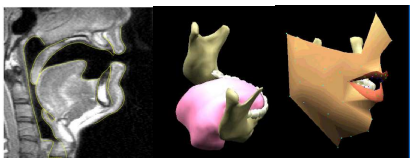
Parole d'homme – Parole de cloneVers une machine parlante anthropomorphique : Données et modèles en production de parole

Pierre Badin

Institut National Polytechnique de Grenoble

Parole d'homme – Parole de clone

Vers une machine parlante anthropomorphique : Données et modèles en production de parole.



HABILITATION A DIRIGER DES RECHERCHES

Présentée le 18 avril 2002

par

Pierre Badin

Ingénieur ENSERG – Docteur-Ingénieur INPG

Devant le Jury composé de :

Monsieur	Bernard GUERIN	Président
Monsieur	Louis-Jean BOË	Rapporteur
Monsieur	Bernard DUBUISSON	Rapporteur
Monsieur	Kevin MUNHALL	Rapporteur
Monsieur	Shinji MAEDA	Examinateur
Monsieur	Joseph MARIANI	Examinateur
Monsieur	Jean-Luc SCHWARTZ	Examinateur

Table des matières

Habilitation à diriger des recherches – Résumé

PAROLE D'HOMME – PAROLE DE CLONE

VERS UNE MACHINE PARLANTE ANTHROPOMORPHIQUE :

DONNEES ET MODELES EN PRODUCTION DE PAROLE.

PIERRE BADIN

ETAT CIVIL

- Nom : BADIN, Pierre, Léon, Fernand
- Naissance : 15 décembre 1955, Nîmes, Gard, France
- Adresse personnelle : Apt. 2305, 60 Place des Géants, 38100 Grenoble, France
- Situation familiale : Marié - Un enfant

SITUATION PROFESSIONNELLE ACTUELLE

- Emploi : Chargé de Recherche 1ère classe au CNRS, à l'*Institut de la Communication Parlée*
- Adresse professionnelle : ICP, 46 avenue Félix Viallet, 38031 Grenoble Cedex, France
- Tél. : 04 76 57 48 26 – Fax : 04 76 57 47 10
- Mél : badin@icp.inpg.fr – Toile : http://www.icp.inpg.fr/~badin

DIPLOMES

- Docteur-Ingénieur INPG (Mars 1983)
- DEA d'Electronique et Radiocommunication de l'ENSERG-INPG (Juin 1980)
- Ingénieur Radioélectricien de l'ENSERG-INPG (Juin 1979)
- Baccalauréat, Série C, Mention Assez Bien (Juin 1973)

FORMATION

- Septembre 1980 – mars 1983 : Doctorant de l'école *Image Signal Parole* de l'INPG, à l'Institut de la Communication parlée (Bourse de la DGRST) (Sujet: Analyse de la Parole, Synthèse à formants)
- Septembre 1979 – juin 1980 : DEA Image Signal Parole de l'INPG, Institut de la Communication parlée
- Septembre 1976 – juin 1979 : Elève-Ingénieur à l'ENSERG-INPG
- Septembre 1973 – juin 1976 : Classes préparatoires au Lycée Thiers, Marseille

CARRIERE

- Octobre 1989 – présent : Chargé de recherche 1ère classe au CNRS, ICP
- Octobre 1984 – septembre 1989 : Attaché de recherche 2e classe au CNRS, ICP
- Avril 1984 – septembre 1984 : assistant de recherche au *Kungliga Tekniska Högskolan* (KTH), Stockholm, Suède
- Avril 1983 – mars 1984 : Stage post-doctoral au KTH, Stockholm, Suède, avec une bourse de recherche de l'INRIA

SEJOURS A L'ETRANGER

- Mars 1996 : Invité par la JSPS (*Japan Society for the Promotion of Science*) pour un projet sur l'imagerie du conduit vocal par IRM aux laboratoires *ATR* (*Advanced Telecommunication Research*), Kyoto, Japan, et au *Nara General Hospital.*
- Août 1994 : chercheur invité à l'*Institute for Applied Electricity*, *Hokkaido University*, Sapporo, Japon, et à la *Faculty of Engineering*, *Hokkai-Gakuen University*, Sapporo, Japon.
- Juillet 1993 : chercheur invité à l'*Institute for Applied Electricity*, *Hokkaido University*, Sapporo, Japon.
- Décembre 1990 – Janvier 1991 : chercheur invité à l'*Institute for Applied Electricity*, *Hokkaido University*, Sapporo, Japon
- Mars 1988 – Août 1989 : séjour sabbatique au KTH, Stockholm, Suède

ADMINISTRATION DE LA RECHERCHE

❖ *ADMINISTRATION*

- Animateur de l'équipe *Acoustique* à l'ICP (1990 - présent)
- Membre du Conseil de Laboratoire de l'ICP (1990 - présent)
- Responsable de la bibliothèque de l'ICP/INPG (1986 – présent)

❖ *ENCADREMENT SCIENTIFIQUE*

- Directeur ou co-directeur de 7 thèses doctorales (Signal, Image, Parole, Télécoms ; Sciences Cognitives ; Sciences du Langage), 11 stages de DEA, 14 stages d'Ingénieurs
- Membre des jurys de thèse de Denis Beautemps (1993), Pham Thi Ngoc (1995), Mawass (1997), Rossato (2000), Vilain (2000)
- *Opposant* de Mats Båvegård à la soutenance de sa *licenciate thesis*, KTH, Stockholm, Suède (1996)

❖ *CONTRATS*

- Co-responsable avec Louis-Jean Boë du projet "Une Tête Parlante Virtuelle : Données et modèles en production de parole" financé par l'ARASSH (Agence Rhône-Alpes pour les Sciences Sociales et Humaines) (1997-1999).
- Coordinateur (avec Christian Abry) du projet européen ESPRIT/BR *Speech Maps* N°6975 *Sound-to-Gesture Inversion in Speech* (1992-1995).
- Co-responsable (avec Bernard Guérin, à l'ICP) de l'action européenne SCIENCE *SC1*0147 C (EDB)*: *Measure, characterisation and modelling of fricative consonants* (1989-1991).
- Coordinateur du "European Laboratory Network", financé par le Ministère de la Recherche et de la Technologie, *Base de données articulatoires et acoustiques pour la parole* (5 membres, France, UK, Suède)

❖ *ORGANISATION DE MANIFESTATIONS SCIENTIFIQUES*

- Co-organisateur (avec Gérard Bailly) des *XXIII*èmes *Journées d'Etude sur la Parole* (Aussois, France, juin 2000)
- Membre du comité scientifique du *Hokkaido Workshop on Speech Production* (Kutchan, Japon, 1998)
- Membre du comité scientifique du séminaire *ETRW / 4th Speech Production Seminar* (Autrans, France, 1996)
- Co-organisateur of the *1*st *Seminar on Speech Production* (Grenoble, 1988)

❖ *EDITION SCIENTIFIQUE*

- Co-fondateur / co-éditeur des *Cahiers de l'ICP* (1990-présent) : *Bulletin de la Communication Parlée*, revue francophone sur la parole avec comité de lecture (5 numéros), *Rapports de Recherche de l'ICP* (7 numéros), *Monographies* (1 numéro)
- Relecteur pour le *Journal of the Acoustical Society of America*, le *Journal of Phonetics*, *Speech Communication*, *IEEE transactions on Speech and Audio Processing*

❖ *PARTICIPATION AUX MANIFESTATIONS SCIENTIFIQUES*

- *Seminar on Speech Production, Data and Models* (Grenoble, 1988 ; Leeds, UK, 1990 ; New Haven, USA, 1993 ; Autrans, France, 1996 ; Munich, Allemagne, 2000)
- *International Congress of Phonetic Sciences* (Tallinn, Estonie, 1987; Aix-en-Provence, 1991 ; Stockholm, Suède, 1995)
- *International Conference on Spoken Language Processing* (Yokohama, Japon, 1994 ; Philadelphie, USA, 1996)
- *IEEE International Conference on Acoustics, Speech and Signal Processing* (Paris, 1982)
- *EuroSpeech* (Paris, 1989 ; Rhodes, Greece, 1997)
- *Meeting of the Acoustical Society of America* (Ottawa, Canada, 1993)
- *Hokkaido Workshop on Speech Production* (Kutchan, Japon, 1998)
- *ESCA Tutorial and Research Workshop on Speech Technology in Language Learning* (Stockholm, Suède, 1998)
- *EUROCALL'99, Satellite workshop on "Speech Technology applications in CALL"* (Besançon, 1999)

RESUME DES TRAVAUX DE RECHERCHE

PAROLE D'HOMME – PAROLE DE CLONE

VERS UNE MACHINE PARLANTE ANTHROPOMORPHIQUE :

DONNEES ET MODELES EN PRODUCTION DE PAROLE.

❖ *LA PAROLE, UN SIGNAL BIOLOGIQUE DE COMMUNICATION*

Le signal de parole est un signal destiné à la *communication orale* entre humains, et donc à *encoder* des *messages linguistiques*. Il possède un certain nombre de propriétés qui en font un type de signal très particulier. C'est un signal produit par un système *biologique*, l'appareil phonatoire humain, et qui reflète donc les propriétés biomécaniques des articulateurs. C'est un signal *audiovisuel*, puisqu'il fait simultanément intervenir le son et l'image du visage du locuteur, pour ne pas mentionner le toucher. C'est un signal *redondant*, aussi bien au niveau du son qu'au niveau de la complémentarité entre les canaux acoustiques et visuels, ce qui lui confère des qualités de robustesse indispensables à un signal de communication. Son degré de redondance est *adaptable* en fonction des conditions environnementales de bruit et de la quantité d'information contenue dans le message à transmettre (liée en particulier au degré de prédictibilité). Cette adaptabilité en fait un signal très *variable*.

Ainsi, le signal de parole est extrêmement complexe du point de vue de sa structure, mais cette complexité peut être lue et interprétée plus facilement si l'on fait référence aux *gestes* des articulateurs qui l'ont produit. Les mécanismes de production de parole font intervenir la coordination des gestes des différents articulateurs – mâchoire, langue et lèvres – qui modulent la forme du conduit vocal et du visage au cours du temps ; les sources d'excitation acoustiques générées par l'écoulement de l'air issu des poumons à travers le conduit vocal sont alors filtrées par les résonances de ce conduit et finalement rayonnées vers l'extérieur. Depuis mon arrivée à l'ICP en 1979, mon travail de recherche a été essentiellement consacré, selon une *approche anthropomorphique*, à modéliser les signaux de parole en tant que conséquences de ces mécanismes biomécaniques et aéroacoustiques qui se produisent dans le conduit vocal humain.

❖ *DONNEES, MODELES, ET TETE PARLANTE AUDIOVISUELLE*

Notre principale approche en modélisation consiste à développer des modèles fonctionnels à partir de données expérimentales, et, dans une moindre mesure, à mettre en œuvre des modèles physiques basés sur des théories pré-établies, en les confrontant aux données. Ainsi, dans tous les cas, modèles et données jouent des rôles fondamentaux et complémentaires.

Données acoustiques et articulatoires – dispositifs expérimentaux. Nous avons utilisé ou développé un certain nombre de techniques expérimentales de mesure de paramètres liés à la production de la parole : banc de mesure de la fonction de transfert acoustique du conduit vocal, masque pneumotachométrique pour la mesure de l'écoulement et des pressions de l'air dans le conduit vocal,

cinéradiographie et articulographie électromagnétique pour l'étude du mouvement, imagerie IRM pour la caractérisation tridimensionnelle des articulateurs, vidéo pour les mesures tridimensionnelles de lèvres et de visage. Un ensemble précieux de données articulatoires et acoustiques complémentaires a ainsi été recueilli, sur quelques sujets de référence prononçant, dans des conditions maîtrisées, les mêmes corpus représentatifs de l'ensemble des articulations de la langue. Cette démarche *orientée sujet* offre ainsi la possibilité de disposer, pour le même phénomène (un sujet et une articulation), de données qui ne peuvent être acquises qu'avec des dispositifs expérimentaux impossibles à mettre en œuvre au cours d'une même expérience, comme par exemple la cinéradiographie et le masque pneumotachographique.

Modèles articulatoires et acoustiques. Nous avons ainsi développé des modèles articulatoires linéaires de conduit vocal, de langue ou de velum, médiosagittaux ou tridimensionnels, pilotés par les degrés de liberté articulatoires extraits par analyse en composantes linéaires des données. Des degrés de liberté tout à fait similaires ont pu être identifiés pour les différents locuteurs, même si ces locuteurs utilisent des stratégies de contrôle parfois assez différentes. La décomposition selon ces degrés de liberté des gestes articulatoires présents dans certaines séquences Voyelle – Consonne – Voyelle (VCV) a dévoilé des stratégies de compensation entre articulateurs qui n'auraient pas été lisibles directement sur les contours sagittaux bruts. Des stratégies de synergies entre langue et mâchoire ont également pu être mises en évidence. Par ailleurs, nous avons mis en œuvre un ensemble de modèles d'écoulement d'air, de sources acoustiques de voisement et de bruit de friction, et de propagation et rayonnement acoustique dans les domaines temporels et/ou fréquentiels. Nous avons ainsi pu étudier la coordination précise des gestes glotte / constriction orale nécessaire à la production des consonnes fricatives, en liaison avec les interactions entre sources et conduit vocal.

Tête parlante audiovisuelle et synthèse articulatoire. Nous avons intégré les modèles mentionnés ci-dessus dans un *robot articulatoire anthropomorphique* : une *tête parlante*. Cette tête parlante est donc contrôlée par des paramètres articulatoires supra-laryngés qui pilotent le modèle articulatoire et par des paramètres de contrôle glottique qui déterminent les sources acoustiques en interaction avec le conduit vocal ; elle est finalement capable de fournir un signal audio-visuel de parole cohérent. Nous avons par ailleurs développé des procédures d'inversion, basées sur le concept de *robotique de la parole*, qui nous ont permis de reconstruire avec une bonne fiabilité les trajectoires des paramètres de contrôle articulatoire à partir de l'acoustique, même si ce problème d'inversion est un problème mal posé *a priori*. Nous avons ainsi pu réaliser une synthèse articulatoire de séquences VCV contenant les fricatives du français[1].

❖ *PERSPECTIVES*

D'un côté, il sera nécessaire de poursuivre le développement et l'amélioration des différents modèles qui constituent la tête parlante. D'autre part, le temps est venu de nous tourner de manière plus approfondie dans le cadre du développement des STIC (*Sciences et Technologies de l'Information et de la Communication*) et du 6e *Programme cadre européen de recherche et de développement technologique* européen, vers des applications comme la synthèse articulatoire audiovisuelle, les clones pour les télécommunications, ou encore l'aide à l'apprentissage des langues.

Données et modèles en production de parole. Le développement de la tête parlante continuera à être basé sur des données expérimentales, l'objectif étant de modéliser tous les articulateurs, afin de générer des fonctions d'aire tridimensionnelles complètes. L'approche de modélisation *linéaire* sera conservée, en explorant ses limites, mais sans exclure des modèles locaux non-linéaires capables de prendre en compte la déformation des organes qui entrent en contact les uns avec les autres. Cette approche *orientée sujet* sera par ailleurs étendue à plusieurs locuteurs afin de comparer les stratégies individuelles, et d'en tirer des principes plus généraux. La nécessaire normalisation inter-sujets sera explorée à deux niveaux : conformation anatomique, et stratégies de synergie / compensation articulatoires. Les modèles aérodynamiques et acoustiques devront être développés pour prendre en compte les modes transversaux nécessaires pour les consonnes fricatives, le couplage avec les cavités nasales pour les voyelles et consonnes nasales, et la génération des bruits de relâchement pour les consonnes occlusives. Par ailleurs, nous explorerons les degrés de liberté des articulateurs en relation avec l'anatomie, et nous déterminerons les *espaces de réalisation* des différents phonèmes sous forme d'*espaces de réalisation de cibles spatio-temporelles* aux

[1] Voir des exemples de synthèse sur les sites http://www.icp.inpg.fr/~badin/ActaAcustica_Sounds.html et http://www.icp.inpg.fr/~badin/VTH-SPS5.html

niveaux articulatoire, géométrique, aérodynamique, et acoustique, pour différentes conditions d'élocution, ce qui nous permettra d'aborder l'étude de la variabilité de la parole.

Têtes parlantes et applications. Un certain nombre d'applications des têtes parlantes peuvent être envisagées. L'un des intérêt de la tête parlante réside dans la possibilité de *réalité augmentée* qu'elle offre : en affichant la peau et certains articulateurs de manière semi-transparente, ou en utilisant des techniques d'*écorché*, il est possible de montrer des articulateurs cachés dans des conditions normales d'élocution. L'*apprentissage de la prononciation des langues étrangères* pourrait bénéficier de ces propriétés : en effet montrer à un apprenant les mouvements articulatoires qu'il doit effectuer pour produire un son fait partie des stratégies pédagogiques intéressantes ; il sera donc nécessaire d'évaluer la tête parlante à ce niveau, en déterminant les modes de présentation les plus efficaces. De manière similaire, nous envisageons d'utiliser la tête parlante dans le cadre de la *réhabilitation des déficients auditifs*. Par ailleurs, la tête parlante et l'ensemble des données articulatoires et acoustiques qui ont été progressivement accumulées permettent d'envisager le développement d'un système de *synthèse articulatoire audiovisuelle à partir du texte*. Enfin, dans le domaine des télécommunications, il sera possible à tout locuteur auquel un *clone* aura été adapté à partir d'un clone générique d'intervenir dans une visioconférence par l'intermédiaire de ce clone, avec les avantages d'une réduction considérable de la bande passante nécessaire à l'image et d'une représentation complète tridimensionnelle de la tête du locuteur.

PUBLICATIONS

❖ *PUBLICATIONS (113)*

- Revues internationales avec comité (10)
- Ouvrages ou chapitres dans un ouvrage (2)
- Colloques internationaux avec comité (45)
- Colloques internationaux sans comité (12)
- Colloques nationaux avec comité (8)
- Colloques nationaux sans comité (4)
- Rapports d'activité (8)
- Divers (7)
- Rapports de contrats (15)
- Vulgarisation (2)

❖ *PRINCIPALES PUBLICATIONS*

Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C. & Savariaux, C. (In press). Three-dimensional articulatory modeling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*.

Beautemps, D., Badin, P. & Bailly, G. (2001). Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. *Journal of the Acoustical Society of America, 109*(5), 2165-2180.

Mawass, K., Badin, P. & Bailly, G. (2000). Synthesis of French fricatives by audio-video to articulatory inversion. *Acta Acustica, 86*(1), 136-146.

Badin, P., Bailly, G. & Boë, L.-J. (1998). Towards the use of a Virtual Talking Head and of Speech Mapping tools for pronunciation training. In *Proceedings of the ESCA Tutorial and Research Workshop on Speech Technology in Language Learning*, pp. 167-170. Stockholm, Sweden, May 1998. ESCA and Dept. Speech, Music and Hearing, KTH, Stockholm.

Badin, P., Beautemps, D., Laboissière, R. & Schwartz, J.-L. (1995). Recovery of vocal tract geometry from speech signal for vowels and fricative consonants using a midsagittal-to-area function conversion model. *Journal of Phonetics, 23*, 221-229.

Badin, P., Motoki, K., Miki, N., Ritterhaus, D. & Lallouache, T.M. (1994). Some geometric and acoustic properties of the lip horn. *Journal of the Acoustical Society of Japan (English), 15*(4), 243-253.

Abry, C., Badin, P. & Scully, C. (1994). Sound-to-gesture inversion in speech: The *Speech Maps* approach. ESPRIT Research Report No. 6975. In *Advanced Speech Applications* (K. Varghese, S. Pfleger & J.P. Lefèvre, Eds.), pp. 182-196. Berlin: Springer Verlag.

Badin, P. (1991). Fricative consonants: acoustic and X-ray measurements. *Journal of Phonetics, 19*, 397-408.

Badin, P., Perrier, P., Boë, L.-J. & Abry, C. (1990). Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *Journal of the Acoustical Society of America, 87*, 1290-1300.

Badin, P. (1989). Acoustics of voiceless fricatives: production theory and data. *Speech Transmission Laboratory - Quarterly Progress Status Report - Stockholm, 3/1989*, 33-55.

Badin, P. & Fant, G. (1984). Notes on vocal tract computation. *Speech Transmission Laboratory - Quarterly Progress Status Report - Stockholm, 2-3/1984*, 53-108.

Une …

… deux …

… trois minutes de silence à la mémoire des nombreuses victimes de la barbarie (in)humaine …

Avant-propos

Commencé il y a très longtemps, plusieurs fois abandonné, remis sur le métier, ré-abandonné, ce mémoire a enfin vu le jour grâce aux pressions répétées, mais néanmoins amicales, de mes chers collègues et amis de l'ICP, et au soutien affectueux de mes japonaises et de ma famille … À vous toutes et à vous tous, MERCI !

Sous l'influence de certains, j'ai désappris à compter les moutons. Je ne ferai donc pas de listes de remerciements, toujours injustes par les oublis inévitables. Que toutes celles et tous ceux qui m'ont aidé, d'une manière ou d'une autre, depuis mon arrivée à l'ICP, en dirigeant le laboratoire, en faisant de la recherche, en me conseillant, en me remettant en cause, en m'encourageant, en entretenant les machines et les locaux, en préparant le café que je ne bois pas, en résolvant les problèmes administratifs, en discutant avec moi dans les couloirs, en faisant des thèses ou des stages, français ou étrangers, … et en relisant ce mémoire, se retrouvent également remerciés ici du fond de mon cœur !

Première partie : Synthèse des travaux de recherche

Parole d'homme – Parole de clone

Vers une machine parlante anthropomorphique : Données et modèles en production de parole.

I. CADRE SCIENTIFIQUE ET MOTIVATIONS

Le langage, ou plus précisément la parole, constitue depuis des temps immémoriaux le moyen privilégié de communication entre les êtres humains. Même si le débat reste ouvert, la parole constitue vraisemblablement, de manière ultime, la faculté qui permet de distinguer l'homme des autres espèces animales. Je suis convaincu que cela constitue la raison profonde de l'intérêt constant porté à l'étude de la parole depuis des siècles.

Depuis mon arrivée à l'ICP en avril 1979, mes travaux de recherche sont essentiellement centrés sur l'étude et la compréhension des phénomènes qui interviennent dans la production de la parole.

A. La parole, un signal biologique de communication

Le signal de parole est un signal destiné à la *communication orale* entre humains, et donc à *encoder* des *messages linguistiques*. Il possède un certain nombre de propriétés qui en font un type de signal très particulier. C'est un signal produit par un système *biologique*, l'appareil phonatoire humain, et qui reflète donc les propriétés biomécaniques des articulateurs. C'est un signal *audiovisuel*, puisqu'il fait simultanément intervenir le son et l'image du visage du locuteur, pour ne pas mentionner le toucher. C'est un signal *redondant*, aussi bien au niveau du son qu'au niveau de la complémentarité entre les canaux acoustiques et visuels, ce qui lui confère des qualités de robustesse indispensables à un signal de communication. Son degré de redondance est *adaptable* en fonction des conditions environnementales de bruit et de la quantité d'information – liée en particulier au degré de prédictibilité – contenue dans le message à transmettre. Cette adaptabilité en fait un signal très *variable*.

Partant de ces considérations, l'une des voies les plus productives pour l'acquisition de connaissance sur la parole paraît être l'étude des phénomènes qui président à sa production. En effet, la parole est un signal extrêmement complexe du point de vue de sa structure acoustique, mais cette complexité peut être lue et interprétée plus facilement si l'on fait référence aux *gestes* des articulateurs qui suffisent à produire ce signal. Cette approche rénovée de la production de la parole a été relancée et confortée depuis une dizaine d'années dans le cadre de la *robotique de la parole*, notion développée en particulier au cours du projet européen *Speech Maps* (Abry, Badin & Scully (1994)).

La chaîne des processus de production de la parole

La Figure 1 schématise les divers niveaux de transformation qui interviennent lors de la génération des signaux de parole audiovisuels à partir du code linguistique que le locuteur souhaite transmettre. Un message conçu au niveau du cerveau génère des influx nerveux qui activent les divers muscles de l'appareil vocal humain. Les gestes coordonnés des différents articulateurs – mâchoire, langue et lèvres – font évoluer la forme du conduit vocal au cours du temps, en même temps que ses sources d'excitation, afin de générer et moduler le son. La parole produite est finalement un signal audio-visuel destiné à être perçu par un interlocuteur dans ses modalités à la fois auditive et visuelle. Je présente très brièvement ci-après les différents maillons de cette chaîne.

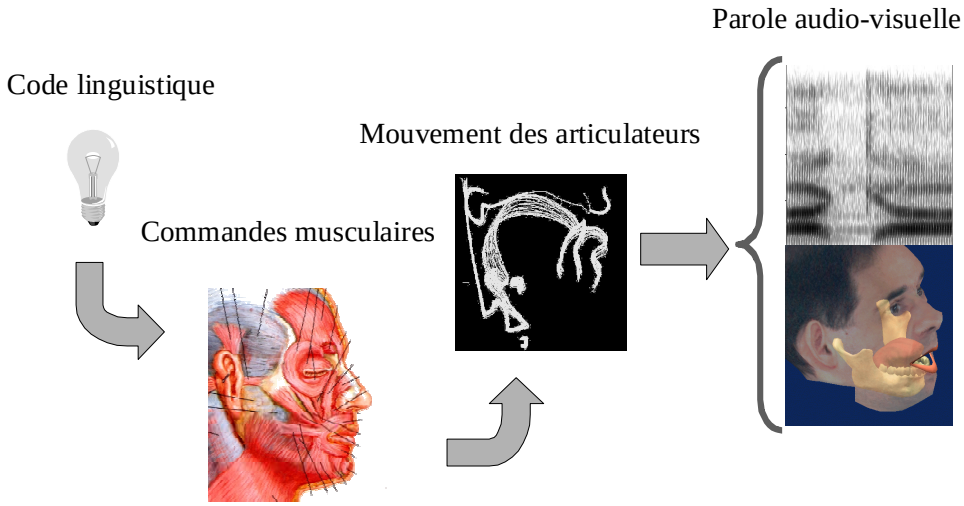


Figure 1 : Schéma de principe des processus de production de la parole

Du code linguistique aux commandes nerveuses. La façon dont le système nerveux central contrôle l'appareil phonatoire fait toujours l'objet d'un large débat (*cf.* le *Bulletin de la Communication Parlée N°4*, Abry & Badin (1998), numéro spécial de commentaires sur l'article cible de Perrier, Ostry & Laboissière (1996) "The Equilibrium Point Hypothesis and its application to speech motor control"). Le modèle de control moteur que Perrier *et al.* (1996) défendent pour la parole, *i.e.* le *modèle* λ, repose sur l'*Hypothèse du Point d'Équilibre* selon laquelle le mouvement des articulateurs résulte de changements des signaux de contrôle nerveux qui ont pour effet de déplacer l'état d'équilibre du système moteur. Les changements induits par le système nerveux sur les potentiels de membrane des neurones moteurs qui innervent les muscles central constituent donc les véritables paramètres de contrôle. L'équilibre du système moteur résulte finalement des interactions entre les commandes nerveuses centrales, les mécanismes réflexes, les propriétés des muscles et les charges externes. Plus précisément, les forces générées par les muscles sont dues aux variations de polarisation des neurones moteurs qui changent le seuil de longueur musculaire λ à partir duquel les neurones moteurs sont recrutés.

Biomécanique des articulateurs. Les forces générées par la contraction des fibres musculaires induite par les neurones moteurs agissent sur l'ensemble des articulateurs qui réagissent suivant les lois de la mécanique classique. Le système articulatoire affiche ainsi des propriétés dynamiques qui dépendent des masses des articulateurs mis en jeu, mais aussi des forces et de divers types de pertes. Les connaissances en biomécanique de la parole commencent à se développer considérablement, comme en témoignent les modèles de plus en plus nombreux qui tiennent compte de ces propriétés dynamiques (Payan, Perrier & Laboissière (1995), Wilhelms-Tricarico & Perkell (1995), Laboissière, Ostry & Feldman (1996)).

Articulateurs et géométrie du conduit vocal. La position de l'ensemble des articulateurs *supra laryngés* permet de définir la forme complète du conduit vocal, c'est-à-dire les dimensions précises de chacune des cavités qui participent à la production des sons. Cette géométrie peut être décrite de manière plus ou moins complète, soit en spécifiant l'ensemble des limites du conduit vocal (description entièrement tridimensionnelle), soit par l'utilisation d'une fonction d'aire (représentation pseudo tridimensionnelle) ou encore simplement par le contour de la coupe médiosagittale du conduit vocal (représentation bidimensionnelle).

