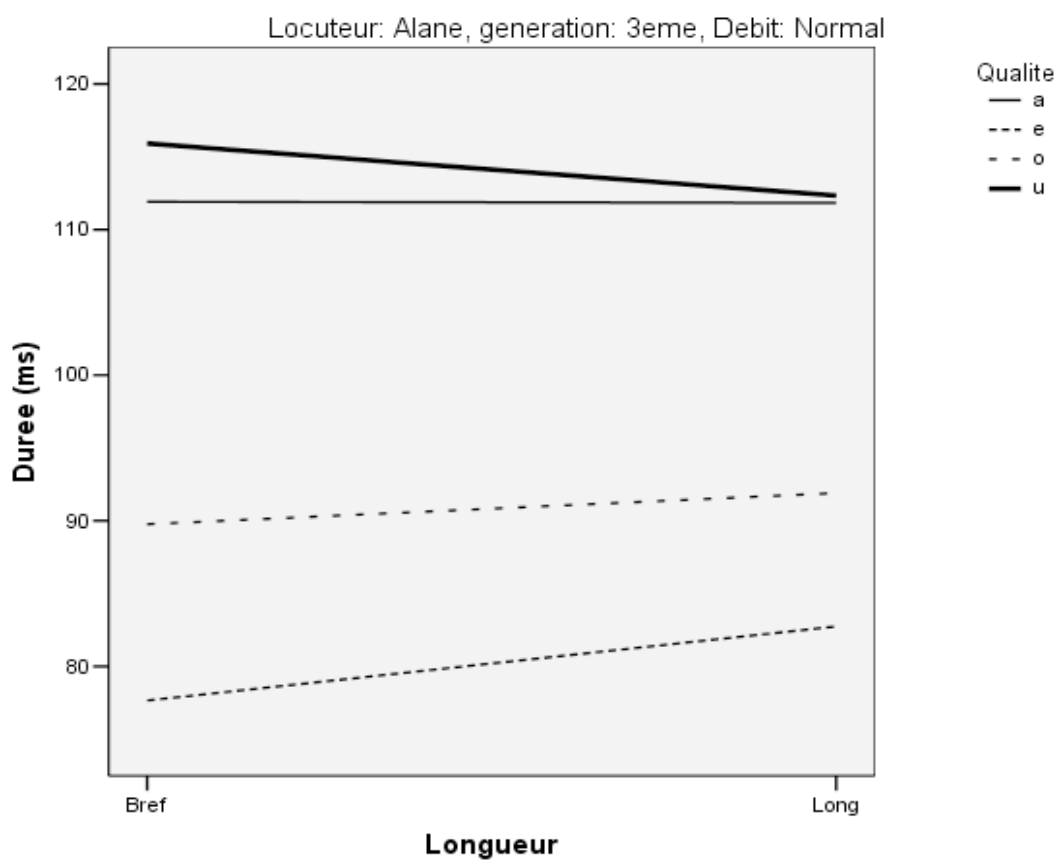
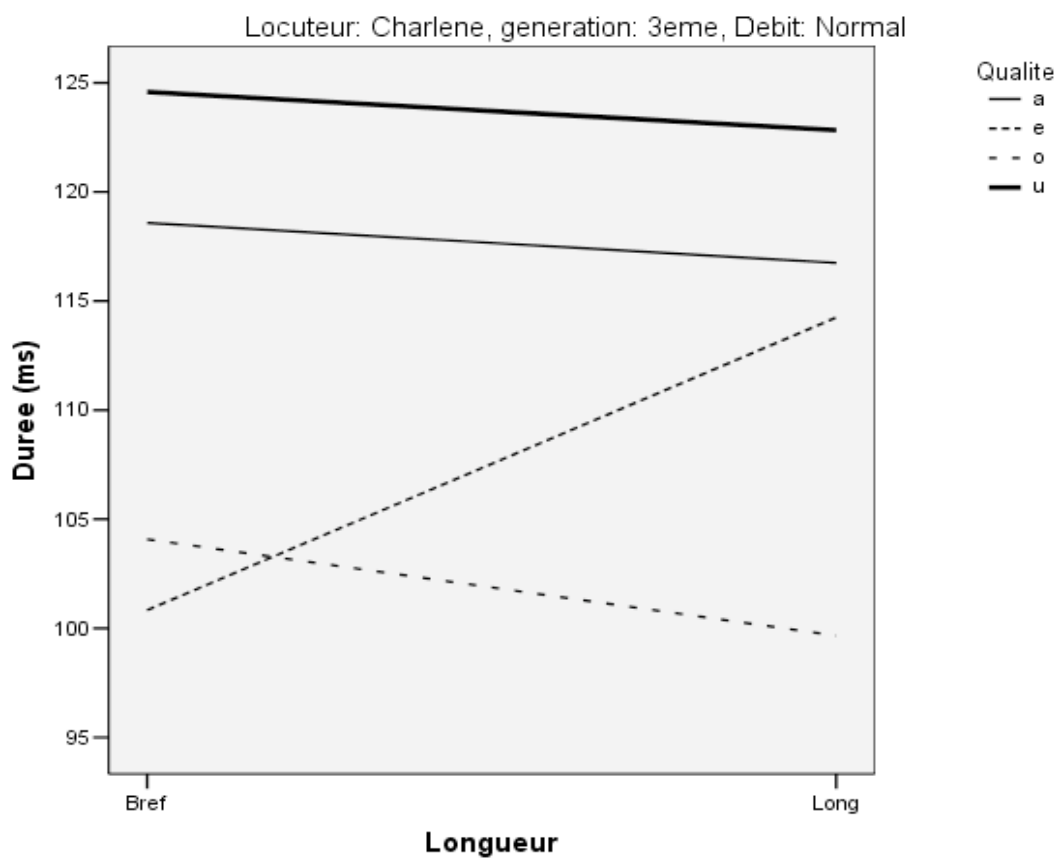
Nous présentons dans cette annexe les résultats concernant la partie qui a porté sur les mesures de la durée des voyelles en français de la région de la Savoie pour chacun des locuteurs et pour chaque débit de parole à part. Cette manière d'exposition des résultats est mieux adaptée pour mettre en évidence les différences statistiques qui sont significatives. Les flèches sur les figures pointent les courbes des qualités vocaliques qui présentent des différences significatives ($p < 0,01$).