Aéroacoustique du conduit vocal et sources d'excitation acoustique. Le processus de production des sons dans le conduit vocal repose sur le principe de la transformation de l'énergie mécanique fournie par les poumons, sous forme d'un certain débit d'air à une certaine pression, en une énergie acoustique qui est finalement rayonnée. Plus précisément, l'écoulement de l'air dans le conduit vocal assure deux fonctions : il permet d'une part de transformer l'écoulement de l'air en excitation acoustique suivant plusieurs modes, et d'autre part il coordonne entre elles les différentes sources acoustiques ainsi créées, comme nous le verrons plus loin. Trois types de sources d'excitation acoustique peuvent être considérés dans des conditions normales de production de parole.

Excitation voisée. Si les cordes vocales sont suffisamment rapprochées, et si la différence de pression à laquelle elles sont soumises est suffisamment élevée, elles entrent alors dans un mouvement d'oscillations de relaxation approximativement périodique qui module le débit d'air qui traverse la glotte et fournit une source de voisement (voir par exemple Pelorson, Hirschberg, Van Hassel, Wijnands & Auregan (1994), Pelorson, Hirschberg, Wijnands & Bailliet (1995)). La forme de l'onde ainsi générée, et par conséquent le spectre et la fréquence fondamentale, dépendent donc d'un certain nombre de paramètres tels que la chute de pression moyenne à travers la glotte, la masse et la tension des cordes vocales, et leur position au repos. Si l'écartement au repos est trop élevé, les vibrations cessent. Si l'écartement est suffisant pour que les cordes vibrent, mais que leur tension est trop forte, elles ne vibrent pas non plus. Les paramètres de contrôle des cordes vocales doivent donc se situer dans un espace limité pour assurer le voisement (*cf.* Laboissière & Pelorson (1995), ou Abry, Badin, Mawass & Pelorson (1998)) : le locuteur joue sur ces dimensions pour obtenir l'opposition phonologique « voisé / sourd » par exemple.

Excitation bruitée. Lorsqu'une constriction, c'est-à-dire un resserrement des parois, se forme en un point du conduit vocal, l'écoulement de l'air, laminaire en amont de la constriction, se transforme en jet plus ou moins turbulent à la sortie de celle-ci. Ce jet génère des sources de bruit acoustiques dans le conduit vocal. On parle de bruit d'*aspiration* lorsque le phénomène se produit à la glotte, comme dans les voyelles chuchotées par exemple, et de bruit de *friction* dans les autres cas. Shadle (1990) distingue trois types de génération de bruit : (1) les sources de surface dues à l'écoulement de l'air le long d'une paroi du conduit vocal (exemple : [x]), (2) les sources d'obstacle créées par la brisée du jet sur un obstacle tel que les incisives (exemple : [s]), et (3) les sources de jet libre dues à la turbulence seule, et qui rayonnent extrêmement peu d'énergie acoustique (exemple : [ɸ]).

De manière similaire à la source vocale, les conditions de génération des sources de friction dépendent de l'état de l'écoulement de l'air dans le conduit vocal (*cf.* Hixon (1966), Stevens (1971), ou Badin (1989)), et en outre de sa géométrie (Shadle (1986)). Ainsi, le bruit de friction nécessite une chute de pression à travers la constriction suffisamment élevée, et une aire de constriction suffisamment faible. Ces mécanismes de coordination entre source vocale et source de bruit seront présentées plus précisément à la section II.C.1.

Excitation de relâchement. Lors d'une fermeture totale du conduit vocal, la pression en amont de l'occlusion s'élève jusqu'à équilibre avec la pression des poumons. Le relâchement plus ou moins brutal de cette occlusion produit une variation transitoire importante de l'écoulement (Pelorson (2000)), qui est à l'origine d'une source acoustique cohérente de type impulsionnel (Maeda (1985)). En utilisant des modèles très simplifiés, Maeda (1985) et Stevens (1993) ont montré que les caractéristiques spectrales et temporelles de cette source acoustique dépendent en particulier de la vitesse de relâchement et de la longueur de l'occlusion au moment du relâchement.

Propagation acoustique dans le conduit vocal et rayonnement. Les sources que je viens de décrire génèrent à leur tour des ondes acoustiques qui se propagent dans le conduit vocal, qui sert de guide d'onde et joue le rôle d'un résonateur qui rehausse l'amplitude de certaines fréquences et en affaiblit d'autres. Aux basses fréquences, on considère traditionnellement que seules des ondes planes se propagent (*cf.* Fant (1960), Portnoff (1973), ou Badin & Fant (1984)). Cependant, comme dans tous les guides d'ondes, il existe une fréquence à partir de laquelle des ondes transverses se propagent également : cette fréquence, qui dépend naturellement des dimensions du conduit vocal, se situe aux alentours de 4-5 kHz (El Masri, Pelorson, Saguet & Badin (1998)). L'importance de ces ondes transverses pour la production de la parole n'est pas encore clairement établie, et fait encore l'objet de recherches. L'objet de l'acoustique du conduit vocal est de prédire, à partir de la forme du conduit, la fonction de transfert acoustique entre source et lèvres, et en particulier les *formants* que l'on définit comme les résonances du conduit vocal.

Le rayonnement des ondes acoustiques établies dans le conduit vocal constitue le processus final de la production de la parole. Ce rayonnement se produit principalement au niveau des lèvres et des narines, mais aussi à travers les parois du conduit, en particulier au niveau du pharynx et des joues (Fant, Nord & Branderud (1976)).

B. Les principes scientifiques préalables

Avant de décrire l'ensemble de mes travaux, il est indispensable de présenter les principes scientifiques qui ont sous-tendu mon approche de l'étude de la production de la parole. Les trois concepts clé de cette recherche sont la *modélisation*, l'*anthropomorphisme*, et les *données expérimentales*.

1. Modéliser pour comprendre et connaître

La modélisation constitue, avec la description méthodique, la méthode par excellence permettant de formuler, et ensuite d'évaluer, notre connaissance d'un phénomène donné. Par la nécessité de formalisation du phénomène étudié qu'elle impose, la modélisation incite à poser les problèmes de manière claire, en particulier en mettant en évidence les paramètres pertinents, tout en ouvrant la voie à la possibilité de *falsification*, c'est-à-dire en permettant de confronter ses prédictions à la réalité mesurable, et ainsi de mettre à l'épreuve le modèle. Modéliser permet ainsi de mieux comprendre, de mieux connaître, et aussi finalement de mieux maîtriser.

C'est ainsi qu'afin de faire progresser notre compréhension du fonctionnement de la production de la parole, je me suis placé dans l'optique du *faire comme*, c'est-à-dire modéliser de manière *physique* ou *fonctionnelle* les divers phénomènes et leurs interactions, plutôt que dans celle du *faire semblant* qui vise à simuler plutôt les signaux eux-mêmes que leurs causes.

2. Modélisation physique et modélisation fonctionnelle

La *modélisation fonctionnelle* permet de rendre compte d'un phénomène de manière globale. Cette approche renvoie au concept de *boîte noire* : on établit les relations entre les paramètres supposés être les paramètres de contrôle du phénomène (les entrées de la boîte noire) et les paramètres qui en sont les conséquences (les sorties de la boîte noire). Même s'il peut s'avérer extrêmement efficace dans des applications technologiques par exemple, ce type de modélisation piloté par le résultat n'apporte pas en retour de connaissance sur le phénomène lui-même. Prenons l'exemple de la relation articulatori-acoustique. Si nous nous intéressons à l'effet de la position de la mâchoire sur les fréquences de résonance du conduit vocal, il est envisageable d'entraîner un système d'interpolation (réseau de neurones, régression multilinéaire, interpolation polynomiale, fonctions de base radiales, etc.) sur un ensemble de données expérimentales acquises pour un sujet donné, et de construire ainsi une boîte noire qui simule le lien entre mâchoire et formants. Quelle est la nature de la connaissance du phénomène ainsi acquise ? Est-t-il possible d'extrapoler le modèle à des données non présentes au départ ? Comment interpréter le phénomène ?

À l'inverse, la *modélisation physique* correspond à une analyse du processus lui-même. Il faut d'abord savoir quelle est l'influence de la position de la mâchoire sur la forme du conduit vocal. Pour répondre à cette question, on peut être amené à établir un modèle biomécanique de la mâchoire et de la langue. Il faut ensuite établir les conséquences acoustiques et aérodynamiques de ce changement de géométrie. Dans ce cas, il s'agit de développer un modèle aéroacoustique du conduit vocal. Ainsi, la modélisation physique permet une description précise du phénomène étudié, et interprétable en terme de paramètres qui peuvent être eux-mêmes mesurés (comme par exemple la forme du conduit vocal ou la pression intra-orale).

Le choix entre modélisation fonctionnelle et modélisation physique dépend d'un équilibre entre le degré de réalisme souhaité et le degré de complexité nécessaire à la modélisation. Il doit rester clair que tous les niveaux intermédiaires existent entre la boîte la plus opaque de la modélisation fonctionnelle et la boîte la plus transparente de la modélisation physique détaillée. La modélisation physique la plus détaillée peut entraîner le chercheur très loin du champ d'investigation de départ : modéliser la structure biomécanique interne des cordes vocales par éléments finis est-t-il utile à la compréhension de la phonation si l'on sait par ailleurs que seuls les deux ou trois premiers modes de vibration sont importants pour l'excitation du conduit vocal ? Réciproquement, est-t-il suffisant de modéliser seulement les ondes planes dans le conduit vocal, quand il est bien connu que pour la plupart des consonnes la masse la plus importante d'énergie se situe aux alentours de la fréquence d'apparition des modes transverses ? Il est donc nécessaire d'avoir en permanence à l'esprit l'objet de la recherche, et d'adapter le niveau de modélisation aux besoins de compréhension, tout en évaluant le coût des suppléments de complexité. Deux exemples de modélisation dans le domaine des sources acoustiques du conduit vocal illustrent mon approche. Les modèles physiques d'interaction entre jet et paroi intervenant dans la génération des bruit de friction dans le conduit vocal étant encore trop complexes et mal connus pour qu'ils constituent une voie exploitable à court terme, j'ai opté pour un modèle fonctionnel basé sur des données, en attendant que des progrès soient réalisés dans ce domaine par les spécialistes de l'aéroacoustique. Par contre, il me

paraissait plus intéressant d'utiliser les modèles de cordes vocales, déjà amplement maîtrisés, que de mettre en œuvre des modèles paramétriques qui décrivent analytiquement la forme du signal d'onde glottique.

3. Modélisation, anthropomorphisme, et robotique de la parole

L'étude du signal de parole peut être envisagée sous deux angles complémentaires :

- une approche de type *traitement du signal*, qui ne considère dans la parole que les propriétés du signal produit, sans référence implicite ou explicite à l'origine de ces signaux,
- une *approche anthropomorphique*, qui vise à modéliser les signaux non pas en tant que tels, mais en tant que conséquences de mécanismes biomécaniques et aéroacoustiques de production.

Le traitement du signal permet de modéliser et de manipuler, et avec beaucoup de succès[1], le signal de parole, mais n'a pas pour ambition de faire avancer la connaissance sur le fonctionnement humain.

Puisque modéliser permet d'avancer dans la connaissance, et puisque les processus de production de la parole constituent mon objet d'étude central, mon approche se place délibérément dans un cadre de modélisation anthropomorphique.

À ces motivations de fond, s'ajoutent des orientations plus technologiques, qui répondent à la demande du secteur des Sciences et Technologies de l'Information et de la Communication (STIC) du CNRS. On peut penser que les systèmes artificiels pourraient ou devraient prendre leur inspiration dans la nature, comme Hermansky & Pavel (1995) le proposent en déclarant (p. 42) :

> « Airplanes do not flap their wings, but their design is based on thorough understanding and use of principles of aerodynamics which allow birds to fly[2]. »

L'approche anthropomorphique consiste plus précisément à construire un système capable de produire de la parole en imitant chacune des fonctions des articulateurs et du conduit vocal, que ce soit par modélisation fonctionnelle ou par modélisation physique : ce système peut être considéré comme un véritable *robot articulatoire anthropomorphique*, une véritable *tête parlante*. Le cadre de la robotique se prêtant parfaitement à ce type d'approche, nous avons développé à l'ICP le concept de *robotique de la parole* (Laboissière, Schwartz & Bailly (1990) ; Abry *et al.* (1994), Bailly (1997)), qui consiste à considérer le signal de parole comme le résultat *audible* et *visible* des mouvements des divers articulateurs du *robot articulatoire* qui sont coordonnés par un *contrôleur*. Les propriétés du signal de parole produit dépendent donc à la fois du *robot* et du *contrôleur*.

L'un des problèmes qui se pose alors est celui de déterminer les responsabilités : est-ce l'appareil phonatoire ou la commande de tel ou tel muscle qui est à l'origine de telle propriété acoustique du signal (*cf.* Perkell (1991), Scully (1991)) ?

Il est utile de rappeler la disproportion entre la quantité d'information contenue dans le signal audiovisuel de parole et celle – *a priori* équivalente – associée aux mouvements relativement lents des articulateurs. Il me semble donc plus intéressant du point de vue de la connaissance – et peut-être plus économique du point de vue du stockage et du contrôle – d'essayer d'encoder cette redondance d'information du signal audiovisuel de parole, naturellement nécessaire à la robustesse de la communication, dans les propriétés de la tête parlante. Pour reprendre cette idée en termes de robotique, le principe général consiste donc à décharger au maximum le contrôleur de tous les problèmes de maintien de la cohérence du signal audiovisuel qui sont, eux, réglés de manière implicite par la tête parlante. L'approche que nous poursuivons à l'ICP consiste donc à élaborer une tête parlante la plus complète et la plus réaliste possible, en y incorporant toutes les connaissances disponibles, de façon à pouvoir ensuite mieux en comprendre, puis en simplifier au maximum les stratégies de contrôle.

Quelques exemples peuvent illustrer l'intérêt de cette approche anthropomorphique. Je montrerai dans la suite que, bien que les fricatives voisées et sourdes soient très différentes du point de vue spectral, y compris dans leurs transitions avec les voyelles adjacentes, il est relativement facile d'inclure un simple geste d'ouverture glottique dans une séquence Voyelle – Consonne fricative voisée – Voyelle, pourvu qu'il soit correctement coordonné avec le mouvement de la constriction orale, afin de transformer cette séquence en séquence Voyelle – Consonne fricative sourde – Voyelle. Dans un autre registre, on peut

[1] Voir par exemple la méthode d'analyse-synthèse par prédiction linéaire, Markel & Gray (1976), ou la méthode temporelle PSOLA, Charpentier & Moulines (1990).

[2] *Les avions ne battent pas des ailes, mais leur conception s'appuie sur une compréhension et une utilisation approfondies des principes aérodynamiques qui permettent aux oiseaux de voler.*

noter qu'il est nettement moins complexe de gérer l'influence des mouvements de mâchoire sur les lèvres et l'ensemble du visage avec un modèle articulatoire anthropomorphique (un seul paramètre suffit) qu'avec un modèle du type de celui développé par Parke (1982) dans lequel le mouvement de la mâchoire n'a pas d'influence sur les lèvres, et qui nécessite donc l'intervention de plusieurs paramètres afin d'obtenir un résultat adéquat. Je peux enfin mentionner le cas de la séquence [izi] : pour passer de la voyelle [i], qui est une voyelle voisée simple, à la consonne [z], qui se trouve dans un mode à la fois voisé et bruité, un léger mouvement de l'apex suffit à créer une constriction adaptée dans le conduit vocal. Ce geste articulatoire est relativement simple et d'amplitude faible, alors que les conséquences acoustiques sont complexes : variation des formants, réduction de l'amplitude de voisement et modification des caractéristiques spectrales de ce voisement (perte des hautes fréquences), apparition d'un bruit de friction, d'abord dans les basses fréquences, puis dans les hautes fréquences. Les commandes qu'il faudrait fournir à un synthétiseur à formants, basé sur une représentation sous forme de filtres du signal de parole (*cf.* II.B), pour obtenir un tel résultat acoustique seraient extrêmement complexes. Nous voyons donc que, de manière générale, le contrôle sera d'autant plus simple que le modèle anthropomorphique sera plus réaliste.

En plus de ces avantages de simplification des stratégies de contrôle, l'approche robotique, dans laquelle les fonctions de contrôle et de réalisation sont séparées, permet d'accéder à des propriétés très intéressantes d'adaptabilité. Il sera en effet beaucoup plus facile, et plus économique en développement, d'adapter le style de la parole produite aux besoins de communication que de travailler directement sur le signal. Les recherches actuelles sur les robots bipèdes (ou multipèdes) constituent une bonne illustration de cette approche : il est plus difficile, au départ, de concevoir un robot avec des jambes articulées et coordonnées qu'un robot doté simplement de quatre roues, mais le robot à quatre roues ne pourra fonctionner que dans l'environnement très limité de lieux plats et sans obstacles, alors que le robot doté de jambes articulées pourra s'adapter à des terrains plus accidentés.

Enfin, un dernier avantage qui doit être mentionné réside dans la garantie que la cohérence entre les signaux acoustiques et visuels est assurée quelle que soit la stratégie de contrôle, puisque ces signaux sont issus du même modèle articulatoire sous-jacent.

4. Pas de modèles sans données … et pas de données sans sujet

Les données expérimentales constituent, de leur côté, la matière première indispensable à la modélisation. Dans le cas de modèles basés sur des théories pré-établies, les données servent à tester et à falsifier ces modèles. En effet, un modèle ne pourra être considéré comme validé que lorsqu'il aura été montré qu'il prédit correctement les résultats d'un certain nombre d'expériences. Dans le cas des modèles fonctionnels, les données constituent la matière première à partir de laquelle le modèle est construit, par exemple à l'aide de méthodes statistiques, qui peuvent être, ou ne pas être, basées sur des hypothèses théoriques préalables.

Notre domaine de recherche porte sur des phénomènes liés à des êtres vivants : il en découle qu'un certain nombre des paradigmes expérimentaux s'appuient sur des expériences *in vivo*. Mon souci a donc été de mettre en place des conditions expérimentales contrôlées de la manière la plus fiable possible. L'une des caractéristiques de ma démarche a en particulier consisté à utiliser, pour un certain nombre de dispositifs expérimentaux, le même sujet de référence *P1X*, entraîné à prononcer le même corpus de manière contrôlée et répétable d'expérience en expérience. De cette stratégie a découlé un ensemble de données articulatoires, géométriques, aérodynamiques, acoustiques qui alimentent mes travaux et ceux d'un certain nombre de collègues. L'un des intérêts principaux de cette démarche centrée sur un sujet spécifique réside dans la possibilité de disposer, sur le même phénomène, de données qui ne peuvent être acquises qu'avec des dispositifs expérimentaux impossibles à mettre en œuvre au cours d'une même expérience, comme par exemple la cinéradiographie et les mesures de débit et de pression à l'aide d'un masque pneumotachographique.

Enfin, dans un certain nombre de cas, lorsque le problème de modélisation n'est pas directement spécifique à la parole, mais correspond simplement à un processus physique mal connu ou pour lequel aucune théorie simple n'existe, nous avons recours à des expériences *in vitro*, avec les avantages et les inconvénients classiques liés à ce type d'expérience. C'est ainsi que j'ai été amené à effectuer des campagnes de mesures acoustiques sur des maquettes en plexiglas.

Une partie très importante des études présentées dans ce mémoire a été réalisée sur un seul sujet, le sujet *P1X*. Un certain nombre d'arguments en faveur du choix d'un sujet unique seront discutés à la

section II.D.3. Mentionnons ici que le fait d'utiliser essentiellement un seul sujet avec des corpus les plus homogènes possible offre l'intéressante possibilité d'accès à des données correspondant à divers niveaux du processus de production de parole : équivalence aérodynamique de l'aire de constriction pour contraindre l'inversion et déterminer les petites aires, fonctions de transfert acoustiques mesurées pour assister les mesures de formants difficiles pour les fricatives sourdes, interprétations des images IRM (Imagerie par Résonance Magnétique) médiosagittales par comparaisons avec celles issues de la cinéradiographie, calages des données EMA (Articulographe ElectroMagnétique) par rapport à la géométrie des structures osseuses mesurées sur les radiographies, comparaisons des écoulements obtenus par simulations après inversion par rapport à ceux mesurés à l'aide du masque pneumotachographique pour des réalisations du même type, et dans un futur proche utilisation de l'ElectroPalatoGraphie (EPG) pour valider les contacts palato-linguaux déterminés à partir des données IRM. De plus, le sujet *P1X* est toujours disponible et motivé pour des enregistrements complémentaires, … puisque qu'il n'est autre que l'auteur de ce mémoire …

Par ailleurs, le fait d'inverser un modèle à partir de données enregistrées sur le même sujet que celui utilisé pour construire le modèle permet d'échapper (provisoirement) aux problèmes de normalisation entre sujets.

Enfin, il est clair que les modèles développés à partir d'un corpus de données sont complètement tributaires du contenu du corpus. Il est donc fondamental de concevoir des corpus adaptés au phénomène que l'on souhaite modéliser : il faut s'efforcer d'inclure dans le corpus les éléments correspondant aux situations extrêmes. Pour développer un modèle de velum, par exemple, il est de première importance d'inclure dans le corpus de base les consonnes occlusives (pour lesquelles le velum est maximalement relevé) de même lieu d'articulation que les consonnes nasales, de façon à assurer que les positions extrêmes du velum soit présentes, et obtenir ainsi une couverture maximale de l'ensemble de ses possibilités de mouvement.

5. Modélisation linéaire

Comme nous le verrons plus loin en détail, mon approche au niveau articulatoire repose sur une modélisation fonctionnelle basée sur des modèles *linéaires* simples. J'ai exclu la modélisation physique de type biomécanique parce qu'il me semble que le nombre de paramètres à déterminer pour obtenir des modèles réalistes est trop élevé, et que ces modèles impliquent des décisions *a priori* sur les degrés de liberté qui doivent être inclus dans le modèle : soit le modèle est trop simplifié, auquel cas il n'est pas très utile ; soit le modèle est trop complexe, et il est alors nécessaire de regrouper les nombreux degrés de liberté en coordinations plus simples, ce qui est fait directement en fin de compte dans les modèles fonctionnels basés sur les données.

La simplicité constitue un attrait évident des modèles linéaires, et permet également de conserver une maîtrise sur la compréhension globale des phénomènes. De plus, nous verrons à quel point les modèles linéaires peuvent représenter des phénomènes non linéaires si le système de coordonnées est judicieusement choisi. Prenons comme exemple un point se déplaçant de manière quelconque sur un cercle de rayon constant : en coordonnées cartésiennes, le déplacement de ce point est régi par deux paramètres qui ne sont pas corrélés linéairement, puisque le sinus et le cosinus d'un même angle ont une intercorrélation nulle. Par contre, dans un système de coordonnées polaires, ce même point est régi par un seul degré de liberté. Cet exemple montre qu'un système de coordonnées adapté peut transformer un système apparemment non-linéaire et complexe en un système linéaire et plus simple, et que la corrélation n'est pas un critère suffisant pour l'indépendance (Cardoso (1998)). Des précautions de ce type permettent donc d'éviter d'attribuer à des non-linéarités – qui sont en règle générale difficiles à caractériser – des propriétés qui peuvent être expliquées par un système linéaire simple. Ceci motive les systèmes de grilles particuliers utilisés dans la suite pour décrire certains articulateurs tels que la langue ou le velum. Nous verrons un résultat intéressant de ce point de vue : lors de l'élévation de l'ensemble de la masse de la langue due à un mouvement d'élévation de la mâchoire dans le modèle 3D de langue, la forme de sa surface supérieure tend à s'adapter naturellement à la forme du palais, et donc à adopter un comportement plutôt non linéaire, alors que le cœur du modèle est purement linéaire.

6. Une recherche pluridisciplinaire

L'approche qui vient d'être décrite conduit à une activité nécessairement pluridisciplinaire. En effet, les divers aspects de l'objet d'étude que constitue la production de la parole m'ont poussé à la recherche des outils d'investigation propres à permettre de comprendre et de modéliser les phénomènes observés, par l'intermédiaire de collaborations locales, nationales ou internationales. Le traitement du signal est

naturellement l'outil quotidien qui permet de manipuler, analyser et visualiser le signal de parole et tous les autres signaux mesurés. L'imagerie médicale joue également un grand rôle pour fournir les informations sur la forme du conduit vocal : téléradiographie, cinéradiographie, imagerie par résonance magnétique nucléaire (IRM), imagerie par échographique ultrasonique. L'autre extrémité de la chaîne articulatori-acoustique, qui se situe au niveau de l'aérodynamique et de l'acoustique, a par ailleurs nécessité la mise en œuvre de techniques de mesures acoustiques et aérodynamiques.

C. Quelques repères dans l'histoire des machines parlantes

Avant d'aborder la synthèse de mes travaux de recherche, et afin de proposer un cadre de référence, je rappellerai les étapes importantes qui ont marqué l'histoire des machines parlantes. Je restreindrai ici la notion de *machine parlante* aux systèmes de production artificielle de parole basés sur une approche *anthropomorphique*.

1. Les machines parlantes mécaniques

Parmi les nombreux automates élaborés dans les siècles précédents, je présente brièvement ici les deux machines parlantes mécaniques qui ont été les plus connues, et les plus convaincantes. Il se trouve de plus, peut-être n'est-ce pas par hasard, que ces machines constituaient de véritables simulations mécaniques des mécanismes de production de la parole.

La machine de Kempelen. Le premier travail qui, semble-t-il, ait réellement abouti à une machine parlante anthropomorphique est celui de celui de von Kempelen (1791) (Dudley, Riesz & Watkins (1939), Dudley & Tarnoczy (1950), Flanagan (1972a), Liénard (1968), Liénard (1977)). La Figure 2 montre les différents composants de cette machine, dont la philosophie a inspiré un grand nombre de chercheurs, jusqu'aux réalisations virtuelles récentes. Un soufflet joue le rôle des poumons; il alimente en air sous pression et mets en vibration une anche – modèle de cordes vocales – qui, à son tour, excite un résonateur en caoutchouc dont la forme est manipulée par la main gauche de l'opérateur pour produire les sons voisés. Les consonnes, y compris les nasales, sont produites par des passages resserrés séparés, contrôlés par les doigts de l'autre main. D'après l'inventeur, la machine, manipulée avec dextérité, pouvait prononcer quelques centaines de mots et quelques phrases.

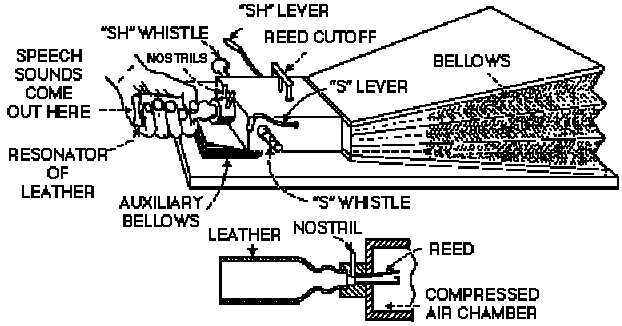


Figure 2 : Reconstruction par Wheatstone de la machine parlante de Kempelen (d'après Flanagan (1972b)*)*

La machine de Faber. Quelques années plus tard, Joseph Faber (1846) présentait une autre machine parlante mécanique, encore plus élaborée, celle-ci commandée par un clavier de quatorze touches. C'est un véritable *modèle articulatoire*, dans lequel la langue et les deux maxillaires sont articulés, et le profil du pharynx modifiable par une série de diaphragmes coulissants (Liénard (1977)). Il semble que cette machine pouvait parler à haute voix, chuchoter, et chanter "God save the Queen" (Dudley & Tarnoczy (1950)).

Les temps modernes. Il est intéressant de noter que depuis quelques années, une équipe de l'université de Waseda à Tokyo développe un conduit vocal entièrement mécanique (Nishikawa, Asama, Hayashi, Takanobu & Takanishi (2000)), dans le cadre d'un grand projet de robot humanoïde. Les principes sont similaires, même si la technologie a évolué, mais la machine réalisée pour l'instant est beaucoup plus simple et ne produit que des voyelles.

On peut aussi mentionner un dispositif mécanique très particulier, développé par Reed, Rabinowitz, Durlach, Braida, Conway-Fithian & Schultz (1985), qui vise à reproduire les mouvements mécaniques des lèvres et de la mâchoire, les vibrations de la région du larynx, et l'écoulement de l'air aux lèvres. Ce dispositif, qui ne produit pas vraiment de son de parole, était destiné à des études sur la méthode *Tadoma*

de communication développée pour les sourds-aveugles : cette méthode repose sur la perception tactile qui permet à l'interlocuteur, qui doit placer sa main sur le visage du locuteur, de suivre les actions associées à la production de parole telles que les mouvements externes du conduit vocal, les vibrations de la région du laryngée liées au voisement, ou les poufs d'air créés aux lèvres lors du relâchement des consonnes occlusives bilabiales.

2. Les analogues électriques de la fonction d'aire du conduit vocal

À ma connaissance, aucun système électrique anthropomorphique capable de produire de la parole dynamique n'a jamais été construit. Le premier système qui permet de produire des sons de parole statiques à partir d'une description géométrique de la forme de conduit vocal a été la *ligne électrique analogue du conduit vocal* de Dunn (1950). Les systèmes précurseurs n'étaient en fait pas des analogues anthropomorphiques : l'analogue électrique de Stewart (1922) se limitait essentiellement à la production de voyelles et diphtongues en simulant les deux premiers formants par deux circuits électriques résonants, tandis que le principe du Voder (Voice Operation DEmonstratoR, Dudley *et al.* (1939)), qui produisait réellement de la parole intelligible, reposait sur le contrôle du spectre du son par l'intermédiaire d'un banc de filtres commandés grâce à un clavier par une opératrice expérimentée.

La *ligne électrique analogue du conduit vocal* de Dunn (1950) était composée de 35 filtres simulant, dans une analogie acoustique-électrique, 35 sections cylindriques de 0,5 cm de long et de 6 cm^2 d'aire transversale, l'ensemble formant un tuyau uniforme de 17,5 cm de long. Trois paramètres de contrôle sont disponibles : le lieu d'insertion, entre deux filtres, d'une bobine qui représente l'effet d'une constriction dans le conduit vocal; la valeur d'inductance de la bobine qui va contrôler l'équivalent de l'aire de la constriction; la valeur d'inductance d'une autre bobine qui charge l'extrémité de la ligne électrique, et qui contrôle l'équivalent de l'aire aux lèvres. Nous sommes en présence du premier modèle à trois paramètres : Xc, Ac, Al ! La réponse du MIT à ce premier prototype issu des Laboratoires Bell est donnée par Stevens, Kasowski & Fant (1953), qui construisent une ligne électrique analogue où l'aire de chacune des sections peut être contrôlée indépendamment; ce système produit des voyelles, mais aussi des fricatives, puisqu'il offre la possibilité de connecter une source de bruit en un point quelconque de la chaîne de filtres. Ce système sera utilisé deux ans plus tard par Stevens & House (1955) pour réaliser les premiers *nomogrammes vocaliques*. Un analogue électrique tout à fait similaire est également développé à Stockholm par Fant (1953). Par ailleurs Fant (1958) mentionne l'utilisation de deux ordinateurs suédois, BESK en 1951, puis BARK de 1951 à 1953, pour la simulation numérique d'un analogue du conduit vocal comprenant 20 sections, ce qui constitue l'une des toutes premières, si ce n'est la première, applications de l'ordinateur à la production de la parole. De nombreux autres modèles, électriques ou numériques, suivront ces réalisations de pionniers.

3. Les machines parlantes simulées

Les travaux menés dans les années 50 à l'aide des analogues électriques ont été fondamentaux pour le développement d'une théorie de la production de la parole qui sera en particulier marquée par l'ouvrage de référence de Fant (1960). Cependant, le travail est extrêmement lourd, puisque toutes les mesures passent par des manipulations manuelles de circuits électriques, et aucun de ces systèmes n'est capable de produire de parole continue. Il faudra attendre l'arrivée de l'informatique dans le domaine de la parole pour que les premières réalisations, simulées cette fois, voient le jour, et que le contrôle *temporel* et *articulatoire* des analogues soit possible. La présente section mentionne les systèmes de synthèse les plus complets et les plus utilisés qui constituent des références dans le développement de la synthèse articulatoire.

Le premier synthétiseur articulatoire complet est vraisemblablement celui développé par Coker (1967) aux Laboratoires Bell : un modèle articulatoire simulé numériquement est contrôlé par six paramètres et modélise de manière paramétrique la coupe médiosagittale du conduit vocal sous forme d'arcs de cercle et de segments de droite, et produit finalement une fonction d'aire. Le son est généré par un synthétiseur à formants (*cf.* II.B) lui-même contrôlé par les formants extraits de la fonction de transfert acoustique calculée à partir de la fonction d'aire. Coker (1967), et Coker, Umeda & Browman (1973) utilisent ce système pour produire de la synthèse de l'anglais américain à partir du texte.

Ce synthétiseur, sous diverses formes, constituera le cœur d'une longue série de travaux aussi bien aux laboratoires Haskins pour l'étude de la production et de la perception de la parole (Mermelstein (1973), Rubin, Baer & Mermelstein (1981), McGowan (1994a), McGowan (1994b), Löfqvist, Koenig & McGowan (1995), McGowan, Koenig & Löfqvist (1995), McGowan & Saltzman (1995), Rubin, Saltzman, Goldstein, McGowan, Tiede & Browman (1996)) qu'aux laboratoires Bell pour le codage à bas débit (Schroeter & Sondhi (1986), Rahim, Goodyear, Klejin, Schroeter & Sonhi (1993)).

Au Japon, les premiers travaux sur la synthèse articulatoire sont vraisemblablement ceux de Shirai & Honda (1976). Deux équipes ont accompli l'essentiel des développements de modèles articulatoires. À NTT, Honda (Masaaki), Kaburagi et leurs collègues s'intéressent surtout à la dynamique des articulateurs ; leur modèle articulatoire est limité aux points de mesures fournis par articulographie électromagnétique ou par imagerie ultrasonique (Kaburagi & Honda (1994)), et les spectres des sons sont directement associés aux articulateurs à l'aide d'une dictionnaire issu de données articulatori-acoustiques expérimentales (Kaburagi & Honda (1998)), mais ils sont capables de synthétiser des trajectoires articulatoires complètes à partir d'une séquence de phonèmes (Kaburagi & Honda (2001)). À ATR, Honda (Kiyoshi), Dang et leurs collègues abordent également la synthèse articulatoire, mais en se basant sur des modèles plus physiques que fonctionnels (Dang & Honda (1998), Dang, Sun, Deng & Honda (1999), Dang & Honda (2000b)).

Notons par ailleurs l'émergence récente d'une équipe chinoise dans le domaine de la synthèse articulatoire (Yu & Zeng (2000)).

En Europe, peu de travaux portent sur la synthèse articulatoire. Le système de synthèse articulatoire de l'allemand à partir du texte de Kröger, Schröder & Opgen-Rhein (1995) est sans doute le système le plus complet. Paul Boersma (1998) a développé un synthétiseur articulatoire basé sur le modèle articulatoire de Mermelstein (1973), mais considérant en outre que les parois de chaque section peuvent bouger sous l'action de forces aérodynamiques et myoélastiques ; son synthétiseur comprend également un modèle aérodynamique complet qui permet de prendre en compte l'interaction entre sources et conduit vocal, et en particulier de produire des clicks.

En Grande Bretagne, Scully et ses collaborateurs ont développé un modèle composite de conduit vocal, pour étudier particulièrement les interactions entre les principales constrictions du conduit vocal (constriction orale et glotte) et les sources acoustiques de voisement et de bruit de friction (Allwood & Scully (1982), Scully (1986)).

En France, si l'on excepte les travaux de l'ICP, Maeda est le seul à avoir développé de manière active et continue des modèles articulatoires (Maeda (1979b), Maeda (1990), Maeda & Honda (1995)). Son modèle a été introduit pour la première fois à l'ICP par Perrier, Boë, Majid & Guérin (1985), et a donné lieu à toute une lignée de travaux et de développement d'environnement (*cf.* le SMIP dans le projet européen *Speech Maps* (Boë, Gabioud & Perrier (1995), Boë, Gahioud & Perrier (1995)). Je présenterai les travaux de synthèse articulatoire auxquels j'ai participé à l'ICP dans la suite de ce mémoire.

Il est enfin nécessaire de mentionner quelques travaux en synthèse articulatoire basés sur des modèles de fonction d'aire : Flanagan, Ishizaka & Shipley (1980)) ont développé un modèle de fonction d'aire commandé par des paramètres du type Xc, Ac, Al à des fins de codage articulatoire à bas débit. Plus récemment, Fant et ses collaborateurs à Stockholm ont également développé un modèle de fonction d'aire dans le cadre d'un projet de synthèse articulatoire (Lin (1990), Båvegård & Fant (1996)).

Notons enfin que Boë, Maeda & Perrier (1994) donnent une description plus détaillée de l'histoire des modèles articulatoires.

À ce point, il est intéressant de remarquer que la science semble fonctionner en spirale : la machine parlante de von Kempelen et les robots parlants virtuels actuels sont basés sur les mêmes principes anthropomorphiques. De ce point de vue, rien de nouveau n'a donc été inventé, et l'on se retrouve au point de départ, après des décennies pendant lesquelles les voies du traitement du signal ont été explorées, et l'anthropomorphisme oublié. Par contre, des progrès considérables ont été accomplis au niveau du réalisme des synthétiseurs articulatoires, et en particulier, de manière essentielle, au niveau de leur contrôle.

II. SYNTHESE DES TRAVAUX

A. Préambule

Je ne peux pas entamer la présentation de l'ensemble de mes travaux, que je vais classer plutôt par sujets que par ordre chronologique, sans évoquer brièvement les paternités majeures dont j'ai eu la chance de bénéficier, et qui ont fortement influencé mes travaux de recherche depuis mes débuts à l'ICP en avril 1979.

Mes premiers pas de chercheur se sont déroulés à Grenoble, sous la conduite bienveillante de Bernard Guérin et de René Carré : ils m'ont encouragé à aborder l'analyse-synthèse de parole dans le cadre de la synthèse à formants, qui n'est pas une approche anthropomorphique, mais fait référence sinon directement au conduit vocal, du moins aux résonateurs acoustiques qui le constituent.

L'étape suivante, suédoise, est celle de mon apprentissage des bases et secrets de l'acoustique du conduit vocal, aux côtés de Gunnar Fant, avec la lecture assidue de la *bible* qu'il a publiée en 1960 : « Acoustic theory of speech production ».

Vient ensuite, à mon retour en France, l'influence anglaise : Celia Scully est l'inspiratrice de tous mes travaux sur les fricatives et sur la coordination entre glotte et constriction.

L'influence japonaise de Shinji Maeda sur mes travaux en modélisation articulatoire est incontestable : elle a toujours guidé ma réflexion sur le fonctionnement du conduit vocal.

La suite est à nouveau grenobloise : Christian Abry m'infuse la *robotique de la parole* tout au long du projet européen *Speech Maps*, et plus récemment Gérard Bailly m'entraîne dans l'aventure des têtes parlantes, sans oublier Louis-Jean Boë et Jean-Luc Schwartz qui éclairent en permanence ma route de leurs lumières plus extérieures au champ de la production de la parole, mais de ce fait tout aussi précieuses.

Pour conclure, il est évident que j'ai bénéficié de l'influence scientifique de bien d'autres chercheurs, dont beaucoup à l'ICP, avec lesquels j'ai eu l'occasion de collaborer de manière plus ou moins proche. Qu'ils me pardonnent l'injustice de ne pas les citer ici.

B. Synthèse à formants

Dans le cadre d'une thèse de Docteur-Ingénieur (Badin (1983)), j'ai abordé l'étude de la parole sous l'angle de la synthèse à formants. Le projet visait à développer un système d'analyse et de re-synthèse du signal de parole, avec une approche de type *production*, dans le double but de mettre en évidence des indices acoustiques de la parole, et de constituer un ensemble de logatomes devant servir de base à un système de synthèse à formants à partir du texte.

Les synthétiseurs à formants reposent sur l'hypothèse de l'indépendance entre les sources d'excitation et les fonctions de filtrage acoustique (Klatt (1980)). La fonction de transfert acoustique du conduit vocal est simulée par un ensemble de quatre ou cinq filtres résonants du second ordre, connectés en cascade ou en parallèle. Une vingtaine de paramètres de commande sont nécessaires pour piloter le synthétiseur de structure parallèle choisi pour sa capacité à reproduire les caractéristiques spectrales des consonnes aussi bien que celles des voyelles : fréquence centrale, bande passante et amplitude pour chacun des cinq formants, commandes de gain des sources voisée et de bruit de friction. J'ai développé, dans le cadre de mon DEA (Badin (1980)), un tel synthétiseur écrit en assembleur et programmé sur une carte de traitement de signal développée à l'ICP par Daniel Degryse (Badin & Degryse (1982)).

Une analyse de type *production* consiste en quelque sorte à *déconvoluer* le signal du parole par rapport à un modèle de production, en l'occurrence le synthétiseur à formants. J'ai donc développé un ensemble de logiciels permettant une stratégie d'*analyse par la synthèse* mise en œuvre en deux étapes (Badin & Murillo (1983a, (1983b)). La première étape, entièrement automatique, permettait d'obtenir formants et bandes passantes en extrayant les pôles complexes conjugués à partir d'une analyse par prédiction linéaire (Markel & Gray (1976)), et la fréquence fondamentale par l'algorithme du SIFT (Markel (1972)). La deuxième étape, l'analyse par la synthèse proprement dite, impliquait une comparaison entre le signal naturel à analyser et le signal produit par le synthétiseur, à la fois sur le plan spectral et sur le plan perceptif (Badin & Murillo (1983a, (1983b)); les trajectoires des paramètres de commande du synthétiseur était ainsi déterminées de manière itérative.

Le travail a clairement démontré la possibilité d'obtenir des copies de synthèse de très bonne qualité, et en particulier jugées par un ensemble d'auditeurs bien plus naturelles que celle obtenues par synthèse par prédiction linéaire. Avec le recul et au vu des connaissances articulatoires acquises aujourd'hui, il est clair

que les trajectoires formantiques déterminées auraient pu être simplifiées, ce que j'avais en partie mis en évidence par des tests perceptifs au cours desquels l'importance de tel ou tel détail de la trajectoire d'un formant était évaluée.

Cette première expérience m'avait déjà convaincu de l'intérêt d'une approche fortement liée à une description des caractéristiques acoustiques du signal de parole. Un séjour post-doctoral d'un an et demi au KTH à Stockholm, sous la direction du Professeur Gunnar Fant, a achevé de me convertir à une approche véritablement basée sur la modélisation des phénomènes de production. Simultanément, mes objectifs de recherche se sont progressivement élargis à une vision plus étendue du domaine, c'est-à-dire non plus seulement à la synthèse paramétrique de la parole, mais à la compréhension et la modélisation des phénomènes de production de la parole en général.

C. Acoustique du conduit vocal : données, modèles et simulations

L'acoustique du conduit vocal concerne les relations entre la forme du conduit vocal, son état aérodynamique, les caractéristiques mécaniques de certains de ses constituants (cordes vocales, lèvres ou parois), et les caractéristiques acoustiques résultantes (résonances ou anti-résonances du conduit, signal acoustique rayonné).

Notre approche vise en particulier à élargir la bande des fréquences utiles au-delà des 5 kHz qui correspondent au mode de propagation plan, et aussi à prendre en compte les phénomènes aéroacoustiques liés à l'écoulement de l'air dans le conduit vocal.

Mon intérêt s'est porté sur les trois fonctions principales du système phonatoire humain : les sources d'excitation acoustique, la propagation des ondes acoustiques dans le conduit vocal, et leur rayonnement. Je vais décrire dans les trois cas les méthodes et modèles développés.

1. Sources de bruit de friction et modélisation de l'écoulement de l'air

Avant de décrire des méthodes de simulation de la propagation des ondes acoustiques dans le conduit vocal, je présente rapidement les travaux portant sur l'excitation acoustique du conduit vocal, en liaison avec l'écoulement de l'air dans le conduit vocal.

Modèle de source de bruit de friction pour les fricatives et les voyelles chuchotées. L'étude théorique de la génération des bruits de friction est très difficile du fait que les résultats dépendent beaucoup de la géométrie exacte du conduit vocal. Nous avons donc choisi une approche de modélisation fonctionnelle basée sur des données expérimentales. Suivant les principes proposés par Hixon, Minifie & Tait (1967) ou Stevens (1971), nous avons établi des relations entre les variables qui caractérisent l'état aérodynamique du conduit vocal (débit de l'écoulement d'air, chute de pression et aire à la constriction) et celles qui définissent le spectre de la source de bruit de friction considérée comme une source de pression localisée (intensité globale de la source, pente spectrale). Le principe de cette approche est de mesurer les variables en question pour un ensemble de productions qui couvre la gamme des valeurs correspondant à de la parole, pour un sujet de référence, en l'occurrence le sujet *P1X*. C'est ainsi que le débit aux lèvres, la pression rayonnée aux lèvres et la pression intra-orale ont été mesurés à l'aide d'un masque pneumotachographique pour des corpus de fricatives soutenues et en contextes vocaliques prononcées à différents niveaux d'effort vocal. Ce travail, démarré à Stockholm (Badin & Fant (1989), Badin (1989)), a ensuite été poursuivi dans le cadre de la thèse de Pham Thi Ngoc (1995) par la mise en œuvre des procédures de filtrage inverse pour déterminer les variations d'intensité et de pente spectrale de la source de bruit en fonction des conditions aérodynamiques (Badin, Shadle, Pham Thi Ngoc, Carter, Chiu, Scully & Stromberg (1994b), Pham Thi Ngoc & Badin (1994)). Il a permis l'élaboration, par régression linéaire multiple, d'un modèle de source de bruit de friction basé sur les données du sujet *P1X* (Badin, Mawass & Castelli (1995c)). Le gain relatif de la source de bruit de friction par rapport à la source de voisement a ensuite été déterminé de manière à ce que le programme de simulation temporelle *SIMOND* (voir ci-dessous) produise les mêmes niveaux que ceux observés dans les signaux issus du sujet *P1X* ; en effet, l'équilibre adéquat entre les contributions des sources acoustiques de voisement et de bruit de friction apporte une contribution importante à la qualité et au naturel de la synthèse (Scully & Allwood (1985)). Nous avons ensuite vérifié de manière informelle, en faisant artificiellement varier ce gain dans des séquences [ava], [aza], [aʒa], que ces niveaux correspondent à une qualité optimale de la synthèse (Mawass, Badin & Bailly (2000))[1].

[1] Ces sons sont accessibles à l'adresse http://www.icp.inpg.fr/~badin/ActaAcustica_Sounds.html

Modélisation basse fréquence de l'écoulement de l'air – Coordination glotte / constriction orale. Il découle des descriptions ci-dessus que les modèles de voisement et de bruit de friction sont contrôlés par des variables communes (le débit d'air), ou liées (les chutes de pression à travers la glotte et la constriction orale). En conséquence, les comportements de ces deux sources d'excitation acoustique du conduit vocal interagissent fortement. Il était donc indispensable de mettre en œuvre un modèle qui prenne en compte ces interactions. Nous avons ainsi défini un modèle simplifié d'écoulement d'air dans le conduit vocal valable pour les basses fréquences (c'est-à-dire en dessous de 200 Hz) (Mawass *et al.* (2000)). Ce modèle considère le conduit vocal comme un ensemble de deux constrictions localisées : la glotte et la constriction orale. Les équations de Bernoulli et de Poiseuille sont utilisées pour exprimer les relations entre les composantes basse fréquence du débit, des aires et des chutes de pression au niveau de ces constrictions. En particulier, la somme des deux chutes de pressions est égale à la pression des poumons, qui constitue ainsi l'un des paramètres de commande de ces modèles de sources. Nous avons finalement montré, grâce à une série de simulations systématiques (Mawass (1997), Mawass *et al.* (2000)), qu'une coordination très précise entre l'aire de la glotte et l'aire de la constriction orale était indispensable pour maintenir un équilibre entre l'intensité du bruit et celle du voisement pour les consonnes fricatives voisées : en effet, l'augmentation de l'intensité du bruit nécessite une augmentation de la chute de pression à la constriction orale, donc une diminution de la chute de pression à la glotte, et donc une diminution de l'intensité du voisement, et vice versa. Cela explique les statistiques d'occurrence des différents types de consonnes dans les langues du monde (Boë, Vallée, Badin, Schwartz & Abry (2000)).

2. Propagation des ondes acoustiques dans le conduit vocal

Simulation acoustique fréquentielle du conduit vocal. Dès le début (Dunn (1950), Fant (1960)), l'acoustique du conduit vocal a reposé sur un certain nombre d'hypothèses simplificatrices : propagation d'ondes planes, impédance de rayonnement limitée aux basses fréquences, pertes par viscosité et dissipation thermique simplifiées, pas de prise en compte de l'écoulement. Pour des raisons historiques, ce sont des ingénieurs en électricité (Dunn, Fant, Stevens, etc.) qui ont développé les premiers modèles, et qui ont donc utilisé des analogies entre grandeurs acoustiques et grandeurs électriques (le plus souvent, pression acoustique / tension électrique et débit acoustique / courant électrique). Dans ces conditions, le conduit vocal peut être assimilé à un tuyau d'aire variable le long de son axe médian, et finalement traité comme une cascade de sections cylindriques dans lesquelles se propagent des ondes acoustiques planes (Badin & Fant (1984)).

Lors de mon séjour au KTH de Stockholm en 1983-84, j'ai donc étudié la littérature dans ce domaine, et implémenté un logiciel de simulation acoustique fréquentielle du conduit vocal (*VCTR*). Ce logiciel permet de calculer la *fonction de transfert acoustique* du conduit vocal, c'est-à-dire la réponse en fréquence du tuyau acoustique que constitue le conduit vocal. Cette réponse est estimée comme le rapport entre l'amplitude (complexe) du débit acoustique aux lèvres sur l'amplitude de la source de débit acoustique à la glotte, ou celle d'une source de pression de bruit de friction insérée en un point quelconque du conduit. Diverses conditions aux limites (ouverture à la glotte, rayonnement aux lèvres, vibration des parois) et pertes (par viscosité et par conduction thermique) ont été testées. Ce logiciel *VCTR* a été constamment maintenu depuis, et continue d'être abondamment utilisé à l'ICP.

Simulation acoustique temporelle du conduit vocal par « ligne analogue à réflexion ». Le logiciel de simulation temporelle par « ligne analogue à réflexion » développé à l'ICP par Degryse (1981) a servi de base à notre synthétiseur articulatoire. Jusqu'en 1987, la variation de longueur du conduit vocal n'était pas prise en compte par ce type de ligne qui suppose constante la longueur des sections. Nous avons alors mis en place un algorithme de changement de fréquence d'échantillonnage, qui permet de résoudre ce problème (Wu, Badin, Cheng & Guérin (1987a, (1987c, (1987b)). À la suite des travaux menés sur les sources de bruit pour les fricatives (voir plus loin), nous avons ensuite implémenté un modèle de source dans *SIMOND*, la version de la ligne analogue à réflexion remaniée par Castelli (1989) pendant sa thèse, et nous avons testé cette implémentation en comparant la réponse impulsionnelle de *SIMOND* à la référence fournie par la simulation fréquentielle *VCTR* (*cf.* Mawass, Badin, Vescovi & Beautemps (1996)).

Évaluation perceptive de l'échantillonnage optimal de la fonction d'aire. Le pas d'échantillonnage optimal de la fonction d'aire pour la simulation acoustique temporelle du conduit vocal est un compromis entre la valeur la plus grande possible qui minimise les temps de calcul, et la valeur maximale au delà de laquelle l'effet de l'échantillonnage apporte une dégradation perceptible. Nous avons déterminé un pas optimal entre 3 et 5 mm par une étude théorique basée sur les résultats de Flanagan (1955) sur les seuils différentiels minimaux, et confirmé ces résultats par un test perceptif (Wu *et al.* (1987b, (1987c)).

Fonctions de sensibilité différentielles. Lors d'une première tentative d'estimation de fonctions d'aire cohérentes à la fois avec la coupe sagittale et les formants ou le spectre du son pour des consonnes fricatives (Badin (1991)), nous avons été amenés à utiliser les fonctions de sensibilité différentielles proposées par Fant & Pauli (1974) et reprises par Mrayati & Carré (1976). En effet, la relation entre fonction d'aire et formants est loin d'être linéaire, et les fonctions de sensibilité différentielles constituent en quelque sorte le Jacobien des formants en fonction des aires des sections. Une fonction de sensibilité peut être définie, pour une fréquence de résonance donnée F_j, comme la variation relative $\Delta F_j / F_j$ du formant F_j induite par une variation donnée $\Delta A_i / A_i$ (pour les sensitivités transversales) ou $\Delta l_i / l_i$ (pour les sensitivités transversales longitudinales) pour chacune des *n* sections de la fonction d'aire. Ainsi, en liaison avec les distributions de pression et de débit acoustique le long du conduit vocal à la même fréquence de résonance, les fonctions de sensibilité constituent une aide pour la détermination des affiliations cavités/formants, c'est-à-dire déterminer si un formant dépend plus particulièrement d'une cavité donnée dans le conduit vocal plutôt que d'une autre. L'utilisation de ces fonctions de sensibilité nous a donc permis de déterminer les affiliations pour les consonnes fricatives, et d'affiner notre connaissance des fonctions d'aire correspondantes (Badin (1991)).

Banc de mesure de la fonction de transfert acoustique du conduit vocal. Afin de valider les résultats des simulations précédentes, de déterminer les limites d'une théorie simplifiée, et de pouvoir caractériser certaines catégories de sons, nous avons entrepris l'élaboration d'un banc de mesure de la fonction de transfert acoustique du conduit vocal utilisable pour des sujets humains. Poursuivant la voie ouverte par Fujimura & Lindqvist (1971), une première version du banc de mesure a été développée (Castelli & Badin (1988) ; Feng & Castelli (1996)). Le principe consistait à exciter le conduit vocal d'un sujet de manière transcutanée au niveau du larynx à l'aide d'un petit haut-parleur alimenté par un bruit blanc, et à enregistrer le son rayonné aux lèvres ou aux narines (voir Figure 3). La fonction de transfert était alors simplement calculée comme le spectre du son rayonné.

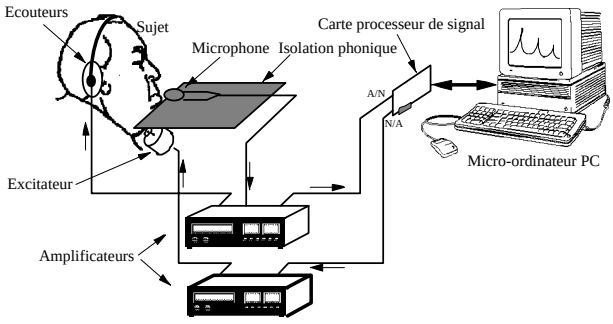


Figure 3 : Schéma du dispositif pour la mesure de fonction de transfert acoustique du conduit vocal développé à l'ICP.

Ce dispositif, qui a fourni nombre de données intéressantes sur les nasales et sur les fricatives, présentait cependant deux inconvénients majeurs : l'obligation de travailler sans phonation, et surtout la nécessité de temps d'acquisition très longs (de 8 à 16 sec.). Une nouvelle méthode a donc été mise au point. Elle consiste à exciter le conduit vocal par une séquence pseudo-aléatoire : la fonction de transfert est alors calculée comme la transformée de Fourrier de l'intercorrélation entre ce signal d'excitation et le signal rayonné (Djéradi, Guérin, Badin & Perrier (1991)). Cette méthode présente l'avantage de réduire la durée de mesure à une centaine de millisecondes, et d'autoriser la phonation pendant la mesure, à la seule condition que son niveau ne soit pas trop élevé.

Nous avons ensuite encore amélioré le dispositif, d'une part au niveau de l'ergonomie, et d'autre part par la possibilité de calculer automatiquement les formants et les bandes passantes des fonctions de transfert, grâce à l'utilisation de méthodes d'identification ARX (Pham Thi Ngoc & Badin (1994), Pham Thi Ngoc (1995)).

Voyelles et consonnes nasales. Le banc de mesure acoustique a permis dans un premier temps de mettre en évidence certaines caractéristiques de la nasalité (Castelli (1989), Castelli & Badin (1989)) : présence d'un zéro à basse fréquence dans les fonctions de transfert nasopharyngales, et aussi dans les voyelles orales pures (d'où l'hypothèse d'un couplage – interne ou externe – entre le conduit oral et le conduit nasal), et présence de zéros supplémentaires qui peuvent correspondre aux résonances des sinus.

Voyelles orales et affiliations. Nous avons pu confirmer les hypothèses sur les affiliations formants/cavités en mesurant l'influence de l'ouverture des cordes vocales sur les bandes passantes des fonctions de transfert mesurées pour les voyelles (Pham Thi Ngoc & Badin (1994), Pham Thi Ngoc (1995)).