Annexe XVII: voies de la transmission de l'information sur les identités des voyelles.

Nous discutons brièvement dans cette annexe certaines des voies de la transmission de l'information sur les identités des voyelles.

1. Les transitions formantiques (le contexte)

La résolution formantique des voyelles, et par conséquent leur identification, est meilleure lorsque celles-ci sont introduites dans un contexte consonantique que lorsqu'elles sont produites sans contexte, un contexte implique la présence de transitions formantiques « *Indeed performance is often better for vowels in consonantal contexts than for vowels spoken in null contexts* » (Strange, 1989). C'est pourquoi les voyelles isolées ont besoin d'une distance plus importante entre elles pour se différencier que les voyelles introduites dans un contexte. En effet, le fait que les formants de deux voyelles aient les mêmes patrons en ce qui concerne les valeurs fréquentielles n'implique pas forcément le fait que ces deux voyelles ont les mêmes transitions formantiques et cette différence qui ne se situe qu'au niveau des transitions garantit la différenciation de deux voyelles. Lindblom (1967) accorde aux transitions formantiques un rôle capital dans l'identification: « *in the recognition of monosyllabic non sense speech the identity of a vowel sound is determined not only by the formant pattern at the point of closest approach to target but also by the direction and rate of adjacent formant transition (...) vowel stimuli that were identical with respect to midpoint formant frequencies but differed with respect to formant transitions were not perceptually equivalent* ».

2. La variation

Une autre source de l'information sur l'identité de la voyelle pourrait se trouver dans la variation lors de l'émission. D'après l'étude faite par Nearey (1989) sur les voyelles de l'anglais canadien, il s'est avéré que les voyelles non co-articulées, donc sans transition, peuvent être identifiées correctement et c'est grâce à la variation de l'information acoustique au cours du temps. Nearey assigne les caractéristiques de cette variation à des changements spectraux inhérents à la structure acoustique de la voyelle sans pour autant que ces variations soient de la diphtongaison. "Thus it appears that even uncoarticulated vowels may be identified on the basis of time varying information (...) vowel-inherent spectral change". Strange (1989) lui aussi, est parvenu à un résultat similaire: il y a toute une dynamique qui est

mise en jeu, cette dynamique est assurée par les différentes configurations des informations contenues dans le signal acoustique, et par les relations que les composantes du signal entretiennent les unes par rapport aux autres. "(...) *Co-articulated vowels are identified on the basis of time-varying acoustic information, rather than solely on the basis of "Target" information contained within a single spectral cross section of an acoustic syllable* " (Strange, 1989).