Modes d'ordre supérieur. Dans le cadre de la fédération de laboratoires ELESA, en collaboration avec Pierre Saguet (Laboratoire d'Electromagnétisme et de Micro-Ondes), nous avons tenté d'élargir la bande spectrale des modélisations acoustiques, en tenant compte des modes de propagation non unidimensionnels en appliquant à l'acoustique la méthode TLM (Transmission Line Matrix) initialement développée pour l'électromagnétisme. Ce travail, mené par Samir El Masri et Xavier Pelorson, est basé sur les mesures que j'ai effectuées au Japon (Motoki, Badin & Miki (1994), El Masri, Pelorson, Saguet & Badin (1996a)). J'ai aussi contribué à la comparaison des premiers résultats de simulation TLM avec la théorie des ondes planes pour des configurations typiques du conduit vocal (El Masri, Pelorson, Saguet & Badin (1996b)).

3. Rayonnement aux lèvres

Un autre problème important en acoustique du conduit vocal est celui du rayonnement aux lèvres. Les questions sont de savoir où se termine le conduit vocal du côté des lèvres, et comment caractériser les impédances acoustique et de rayonnement. Une première étude, en collaboration avec deux laboratoires japonais, m'a permis de déterminer les relations entre un équivalent basse fréquence de l'impédance de rayonnement et une caractérisation géométrique du pavillon labial, grâce à des mesures acoustiques et géométriques faites sur des moulages en plâtre des lèvres du sujet *P1X* (Badin, Motoki, Miki, Ritterhaus & Lallouache (1994a)).

Nous avons ensuite établi expérimentalement, pour des fréquences plus élevées, des cartographies du champ de pression pour des conduits de formes géométriques simples (Motoki *et al.* (1994)). Ces mesures ont finalement permis de guider et de valider les modèles théoriques de rayonnement développés à l'ICP par Xavier Pelorson (Pelorson, Badin, Motoki, Miki & Plicque (1995)).

D. Articulation : données, modèles et simulations

Après la description, dans la section précédente, de mes travaux sur l'acoustique du conduit vocal et les mécanismes de génération de sources, je présente maintenant mes travaux sur le fonctionnement des articulateurs qui façonnent ce conduit et en font évoluer la forme au cours du temps afin de générer le signal audiovisuel de parole.

1. Robotique de la parole et degrés de liberté

Nous avons vu que, dans le cadre d'une *robotique de la parole*, nous considérons l'appareil phonatoire comme un *robot* piloté par un *contrôleur* de manière à recruter les articulateurs et en coordonner les mouvements qui ont des conséquences simultanément audibles et visibles. Ce concept implique la notion d'un nombre relativement limité de *degrés de liberté* pour le robot. Ces degrés de liberté correspondent à la spécification, pour chaque articulateur, de l'ensemble limité de mouvements qu'il peut exécuter indépendamment des autres articulateurs. Pour Kelso, Saltzman & Tuller (1986), l'appareil phonatoire est constitué d'un large nombre de composants musculaires qui offrent une dimensionalité potentiellement gigantesque et qui doivent donc être fonctionnellement couplés pour produire des gestes relativement simples (cette vision forme la base du concept de *structures coordinatives de la parole* défendu en particulier par Fowler & Saltzman (1993)). Maeda (1991) fait référence à un concept similaire en termes d'*articulateurs élémentaires*.

Un *degré de liberté* pourrait être plus précisément défini pour un articulateur donné comme une variable qui peut contrôler entièrement une variation spécifique de la forme et de la position de cet articulateur, et qui est statistiquement décorrélée des autres degrés de liberté pour un ensemble de tâches. Ces degrés de liberté peuvent être déterminés à partir de l'observation des corrélations linéaires existant entre les divers paramètres qui constituent la description géométrique précise des formes et des positions des articulateurs. Ces corrélations peuvent être issues de trois niveaux de contraintes implicites ou explicites : (1) la continuité physique des articulateurs et leur constitution (la langue, qui est un hydrostat, ne peut pas prendre une forme en dent de scie, par exemple), (2) les contraintes biomécaniques (la gamme des formes et positions possible pour les articulateurs est limitée par les propriétés physiologiques des structures osseuses et des muscles), et (3) la nature de la tâche en relation avec le contrôle (la mastication fait intervenir des mouvements latéraux de la mâchoire, mais pas la production de parole : ainsi, la mâchoire possède des ensembles de degrés de liberté indépendants liés à la tâche). Les corrélations observées sur les

mesures articulatoires ne peuvent pas toujours être attribuées avec certitude soit aux contraintes biomécaniques soit aux stratégies de contrôle liées à la tâche. Par exemple, il est bien établi que la hauteur du larynx et la protrusion des lèvres sont inversement corrélées pour les voyelles (Riordan (1977), Barbier (1978), et plus récemment Hoole & Kroos (1998)) : cette corrélation ne peut évidemment pas s'expliquer par des liens biomécaniques entre les lèvres et le larynx, mais devrait être attribuée à une stratégie de contrôle liée à la tâche (la descente du larynx et la protrusion des lèvres constituent une stratégie de synergie qui vise à baisser les fréquences de résonance du conduit vocal). Il est donc facile de comprendre que le tri entre les propriétés qui doivent être attribuées au *robot* et celles qui doivent l'être au *contrôleur* constitue toujours un enjeu important en contrôle de la parole (*cf.* Perkell (1991), Scully (1991), Abry *et al.* (1994)).

En robotique, on appelle paramètres *proximaux* les paramètres de commande du robot, tandis que les paramètres produits par le robot sont appelés paramètres *distaux*. Dans le cas de la robotique de la parole, les paramètres proximaux sont les paramètres de contrôle des degrés de liberté articulatoires, tandis que les paramètres distaux correspondent aux caractéristiques géométriques du conduit vocal (lieux et taille des constrictions, aire et protrusion aux lèvres), et aux caractéristiques acoustiques qui en découlent (formants, voisement, bruit de friction). Il est remarquable que le système des articulateurs possède des degrés de liberté en excès (plusieurs combinaisons différentes de paramètres proximaux peuvent aboutir à la même combinaison de paramètres distaux), et que le locuteur humain exploite cette redondance, par le jeu de stratégies de coarticulation / compensation. Par exemple, une fermeture aux lèvres peut être réalisée avec différentes hauteurs de la mâchoire, selon l'influence du contexte, les lèvres permettant de compenser pour atteindre cette fermeture.

2. Données articulatoires et dispositifs expérimentaux

Mon approche de la modélisation articulatoire est basée depuis plus d'une dizaine d'années sur un travail systématique de collection de données variées et complémentaires sur quelques sujets de références prononçant des corpus bien définis et adaptés aux problèmes posés.

Le système de mesure articulatoire idéal en phonétique expérimentale serait celui qui fournirait, à une cadence de plusieurs centaines de Hertz, une image complète tri-dimensionnelle de tous les articulateurs – visibles et invisibles – d'un sujet parlant librement, sans aucune contrainte invasive, ni de risque pour sa santé. Malheureusement, un tel système n'existe pas aujourd'hui, et ne semble pas devoir être développé dans la décennie qui débute, malgré les progrès importants réalisés dans certains domaines, en particulier en Imagerie par Résonance Magnétique. Il est donc nécessaire d'employer plusieurs méthodes expérimentales complémentaires et d'en fusionner ensuite les résultats à l'aide de modèles, afin d'obtenir des données cohérentes pour un corpus donné. C'est l'approche que j'ai adoptée depuis mes premières mesures de débit et pression d'air dans le conduit vocal pour l'étude des consonnes fricatives à Stockholm. Je passerai donc en revue les différentes techniques que j'ai utilisées et / ou mises en œuvre dans ce domaine. Globalement, ces techniques diffèrent par leur résolution spatiale, leur résolution temporelle, leur caractère plus ou moins invasif pour le locuteur, et leur niveau de risque pour la santé du sujet. Le Tableau I récapitule les méthodes que j'ai utilisées, et leurs principales caractéristiques, en particulier en termes de résolution temporelle et spatiale. Le Tableau II précise les publications directement liées à l'utilisation ou au développement des ces méthodes.

Téléradiographie. La téléradiographie est mentionnée ici pour mémoire. Elle utilise une source de rayons X placée à environ 5 mètres du sujet placé de profil, et qui impressionne une plaque sensible placée immédiatement derrière la tête du sujet, introduisant ainsi une distorsion optique minimale. Les images obtenues – statiques – correspondent à une projection de l'ensemble des structures et tissus de la tête, et doivent donc être interprétées de manière soigneuse pour obtenir en particulier les contours du conduit vocal dans le plan médiosagittal. Cette technique, que j'ai eu la possibilité d'utiliser au Département de Radiologie Faciale du CHRU de Grenoble en 1990, très limitée au niveau du nombre de clichés réalisables à cause des risques de santé qu'elle fait courir au sujet, est maintenant avantageusement remplacée par l'IRM (voir ci-dessous). Elle m'a cependant été très utile pour obtenir une dizaine de coupes médiosagittales pour des voyelles et des fricatives artificiellement soutenues par le sujet *P1X*, et pouvoir ainsi déterminer, en conjonction avec la technique de mesure de fonction de transfert acoustique du conduit vocal (*cf.* Djéradi *et al.* (1991)), les fonctions d'aire et les affiliations formants / cavités des consonnes fricatives pour ce même sujet (Badin (1991)). Notons que les méthodes radiographiques présentent l'avantage de visualiser les structures osseuses, ce que ne permettent pas les méthodes magnétiques.

Cinéradiographie. Cette technique, basée elle aussi sur une illumination du sujet par des rayons X, permet d'obtenir de véritables films en format 35 mm ou en vidéo à une cadence de cinquante images par seconde. Couplée et synchronisée avec une prise de vue (film ou vidéo) de face (voir ci-dessous la section sur la labiométrie), elle fournit des données qui offrent, encore aujourd'hui, le meilleur compromis entre résolution spatiale (mieux que 3 à 4 pixels par mm, mais dans le seul plan médiosagittal) et résolution temporelle (50 Hz). Les deux films que nous avons réalisés en collaboration avec l'Institut de Phonétique et l'Hôpital Schiltigheim de Strasbourg, nous ont permis de construire deux bases de données qui constituent la source d'une bonne partie de mes travaux de modélisation articulatoire (voir Badin, Gabioud, Beautemps, Lallouache, Bailly, Maeda, Zerling & Brock (1995b) pour le film sur le sujet *P1X*, et Badin, Baricchi & Vilain (1997) pour le film sur un autre sujet de référence, *J1X*). Les images sont du même type que celles obtenues par téléradiographie, et donc délicates à interpréter. L'usage de cette technique doit aussi être extrêmement limité, à cause de la nocivité des rayons X. Des détails sur le dispositif expérimental et sur le traitement des contours peuvent être trouvés dans Beautemps, Badin & Bailly (2001). Pour chacun des deux sujets, nous disposons donc d'environ 1200 images et mesures du type de celle représentée à la Figure 4.

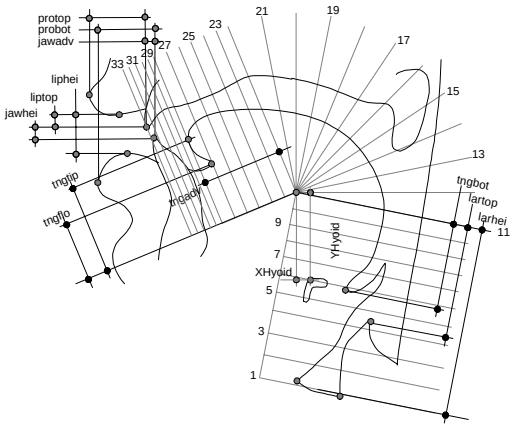


Figure 4 : Exemple de tracé manuel de contour du conduit vocal, de grille de lecture et de mesures articulatoires automatiques associées.

Articulographie électromagnétique. L'articulographe électromagnétique (Perkell, Cohen, Svirsky, Matthies, Garabieta & Jackson (1992; Schönle (1992)) permet de déterminer, avec une bonne résolution temporelle (plusieurs kHz), un petit nombre de points cutanés auxquels sont attachées de petites bobines réceptrices du champ électromagnétique calibré émis par de grosses bobines de référence. L'articulographe dont nous disposons à l'ICP permet l'acquisition de 5 points dans le plan médiosagittal, avec une résolution spatiale de l'ordre du millimètre. Cet outil est complémentaire de la cinéradiographie, car il peut fournir un grand nombre de données à des fréquences d'échantillonnage élevées, mais ne permet pas une description précise de la géométrie des divers articulateurs. Nous avons cependant montré (Badin *et al.* (1997)) qu'il est possible, dans certaines conditions, de retrouver la forme complète de la langue à partir des coordonnées de trois points sur celle-ci et de la connaissance de la hauteur de la mâchoire, par inversion d'un modèle articulatoire. Nous avons également pu effectuer grâce à l'articulographe, une étude des mouvements du velum pour les nasales dans le cadre du travail de thèse de Solange Rossato (Rossato, Badin & Feng (2000), Rossato (2000)).

Imagerie par Résonance Magnétique. L'une des toutes premières utilisations de l'IRM en parole remonte à Rokkaku, Hashimoto, Imaizumi, Niimi & Kiritani (1986), alors que l'utilisation de l'imagerie par rayons X date des années vingt (*cf.* Russell (1928)). L'IRM présente le très grand avantage de fournir des images qui sont de véritables *coupes* de la tête du sujet, et non pas des projections comme dans le cas d'images obtenues par cinéradiographie. Par ailleurs, moyennant des temps d'acquisition de l'ordre de 40-50 sec., il est possible d'obtenir des piles d'images parallèles à partir desquelles on peut ensuite effectuer des reconstructions tri-dimensionnelles. L'inconvénient majeur de l'IRM reste les temps d'acquisition relativement longs, même s'il devient possible de réaliser 10 à 12 coupes simples par seconde (Demolin, Metens & Soquet (2000)). Par contre, jusqu'à présent, aucune nocivité particulière n'a pu être mise en évidence pour les sujets, ce qui permet d'obtenir des quantités de données plus importantes.

J'ai ramené mes premières images IRM des laboratoires ATR à Kyoto, Japon, en 1996. Une collaboration plus locale a ensuite pu être mise en place avec le service de radiologie du CHRU de Grenoble et l'*UM Université Joseph Fourier / INSERM U438, LRC CEA*. J'ai alors développé un ensemble de logiciels permettant la détection semi-automatique de contours, l'édition et la correction manuelle, le calibrage et le recalage des données entre elles, afin de pouvoir traiter les images brutes et obtenir des données exploitables pour la modélisation. Ces outils ont également constitué la base des travaux de deux étudiants analysant le même type de données (Apostol, Perrier, Baciu, Segebarth & Badin (2000), Engwall & Badin (1999), Engwall & Badin (2000)).

Le principe de traitement des données, qui s'applique aussi bien aux contours du conduit vocal (Badin, Bailly, Raybaudi & Segebarth (1998b)) qu'aux contours de la langue (Badin, Bailly, Revéret, Baciu, Segebarth & Savariaux (In revision)), peut se résumer de la manière suivante : les contours sont repérés et tracés à l'aide d'un logiciel interactif d'édition de contours sur chacune des images originales (voir exemples à la Figure 5). Les contours sont ensuite corrigés à l'aide de reconstruction de moulages des parties rigides (l'IRM ne permet pas de faire apparaître les structures osseuses), et ré-interpolés suivant le système de grille semi-polaire adapté à la morphologie du conduit vocal (Figure 5). Les divers modèles issus de ces données sont présentés et discutés dans la section II.D.3.

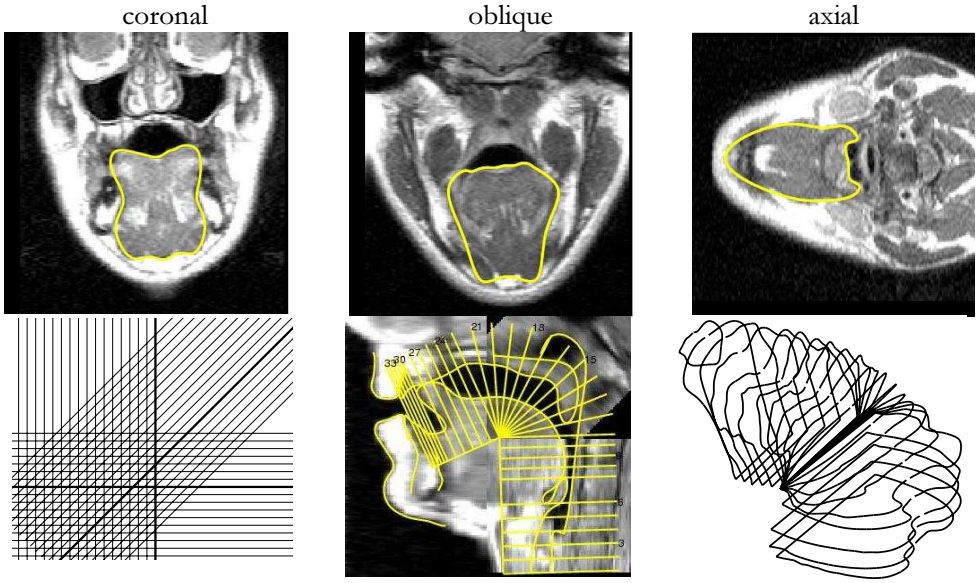


Figure 5 : Exemples de contours de langue superposés à des images IRM pour [l[a]] (en haut); trace dans le plan médiosagittal des trois piles d'image (en bas à gauche; les traces en gras correspondent aux images du haut); lignes de grille et contour médiosagittaux superposés à l'image médiosagittale reconstruite à partir des trois piles initiales (en bas au milieu) ; contours planaires 3D finaux dans le système de coordonnées semi-polaires (en bas à droite).

	Méthode	Résolution temporelle	Résolution spatiale	Commentaires
Aérodynamique / Acoustique	Microphone acoustique			Mesure du signal acoustique.
	Mesure de fonctions de transfert acoustique			Banc de mesure par excitation transcutanée au niveau du larynx et mesure de pression aux lèvres.
	Capteur de pression			Mesure de pression basse fréquence.
	Pneumotachographe de Rothenberg			Mesure du débit acoustique grâce à un masque facial percé de trous recouverts d'une grille métallique. Perturbe légèrement l'articulation.
Géométrique / Articulatoire	Imagerie par RMN (IRM)	~ 0.01 Hz (~45 sec. / articulation)	plans 2D continus	25-50 coupes 2D. Reconstruction 3D possible. Durée d'acquisition encore longue. Structures osseuses non visibles. La position couché sur le dos perturbe légèrement l'articulation.
	Téléradiographie	~ 0.5 Hz (~2 sec. / articulation)	continu 2D	Projection sur un plan de l'ensemble des structures de la tête. Complet, mais difficile à tracer. Dangereux pour la santé.
	Labiométrie vidéo	50 / 400 Hz	continu 2D ½ / ~200 points 3D	Enregistrement vidéo mono/multi caméra du visage avec marqueurs (maquillage lèvres, billes collées sur le visage). Reconstruction 3D.
	Cinéradiographie	50 Hz	continu 2D	*Cf.* téléradiographie, mais en mouvement. Films très long à dépouiller manuellement.
	Articulographie Electromagnétique (EMA)	400 Hz	5 points 2D	Bobines électromagnétiques collées sur la langue, les dents, le velum. Perturbe légèrement l'articulation.
	ElectroPalatoGraphie (EPG)	50 Hz	~64 points 3D	Détecte les contacts entre la langue et un palais artificiel porté par le sujet. Perturbe légèrement l'articulation.
	Imagerie par échographie ultrasonique de la langue	50 Hz	continu 2D	Principe de l'échographie. Seule la partie centrale de la langue est observable. Calage par rapport aux structures osseuses non maîtrisé.
	Vidéo rapide des cordes vocales	4000 Hz	continu 2D	Images des cordes vocales vues de dessus par l'intermédiaire d'une fibre optique nasale.

Tableau I : Méthodes de mesures en production de parole.

Propagation acoustique	**Rayonnement**	• Badin, P., Motoki, K., Miki, N., Ritterhaus, D. & Lallouache, T.M. (1994) Some geometric and acoustic properties of the lip horn.
	Mesure de fonction de transfert acoustique	• Castelli, E. & Badin, P. (1988) Vocal tract transfer functions measurements with white noise excitation. Application to the naso-pharyngeal tract. • Djéradi, A., Guérin, B., Badin, P. & Perrier, P. (1991) Measurement of the acoustic transfer function of the vocal tract : a fast and accurate method. • Pham Thi Ngoc, Y. & Badin, P. (1994) Vocal tract acoustic transfer function measurements : further developments and applications.
Aérodynamique / acoustique	**Capteurs de pression – Masque pneumo-tachographique de Rothenberg**	• Badin, P. (1989) Acoustics of voiceless fricatives : production theory and data. Badin, P., Hertegård, S. & Karlsson, I. (1990) Notes on the Rothenberg mask. • Stromberg, K., Scully, C., Badin, P. & Shadle, C.H. (1994) Aerodynamic patterns as indicators of articulation and acoustic sources for fricatives produced by different speakers. • Pelorson, X., Lallouache, T.M., Tourret, S., Bouffartigue, C. & Badin, P. (1994) Modeling the production of bilabial plosives : aerodynamical, geometrical and mechanical aspects. • Badin, P., Mawass, K. & Castelli, E. (1995) A model of frication noise source based on data from fricative consonants in vowel context. • Mawass, K., Badin, P. & Bailly, G. (2000) Synthesis of French fricatives by audio-video to articulatory inversion.
Articulatoire 2D	**Télé- et ciné-radiographie**	• Badin, P. (1991) Fricative consonants : acoustic and X-ray measurements. • Beautemps, D., Badin, P. & Laboissière, R. (1995) Deriving vocal-tract area functions from midsagittal profiles and formant frequencies : A new model for vowels and fricative consonants based on experimental data. • Badin, P., Gabioud, B., Beautemps, D., Lallouache, T.M., Bailly, G., Maeda, S., Zerling, J.P. & Brock, G. (1995) Cineradiography of VCV sequences : articulatory-acoustic data for a speech production model. • Badin, P. & Abry, C. (1996) Articulatory synthesis from X-rays and inversion for an adaptive speech robot. • Beautemps, D., Badin, P. & Bailly, G. (2001) Linear degrees of freedom in speech production : Analysis of cineradio- and labio-film data and articulatory-acoustic modelling.
	Articulographie électromagnétique	• Badin, P., Baricchi, E. & Vilain, A. (1997) Determining tongue articulation : from discrete fleshpoints to continuous shadow. • Rossato, S., Badin, P. & Feng, G. (2000) Estimation des mouvements du voile du palais à partir du signal de parole pour les voyelles nasales du Français. • Rossato, S., Feng, G., Laboissière, R. & Badin, P. (in preparation) Acoustic effects of velum movements in French nasal vowels.
Articulatoire 3D	**Vidéo + IRM**	• Badin, P., Borel, P., Bailly, G., Revéret, L., Baciu, M. & Segebarth, C. (2000) Towards an audiovisual virtual talking head : 3D articulatory modeling of tongue, lips and face based on MRI and video images. • Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C. & Savariaux, C. (2001) Three-dimensional articulatory modeling of tongue, lips and face, based on MRI and video images.
	IRM	• Badin, P., Bailly, G., Raybaudi, M. & Segebarth, C. (1998) A three-dimensional linear articulatory model based on MRI data. • Engwall, O. & Badin, P. (1999) Collecting and analysing two- and three-dimensional MRI data for Swedish.
	Video	• Mawass, K., Badin, P. & Bailly, G. (2000) Synthesis of French fricatives by audio-video to articulatory inversion. • Borel, P., Badin, P., Revéret, L. & Bailly, G. (2000) Modélisation articulatoire linéaire 3D d'un visage pour une tête parlante virtuelle. • Revéret, L., Bailly, G., Borel, P. & Badin, P. (2000) Analyse par la synthèse d'un visage 3D parlant : inversion optico-articulatoire.

Tableau II : Différentes méthodes de mesures développées ou mises en oeuvre dans les publications.

Labiométrie et mesure du visage. Un premier dispositif de labiométrie a été développé à l'ICP par Lallouache (1991). Ce système, basé sur une détection automatique des contours labiaux sur des images vidéos de face et de profil du visage du locuteur dont les lèvres ont été maquillées en bleu, m'a permis d'obtenir plusieurs ensemble de données labiales très importantes : pour l'acoustique du pavillon labial (Badin *et al.* (1994a)), pour la modélisation articulatoire (Badin *et al.* (1995b), Beautemps *et al.* (2001)), et pour l'inversion de paramètres articulatoires à partir du signal audio-visuel (Mawass *et al.* (2000)). Ce système permettait d'obtenir un nombre limité de paramètres labiaux (aperture, étirement, protrusion), mais ne permettait pas une reconstruction 3D complète et fidèle de la géométrie du pavillon labial. Afin de pallier ce problème, Revéret (1999) a développé un modèle générique 3D de lèvres (voir aussi Revéret & Benoît (1998)). C'est cette méthode que nous avons utilisée pour analyser un corpus complet d'images vidéo de face et de profil (Borel (1999), Badin *et al.* (In revision)). Nous avons également complété cette méthode par l'adjonction de petites billes collées sur divers points du visage, et dont les coordonnées 3D peuvent ensuite être reconstruites à partir des images de face et de profil (voir Figure 6). Notons que ce corpus a été enregistré une deuxième fois avec une éclisse mandibulaire (voir Figure 6) afin de nous permettre de retrouver la position exacte de la mâchoire dans le corpus original, à l'aide d'un modèle de prédiction à partir des coordonnées d'un certain nombre de points cutanés.

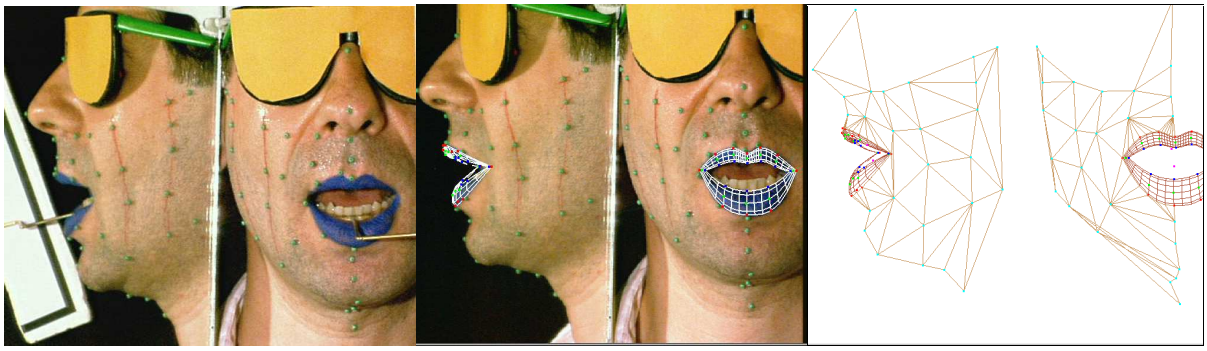


Figure 6 : Exemple d'image vidéo pour un /a/ : sujet équipé avec une éclisse mandibulaire (à gauche); sujet avec superposition du maillage du modèle de lèvres adapté (au milieu); maillage complet dont les coordonnées 3D des sommets constituent les mesures articulatoires.

3. Modèles

Grandes classes de modélisation articulatoire. Parmi les nombreuses études consacrées à la modélisation articulatoire depuis les années soixante-dix, deux approches principales peuvent être identifiées : la *modélisation articulatoire fonctionnelle*, selon laquelle la position et la forme des articulateurs sont des fonctions algébriques d'un petit nombre de paramètres articulatoires, et la *modélisation biomécanique explicite*, selon laquelle la position et la forme des articulateurs sont calculées à partir de simulations physiques des forces générées par les muscles et par leurs conséquences sur les articulateurs.

Dans les modèles articulatoires *linéaires*, les relations entre les positions et formes des articulateurs et les paramètres de contrôle peuvent être définies soit en termes géométriques explicites, auquel cas les degrés de liberté du robot articulatoire sont décidés *a priori* et ajustés aux données *a posteriori* (*cf.* Coker & Fujimura (1966), Liljencrants (1971), Mermelstein (1973)), soit basé sur les données articulatoires mesurées sur un ou plusieurs sujets, auquel cas les degrés de liberté du robot émergent des données (*cf.* Lindblom & Sundberg (1971), Maeda (1979a), Maeda (1990), Maeda (1991), Stark, Lindblom & Sundberg (1996)).

L'approche générale des modèles biomécaniques consiste à modéliser les forces musculaires et les structures articulatoires au moyen de méthodes inspirées de l'analyse mécanique et de la simulation numérique (*cf.* Perkell (1974), Wilhelms-Tricarico (1995), Payan (1996), Laboissière *et al.* (1996), Payan & Perrier (1997)). Ces modèles présentent l'avantage de la modélisation physique, qui inclut donc une dynamique intrinsèque – même si elle est nécessairement simplifiée –, mais leur contrôle reste très complexe, en particulier à cause du nombre élevé de degrés de liberté représentés par les commandes individuelles de chaque muscle. Le travail de Sanguineti, Laboissière & Ostry (1998) constitue une bonne illustration de cette situation. Ils ont ajusté, avec leur modèle biomécanique de mâchoire, os hyoïde et langue, les formes et les positions des articulateurs mesurés sur une base de données issue du film

cinéradiographique déjà utilisée par Maeda (1990), et ils ont déterminé par optimisation les commandes des dix-sept muscles impliqués dans leur modèle. Ils ont ensuite identifié, par analyse en composantes principales appliquée à l'espace des commandes-λ correspondant à l'espace des paramètres de contrôle biomécanique de la langue et de la mâchoire, les synergies entre ces commandes, et ont montré que six composantes décorrélées pouvaient rendre compte de la majeure partie de la variance de la forme de langue dans le plan médiosagittal. Or ces six premières composantes sont très liées aux degrés de liberté qui peuvent être extraits directement des contours originaux du film cinéradiographique. Il apparaît donc que, du point de vue des degrés de liberté, il n'est pas nécessaire de recourir à des modèles biomécaniques aussi complexes pour développer des descriptions précises de l'articulation statique en parole.

Notre approche : émergence des degrés de liberté des données. Ainsi, plutôt que de développer des modèles biomécaniques complexes dotés d'un large excès de degrés de liberté et de réduire ensuite cette dimensionalité élevée sur la base de données articulatoires, nous avons adopté l'approche duale. Un de mes objectifs était donc de déterminer les degrés de liberté des articulateurs d'un sujet et de construire un *robot articulatoire* qui puisse être considéré comme une représentation fidèle des capacités articulatoires du sujet. Cette recherche a fait l'objet d'un nombre important de publications (Badin *et al.* (1998b), Bailly, Badin & Vilain (1998), Vilain, Abry & Badin (1998a), Beautemps *et al.* (2001), Badin *et al.* (In revision)).

Principes de la détermination des degrés de liberté : analyse en composantes linéaires. Avant de décrire de manière plus détaillée les différents modèles que j'ai développés, je présente ici les principes généraux de cette approche, qui sont les mêmes, quelques soient les données analysées.

Le sujet. Le choix du sujet pose toujours un dilemme : une étude portant sur un sujet unique réduit la généralité du travail, mais permet de collecter un ensemble riche et détaillé de données, tandis qu'une étude faisant intervenir un large éventail de sujets peut éventuellement permettre de tirer des conclusions plus générales, mais limite l'étendue des données qui peuvent être réellement analysées en pratique. Le mélange au niveau des statistiques des données de plusieurs sujets présente aussi le risque de rendre floues les stratégies articulatoires utilisées par chaque sujet, et d'empêcher ainsi de trouver ni résultats généraux ni résultats spécifiques à un sujet. Notre approche a donc consisté à étudier de manière très détaillée un seul sujet. Plus récemment, nous avons en outre commencé à étudier d'autres sujets selon la même approche, et à les comparer deux à deux (Vilain *et al.* (1998a), Bailly *et al.* (1998), Engwall & Badin (1999), Engwall & Badin (2000)). Cette approche *non normalisatrice* de la modélisation, que l'on pourrait qualifier d'*orientée sujet*, permet d'extraire pour chaque sujet les degrés de liberté des articulateurs tout en préservant la conformation et les stratégies de synergie individuelles de chacun d'entre eux, sans brouiller les pistes.

Le corpus. Comme nous l'avons vu plus haut, l'un des principaux enjeux de la modélisation articulatoire consiste à déterminer les degrés de liberté d'un sujet pour la parole, en excluant les mouvements qui ne correspondent pas à de la parole, tels que la mastication par exemple. L'atteinte de ce but nécessiterait, de manière idéale, un très large corpus de matériau de parole, qui contienne toutes les combinaisons possible des phonèmes. Mais ceci n'est évidemment pas réalisable en pratique, que ce soit en cinéradiographie, en IRM, ou même avec des méthodes moins lourdes ou nocives telles que la labiométrie par vidéo par exemple. Il est donc indispensable de concevoir des corpus condensés, qui soient à la fois compatibles avec les contraintes des méthodes expérimentales utilisées, et qui assurent en même temps une couverture maximale de l'espace des réalisations possibles pour la parole. Nous avons évalué cette hypothèse lors de l'élaboration d'un modèle articulatoire à partir de contours cinéradiographiques (Badin *et al.* (1998b)). Nous avons montré qu'il était possible de construire un modèle articulatoire à partir d'un sous-ensemble du corpus ne contenant que les cibles vocaliques et consonantiques qui soit quasiment aussi précis dans la reconstruction des données que le modèle établi sur l'ensemble du corpus (erreurs RMS de reconstruction de 0,11cm et 0,09cm respectivement pour des corpus de 20 et 1222 trames).

Principes : Identification des degrés de liberté. Notre approche, qui vise à déterminer les degrés de liberté des divers articulateurs de la parole, doit être basée sur les données articulatoires mesurées sur *un sujet* donné produisant *un corpus* donné, dans *une langue* donnée.

En général, les articulateurs de la parole possèdent des degrés de liberté en excès, c'est-à-dire qu'une articulation donnée peut être réalisée au moyen de différentes combinaisons des degrés de liberté disponibles pour les articulateurs (*cf.* les expériences de *bite-blocks* de Lindblom, Lubker & Gay (1979), ou les tubes labiaux de Savariaux, Perrier & Orliaguet (1995)). En fin de compte, les stratégies de contrôle visent à recruter ces degrés de liberté quand ils sont nécessaires pour atteindre des cibles articulatoires / acoustiques / visuelles données, et à les laisser libres d'anticiper ou de maintenir d'autres cibles à chaque

fois que possible (ceci est le principe même de la *coarticulation*). Dans notre approche basée sur les données, se pose le problème crucial de décider de la répartition de la variance des variables articulatoires mesurées entre les différentes variables associées aux degrés de liberté. Nos travaux s'appuient donc sur le principe suivant, classique en modélisation du contrôle moteur : ce qui est expliqué par la biomécanique du robot articulatoire n'a pas besoin d'être généré par le contrôleur (Abry *et al.* (1994), Perrier *et al.* (1996), Perrier, Payan, Perkell, Jolly, Zandipour & Matthies (1998)). En d'autres termes, toute corrélation observée entre les variables articulatoires devrait être utilisée pour réduire le nombre de degrés de liberté des articulateurs. Cependant, cette approche doit être soigneusement contre-balancée par un autre critère, la *vraisemblance biomécanique*. Par exemple, si le larynx et la protrusion labiale sont inversement corrélés, suite à une stratégie de contrôle articulatoire délibérée (*cf.* II.D.1, Hoole & Kroos (1998), à titre d'exemple), deux degrés de liberté séparés devraient cependant être considérés, afin d'assurer la *vraisemblance biomécanique*, même au prix d'une certaine corrélation résiduelle entre les paramètres correspondants.

Principes : Analyse en composantes linéaires. La *linéarité* de l'analyse et du modèle associé constitue une autre hypothèse importante : les vecteurs des données de forme *DT* sont décomposés en une forme moyenne (neutre) sur le corpus, $\overline{DT}$, à laquelle s'ajoutent les combinaisons linéaires d'une série de vecteurs de formes de base *BV* pondérés par des prédicteurs *LF* :

$$DT = \overline{DT} + LF \cdot BV \ .$$

Chaque prédicteur LF_i correspond à une composante linéaire si ses corrélations croisées avec chacun des autres prédicteurs sont nulles sur l'ensemble des données du corpus. La dimensionalité des formes et des positions des articulateurs peut donc être explorée par des techniques classiques d'analyse linéaire comme l'ACP (Analyse en Composantes Principales, *cf.* Lebart, Morineau & Piron (1995)) et l'analyse par régression linéaire multiple, suivant les travaux de Maeda (1990, (1991) qui ont très largement inspiré notre approche.

La méthode générale pour cette décomposition consiste à déterminer de manière itérative chaque composante linéaire de la façon suivante :

1. Le prédicteur LF_i est déterminé à partir de l'ensemble ou d'un sous-ensemble des variables du résidu courant des données;
2. Le vecteur de base associé est déterminé par la régression linéaire, par rapport au prédicteur LF_i, des résidus courants des données pour l'ensemble du corpus ;
3. La contribution correspondante de la composante, calculée comme le produit des prédicteurs par ce vecteur de base, est soustraite au résidu courant pour fournir le prochain résidu.

Pour certaines des composantes, les prédicteurs sont choisis arbitrairement comme les valeurs centrées et normalisées de mesures géométriques spécifiques extraites des données, telles que la position de la mâchoire ou la hauteur du larynx. Pour les autres composantes linéaires, les prédicteurs sont déterminés par une ACP standard appliquée à des régions spécifiques de l'ensemble des contours, comme la pointe de la langue par exemple.

Il faut noter que la solution à ce type de décomposition linéaire n'est pas unique en général : l'ACP fournit une explication maximale de la variance des données avec un nombre minimal de composantes. L'approche présentée ici laisse une certaine liberté de manœuvre pour contrôler la nature et la répartition de la variance expliquée par les composantes (pour les rendre plus interprétables en termes de contrôle par exemple), au prix d'une explication de la variance sous-optimale.

Dans les sections ci-dessous, sont décrits un peu plus en détails un certain nombre de modèles articulatoires que j'ai développés, ainsi que les résultats auxquels ils ont permis d'aboutir : modèles médiosagittaux pour deux sujets (*P1X* et *J1X*), modèles 3D de conduit vocal, de langue, de velum, de lèvres et de visage pour le sujet *P1X*.

Modèles 2D. À partir des données extraites du film cinéradiographique réalisé sur le sujet *P1X*, nous avons développé un modèle articulatoire médiosagittal linéaire. Le corpus était conçu pour inclure autant de combinaisons de séquences VCV que possible dans un temps très limité. Les consonnes voisées occlusives et fricatives du Français C = [v z ʒ b d g] étaient insérées dans six contextes vocaliques différents impliquant les quatre voyelles françaises extrêmes V = [a i u y] : aCa, aCi, aCu, iCi, iCu, iCy. De plus, une série de voyelles connectées [a ɛ e i y u o ø] était incluse dans le corpus pour tester des hypothèses sur les affiliations formant / cavité (*cf.* Bailly (1995)). La durée du corpus était finalement

d'environ 25 secondes (1222 images), avec la statistique de répartition suivante : 30 [a i], 12 [u], 6 [y], 6 [v z ʒ b d g], 18 [p].

Nous avons montré que 96% de la variance de la langue peuvent être expliqués par quatre composantes. En tant qu'objet rigide, la mâchoire possède potentiellement six degrés de liberté, mais pour des tâches de parole, au maximum deux ou trois sont réellement impliqués (Ostry, Vatikiotis-Bateson & Gribble (1997)). Nous caractérisons ses mouvements dans le plan médiosagittal par les coordonnées du bord supérieur des incisives inférieures, et nous avons trouvé que 97% de la variance étaient expliqués par le mouvement vertical seul. Nous avons en outre montré que l'influence du mouvement horizontal sur la forme de la langue était tout à fait négligeable. La mâchoire étant le principal articulateur porteur de la langue, nous avons imposé son mouvement comme premier prédicteur de la forme de la langue. Après avoir soustrait cette première contribution, nous avons effectué une ACP sur la partie de la langue située entre la racine, au dessus de l'épiglotte, et l'apex, environ 1,5 cm en arrière de l'extrémité. Deux composantes émergent très clairement : une composante liée à un mouvement arrière-bas / avant-haut du corps, tandis que l'autre correspond à un mouvement d'arrondissement plus ou moins important du corps de la langue. Les nomogrammes de la Figure 7 illustrent bien ces caractéristiques. Les contributions de ces deux composantes à l'ensemble des données (c'est-à-dire y compris les parties de la langue qui ont été exclues de l'analyse qui a déterminé ces composantes) sont ensuite extraites des données, et une ACP est alors appliquée à l'apex seulement. Une nouvelle composante est ainsi déterminée (voir Figure 7). Ces résultats sont publiés dans Beautemps *et al.* (2001) (voir l'article annexé en fin de mémoire).

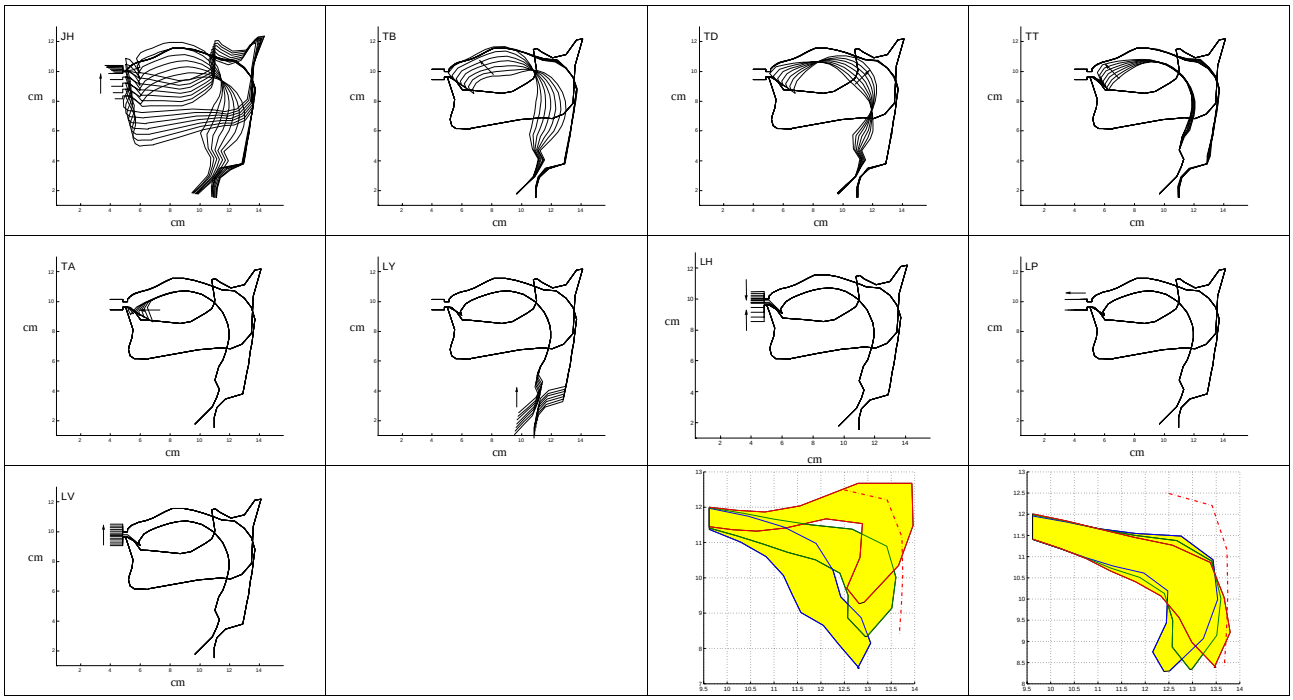


Figure 7 : Nomogrammes du modèle mâchoire / langue / lèvres Bergame *du sujet P1X : variations de –3 à +3 par pas de +1 des différents paramètres de contrôle du modèle. Notons que les mouvements verticaux des lèvres inférieure et supérieure s'effectuent en sens inverse pour le nomogramme de* LH, *mais dans le même sens pour le nomogramme de* LV. *En bas à droite, nomogrammes du velum (valeurs –3, 0, +3 des paramètres* VL *(gauche) et* VV *(droite).*

Afin de compléter ce modèle, nous avons utilisé un ensemble d'images médiosagittales obtenues par IRM, et nous avons développé un modèle de velum de la même manière. Le velum est décrit par une grille polaire de même centre que la grille semi-polaire utilisée pour le reste du conduit vocal. Cette grille se déploie en tournant dans le sens inverse au sens trigonométrique d'une position verticale fixe à une droite proche de l'horizontale astreinte à passer par l'extrémité de la luette, suivant le même principe que pour la pointe de la langue ; l'angle de cette dernière *VelRot* détermine donc l'élévation globale du velum. Les faces interne et externe du velum sont alors échantillonnées selon les différentes droites de la grille mobile. Une composante suffit à expliquer 84,7 % de la variance de la face interne du velum dans le plan médiosagittal. Une deuxième composante explique 11,3 % supplémentaires, amenant à un total de 96 %,

mais semble être liée à un mouvement du velum dû à un recul de la langue qui entraîne la luette. La Figure 7 illustre ces deux mouvements.

Modèles de passage 2D-3D. La section transversale du conduit vocal étant loin d'être circulaire, la donnée du seul diamètre dans le plan médiosagittal (la *distance sagittale*) ne permet pas de prédire l'aire transversale correspondante. Il est donc nécessaire de déterminer une loi qui permette une prédiction de cette aire de manière approchée. Le modèle « α–β » de Heinz & Stevens (1965), largement employé dans la littérature, suppose une relation du type $A = \alpha \cdot d^{\beta}$ entre l'aire A et la distance sagittale d. Les coefficients α et β sont fixés de manière plus ou moins *ad hoc*, et peuvent être adaptés aux régions du conduit vocal auxquelles ils sont appliqués (*cf.* Perrier, Boë & Sock (1992)). Dans le cadre de la thèse de Denis Beautemps, nous avons développé une loi de type $A(x, d) = \alpha(d, x) \cdot d^{\beta}$ où la fonction α a été optimisée à partir des mesures simultanées de coupes sagittales et de formants effectuées sur le sujet *P1X* (Beautemps, Badin & Laboissière (1995), Badin, Beautemps, Laboissière & Schwartz (1995a)). Plus récemment, nous avons développé un modèle de type $A(x, d) = \alpha_1(x) \cdot d + \alpha_2(x) \cdot d^{1,5} + \alpha_3(x) \cdot d^2 + \alpha_4(x) \cdot d^{2,5}$ basé sur le même principe d'optimisation et qui fonctionne pour *P1X* et *J1X* (Beautemps *et al.* (2001)).

Modèle 3D de conduit vocal. Les modèles médiosagittaux sont limités par un certain nombre de problèmes : (1) il est nécessaire d'inférer la fonction d'aire à partir de la coupe sagittale, (2) les consonnes latérales, qui présentent une occlusion dans la zone médiosagittale mais aussi (*cf.* ci-dessous) des canaux latéraux, ouverts ne peuvent pas être traitées par ces modèles, et (3) les modes acoustiques transverses qui se propagent à partir de 4-5 kHz ne peuvent pas être pris en compte. L'intérêt de disposer de modèles réellement tridimensionnels apparaît donc clairement. En outre, ces modèles ouvrent des possibilités d'accéder à la réalité virtuelle par l'élaboration de têtes parlantes audiovisuelles utiles pour la synthèse audiovisuelle, l'aide à l'apprentissage des langues, etc.

Le premier modèle 3D que j'ai développé a été un modèle de *conduit* vocal, c'est-à-dire un modèle des parois du conduit vocal considéré comme un tuyau présentant une section transversale fermée. La méthodologie décrite plus haut a été adoptée pour construire ce modèle. Le corpus se composait de 20 articulations produites par le sujet *P1X*. La partie du conduit en aval de la pointe de la langue était exclue, de façon à assurer que toutes les coupes transversales soient fermées et puissent être analysées statistiquement sans problème. Le conduit était échantillonné selon 28 contours planaires déterminés dans les plans de la grille semi-polaire décrite plus haut, chaque contour étant défini par 51 paires de coordonnées X/Y ; la coordonnée X correspond à la *dimension latérale* du contour, tandis que la coordonnée Y correspond à un déplacement dans la direction des droites de la grille de référence, que nous appelons *dimension sagittale*.

Ces contours sont décomposés en composantes linéaires suivant la méthodologie décrite plus haut. Afin que le modèle 3D puisse constituer une extension du modèle 2D développé pour le sujet *P1X*, nous imposons comme premiers prédicteurs de l'ensemble des coordonnées X/Y les paramètres JH, TB, TD, TT et TA. Pour chaque articulation, ces paramètres sont obtenus par inversion à partir du contour intérieur 2D de la langue constitué par les points d'intersection des contours planaires avec le plan médiosagittal. Le pourcentage de variance des données expliquée par l'ensemble des prédicteurs JH, TB, TD, TT et TA atteint 75 %. Les quatre facteurs supplémentaires obtenus par l'ACP des résidus de cette première analyse font monter l'explication de la variance à 94 %, mais aucun de ces facteurs n'est clairement interprétable : ils correspondent plus vraisemblablement à une reconstruction du bruit de mesure.

Les modèles 3D du conduit vocal résolvent intrinsèquement le problème du passage des contours sagittaux à la fonction d'aire : l'aire transversale d'un contour planaire se calcule en effet de manière triviale à partir des coordonnées X/Y du contour, et ne nécessite plus de formule de passage approchée. Une première évaluation a montré des erreurs RMS la plupart du temps en dessous de 1 cm^2, erreurs qui peuvent être imputées au fait que seulement 75 % de la variance des contours est reconstruite par les cinq paramètres. Les erreurs induites sur les formants par ces erreurs de fonctions d'aire se situent au dessus des seuils de sensibilité différentielle déterminés par Flanagan (1955) (5 % pour F1 et 3 % pour F2) même si elles sont limitées (17 % pour F1, 14 % pour F2, 8 % pour F3). Par ailleurs, il est apparu que les différences entre les aires prédites par le modèle 2D et celles prédites par le modèle 3D était elles aussi la plupart du temps inférieures à 1 cm^2, avec des erreurs induites sur les formants tout à fait comparables entre elles : cela prouve la qualité de formules de passage de la coupe sagittale à la fonction d'aire, mais cela suggère aussi le fait que les caractéristiques 3D du conduit vocal pourraient être largement prédictibles à partir de caractéristiques médiosagittales (voir ci-dessous le modèle 3D de langue).

Notons finalement que ces résultats ont été présentés à plusieurs conférences (dont Badin, Pouchoy, Bailly, Raybaudi, Segebarth, Lebas, Tiede, Vatikiotis-Bateson & Tohkura (1998c), Badin *et al.* (1998b)). À ma connaissance, aucun autre modèle de ce type n'a jamais été développé.

Modèle 3D de langue. La complexité de la forme du conduit vocal due aux articulateurs divers et variés qui le constituent limite l'approche précédente de modélisation globale 3D du conduit vocal. Ce cadre rendait en effet impossible la représentation correcte de la cavité sous-linguale ou du velum, par exemple. J'ai donc tenté une approche qui prend en compte séparément chacun des organes, et commencé par bâtir un modèle 3D de langue (Badin *et al.* (In revision), *cf.* annexe).

Le système d'échantillonnage de la forme de la langue a été décrit plus haut. Suivant la même méthode que pour le modèle 3D de conduit vocal, le modèle 3D de langue s'appuie sur le modèle médiosagittal élaboré à partir des mêmes données.

Les composantes n'ayant pas été déterminées par pure ACP, elles sont légèrement corrélées et la variance expliquée par ces cinq paramètres n'est pas optimale, mais seulement 8 % en dessous de la variance expliquée par les cinq premières composantes ACP orthogonales. Les 72 % de taux d'explication de la variance de la langue sont proche de 75 % trouvés pour le conduit vocal pour le même nombre de paramètres. L'ajout de quatre paramètres supplémentaires augmente l'explication jusqu'à 87 %, mais ces paramètres n'offrent aucune interprétation convaincante.

Les effets des cinq commandes du modèle de langue[1] pour les valeurs extrêmes (–3 et +3 écarts-types) de chacun des paramètres, tous les autres étant maintenus à zéro, peuvent être observés dans l'article Badin *et al.* (In revision) joint en annexe.

JH contrôle l'influence de la hauteur de la mâchoire sur la langue. Le déplacement *avant* / *arrière* de la masse de la langue est associé à TB. Nous avons observé qu'une grosse partie du sillon lingual caractéristique de la consonne [s] est obtenue grâce à ce paramètre. La propriété *plat* / *en arc* de la langue est prise en compte par TD, et se trouve aussi associée à un certain degré de creusement du sillon. Deux paramètres contrôlent la forme de la pointe de la langue : TT prend en compte les mouvements *haut* / *bas* des quatre dernières sections de la langue. TT est particulièrement actif pour la consonne [l^a] pour laquelle le corps de la langue est abaissé par l'action conjointe de JH et TB et le contact pointe de langue / maxillaire est assuré par une valeur élevée de TT. Finalement, le paramètre TA représente le résidu du geste d'avancée de la langue après soustraction des contributions de JH, TB, TD et TT : il prend en particulier en compte la surface inférieure de la pointe de la langue qui peut se trouver en contact avec, et donc déformé par la mandibule, les incisives inférieures et le plancher de la bouche en relation avec l'avancée de la langue.

L'explication totale de variance de la langue en 3D est plus faible que nous aurions pu souhaiter : ceci est dû à une reconstruction imparfaite de la région sous-linguale (63 % pour les fibres les plus basses). Cependant la surface supérieure est reconstruite à 89 %, proche des 96 % que nous obtenons pour la coupe médiosagittale du corpus cinéradiographique. Il apparaît donc que la partie de la langue la plus importante du point de vue de l'acoustique du conduit vocal est la mieux reconstruite : le modèle est donc plutôt satisfaisant.

J'avais démarré cette étude avec le présupposé qu'un ou deux paramètres supplémentaires seraient cruciaux pour prendre en compte les canaux latéraux de la consonne latérale et le creusement caractéristique de certaines articulations comme le [s]. De manière tout à fait inattendue, il est apparu que les caractéristiques géométriques 3D en dehors du plan médiosagittal pouvaient être approximativement prédites par les mêmes paramètres que celles du plan médiosagittal. En d'autres termes, la géométrie 3D complète de la langue pourrait être en grande partie prédite à partir du seul contour médiosagittal, du moins pour la parole. Ces résultats sont important du point de vue de la réduction du nombre de paramètres de contrôle, mais n'effacent pas l'intérêt des modèles 3D. En effet, bien que la connaissance acquise depuis de nombreuses années grâce aux données médiosagittales et aux traditionnels modèle 2D soit loin d'être dépassée, il est clair que seul les modèles 3D peuvent fournir la description exhaustive du conduit vocal à partir duquel les fonctions d'aire peuvent être déterminées pour les latérales comme pour d'autres articulations impliquant un sillon lingual important.

Remarquons que le seul autre modèle de langue construit à partir de données 3D IRM est celui développé pour le suédois par Olle Engwall (Engwall (2000a), Engwall (2000b)) à partir de données enregistrées à Grenoble. Pendant son séjour de trois mois à l'ICP en 1999, Olle Engwall a hérité de tous

[1] Une animation du modèle 3D peut être trouvée à l'adresse http://www.icp.inpg.fr/~badin/Modart3D.html

les logiciels que j'ai développés pour le traitement des données IRM et pour le développement de modèles 3D (Engwall & Badin (1999)). Il a suivi pratiquement la même méthodologie d'analyse en composantes linéaires guidée, et trouve des taux d'explication de la variance tout à fait comparables aux nôtres (Engwall (2000a), Engwall (2000b)).

Modèle 3D de velum. Tout récemment, lors d'un stage d'élève-ingénieur à l'ICP, Guillaume de Penguern a développé une première ébauche de modèle 3D de velum (de Penguern (2001)). Le velum 3D est décrit par la même grille polaire que le velum 2D. Les faces interne et externe sont ainsi échantillonnées dans les plans perpendiculaires au plan sagittal et passant par les différentes droites de la grille mobile. Le premier facteur déterminé par une ACP standard explique 64 % de la variance globale 3D du velum, et correspond clairement à un mouvement d'ouverture / fermeture du port vélopharyngé. Un deuxième facteur accroît l'explication de la variance jusqu'à 68 %, mais semble lié à un recul de la langue. Notons que ce taux d'explication de la variance est proche de celui obtenu pour la langue.

Modèles de lèvres / visage. Pascal Borel a développé le premier modèle 3D de lèvres et de visage basé sur un corpus de données géométriques 3D (Borel (1999), Borel, Badin, Revéret & Bailly (2000), Badin, Borel, Bailly, Revéret, Baciu & Segebarth (2000), Badin *et al.* (In revision)). Le corpus et les données 3D ont été décrits plus haut ; on dispose en outre de mesures articulatoires dans le plan médiosagittal à partir des images de profil du même corpus selon la méthodologie utilisée pour les films cinéradiographiques : hauteur des lèvres *LipHei*, protrusion de la lèvre supérieure *ProTop*, et élévation de la lèvre supérieure *LipTop*, en sus de la hauteur et de l'avancée de la mâchoire *JawHei* et *JawAdv*. Nous avons développé deux modèles, suivant le même principe d'analyse en composantes linéaires décrit plus haut : l'un dont les prédicteurs sont basés sur les mesures géométriques médiosagittales et qui est donc compatible avec le modèle médiosagittal développé à partir du film cinéradiographique de *P1X*, et l'autre basé sur une analyse directe des données 3D.