3. Un rapport et non pas une valeur absolue

Dans une expérience qui visait la caractérisation des voyelles de l'anglais des Etats-Unis, Hillenbrand *et al.* (1995) traitaient les paramètres suivants: Durée, F0, F1, F2 et F3. Ils ont trouvé les valeurs suivantes:

	i	a	u
Durée(ms)			
H	243	267	237
F	306	323	303
E	297	311	278
F0			
H	138	123	143
F	227	215	235
E	246	229	249
F1			
H	342	768	378
F	437	936	459
E	452	1002	494
F2			
H	2322	1333	997
F	2761	1551	1105
E	3081	1688	1345
F3			
H	3000	2522	2343
F	3372	2815	2735
E	3702	2950	2988

Tableau 1. H= homme, F= femme et E= enfant.

Dans le tableau (1)²², il est clair que les valeurs formantiques changent significativement par sexe et par tranche d'âge. Cela met en évidence le fait qu'il n'y ait pas de valeurs précises et uniques qui caractérisent les voyelles et permettent leur identification. Mais il y a, par exemple, pour une voyelle donnée, des relations équivalentes entre formants qui feront qu'on parviendra à catégoriser cette voyelle, qu'elle provienne d'un conduit vocal d'un homme, de celui d'une femme ou de celui d'un enfant.

Nous remarquons que les valeurs changent considérablement d'un locuteur à un autre, et que les rapports entre les formants $\frac{F_1}{F_2}$, $\frac{F_2}{F_3}$ ou $\frac{F_1}{F_3}$ ne sont pas non plus les mêmes d'un

²² Avec H = homme, F = femme et E = enfant (entre 10 et 12 ans).

locuteur à un autre. Nous proposons l'existence d'un rapport relatif, $\Delta F/F$, qui est constant d'un sexe à l'autre et d'un âge à l'autre. Nous faisons les calculs ci-dessous pour le confirmer:

Variation relative($\Delta F/F$ en %)			i	a	u
Homme	F1 => F2	$((F2-F1)/F1)*100$	579	74	164
	F2 => F3	$((F3-F2)/F2)*100$	29	89	135
	F1 => F3	$((F3-F1)/F1)*100$	777	228	520
Femme	F1 => F2	$((F2-F1)/F1)*100$	532	66	141
	F2 => F3	$((F3-F2)/F2)*100$	22	81	148
	F1 => F3	$((F3-F1)/F1)*100$	672	201	496
Enfant	F1 => F2	$((F2-F1)/F1)*100$	582	68	172
	F2 => F3	$((F3-F2)/F2)*100$	20	75	122
	F1 => F3	$((F3-F1)/F1)*100$	719	194	505

Tableau 2. Rapport entre formants.

Interprétation: La forte ressemblance entre les rapports des trois premiers formants des différents locuteurs, nous a permis l'interprétation suivante:

Une voyelle n'est pas caractérisée par des valeurs précises, les valeurs pouvant changer significativement, mais en même temps, ces valeurs ne peuvent pas être n'importe lesquelles.

Une voyelle est caractérisée par des rapports relatifs entre ses formants, et non pas des rapports absolus ($\frac{F_1}{F_2} \times 100$) est un rapport absolu $\neq ((F_2 - F_1) \div F_1) \times 100$ est un rapport relatif. Le premier exprime en pourcentage la valeur que F1 représente de la valeur de F2, alors que le deuxième exprime en pourcentage de la valeur de F1, de combien il faut augmenter F1 pour que cette valeur soit égale à celle de F2. Par exemple, pour la voyelle [i] avec une voix d'homme F1 = 342 et F2 = 2322, le rapport entre F1 et F2 est $((F2-F1)/F1)*100 = 579$, veut dire qu'il faut augmenter la valeur de F1 de 5,79 fois sa valeur pour atteindre la valeur de F2; c'est cette hausse, pour nous, qui est constante dans le rapport entre les formants des voyelles, les valeurs elles mêmes étant variables (dans certaines limites).