La différence entre les deux approches se situe au niveau de la détermination des composantes non liées directement à la mâchoire : pour le premier modèle, les trois paramètres principaux *LP*, *LH*, *LP*, sont directement mesurés dans le plan médiosagittal, tandis que pour le deuxième modèle ils sont extraits par ACP appliquée aux données 3D de lèvres (*lips1*, *lips2*, *lips2*). Les résultats sont très similaires en termes d'explication de la variance des données complètes ; de plus, de fortes corrélations ont été trouvées entre les paramètres *lips1*, *lips2*, *lips3* et *LP*, *LH*, *LV* respectivement (coefficients : 0.98, 0.89 et 0.83). On peut conclure qu'il est donc possible de contrôler l'ensemble complet du modèle 3D à partir de paramètres mesurés dans le plan médiosagittal. Il est également très intéressant de constater que ces paramètres correspondent au traditionnel système de traits phonétiques de la labialité (Abry & Boë (1986)) : *LP* / *lips1* contrôle le geste de protrusion – arrondissement; *LH* / *lips2* contrôle l'aperture; *LV* / *lips3* contrôle le mouvement vertical quasi-simultané des deux lèvres utilisé par le sujet *P1X* pour la réalisation des consonnes labiodentales (et aussi pour les lèvres ouvertes et protrues dans le cas des consonnes [ʃ ʒ]).

Modèle de lèvres / visage avec opposition neutre / sourire. Lors de son stage de DEA à l'ICP, Yan Morvan a doté le modèle lèvres–visage d'un système d'opposition neutre / sourire (Morvan (2000)). Plus précisément, un enregistrement supplémentaire du corpus utilisé pour le modèle de lèvres–visage avait été réalisé avec le sujet *P1X*, avec la consigne de produire les diverses articulations en affichant un large sourire. Nous disposions donc pour chacune des articulations d'une version *neutre* et d'une version *sourire*. Une analyse de type PARAFAC (Harshman & Lundy (1984)) a permis d'établir un modèle articulatoire commun aux deux modes, et un jeu de coefficients qui pondèrent les commandes pour chacun des deux modes. On peut ensuite passer progressivement d'un mode à l'autre en interpolant linéairement ces coefficients entre leurs valeurs pour les deux modes. Ces premiers résultats se sont révélés convaincants, mais restent à établir de manière plus systématique.

Conclusions sur les données et modèles articulatoires. Une grande partie des données articulatoires décrites dans les sections précédentes a été accumulée dans le but de couvrir au maximum le champ des possibilités articulatoires du sujet *P1X* pour la parole. Les modèles articulatoires qui en sont extraits constituent donc une sorte de boîte à outils qui permet de représenter le plus fidèlement possible les capacités d'articulation de ce sujet. Ces modèles, qui permettent de manipuler de manière cohérente des formes de conduit vocal réalistes, peuvent donc être utilisés d'une part pour étudier les stratégies articulatoires employées par le locuteur sur lesquels ils sont basés, et forment d'autre part le cœur des têtes parlantes audiovisuelles développées à l'ICP.

4. Etudes de stratégies de contrôle articulatoire

Je présente ici deux études de stratégies de contrôle articulatoires qui ont pu être menées grâce à la faculté de décomposition en composantes linéaires qu'offre la modélisation articulatoire décrite plus haut.

Synergie mâchoire-langue. La démarche utilisée pour développer le modèle médiosagittal de *P1X* a été appliquée à deux autres sujets. Ainsi le modèle *Patricia* est issu du sujet *B1X* (hérité de Maeda (1979a)), et le modèle *Gentiane* du sujet *J1X* (Vilain *et al.* (1998a)). Ces données et ces modèles nous ont donc permis de comparer les stratégies de synergie mâchoire-langue pour trois sujets français (Bailly *et al.* (1998)). Nous avons en particulier observé que pour les sujets *J1X* et *B1X* la composante du mouvement de la langue linéairement corrélée au mouvement d'élévation de la mâchoire (mesuré comme la coordonnée verticale de l'incisive inférieure) pouvait atteindre quasiment deux fois l'amplitude du mouvement de la mâchoire : ce phénomène ne peut s'expliquer en considérant seulement l'aspect *organe porteur* de la mâchoire pour la langue, mais dénote une stratégie active de synergie entre langue et mâchoire visant à faire partager à ces deux organes le « travail » de déplacement de la surface de la langue, pour venir en contact avec le palais dans le cas de consonnes coronales par exemple. Cette observation doit être prise en compte lors de l'élaboration des modèles articulatoires, en particulier au niveau de la répartition de la variance des données entre les composantes mâchoire et les composantes corps et dos de la langue (pour plus de détails, voir Bailly *et al.* (1998)). Notons que le sujet *P1X* n'utilise pratiquement pas cette stratégie systématique de synergie mâchoire-langue, mais que *B1X* et *J1X* utilisent peu leur mâchoire.

Compensations articulatoires. À l'inverse de ces synergies, nous avons pu observer, grâce aux modèles articulatoires développés pour les locuteurs *P1X* et *J1X*, certains phénomènes de compensation entre *articulateurs* (au sens des *degrés de liberté* définis plus haut).

Bien que l'impact des phénomènes de coarticulation sur les variations de déplacement de la mâchoire, de la langue ou des lèvres puisse être clairement observé sur les seules coupes médiosagittales du conduit vocal, cette observation globale ne permet pas d'identifier les contributions de chaque articulateur individuel : en particulier dans une zone où la stabilité est nécessaire pour assurer une constriction, un contour global ne donne pas toujours les informations sur les combinaisons qui ont été mises en œuvre pour concilier les perturbations induites par des gestes antagonistes, c'est-à-dire en d'autres termes pour atteindre l'*équifinalité* pour l'ensemble des réalisations d'un son dans tous les contextes de coarticulation. La modélisation articulatoire *orientée sujet* développée pour *P1X* et *J1X* permettant d'extraire les degrés de liberté des articulateurs tout en préservant la conformation et les stratégies de synergie individuelles de chaque locuteur, a donc constitué un outil essentiel pour mettre en évidence des stratégies de préservation de la configuration vocalique pour certaines consonnes pour *J1X*, et aussi certaines différences importantes entre les stratégies employées par les deux locuteurs.

Ce travail, mené dans le cadre de la thèse d'Anne Vilain, est décrit de manière plus exhaustive dans Vilain, Abry & Badin (1998b), Vilain *et al.* (1998a), Vilain, Abry & Badin (1999), Vilain, Abry & Badin (2000), et Vilain (2000)).

E. Relations articulatori-acoustiques

Lorsque sont maîtrisés les deux domaines que constituent la modélisation articulatoire et la modélisation acoustique du conduit vocal, il est alors possible de s'intéresser aux *relations articulatori-acoustiques*, c'est-à-dire aux relations entre l'espace des commandes articulatoires et l'espace acoustique, le plus souvent caractérisé par les formants. Ces relations, fondements du robot articulatoire évoqué plus haut, constituent l'un des éléments essentiels de la communication parlée. Elles permettent en effet de spécifier les liens entre la production, caractérisée en particulier dans l'espace des articulateurs, et la perception, qui peut être reliée à l'espace acoustique.

1. Nomogrammes et points focaux

Dans la lignée des travaux de Fant (1960), nous avons étudié les relations articulatori-acoustiques d'un modèle articulatoire simplifié, le célèbre *modèle à quatre tubes*, appelé aussi modèle à trois paramètres (X_c, abscisse de la constriction ; A_c, aire, de la constriction ; A_l, aire aux lèvres). Les quatre tubes constituent des approximations de la cavité arrière, de la constriction orale, de la cavité avant, et de la section labiale. Ces travaux nous ont permis de mieux comprendre les relations entre forme du conduit vocal et formants, en de mettre en évidence des *affiliations formants / cavités :* lorsque la constriction orale est suffisamment resserrée, chaque formant (défini, rappelons-le, comme une résonance du conduit vocal) peut être attribué à une cavité particulière qui résonne selon un mode spécifique (résonance en multiples impairs d'un quart d'onde pour les tubes ouverts à une extrémité et fermés à l'autre, résonateur de

Helmholtz constitué d'une cavité fermée par une constriction, ou encore résonance en multiples pairs de la demi-onde pour les tuyaux ouverts aux deux extrémités). Les *nomogrammes* qui affichent les résonances du conduit vocal en fonction de l'abscisse du centre de la constriction mettent en évidence des croisements de formants, points que nous avons appelés *points focaux* (Boë & Abry (1986)), liés à une position particulière de la constriction induisant la coïncidence de deux formants associés l'un à la cavité arrière et l'autre à la cavité avant. Dans ces zones, les formants sont peu sensibles à la variation de la position de la constriction, alors qu'ils le sont beaucoup plus dans les zones éloignées des points focaux : ceci met en particulier en évidence le fait que les relations articulatori-acoustiques ne sont pas linéaires.

Ces résultats, présentés en détail dans Badin, Perrier, Boë & Abry (1990), ont constitué l'une des bases importantes d'un certain nombre d'autres travaux à l'ICP : théorie acoustique de la production des consonnes fricatives (Badin (1989), interprétation des données de nomogrammes humains de Ladefoged & Bladon (1982) (*cf.* Badin *et al.* (1990)), interprétation du double locus des consonnes occlusives vélaires (Bailly (1995)), théorie de la dispersion-focalisation (Vallée, Schwartz & Escudier (1999)), et normalisation entre locuteurs (Apostol *et al.* (2000)).

2. Macro-sensibilités articulatori-acoustiques

Il semblait intéressant d'étendre la notion d'affiliation à des formes de conduit vocal plus réalistes : les *fonctions de sensibilité acoustique* de Fant & Pauli (1974), reprises pour les voyelles du français par Mrayati & Carré (1976), permettaient de prédire les variations de formants induites par de *petites* perturbations de la fonction d'aire. Ces fonctions présentaient donc l'intérêt certain de linéariser les relations articulatori-acoustiques. Malheureusement, des simulations plus systématiques ont montré que ces fonctions de sensibilité n'avaient de valeur que locale, et ne pouvaient pas prédire les variations de formants pour des variations d'aire plus importantes que quelques 10 ou 20 %. Dans certains cas, le signe même de la fonction peut s'inverser entre petites variations et grandes variations. C'est la raison pour laquelle les *fonctions de macro-sensibilités articulatori-acoustiques* ont été développées (Boë, Badin & Perrier (1995)) : des nomogrammes articulatoires systématiques ont permis de déterminer les variations de formants induites pour de grandes variations des paramètres de contrôle du modèle articulatoire, et d'établir des sortes de cartes qui permettent de prédire, qualitativement du moins, l'influence des mouvements d'articulateurs sur les formants.

3. Inversion articulatori-acoustique par optimisation sous contrainte

L'*inversion* de la relation articulatori-acoustique, en d'autres termes la récupération des gestes articulatoires à partir du signal de parole, répond à trois motivations principales : contrôler, coder et percevoir. Dans le cadre du développement de systèmes de synthèse articulatoire, il est nécessaire de disposer de nombreuses données de trajectoires articulatoires qui permettront d'établir les stratégies de contrôle du synthétiseur : les dispositifs d'acquisition directe de données articulatoires étant relativement lourds, il est intéressant de tenter d'inférer ces paramètres articulatoires par inversion à partir du signal de parole. Par ailleurs, les paramètres articulatoires, étant liés à des mouvements biomécaniques, sont des signaux relativement lents (leur bande passante se chiffre en quelques dizaines de Hz pour la plupart d'entre eux) : ils offrent ainsi une forme de *codage* particulièrement économique pour le signal de parole qui, de son côté, requiert au minimum une dizaine de kHz de fréquence d'échantillonnage. Enfin, selon la *théorie motrice de la perception* (Liberman, Cooper, Harris & MacNeilage (1962), Liberman & Mattingly (1985)), la récupération des gestes articulatoires peut être considérée comme un moyen nécessaire pour la perception de la parole. Dans le cadre du projet européen *Speech Maps* dont l'un des objectifs majeurs était l'inversion de la parole (Abry *et al.* (1994)), nous avons donc développé des procédures d'inversion par optimisation sous contrainte (Badin *et al.* (1995a), Mawass *et al.* (2000)).

L'inversion : un problème mal posé. Une propriété de la relation articulatori-acoustique que je n'ai pas encore évoquée est la surjectivité : un même jeu de formants peut en effet être produit par plusieurs combinaisons différentes de paramètres articulatoires (Atal, Chang, Mathews & Tukey (1978)), même si ces combinaisons présentent dans la majorité des cas des voyelles des propriétés communes, à savoir en particulier une constriction orale située dans la même région du conduit vocal (Boë, Perrier & Bailly (1992)). L'inversion en parole est donc un problème *mal posé* au sens mathématique puisqu'il n'est pas certain que la solution existe et qu'elle soit unique, (Hadamard (1923), cité par Abry *et al.* (1994)) : il ne peut donc pas être résolu de manière explicite et directe. Le concept de robotique de la parole évoqué à la section D.1 offre alors un cadre propice à la résolution de ce problème. Le synthétiseur articulatoire est considéré comme un robot contrôlé par des commandes *proximales* et produisant des paramètres *distaux* (Jordan (1990), Bailly, Laboissière & Schwartz (1991)). Ces commandes, qui présentent des degrés de

liberté en excès (puisque plusieurs combinaisons articulatoires peuvent donner le même résultat acoustique), pourront être déterminées de manière unique à condition d'apporter des informations supplémentaires, sous forme de *contraintes a priori* telles que des trajectoires proximales les plus lisses possible ou des plages de variations limitées pour certains articulateurs.

Optimisation sous contraintes. Dans le cadre de la thèse de Denis Beautemps, nous avons tenté une première inversion de transitions voyelle – consonne fricative basée sur ce type de principes (Badin *et al.* (1995a)). En l'absence de modèle articulatoire, l'espace proximal était celui des 50 valeurs de la fonction sagittale ; une contrainte de lissage spatial était appliquée aux fonctions sagittales, le lissage temporel étant assuré indirectement par l'initialisation de l'algorithme pour un instant donné par les valeurs trouvées à l'instant précédent. La recherche de la solution était implémentée par un algorithme d'optimisation basé sur une descente de gradient. Cette méthode a permis de traiter avec succès quelques transitions, et a montré l'intérêt de l'utilisation de contraintes pour l'inversion.

Dans le cadre de la thèse de Khaled Mawass, nous avons repris ces principes de manière plus systématique (Mawass *et al.* (2000)), en utilisant un modèle articulatoire complet et des signaux distaux à la fois acoustiques et visuels. Les paramètres articulatoires du modèle développé pour le sujet *P1X* constituaient les paramètres proximaux. Les paramètres distaux résultants se composaient des quatre premiers formants, auxquels étaient adjoints deux paramètres géométriques, l'aire de constriction orale et l'aire intérolabiale. L'algorithme d'inversion est basé sur la méthode classique de descente du gradient sous contraintes : il utilise la rétropropagation de l'erreur configurationnelle entre les paramètres distaux calculés et mesurés (Jordan (1990)). L'algorithme utilise aussi une contrainte de lissage temporel : la minimisation de l'accélération des paramètres proximaux (puisque les articulateurs ont une certaine inertie, *cf.* plus haut). Finalement, l'erreur à minimiser par l'algorithme est la somme pondérée de : (1) la distance quadratique entre les six paramètres géométriques et acoustiques mesurés et les six paramètres distaux cumulée sur l'ensemble des trames de la séquence, et (2) l'accélération des paramètres articulatoires cumulée sur les mêmes trames.

Cette procédure d'inversion a été évaluée de manière systématique sur un corpus de séquences Voyelle – Consonne fricative – Voyelle. La racine carrée de l'erreur quadratique moyenne (RMS) relative peut atteindre, pour certaines séquences, 12 % pour F1, et approximativement 5 % pour F2, F3 et F4[1], avec des moyennes de 6,5 %, 3,4 %, 3,1 %, et 3.1 % respectivement ; l'aire intérolabiale présente une RMS relative moyenne de 16.8 % ; l'aire de constriction orale n'étant pas explicitement mesurée, son erreur de reconstruction ne peut pas être évaluée, mais seulement encadrée par des seuils qui dépendent du contexte (voir l'article annexé pour plus de détails).

Nous avons également montré que le paramètre d'aire intérolabiale n'est pas crucial en général pour l'inversion de l'acoustique vers l'articulatoire. Cependant, nous avons remarqué que la procédure d'inversion était sensible dans certains cas aux paramètres initiaux, et en particulier aux paramètres *LP* et *LV*, l'absence d'initialisation pouvant aboutir à des formes de langues compensatoires ne correspondant pas à la réalité. Il est donc utile de disposer d'informations sur l'arrondissement des lèvres pour l'inversion, du moins dans certains cas.

Enfin, l'un des intérêts de l'approche robotique est de pouvoir ne spécifier une cible que par des limites (inférieures et supérieures, par exemple), sans nécessité de définir de valeur précise. Cela permet de tenir compte des éventuelles imprécisions sur les mesures, ou de préciser simplement des zones définies de manière qualitative seulement, comme la présence ou l'absence d'une constriction orale qui peut être inférée approximativement à partir d'une mesure de l'intensité globale du signal.

4. Reconstruction de la forme de la langue à partir de points

Le lecteur a compris que la cinéradiographie pourrait être la solution idéale pour obtenir des données articulatoires alliant de bonnes résolutions temporelle et spatiale (en attendant que l'IRM dynamique dépasse la dizaine d'images par seconde, *cf.* Demolin *et al.* (2000)), si elle n'était pas exclue à cause de sa nocivité. La modélisation articulatoire associée à l'articulographie électromagnétique offre une voie alternative intéressante que nous avons explorée (Badin *et al.* (1997)). Le problème initial consiste à reconstruire la coupe médiosagittale du conduit vocal à partir des coordonnées dans le plan médiosagittal de quelques points de chair (c'est-à-dire attachés aux articulateurs) déterminées à l'aide d'un articulographe électromagnétique (*cf.* D.2). Le principe consiste à utiliser un modèle articulatoire pour prendre en compte

[1] Rappelons que le seuil différentiel déterminé par Flanagan (1955) se situe aux alentours de 5 % pour F1 et de 3 % pour F2.

les propriétés de continuité physique de la langue et reconstruire ainsi la forme de celle-ci à partir de quelques points seulement, et de la position de la mâchoire.

Les données cinéradiographiques enregistrées pour le sujet *J1X* nous ont permis d'évaluer cette méthode de manière exhaustive : en effet, lors de la séance d'enregistrement, trois petites billes de plomb (donc opaques aux rayons X et facilement identifiables sur les images radiographiques) étaient collées sur la langue du sujet : ces données présentent l'immense et rare avantage d'associer de manière synchrone des contours de langue et des coordonnées de marqueurs attachés à des points de chair (Kaburagi & Honda (1994) disposent également d'une combinaison de ce type de données, grâce à l'association de l'imagerie ultrasonique de la langue et de l'articulographe électromagnétique). En pratique, il s'agit d'inverser le modèle articulatoire de langue, et donc de déterminer les paramètres *JH*, *TB*, *TD*, *TT*, *TA*, et *LY* à partir des coordonnées X/Y des trois billes sur la langue et de la hauteur de mâchoire *JawHei*.

Le paramètre *JH* est directement déterminé à partir de *JawHei*. Le paramètre de hauteur de larynx *LY* ne peut pas être déterminé à partir des coordonnées des trois points sur la langue. Les paramètres restants sont déterminés simultanément par un algorithme d'optimisation d'atteinte de cibles multiples par optimisation, les cibles étant les distances entre les centres des billes et le contour de langue modélisé. Nous avons trouvé une erreur RMS de reconstruction de 1,80 mm, qui chute à 1,26 mm si l'on exclut les deux points les plus bas dans la région du larynx, à comparer avec les 1,24 mm trouvés par Kaburagi & Honda (1994) pour une expérience similaire, mais n'utilisant pas de modèle articulatoire. Notre méthode permet donc une reconstruction relativement précise d'une majeure partie de la forme de la langue dans le plan médiosagittal à partir de trois points de chair et de la hauteur de la mâchoire. Il est donc envisageable de combiner les avantages de la cinéradiographie et de l'articulographe électromagnétique pour acquérir de grandes quantités de données articulatoires pour un sujet. Toutefois, il sera nécessaire de résoudre préalablement un problème de cadrage lié à cette méthode : l'angle entre la paroi arrière du conduit vocal (pharynx et nasopharynx) et le plan occlusal est difficilement maîtrisable lors des enregistrements articulographiques, donc la forme moyenne de la langue associée, dont dépend en fin de compte la précision de l'inversion. Il est donc indispensable de trouver une méthode de normalisation qui permette de résoudre ce problème, qui se pose également, même si c'est de manière moins cruciale, pour l'intégration de données issues de diverses séances d'enregistrement en IRM et en cinéradiographie.

F. Synthèse articulatoire

Les différents modèles présentés dans les sections précédentes peuvent être intégrés de manière à former un synthétiseur articulatoire complet. L'objectif de la présente section est de décrire le synthétiseur avec les connections entre les différents modules, et notre approche de la synthèse articulatoire des consonnes fricatives. Une présentation beaucoup plus détaillée de ce travail se trouve dans l'article annexé Mawass *et al.* (2000), ainsi que dans le manuscrit de thèse de Mawass (1997).

1. Le synthétiseur articulatoire et ses modules

Comme l'illustre la Figure 8, le synthétiseur articulatoire est constitué de l'interconnexion d'un certain nombre de modèles : modèle articulatoire, modèle aérodynamique, modèles de sources acoustiques, et modèle de propagation acoustique et de rayonnement. Rappelons que le modèle articulatoire, le modèle de source de bruit de friction sont des modèles fonctionnels basés sur des données issues du sujet *P1X*, et que le modèle à deux masses de cordes vocales a été ajusté pour représenter au mieux le comportement du sujet *P1X* (Vescovi, Castelli & Pelorson (1995)) ; une certaine cohérence est ainsi assurée entre les différents éléments.

Le synthétiseur articulatoire complet est globalement contrôlé par deux jeux de paramètres articulatoires : les paramètres supra-laryngés qui commandent le modèle articulatoire, et un jeu de paramètres qui contrôlent les cordes vocales (pression sous-glottique *PS*, longueur des cordes vocales *LG* et hauteur de la glotte au repos *H0*), qui doivent être soigneusement coordonnés pour pouvoir générer de la parole de haute qualité.

Le modèle d'écoulement, qui considère le conduit vocal comme deux constrictions localisées, A_g à la glotte et A_c à la constriction orale (*cf.* C.1), est un module central interconnecté avec presque tous les autres modules (voir Figure 8). Ce module est essentiel pour la coordination et l'interaction entre articulation supra-laryngée et sources acoustiques. D'une part, le signal d'aire glottique est filtré passe-bas de manière à ne retenir que les variations lentes à la fréquence fondamentale et réinjecté dans le modèle d'écoulement, ce qui a pour conséquence une variation de la chute de pression à la constriction ΔP_c, et induit une modulation de la source de bruit synchrone avec la fréquence fondamentale, telle qu'on l'observe dans le cas des consonnes fricatives voisées (*cf.* Stevens (1971)). D'autre part, l'aire de

constriction orale A_c est l'un de paramètres de contrôle du modèle d'écoulement, ce qui assure une influence du système supra-laryngé sur l'écoulement dans le conduit vocal et donc en particulier sur l'amplitude des sources aussi bien de voisement que de bruit.

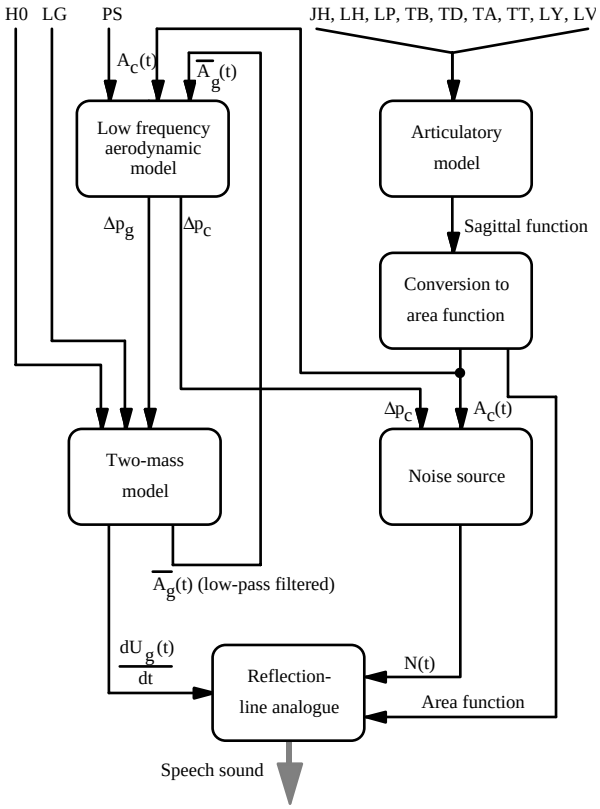


Figure 8 : Description schématique du synthétiseur articulatoire.

2. Synthèse articulatoire par copie

À ce stade, l'un de nos objectifs majeurs était d'évaluer les possibilités de notre synthétiseur articulatoire, en particulier pour les consonnes fricatives du français, et d'aller aussi loin que possible dans l'imitation du sujet *P1X*. Notre approche a donc été basée sur la synthèse articulatoire par copie, qui consiste à construire des trajectoires articulatoires aussi proches que possible de celles du sujet.

Nous avons réalisé un enregistrement vidéo du sujet *P1X* prononçant les 27 combinaisons V_1CV_2 des consonnes fricatives françaises [v z ʒ] dans tous les contextes vocaliques possibles, V_1 et V_2 étant choisis parmi [i a u]. Les formants étant plus faciles à détecter dans les fricatives voisées que dans les sourdes, seules les fricatives voisées ont été enregistrées.

Les trajectoires des commandes supra-laryngées ont été déterminées en utilisant la méthode d'inversion décrite à la section E.3.

Les trois paramètres de contrôle des sources acoustiques, *PS*, *H0* et *LG*, ont été déterminés par une stratégie commune à toutes les séquences voisées qui consistait à : (1) maintenir la pression sous-glottique *PS* à 10 cm H_2O et la longueur des cordes vocales *LG* à 1,6 cm pendant toute la séquence; (2) affecter à la hauteur de la glotte au repos *H0* une valeur de 0.03 cm pendant les voyelles, une valeur de 0.035 cm pendant les fricatives, une sigmoïde assurant l'interpolation entre ces cibles.

Nous avons également généré une version de ces séquences contenant les fricatives sourdes associées, en utilisant les mêmes trajectoires pour les articulateurs supra-laryngés. Une valeur de 0,1 cm était affectée à *H0* à l'instant de minimum d'intensité du son dans le segment fricatif, afin d'assurer la cessation du voisement pendant la consonne. Ce geste glottal doit être soigneusement coordonné avec la trajectoire de la constriction orale afin d'obtenir des fricatives sourdes réalistes (Scully & Allwood (1985), McGowan *et al.* (1995)). La Figure 9 illustre cette coordination pour les séquences [aʃa] et [aʒa].

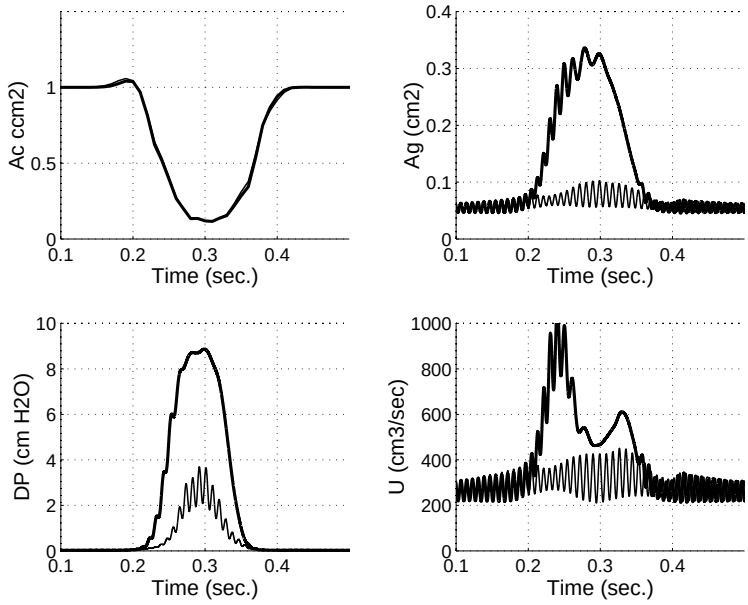


Figure 9 : Exemple de trajectoires des paramètres Ac, Ag, ΔPg et U pour la fricative sourde [aʃa] (trait gras) et pour la fricative voisée [aʒa] (trait fin).

3. Evaluation perceptive

Un test d'évaluation perceptive nous a permis d'évaluer la qualité des sons synthétisés. Dix auditeurs français naïfs ont été soumis à un test à choix forcé unique afin d'évaluer trois jeux de stimuli : (1) le son des séquences de fricatives voisées enregistrées en vidéo, (2) les copies voisées obtenues par synthèse articulatoire, et (3) les versions sourdes de ces mêmes stimuli[1]. Les auditeurs devaient identifier pour chaque stimulus la consonne parmi six réponses possibles : [v z ʒ f s ʃ]. Les taux d'identification résultants sont quasiment identiques pour les stimuli originaux (98.8 %) et les stimuli synthétiques (98.6 %). Ces résultats, qui ne sont pas vraiment surprenants, sont dus à la fois à la bonne qualité de la synthèse, et au fait que [s ʃ] sont les fricatives les plus facilement identifiables dans n'importe quelle langue, et que [f] est toujours bien distingué de [s ʃ].

4. Conclusion

Les différents modules de notre synthétiseur articulatoire ont été évalués indépendamment les uns des autres par comparaison avec les données sur lesquelles ils sont basés, et qu'ils doivent être capables de représenter. Le travail de synthèse articulatoire par copie présenté ci-dessus constitue une sorte de validation ultime du synthétiseur, de l'intégration de tous ses modules, et de la possibilité de déterminer des commandes qui produisent une synthèse articulatoire de parole d'excellente qualité. Il valide également notre choix d'une approche anthropomorphique.

Notons qu'en fait, bien que la synthèse articulatoire ait été proposée et étudiée depuis de très nombreuses années (Coker (1967), ou Scully (1986)), le nombre de synthétiseurs articulatoires utilisés actuellement est assez limité (Kröger *et al.* (1995), Rubin *et al.* (1996), Dang *et al.* (1999)). De plus, malgré un certain nombre d'études intéressantes consacrées à la modélisation articulatoires des consonnes fricatives (Flanagan & Ishizaka (1976), Scully & Allwood (1985), McGowan *et al.* (1995), ou Sinder, Krane & Flanagan (1998)), il semble qu'aucun travail systématique sur la synthèse de séquences Voyelle – Consonne fricative – Voyelle n'ait été mené et évalué jusqu'au niveau perceptif.

[1] Ces sons sont accessibles à l'adresse http://www.icp.inpg.fr/~badin/ActaAcustica_Sounds.html

III. BILAN ET PERSPECTIVES

À ce niveau du mémoire, avant de proposer des pistes pour l'avenir, il est temps d'effectuer un bilan global des travaux présentés.

A. Les acquis

Il est toujours stimulant de regarder vers l'avenir, et indispensable de rester modeste devant la complexité d'un objet d'étude tel que la parole et devant le chemin qui reste à parcourir pour prétendre à le maîtriser tant soit peu. Il est cependant utile – et réconfortant – de jeter un coup d'œil en arrière, et de mesurer la distance parcourue entre les premiers modèles de voyelles statiques en ondes planes et les modèles tridimensionnels actuels capables de reproduire des transitions VCV, ou la synthèse articulatoire des consonnes fricatives.

Les acquis se résument souvent à des chiffres dans un ordinateur (données, programmes), mais le plus important réside bien sûr dans la méthodologie et le savoir-faire, et aussi dans les principes qui sous-tendent la démarche du chercheur. Je présente d'abord un bilan rapide des acquis concrets de mon travail de recherche dans la présente section, en tentant d'éviter de transformer le texte en liste de ratons laveurs[1]…

1. Données et dispositifs expérimentaux

Comme je l'ai longuement décrit plus haut, je me suis intéressé à un certain nombre de techniques expérimentales de mesures de paramètres liés aux niveaux périphériques de la production de la parole. La plupart des dispositifs sont conçus pour acquérir des mesures *in vivo :* banc de mesure de fonctions de transfert acoustique du conduit vocal, masque pneumotachométrique pour la mesure de l'écoulement et des pressions dans le conduit, cinéradiographie et articulographie électromagnétique pour le mouvement, et imagerie IRM pour la caractérisation tridimensionnelle des articulateurs, sans compter la vidéo pour les mesures tridimensionnelles de lèvres et de visage. Nous disposons ainsi d'un ensemble de données articulatoires et acoustiques complémentaires, pour la majorité d'entre elles acquises sur le sujet *P1X* prononçant, dans des conditions très maîtrisées, les mêmes corpus soigneusement conçus. Ces corpus comprennent, suivant les techniques utilisées, les voyelles et les consonnes du français artificiellement soutenues, ainsi que des séquences VCV faisant intervenir au maximum les articulations vocaliques extrêmes /a i u y/.

Nous avons donc acquis un savoir-faire reconnu dans ce domaine expérimental, ce qui nous a en particulier aidé à convaincre le CNRS d'affecter un poste d'ingénieur de recherche à l'ICP pour maintenir et développer cette activité de mesures. Mais aussi, et surtout, ce savoir-faire fonde le développement de nos modèles de production.

2. Modèles

Nous disposons aujourd'hui d'une panoplie de modèles articulatoires linéaires, médiosagittaux et tridimensionnels. Au cours du développement de ces modèles, nous avons pu identifier les degrés de liberté articulatoires associés aux sujets à l'origine de ces modèles. Des degrés de liberté tout à fait similaires ont pu être identifiés pour les différents locuteurs, même si ces locuteurs peuvent utiliser des stratégies de contrôle parfois très différentes. Rappelons que les corpus utilisés ne contiennent que des tâches de parole, en expression neutre de manière générale, mais à l'exclusion de tous autres mouvements tels que mastication ou déglutition qui impliquent d'autres degrés de liberté. Il s'est aussi avéré que, dans une large mesure, les degrés de liberté des organes tridimensionnels sont prédictibles à partir des degrés de liberté dans le plan médiosagittal.

Nous possédons également un ensemble de modèles d'écoulement d'air, de sources acoustiques de voisement et de bruit de friction, et de propagation et rayonnement acoustique dans les domaines temporels et/ou fréquentiels.

Nous avons ainsi pu étudier les relations articulatori-acoustiques, en mettant en particulier en évidence les affiliations formants / cavités et les macro-sensibilités. La décomposition de transitions VCV en degrés de liberté a dévoilé des stratégies de compensation entre articulateurs qui ne sont pas lisibles directement sur les contours sagittaux.

[1] Jacques Prévert, *Inventaire*, dans *Paroles* (1946)

Nous avons par ailleurs montré la précision nécessaire de la coordination des gestes glotte / constriction orale lors de la production des consonnes fricatives voisées, en liaison avec les interactions entre sources et conduit vocal.

Enfin, nous avons pu développer des procédures d'inversion, basées sur le concept de robotique de la parole, qui permettent de reconstruire avec une bonne fiabilité les trajectoires des paramètres de contrôle articulatoire à partir de l'acoustique, même si ce problème d'inversion est un problème mal posé *a priori*.

Cet ensemble de données et de modèles constitue une étape importante dans nos efforts de compréhension et de modélisation de la production de la parole.

3. Une première tête parlante audiovisuelle tridimensionnelle

Par *tête parlante*, nous entendons l'intégration dans un *robot articulatoire anthropomorphique* de l'ensemble des modèles qui interviennent dans les processus périphériques de production de la parole tels que nous les avons décrits dans les sections précédentes de ce mémoire. Comme le synthétiseur articulatoire, la tête parlante est contrôlée par des paramètres articulatoires supra-laryngés et des paramètres de contrôle glottique ; mais de manière beaucoup complète, elle donne à *entendre* et aussi à *voir* les conséquences de ces gestes articulatoires, et ce de manière intrinsèquement cohérente, puisque les signaux acoustiques et optiques produits sont les conséquences des mêmes phénomènes articulatoires. L'intérêt de l'approche anthropomorphique se situe en particulier à ce niveau. À l'inverse, les *visages* parlants ne possèdent pas d'articulateurs internes ni de modélisation articulatori-acoustique, et ne produisent pas intrinsèquement de signal acoustique de parole ; ils peuvent simplement être synchronisés de manière plus ou moins sophistiquée avec des systèmes de synthèse de parole.

Nous avons donc puisé dans notre panoplie de modèles pour construire une première tête parlante audiovisuelle tridimensionnelle, ou plutôt différentes versions de tête parlante. La Figure 10 illustre les différentes combinaisons que nous avons explorées. Le cœur du système, pour l'instant simulé hors temps réel, est toujours constitué d'un modèle articulatoire, 2D ou 3D ; les formes et positions des différents articulateurs délivrés par le modèle articulatoire servent ensuite d'une part à générer des images des articulateurs et de la tête parlante, et d'autre part à calculer pour chaque instant la fonction d'aire associée, les sources acoustiques et le son rayonné, en tenant compte à la fois de la coordination entre glotte et conduit vocal et aussi de l'interaction entre sources et constrictions. Diverses séquences audiovisuelles peuvent ensuite être assemblées et jouées[1].

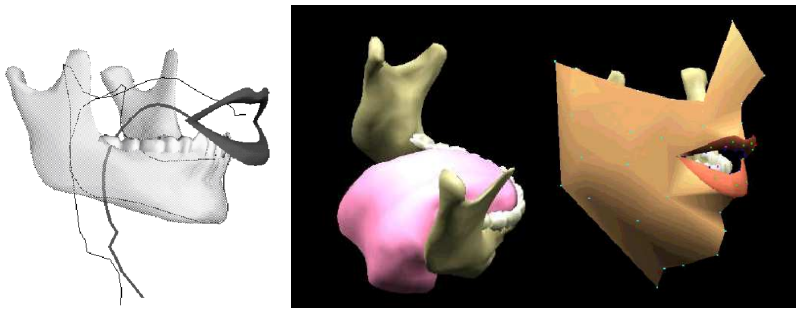


Figure 10 : Exemples de diverses modalités de tête parlante : (a) modèle médiosagittal de langue, modèle 3D de lèvres, mâchoire ; (b) modèle 3D de langue, mâchoire ; (c) modèle 3D de lèvres / visage, mâchoire.

4. Quelques références à l'état de l'art

Un certain nombre des travaux de recherche présentés dans ce mémoire n'a pas d'équivalent dans le monde. Quelques rares modèles tridimensionnels de langue basés une modélisation physique ont été développés (Wilhelms-Tricarico (1995), Dang & Honda (2000a)). Cohen, Beskow & Massaro (1998) ont développé un modèle linéaire basé sur des courbes splines, mais ce modèle est contrôlé par un nombre très élevé de paramètres et ne suit pas l'approche qui consiste à déterminer les vrais degrés de liberté de la langue pour la parole. Le modèle langue que j'ai présenté dans ce mémoire constitue donc le premier modèle 3D développé par analyse en composantes linéaires à partir de données volumétriques. Notons que Engwall (2000b) a développé un modèle similaire en suivant la même méthode à partir des données obtenues en collaboration (Engwall & Badin (1999)).

[1] Quelques-unes d'entre elles sont accessibles sur le site http://www.icp.inpg.fr/~badin/VTH-SPS5.html

La situation est semblable pour les modèles de lèvres et de visage. Un nombre beaucoup plus importants de modèles a été développé (*cf.* Bailly, Vatikiotis-Bateson, Badin, Revéret & Yehia (In preparation) pour une analyse exhaustive), mais seuls ceux de Bateson et ses collègues (*cf.* Vatikiotis-Bateson, Kuratate, Kamachi & Yehia (1999)) sont basés sur l'analyse en composantes linéaires de données extraites d'un sujet de référence, et ainsi comparables aux nôtres.

Bien que la synthèse articulatoire ait débuté il y a plusieurs décennies (*cf.* Coker (1967), Scully & Allwood (1985)), le nombre de synthétiseurs articulatoires disponibles aujourd'hui est assez limité (*cf.* Kröger *et al.* (1995), Stevens, Blumstein, Glicksman, Burton & Kurowski (1992), P. Boersma (1995)). En particulier, aucune étude systématique sur la synthèse articulatoire des consonnes fricatives n'a été menée jusqu'à présent. Notons cependant deux approches qui, bien que n'étant pas entièrement articulatoires, doivent être cependant citées: (1) celle de Maeda (1996), basée sur l'interpolation de fonctions d'aire entre deux cibles vocaliques et une cible consonantique, qui n'a pas été formellement évaluée; (2) celle de Stevens *et al.* (1992) qui utilisent aussi une approche composite dans laquelle le synthétiseur est en fait un synthétiseur à formants piloté par des trajectoires formantiques. Enfin, un travail mené aux laboratoires Haskins devrait également être cité : McGowan *et al.* (1995) ont tenté une approche articulatoire de la production de syllabes [aCa], mais d'une manière très limitée, et sans évaluation perceptive de la qualité des sons synthétisés, et en particulier des fricatives. Le travail présenté dans ce mémoire n'a donc pas d'équivalent dans la littérature.

Pour finir ce rapide survol, je dois mentionner que même si les bases de données en production de parole sont nombreuses, je ne connais aucune base orientée sujet qui rassemble une grande variété de mesures sur les mêmes corpus.

B. Les perspectives

Dans la lignée des travaux que je viens de présenter, différentes voies doivent être poursuivies ou explorées. D'une part, il me semble indispensable de continuer à développer et à améliorer les modèles qui constituent la tête parlante, en se basant toujours sur des données expérimentales et / ou en faisant appel aux progrès dans le domaine de l'aéroacoustique. D'autre part, grâce aux méthodes et protocoles de mesure expérimentale que nous maîtrisons, j'envisage maintenant l'enregistrement de plus grands corpus pour étudier de manière plus systématique les stratégies de production de parole d'un ou plusieurs locuteurs. Enfin, le temps est venu de me tourner de manière plus concrète, en particulier dans le cadre du développement des STIC, vers des applications comme la synthèse audiovisuelle à partir du texte, les clones pour les télécommunications, ou l'aide à l'apprentissage des langues. Cette activité se trouve également en phase avec l'une des sept thématiques prioritaires du 6e *Programme cadre européen de recherche et de développement technologique* européen (6e PCRD, 2002-2006), les *Technologies pour la société de l'information*, et plus particulièrement le sous-thème des *interfaces multi-sensorielles* capables de comprendre et d'interpréter l'expression naturelle de l'homme à travers les paroles, les gestes et les différents sens.

1. Données et modèles en production de parole

Rappelons que le contrôle de la tête parlante sera d'autant plus simple que notre modélisation des phénomènes articulatoires et aéroacoustiques impliqués dans la production de la parole sera plus réaliste.

Modélisation articulatoire. Nous disposons d'ores et déjà de modèles 3D de langue, de lèvres, de visage, et d'une première ébauche de modèle de velum. Il est cependant indispensable d'améliorer et / ou d'étendre la modélisation de certains organes pour aboutir à une modélisation articulatoire complète. Au niveau des organes, nous nous intéresserons à la partie interne des lèvres (celles-ci sont pour l'instant seulement modélisées sous leur aspect visible de l'extérieur) ; la modélisation de la pointe de la langue sera raffinée, en particulier pour nous permettre dans le futur de traiter des articulations rétroflexes, sans oublier de traiter la cavité sous-linguale, importante pour certains sons comme /ʃ/ ; un modèle de velum est en cours de développement ; un modèle de parois externes du conduit vocal (joues, parois des larynx, pharynx, nasopharynx) sera mis en œuvre ; l'épiglotte pourra également être modélisée, en espérant obtenir des formants plus précis pour les articulations arrière. Cette approche de modélisation de chaque articulateur un par un présente l'avantage, par rapport à mon approche initiale de modélisation globale du *tuyau* vocal, de permettre à la fois une modélisation plus détaillée des différentes cavités délimitées par les articulateurs, et la détermination d'une fonction d'aire éventuellement tridimensionnelle qui servira ensuite de point de départ à la modélisation aéroacoustique.

Mon approche globale en modélisation articulatoire sera donc de continuer à développer des modèles *linéaires*, en explorant leurs limites, mais sans exclure la possibilité de mettre en œuvre des non-linéarités, mais de manière locale, et sans nier l'intérêt des modèles biomécaniques, en particulier pour leur aptitude

intrinsèque à la dynamique. Ainsi, j'envisage de développer des modèles locaux de contact capables de prendre en compte la déformation des organes qui entrent en contact les uns avec les autres avec une force plus ou moins grande.

Le développement de la modélisation articulatoire continuera à nécessiter l'acquisition et / ou le traitement de données IRM complémentaires. Il est ainsi indispensable d'améliorer les techniques de segmentation de contours 3D en utilisant des surfaces splines, et non plus simplement des courbes splines dans des contours plans, en même temps que de suivre les progrès rapides réalisés en imagerie médicale pour bénéficier d'images de la meilleure qualité possible et de temps d'acquisition les plus faibles possible.

Modélisation (aéro)acoustique. La méthode TLM de modélisation acoustique, en conjonction avec la description détaillée de la géométrie tridimensionnelle du conduit vocal, permettra d'évaluer les modes transverses dans la gamme 5-15 kHz, et de déterminer le degré de finesse de la description géométrique du conduit vocal réellement nécessaire pour un modèle de production de la parole de haute qualité, en particulier pour les consonnes fricatives pour lesquelles les sources d'excitation sont localisées en des régions bien spécifiques du conduit vocal. Par ailleurs, des mesures complémentaires sur des maquettes en plexiglas et sur quelques sujets devraient ouvrir la voie à une meilleure modélisation du couplage entre conduit oral et conduit nasal, que ce couplage se réalise par le passage vélo-pharyngé ou par l'extérieur entre lèvres et narines.

Nous avons vu la grande importance de la coordination glotte – constriction orale ; mon objectif dans ce domaine est d'intégrer les récents développements des modèles d'écoulement et d'interaction fluide – parois, d'une part, et d'effectuer des mesures de mouvement des cordes vocales par vidéo rapide d'autre part, afin d'améliorer notre modélisation des consonnes fricatives et occlusives (un séjour d'une semaine au *Research Institute for Logopedics and Phoniatrics* de l'Université de Tokyo m'a déjà donné l'occasion d'utiliser une technique de visualisation à 4000 images / secondes pour les cordes vocales).

Extension à différents sujets – normalisation inter-sujets. Il est évident qu'il est impossible de tirer des conclusions universelles de résultats obtenus pour un sujet. Il est donc indispensable, à partir de l'expérience importante acquise pour un sujet, d'étudier d'autres sujets. Cette approche a déjà été amorcée, avec l'étude des sujets *J1X* et *B1X* en cinéradiographie. De même, un sujet suédois a été enregistré en IRM 3D, tandis que nous allons élaborer un modèle articulatoire médiosagittal d'un sujet italien à partir d'images IRM médiosagittales. Ces travaux sont naturellement menés dans le cadre de l'approche orientée sujet discutée plus haut, avec l'objectif de comparer les stratégies individuelles, et d'en tirer des principes plus généraux.

Un problème crucial dans le futur est la normalisation inter-sujets. Cette normalisation devra vraisemblablement être traitée à deux niveaux au moins. L'anatomie de chaque sujet est spécifique, et l'on pourrait donc définir un modèle articulatoire générique adaptable à chaque sujet grâce à des opérations géométriques simples portant sur le système de grilles servant à décrire les articulateurs. Un deuxième niveau de normalisation serait celui de la stratégie articulatoire mise en œuvre par le locuteur : en fonction de son anatomie, et très probablement en fonction d'autres critères individuels plus difficiles à caractériser (tonus musculaire par exemple), chaque locuteur va adopter des stratégies articulatoires qui peuvent être très différentes. Les synergies plus ou moins fortes entre mâchoire et langue ont déjà été évoquées dans ce mémoire ; nous avons également observé les différences de jeu de lèvres entre *P1X* et *J1X* (*J1X* présente une variabilité extrêmement faible sur la forme des lèvres pour les consonnes fricatives labiodentales, alors que pour *P1X* le contexte vocalique influe largement sur la forme et la position des lèvres de la consonne). L'analyse PARAFAC (Parallel Factor Analysis), qui constitue une extension de l'ACP basée sur l'introduction d'une matrice intermédiaire supplémentaire dans la décomposition (Harshman & Lundy (1984)), déjà été explorée par Nix, Papcun, Hogden & Zlokarnik (1996), et plus récemment par Hoole (1999), pourrait se révéler une piste intéressante.

Variabilité et espaces spatio-temporels multi-paramétriques. La production de la parole est un phénomène éminemment variable, que ce soit au niveau segmental avec la coarticulation, ou au niveau suprasegmental avec la variation du style de voix et d'élocution. Il est donc naturel de s'intéresser à cette variabilité et d'essayer de la caractériser dans les différents espaces auxquels nous avons accès, aussi bien sur le plan expérimental que sur celui de la modélisation. Divers paradigmes peuvent induire de la variabilité en production de la parole : l'étude d'un certain nombre d'entre eux fait partie de mes projets dans les années à venir.

Jusqu'à présent, les corpus de données utilisés pour le développement des modèles étaient enregistrés dans des conditions d'élocution parfaitement contrôlées, dans un style neutre, plutôt hyper-articulé et à une vitesse d'élocution plutôt lente (afin d'assurer d'atteindre les extrêmes des possibilités des articulateurs),

sans mentionner certains corpus, comme ceux utilisés pour l'IRM, constitués d'articulations artificiellement soutenues. L'un de mes objectifs est maintenant de susciter de la variabilité contrôlée en donnant au sujet des consignes d'élocution différentes : parole hyper- ou hypo-articulée grâce à l'effet Lombard (parole produite dans un environnement fortement bruité), avec éventuellement des consignes d'insistance, différents styles de voix (chuchotée, criée), et aussi en tentant d'aborder certains types d'émotion ou d'attitude (sourire, peur, dégoût, etc.).

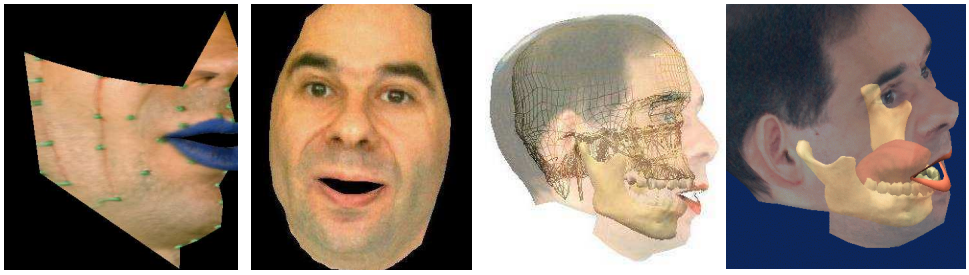
Pour étudier ces phénomènes, il est donc indispensable de disposer de masses plus importantes de données articulatoires, aérodynamiques et acoustiques. En combinant les méthodes directes d'acquisition de données et les méthodes d'inversion mises au point ces dernières années, il est maintenant envisageable de recueillir des données cohérentes, synchronisées entre les différents espaces concernés (en utilisant en particulier les méthodes de fusion de données), pour un grand nombre de contextes.

Ces bases de données pourront alors servir à déterminer les *espaces de réalisation* des différents phonèmes, sous forme d'*espaces de réalisation de cibles spatio-temporelles* dans plusieurs domaines : articulatoire, géométrique (description de la géométrie du conduit vocal, en particulier taille et emplacement des constrictions), aérodynamique (pressions et vitesse d'écoulement de l'air), et acoustique (formants). Il s'agit donc, pour chaque corpus représentatif d'une condition particulière d'élocution, d'étudier la répartition statistique des réalisations dans les différents domaines, ainsi que leurs interactions. Rappelons par exemple que les voyelles possèdent la plupart du temps des formants peu dispersés, mais que la position des articulateurs correspondant peut présenter une dispersion beaucoup plus grande, en particulier pour les voyelles basses ; à l'inverse, les formants d'une consonne peuvent être très variables, même si le lieu géométrique d'articulation est très finement spécifié. L'étape suivante consistera ensuite à étudier les variations, en fonction des différentes conditions d'élocution, de ces espaces multi-paramétriques que Bailly (1997) nomme *cartes sensori-motrices*. Bailly a déjà proposé un cadre à ce travail, et réalisé une première étude sur les consonnes plosives ; de mon côté, j'ai établi pour les consonnes fricatives voisées les conditions de maintien simultané du voisement et du bruit de friction. Il s'agit maintenant de poursuivre ce type de travail de manière beaucoup plus extensive, en l'étendant à toutes les catégories de phonèmes.

2. Têtes parlantes et applications

Cette section présente un certain nombre d'applications des têtes parlantes, qui correspondent soit à des projets qui viennent de démarrer, soit à des idées qui demandent à être concrétisées.

Intégration et développement d'une tête parlante audiovisuelle douée de réalité augmentée. Nous avons déjà évoqué une première tête parlante audiovisuelle, qui pouvait être présentée sous diverses apparences. L'intégration dans ce cadre des modèles évoqués plus haut devrait conduire à une tête parlante de plus en plus réaliste. Plus spécifiquement, au delà de la validité des modèles articulatoires et acoustiques constituant la base de la tête parlante, les aspects de présentation visuelle sont très importants dans ce domaine. Des avancées déjà très importantes ont été récemment réalisées à l'ICP au niveau de la présentation de la tête et de la texture du visage (Revéret, Bailly, Borel & Badin (2000b), Bailly, Revéret, Borel & Badin (2000), Revéret, Bailly & Badin (2000a), Elisei, Odisio, Bailly & Badin (2001)), comme en témoignent les exemples de la Figure 11. L'existence des modèles articulatoires d'organes tels que la langue ou le velum offre clairement des capacités de *réalité augmentée :* en effet il est maintenant possible d'afficher ces articulateurs, qui sont en général partiellement ou totalement cachés. Deux pistes principales seront explorées dans ce sens : l'utilisation de la *transparence*, c'est-à-dire rendre plus ou moins transparents certains organes qui masquent ceux que l'on souhaite plus particulièrement observer (voir Figure 11c,d), et l'utilisation de l'*écorché*, technique bien connue dans le domaine médical qui consiste à découper et retirer (de manière virtuelle !) certaines parties de certains organes afin de permettre de découvrir ceux qui sont situés plus en profondeur. Une nouvelle voie s'ouvre donc vers l'animation d'une tête parlante virtuelle tridimensionnelle dans laquelle il est possible de faire apparaître les organes souhaités, et qui peut naturellement être également affichée sous l'angle désiré. Les différentes versions de tête parlante qui seront développées devront naturellement faire l'objet d'une évaluation perceptive systématique, qui permette de quantifier les gains ou les pertes liés à telle ou telle caractéristique particulière ; il sera en particulier très important de quantifier l'apport de la réalité augmentée.



(a) (b) (c) (d)

Figure 11 : Exemples de développements récents de la tête parlante : (a) modèle partiel de visage texturé, (b) modèle de visage texturé avec une texture plus naturelle, (c) et (d) visages semi-transparent laissant apercevoir les organes internes en mode réalité augmentée[1].

Synthèse articulatoire audiovisuelle. La tête parlante et l'ensemble des données articulatoires et acoustiques qui sont progressivement accumulées permettent d'envisager le développement d'un système de synthèse articulatoire audiovisuelle à partir du texte. Revéret *et al.* (2000a) ont élaboré une première ébauche à l'ICP, hybride en ce sens que seuls le visage et les lèvres sont représentés par une modélisation articulatoire alors que le son est produit par une concaténation de polyphones manipulés par la méthode PSOLA. Pour élaborer un système de synthèse articulatoire complet, il faudrait acquérir les données articulatori-acoustiques nécessaires à la description de tous les phonèmes d'une langue sous la forme de patrons décrivant les cibles vocaliques et consonantiques dans un espace multiple articulatoire / géométrique / acoustique. Dans le cadre d'approche *robotique de la parole*, le système pourrait alors offrir une souplesse telle qu'il soit possible de reproduire les stratégies utilisées par les humains pour s'adapter aux nécessités de communication et aux conditions d'environnement, propriétés éminemment intéressantes au niveau des applications technologiques.

Aide à la prononciation des langues étrangères. Les applications de synthèse audiovisuelle basées sur la tête parlante ne font pas appel aux possibilités de réalité augmentée évoquées plus haut. Par contre, cette caractéristique, associée aux possibilités d'expérimentation de la relation articulatori-acoustique, ouvre la voie à des applications originales dans le domaine de l'aide à l'apprentissage de la prononciation (Badin, Bailly & Boë (1998a)).