Annexe XVIII: Quelques exemples du rôle de la quantité vocalique dans la morphologie de l'arabe.

Du moment où la quantité vocalique est significative dans un système phonologique donné (une langue), on peut observer son importance à travers les différents procédés morphologiques responsables des changements des catégories des unités lexicales. Ainsi la morphologie ne s'intéressera pas seulement à la description des règles de la formation des mots, les règles qui régissent la combinaison des morphèmes pour constituer les mots, mais aussi à la description des formes que ces mots pourraient prendre selon la catégorie de nombre, de genre, de personne, de temps, et selon le cas suite à une modification phonique.

Nous remarquons que la morphologie et la phonologie d'une langue sont si étroitement liées qu'elles donnent naissance à une nouvelle discipline qui s'occupe de l'étude de l'emploi, en morphologie d'une langue, des moyens phonologiques, c'est la morphophonologie (ou la morphonologie), elle étudie la structure phonologique des morphèmes, les modifications combinatoires dans les groupes des morphèmes et la mutation phonique qui joue un rôle morphologique. Nous cherchons à savoir à quel niveau se situent les modifications des catégories grammaticales dues aux modifications de la quantité vocalique en arabe standard ou dialectal de Mayadin, c'est-à-dire le passage d'une voyelle brève à une voyelle longue ou inversement au sein du même contexte phonique.

En ce qui suit, nous essayerons de donner quelques aspects de l'influence de la quantité vocalique sur la morphologie de l'arabe.

Dans le processus de génération des unités lexicales à partir d'autres unités qui présentent, les unes par rapport aux autres, une seule différence se situant sur le plan phonique; nous avons pu discerner, **deux catégories** de rapports entre les deux termes des paires de mots:

1- le champ sémantique est différent: la première catégorie est caractérisée par un passage d'une unité lexicale à une autre sans qu'il y ait entre elles une motivation sémantique commune, les deux unités n'appartiennent pas au même champs sémantique, au sein de cette même catégorie, on peut distinguer:

A) les 2 unités appartiennent à 2 catégories syntaxiques différentes, la prolongation de la durée de la brève permet le passage:

A-a) d'un verbe à un adjectif:

sud: l'impératif du verbe de sens " fermer hermétiquement, murer " pour la 2e personne de masculin singulier.

su:d: adjectif de sens " noir " masculin pluriel.

On remarque que le passage d'une unité lexicale à une autre est accompagné par une modification de la catégorie syntaxique.

A-b) d'un adjectif à un substantif:

ħur (libre) ~ *ħu:r* (nymphe).

B- les deux unités appartiennent à la même catégorie syntaxique, la prolongation de la durée de la brève permet le passage:

B-a) d'un verbe à un verbe:

zar (boutonner) ~ *za:r* (visiter).

B-b) d'un substantif à un substantif:

sin (dent) ~ *si:n* (lettre numéro 12 de l'alphabet arabe).

2- le champ sémantique est le même: la 2e catégorie est caractérisée par la présence d'une motivation sémantique entre les deux unités lexicales, la structure dérivationnelle de cette catégorie fait en sorte que le thème sémantique peut recevoir, par la modification de la quantité, de différentes nuances qui se rajoutent à sa valeur fondamentale, le champ sémantique est, cependant, le même.

A) d'une catégorie syntaxique à une autre:

A-a) d'un substantif à un adjectif:

ħar (chaud- substantif) ~ *ħa:r* (chaud- adjectif).

madʒid (gloire) ~ *madʒi:d* ou *ma: dʒid* (glorieux).

B) au sein de la même catégorie syntaxique:

B-a) d'un verbe à un verbe:

qatala (tuer) ~ *qa:tala* (lutter contre, faire la guerre à).

sabaqa (dépasser, devancer) ~ *sa:baqa* (lutter de vitesse, chercher à devancer).

B-b) d'un substantif à un substantif:

rʕai (pacage) ~ ra:ʕi (berger).

ħariq (brûlure) ~ ħari:q (incendie).