Il est légitime de supposer que, pour l'apprenant d'une langue étrangère, une bonne perception et une bonne reconnaissance des sons / articulations nouveaux constitue un pré-requis à l'apprentissage de leur production. En d'autres termes, il est difficile d'apprendre à prononcer des sons que l'on ne sait pas reconnaître ou discriminer correctement. Cette situation est bien connue chez les enfants sourds qui présentent des retards de développement du langage liés à leur impossibilité de percevoir *auditivement* les sons. C'est ainsi qu'un enseignant comme Jean-Guy Le Bel dans son *Traité de correction phonétique ponctuelle* (Le Bel (1990)) prône la *méthode articulatoire*, ou *approche cognitive*, basée sur l'idée que l'apprenant doit acquérir une connaissance explicite du fonctionnement de son conduit vocal – autrement dit des relations articulatori-acoustiques – pour l'apprentissage de la prononciation. Certains des *grands moyens* qu'il propose dans le domaine de la correction phonétique sont directement liés à la perception et à la production : la *discrimination auditive* (on ne peut prononcer bien que ce que l'on perçoit bien), la *composition articulatoire et acoustique* (le processus d'apprentissage sera d'autant plus efficace que l'apprenant sera mieux conscient de l'articulateur auquel il doit prêter attention afin de résoudre un problème spécifique), et la *phonétique combinatoire* (divers effets de coarticulation peuvent être utilisés pour conduire au bon geste articulatoire pour un phonème nouveau donné).

Il est donc important pour l'enseignant d'une deuxième langue (L2) d'être capable de mettre en évidence, et pour l'apprenant d'*internaliser*, les *relations entre les gestes articulatoires et les sons résultants*. Les principales tâches de l'enseignant sont donc : (1) d'évaluer et d'améliorer les capacités de l'apprenant à percevoir les voyelles et les consonnes de la L2, et (2) d'élaborer une méthodologie visant à aider l'apprenant à découvrir les stratégies articulatoires appropriées à la production des sons qu'il / elle a appris à percevoir.

Puisqu'elle constitue un modèle des relations articulatori-acoustiques, la tête parlante constitue un outil de choix dans le domaine de l'apprentissage de la prononciation. L'enseignant va pouvoir en effet l'utiliser pour manipuler de manière contrôlée les articulations et les signaux audiovisuels de parole associés, et

[1] Des animations peuvent être trouvées sur le site http://www.icp.inpg.fr/~badin/TP_Applications.html

générer les stimuli nécessaires à l'apprentissage perceptif de l'apprenant. Il pourra aussi mener des expériences visant à développer les stratégies facilitantes prônées par la phonétique combinatoire, en éditant les trajectoires des paramètres de contrôle de la tête parlante qui auront été au préalable extraits de bases de données, obtenus par inversion ou encore générés par un système de synthèse à partir du texte. Un dictionnaire de *bonnes idées* pourrait ainsi être progressivement construit par l'enseignant.

Dans ce cadre, les possibilités de réalité augmentée de la tête parlante apportent une contribution supplémentaire très intéressante. Le cas du /y[en français illustre cet intérêt : le /y[ne figurant pas dans l'inventaire phonologique d'un certain nombre de langues (italien, japonais, arabe …), nombre d'étrangers ont des difficultés à le prononcer correctement. Cette difficulté est très vraisemblablement due en partie au fait que la différence entre /y[et /u[n'est pas visible, puisque due seulement à un mouvement avant-arrière de la langue. Pouvoir *montrer* les mouvements de langue associés devrait faciliter considérablement cette apprentissage. L'exemple d'une expérience de compensation de l'effet d'un tube labial conforte l'idée que des indications sur les bonnes manœuvres articulatoires à mettre en œuvre peuvent fortement aider l'apprenant dans son apprentissage. En effet, à l'occasion d'un travail sur la nature des représentations internes de la parole pour les locuteurs, Savariaux *et al.* (1995) ont mené une expérience au cours de laquelle des locuteurs équipés d'un tube labial de diamètre conséquent devaient prononcer la voyelle française /u[: la plupart des sujets ne réalisèrent le fort mouvement de recul de la langue nécessaire pour produire la voyelle que lorsqu'on leur donna clairement la piste vers la solution, à savoir prononcer un /o[et tendre ensuite graduellement vers un bon /u/[compensé. La mise en évidence et la démonstration de gestes facilitants semblent donc pouvoir jouer un rôle important dans l'apprentissage de certains sons.

La tête parlante audiovisuelle virtuelle semble donc être un outil qui peut se révéler intéressant dans le domaine de l'apprentissage de la prononciation, dans le cadre des nouvelles technologies pour l'éducation. L'un de mes objectifs sera de tester ces possibilités en collaboration avec des enseignants.

Réhabilitation des déficients auditifs. On peut supposer que l'interlocuteur cherche, de manière générale, à retrouver, par tous les canaux sensoriels à sa disposition, la réalité motrice de la parole, c'est-à-dire les mouvements des articulateurs. La vision, qui sert en général seulement à compléter les informations auditives, doit carrément les remplacer, lorsque l'audition est déficiente, comme c'est le cas pour les sourds avec la lecture labiale (Cathiard (1989)). Cependant la lecture labiale seule est loin de pouvoir fournir suffisamment d'information pour permettre à l'auditeur (ou plutôt au *visionneur* !) de percevoir ou de reconstituer l'ensemble du flux de parole. Il peut donc être intéressant d'associer au système traditionnel de la communication parlée d'autres signaux gestuels. Le projet « Tête parlante audiovisuelle virtuelle : Réalité augmentée et Langage Parlé Complété (LPC) pour la réhabilitation des déficients auditifs » vise à évaluer deux modalités complémentaires, la réalité augmentée et le LPC, comme outils d'aide à la réhabilitation des déficients auditifs. Les signaux visuels du LPC (Cornett (1967)) sont produits par les mouvements de la main et des doigts du locuteur, placés près du visage, et fournissent les informations manquantes car non visibles (sur le voisement, la nasalité, ou les mouvements cachés de la langue). Le projet devrait montrer que la tête parlante en *mode réalité augmentée*, associée à une main de synthèse qui code des informations manquantes selon la *modalité LPC*, peut constituer un outil facilitant l'accès à l'oral chez les déficients auditifs, dès le plus jeune âge, en permettant d'acquérir des stratégies articulatoires aussi bien pour la langue maternelle que pour une langue seconde.

Clones et télécommunications. La tête parlante constitue un véritable clone, ou avatar, du locuteur qui a servi de base à son développement. Les méthodes que nous avons utilisées pour le tout premier locuteur vont à l'évidence s'améliorer de manière considérable, en particulier dans le sens de l'automatisation du travail. La création de clones pour de nouveaux locuteurs devient ainsi de plus en plus facile et rapide, ce qui a permis de démarrer des projets de transmission de parole audiovisuelle dans le domaine des télécommunications, *Tempo-Valse* et *ARTUS*. Les clones de locuteurs particuliers pourront être aussi construits comme des adaptations d'un *clone générique*, sorte de clone représentatif d'un locuteur moyen, aux caractéristiques morphologiques et aux habitudes articulatoires de ces locuteurs.

Le projet *Tempo-Valse (Terminal Expérimental MPEG4 POrtable de Visiophonie et Animation Labiale ScalablE)* est un projet de transmission audiovisuelle dont la qualité de transmission peut être adaptée à la capacité de transmission du réseau téléphonique. Il repose sur un système d'analyse et de synthèse de visages parlants qui bénéficie des récents progrès réalisés dans le domaine de la modélisation articulatoire des lèvres et du visage. La phase d'analyse, dont le rôle est d'extraire un petit nombre de paramètres décrivant le visage, repose sur le principe de l'analyse par la synthèse : un algorithme itératif cherche les six paramètres articulatoires du modèle de visage qui minimise une mesure de différence entre l'image vidéo originale à

transmettre et l'image reconstruite à partir du modèle texturé. Une fois les paramètres transmis, la synthèse utilise le même modèle. L'avantage de cette méthode d'analyse par la synthèse est qu'elle réduit considérablement la quantité d'information de l'image de manière *intelligente*, c'est-à-dire en extrayant les caractéristiques importantes de la forme géométrique du visage, et donc en minimisant le bruit de transmission (*cf.* Elisei *et al.* (2001)). Ce système nécessite en revanche la transmission préalable de l'ensemble des données du modèle avant de pouvoir fonctionner. Mais cette approche, qui effectue en fin de compte un suivi des commandes articulatoires d'un clone 3D du locuteur, offre également l'avantage, par rapport aux méthodes de codage classique d'images 2D, de pouvoir reconstruire, dans le cadre de la visioconférence, un espace virtuel unique dans lequel chaque participant peut intervenir simultanément par l'intermédiaire de son clone.

Le projet *ARTUS* (*Animation Réaliste par Tatouage audiovisuel à l'Usage des Sourds*) constitue un autre application des têtes parlantes, au carrefour du domaine des télécommunications et du monde des déficients auditifs. Ce projet, qui doit débuter très prochainement, vise à substituer à l'affichage du télétexte classique de certaines émissions télévisées l'incrustation dans l'image vidéo d'une tête parlante virtuelle associée à une main de synthèse qui code le Langage Parlé Complété. Il repose donc très clairement sur les avancées en cours dans le développement des têtes parlantes.

C. Conclusions

En résumé, j'espère avoir montré que mon approche de l'étude des mécanismes de production de la parole, appuyée sur l'accumulation de connaissances sous forme de modèles *falsifiables*, basés sur des corpus de données articulatori-acoustiques soigneusement conçus et confrontés en permanence à ces données, a permis d'accroître notre compréhension de l'objet *parole,* c'est-à-dire à la fois notre connaissance des propriétés intrinsèques du signal audiovisuel et celle des stratégies articulatoires mises en œuvre par les locuteurs. Citons à titre d'exemples la notion de point focal, en relation avec les affiliations formants-cavités, qui joue un rôle important dans la théorie de la Dispersion Focalisation (Schwartz, Boë, Vallée & Abry (1997)), l'influence des stratégies de coordination glotte / constriction sur l'utilisation des consonnes dans les langues du monde (Boë *et al.* (2000)), ou les stratégies de coarticulation / compensation. Notre approche anthropomorphique se trouve donc confortée par ces résultats. À cela on peut ajouter le succès de l'approche anthropomorphique orientée données que représente l'intérêt manifesté par France Télécom R&D pour nos modèles dans le cadre de contrats.

Cette approche, jusqu'à présent focalisée sur un très petit nombre de sujets, est encore loin de pouvoir prétendre à l'universalité : mon objectif majeur sera désormais d'étendre le champ d'observation à d'autres sujets, et à des corpus plus diversifiés – en particulier au niveau du style (parole hyper ou hypo-articulée, différentes attitudes ou expressions faciales, etc.).

Par ailleurs, il faut rappeler que mes travaux, jusqu'à présent de caractère assez fondamental, avaient comme objectif majeur de *comprendre et modéliser* la production de la parole. Aujourd'hui, je suis convaincu que nos modèles sont suffisamment avancés pour aborder avec confiance le champ des applications. C'est ainsi que j'espère consacrer une partie de mon activité à des projets plus appliqués, en particulier dans le domaine de l'aide à l'apprentissage de la prononciation, aussi bien pour les apprenants de langues étrangères que pour les déficients auditifs. En outre, les modèles plus complets et plus divers qui résulteront de mon activité de modélisation pourront servir de base à des applications de synthèse audiovisuelle à partir du texte et de télécommunications, en liaison avec nombre d'autres projets à l'ICP.

Il est enfin utile de mentionner que cette approche pluridisciplinaire, qui fait appel à des domaines aussi divers que l'imagerie médicale, le traitement de signal, la phonétique expérimentale, ou l'analyse statistique, s'intègre complètement dans le projet global du nouveau département STIC (Sciences et Technologies de l'Information et de la Communication) du CNRS, à la fois dans ses aspects de recherche fondamentale sur l'homme, le langage et la cognition, et les applications technologiques nécessaires à l'enseignement assisté par ordinateur, aux télécommunications audiovisuelles, et aux *agents et objets communicants*, de manière plus large.

IV. REFERENCES

Abry, C. & Badin, P. (1998). Foreword to comments on "The Equilibrium Point Hypothesis and its application to speech motor control". *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, **4**, 3-4.

Abry, C., Badin, P., Mawass, K. & Pelorson, X. (1998). The Equilibrium Point Hypothesis and control spaced for relaxation movements or "When is movement actually needed to control movement ?". *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, **4**, 27-33.

Abry, C., Badin, P. & Scully, C. (1994). Sound-to-gesture inversion in speech: The *Speech Maps* approach. ESPRIT Research Report No. 6975. In *Advanced Speech Applications* (K. Varghese, S. Pfleger & J.P. Lefèvre, editors), pp. 182-196. Berlin: Springer Verlag.

Abry, C. & Boë, L.-J. (1986). 'Laws' for Lips. *Speech Communication*, **5**, 97-104.

Allwood, E. & Scully, C. (1982). A composite model of speech production. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 932-935.

Apostol, L., Perrier, P., Baciu, M., Segebarth, C. & Badin, P. (2000). Using the formant/cavity affiliation to study the inter-speaker variability: assessment from MRI data. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 213-216. Kloster Seeon, Germany.

Atal, B.S., Chang, J.J., Mathews, M.V. & Tukey, J.W. (1978). Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer-sorting technique. *Journal of the Acoustical Society of America*, **63**, 1535-1555.

Badin, P. (1980). *Simulation numérique en temps réel d'un synthétiseur à formants*. Unpublished Rapport de D.E.A., ENSERG, Institut National Polytechnique de Grenoble, Grenoble.

Badin, P. (1983). *Analyse de la parole - synthèse à formants. Application à la synthèse des contoïdes constrictives voisées du Français. Discussion d'une méthode en vue de la détermination des indices acoustiques de la parole*. Unpublished Thèse de Docteur Ingénieur, Institut National Polytechnique de Grenoble.

Badin, P. (1989). Acoustics of voiceless fricatives: production theory and data. *Speech Transmission Laboratory - Quarterly Progress Status Report - Stockholm*, **3/1989**, 33-55.

Badin, P. (1991). Fricative consonants: acoustic and X-ray measurements. *Journal of Phonetics*, **19**, 397-408.

Badin, P., Bailly, G. & Boë, L.-J. (1998a). Towards the use of a Virtual Talking Head and of Speech Mapping tools for pronunciation training. In *Proceedings of the ESCA Tutorial and Research Workshop on Speech Technology in Language Learning*, pp. 167-170. Stockholm, Sweden, ESCA and Dept. Speech, Music and Hearing, KTH, Stockholm.

Badin, P., Bailly, G., Raybaudi, M. & Segebarth, C. (1998b). A three-dimensional linear articulatory model based on MRI data. In *Proceedings of the Third ESCA / COCOSDA International Workshop on Speech Synthesis*, pp. 249-254. Jenolan Caves, Australia.

Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C. & Savariaux, C. (In revision). Three-dimensional articulatory modeling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*.

Badin, P., Baricchi, E. & Vilain, A. (1997). Determining tongue articulation: from discrete fleshpoints to continuous shadow. In *Proceedings of the 5th EuroSpeech Conference*, vol. 1, pp. 47-50. Rhodos, Greece, University of Patras, Wire Communication Laboratory, Patras, Greece.

Badin, P., Beautemps, D., Laboissière, R. & Schwartz, J.-L. (1995a). Recovery of vocal tract geometry from speech signal for vowels and fricative consonants using a midsagittal-to-area function conversion model. *Journal of Phonetics*, **23**, 221-229.

Badin, P., Borel, P., Bailly, G., Revéret, L., Baciu, M. & Segebarth, C. (2000). Towards an audiovisual virtual talking head: 3D articulatory modeling of tongue, lips and face based on MRI and video images. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 261-264. Kloster Seeon, Germany.

Badin, P. & Degryse, D. (1982). Speech Communication Hardware. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 514-516. Paris, France.

Badin, P. & Fant, G. (1984). Notes on vocal tract computation. *Speech Transmission Laboratory - Quarterly Progress Status Report - Stockholm*, **2-3/1984**, 53-108.

Badin, P. & Fant, G. (1989). Fricative production modelling: aerodynamic and acoustic data. In *Proceedings of the 1st EuroSpeech Conference*, vol. 2, pp. 23-26. Paris, France.

Badin, P., Gabioud, B., Beautemps, D., Lallouache, T.M., Bailly, G., Maeda, S., Zerling, J.P. & Brock, G. (1995b). Cineradiography of VCV sequences: articulatory-acoustic data for a speech production model. In *Proceedings of the 15th International Conference on Acoustics*, vol. IV, pp. 349-352. Trondheim, Norway.

Badin, P., Mawass, K. & Castelli, E. (1995c). A model of frication noise source based on data from fricative consonants in vowel context. In *Proceedings of the 13th International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors), vol. 2, pp. 202-205. Stockholm, Sweden, Arne Strömbergs Grafiska Press.

Badin, P., Motoki, K., Miki, N., Ritterhaus, D. & Lallouache, T.M. (1994a). Some geometric and acoustic properties of the lip horn. *Journal of the Acoustical Society of Japan (English)*, **15**(4), 243-253.

Badin, P. & Murillo, G. (1983a). An analysis method for high quality formant synthesis. In *Abstract of the 10th International Congress of Phonetic Sciences*, pp. 378. Utrecht, The Netherlands.

Badin, P. & Murillo, G. (1983b). Méthode d'analyse des sons en vue d'une synthèse de très haute qualité et de la détection des indices acoustiques de la parole. In *Proceedings of the 11th International Conference on Acoustics*, pp. 85. Paris, France.

Badin, P., Perrier, P., Boë, L.-J. & Abry, C. (1990). Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *Journal of the Acoustical Society of America*, **87**, 1290-1300.

Badin, P., Pouchoy, L., Bailly, G., Raybaudi, M., Segebarth, C., Lebas, J.-F., Tiede, M.K., Vatikiotis-Bateson, E. & Tohkura, Y.I. (1998c). Un modèle articulatoire tridimensionnel du conduit vocal basé sur des données IRM. In *Actes des 22èmes Journées d'Etude sur la Parole*, pp. 283-286. Martigny, Suisse.

Badin, P., Shadle, C.H., Pham Thi Ngoc, Y., Carter, J.N., Chiu, W., Scully, C. & Stromberg, K. (1994b). Frication and aspiration noise sources: contribution of experimental data to articulatory synthesis. In *Proceedings of the 3rd International Conference on Spoken Language Processing*, vol. 1, pp. 163-166. Yokohama, Japan.

Bailly, G. (1995). Characterisation of formant trajectories by tracking vocal tract resonances. In *Levels in Speech Communication, Relations and interactions* (C. Sorin, J. Mariani, H. Meloni & J. Schoentgen, editors), pp. 91-102. Amsterdam, The Netherlands.: Elsevier.

Bailly, G. (1997). Learning to speak. Sensori-motor control of speech movements. *Speech Communication*, **22**, 251-267.

Bailly, G., Badin, P. & Vilain, A. (1998). Synergy between jaw and lips/tongue movements: Consequences in articulatory modelling. In *Proceedings of the 5th International Conference on Spoken Language Processing* (R.H. Mannell & J. Robert-Ribes, editors), vol. 5, pp. 1859-1862. Sydney, Australia, Australian Speech Science and Technology Association Inc.

Bailly, G., Laboissière, R. & Schwartz, J.-L. (1991). Formant trajectories as audible gestures: an alternative for speech synthesis. *Journal of Phonetics*, **19**, 9-23.

Bailly, G., Revéret, L., Borel, P. & Badin, P. (2000). Hearing by eyes thanks to the "labiophone": exchanging speech movements. In *COST254 Workshop Friendly Exchanging Through The Net* (C. Germain, E. Grivel & O. Lavialle, editors), pp. 67-72. Bordeaux, France, COST254 (Intelligent Processing and Facilities for Communication Terminals).

Bailly, G., Vatikiotis-Bateson, E., Badin, P., Revéret, L. & Yehia, H. (In preparation). Visible characteristics of speech production. In *Audiovisual speech* (E. Vatikiotis-Bateson, G. Bailly & P. Perrier, editors), Cambridge, MA, USA: MIT Press.

Barbier, P. (1978). Les mouvement du larynx dans la chaîne parlée en français. *Travaux de l'Institut Phonétique de Strasbourg*, **10**, 98-119.

Båvegård, M. & Fant, G. (1996). Parameterized VT area function inversion. In *Proceedings of the 4th International Conference on Spoken Language Processing*, pp. 961-964. Philadelphia, PA, USA, University of Delaware & Alfred I. du Pont Institute.

Beautemps, D., Badin, P. & Bailly, G. (2001). Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. *Journal of the Acoustical Society of America*, **109**(5), 2165-2180.

Beautemps, D., Badin, P. & Laboissière, R. (1995). Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data. *Speech Communication*, **16**, 27-47.

Boë, L.-J. & Abry, C. (1986). Nomogrammes et systèmes vocaliques. In *Actes des 15èmes Journées d'Etude sur la Parole*, pp. 303-306.

Boë, L.-J., Badin, P. & Perrier, P. (1995). From sensitivity functions to macro-variations. In *Proceedings of the 13th International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors), vol. 2, pp. 234-237. Stockholm, Sweden, Arne Strömbergs Grafiska Press.

Boë, L.-J., Gabioud, B. & Perrier, P. (1995). Speech Maps interactive plant 'SMIP'. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors), vol. 2, pp. 426-429. Stockholm, Sweden, Arne Strömbergs Grafiska Press.

Boë, L.-J., Gahioud, B. & Perrier, P. (1995). The SMIP : An interactive articulatory-acoustic software for speech production studies. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, **3**, 137-154.

Boë, L.-J., Maeda, S. & Perrier, P. (1994). La modélisation articulatoire : un demi siècle d'évolution entre fonctionnel, physique et biomécanique. In *XX^{emes} Journées d'Etude sur la Parole*, pp. 41-54. Trégastel, France, TSS, France Télécom CNET/LAA. GFCP-SFA.

Boë, L.-J., Perrier, P. & Bailly, G. (1992). The geometric vocal tract variables controlled for vowel production: proposals for constraining acoustic-to-articulatory inversion. *Journal of Phonetics*, **20**, 27-38.

Boë, L.-J., Vallée, N., Badin, P., Schwartz, J.-L. & Abry, C. (2000). Tendencies in phonological structures : the influence of substance on form. Current Trends in Phonology and Phonetics II : Relationship between phonetics and phonology. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, **5**, 35-55.

Boersma, P. (1995). Interaction between glottal and vocal-tract aerodynamics in a comprehensive model of the speech apparatus. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors), vol. 2, pp. 430-433. Stockholm, Sweden, Arne Strömbergs Grafiska Press.

Boersma, P. (1998). *Functional phonology*. Unpublished Ph. D. Dissertation, Amsterdam University, Amsterdam.

Borel, P. (1999). *Modélisation articulatoire linéraire d'un visage incluant des lèvres*. Unpublished Rapport de DEA Signal Image Parole Télécom, Institut National Polytechnique de Grenoble, Grenoble.

Borel, P., Badin, P., Revéret, L. & Bailly, G. (2000). Modélisation articulatoire linéaire 3D d'un visage pour une tête parlante virtuelle. In *Actes des 23èmes Journées d'Etude de la Parole*, pp. 121-124. Aussois, France.

Cardoso, J.-F. (1998). Blind signal separation: statistical principles. *Proceedings of the IEEE,* **9**(10), 2009-2025.

Castelli, E. (1989). *Caractérisation acoustique des voyelles nasales du français. Mesures, modélisation et simulation temporelle*. Unpublished Thèse doctorale : Signal, Image, Parole, Institut National Polytechnique, Grenoble.

Castelli, E. & Badin, P. (1988). Mesures de fonctions de transfert du conduit vocal - Application à la détermination des fonctions de transfert du conduit nasopharyngal. In *Actes des 17èmes Journées d'Etude sur la Parole*, pp. 189-193. SFA.

Castelli, E. & Badin, P. (1989). Nasopharyngeal tract transfer functions measurements with white noise excitation. In *Proceedings of the 13th International Conference on Acoustics*, vol. 2, pp. 511-514.

Cathiard, M.-A. (1989). La perception visuelle de la parole : aperçu de l'état des connaissances. *Bulletin de l'Institut de Phonétique de Grenoble,* **17-18**, 109-193.

Charpentier, F. & Moulines, E. (1990). Pitch-synchronous waveform processing techniques for text-to-speech using diphones. *Speech Communication,* **9**(5-6), 453--467.

Cohen, M.M., Beskow, J. & Massaro, D.W. (1998). Recent developments in facial animation: an inside view. In *Proceedings of the International Conference on Auditory-Visual Speech Processing / Second ESCA ETRW on Auditory-Visual Speech* (D. Burnham, J. Robert-Ribes & E. Vatikiotis-Bateson, editors), pp. 201-206. Terrigal-Sydney, Australia.

Coker, C.H. (1967). Synthesis by rule from articulatory parameters. In *Proceedings of the 1967 Conference on Speech Communication Processes*, pp. 52-53. IEEE.

Coker, C.H. & Fujimura, O. (1966). A model for specification of vocal tract area function. *Journal of the Acoustical Society of America,* **40**(5), 1271.

Coker, C.H., Umeda, N. & Browman, C. (1973). Automatic synthesis from ordinary English text. In *Speech Synthesis* (J.L. Flanagan & L.R. Rabiner, editors), pp. 400-411. Stroudsburg, Pennsylvania: Dowden, Hutchinson and Ross, Inc.

Cornett, O. (1967). Cued Speech. *American Annals of the Deaf,* **112**, 3-13.

Dang, J. & Honda, K. (1998). Speech production of vowel sequences using a physiological articulatory model. In *Proceedings of the 5th International Conference on Spoken Language Processing* (R.H. Mannell & J. Robert-Ribes, editors), vol. 5, pp. 1767-1770. Sydney, Australia, Australian Speech Science and Technology Association Inc.

Dang, J. & Honda, K. (2000a). Estimation of vocal tract shape from speech sounds via a physiological articulatory model. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 233-236. Kloster Seeon, Germany.

Dang, J. & Honda, K. (2000b). Improvement of a physiological articulatory model for synthesis of vowel sequences. In *Proceedings of the 6th International Conference on Spoken Language Processing* (B. Yuan, T. Huang & X. Tang, editors), vol. I, pp. 457-460. Beijing, China.

Dang, J., Sun, J., Deng, L. & Honda, K. (1999). Speech synthesis using a physiological articulatory model with feature-based rules. In *Proceedings of the 14th International Congress of Phonetic Sciences*, vol. 3, pp. 2267-2270. San Francisco, California, USA.

de Penguern, G. (2001). *Modélisation articulatoire 2D / 3D et acoustique des nasales pour une tête parlante audiovisuelle*. Unpublished Rapport de stage Ingénieur, ENSERG, Institut National Polytechnique de Grenoble, Grenoble.

Degryse, D. (1981). Temporal simulation of wave propagation in the lossy vocal tract. In *Proceedings of the 4th FASE Symposium*, pp. 193-196.

Demolin, D., Metens, T. & Soquet, A. (2000). Real time MRI and articulatory coordinations in vowels. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 93-96. Kloster Seeon, Germany.

Djéradi, A., Guérin, B., Badin, P. & Perrier, P. (1991). Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method. *Journal of Phonetics,* **19**, 387-395.

Dudley, H., Riesz, R.R. & Watkins, S.S.A. (1939). A synthetic speaker. *Journal of the Franklin Institute,* **227**, 739-764.

Dudley, H. & Tarnoczy, T.H. (1950). The speaking machine of Wolfgang von Kempelen. *Journal of the Acoustical Society of America,* **22**(2), 151-166.

Dunn, H.K. (1950). The calculation of vowel resonances, and an electrical vocal tract. *Journal of the Acoustical Society of America,* **22**, 740-753.

El Masri, S., Pelorson, X., Saguet, P. & Badin, P. (1996a). Etude et analyse par la méthode TLM de la propagation acoustique dans le conduit vocal: effet des modes d'ordre supérieur. In *Actes des 21èmes Journées d'Etude sur la Parole*, pp. 243-246. Avignon, France.

El Masri, S., Pelorson, X., Saguet, P. & Badin, P. (1996b). Vocal tract acoustics using the transmission line matrix (TLM) method. In *Proceedings of the 4th International Conference on Spoken Language Processing*, vol. 2, pp. 953-956. Philadelphia, PA, USA, University of Delaware & Alfred I. duPont Institute.

El Masri, S., Pelorson, X., Saguet, P. & Badin, P. (1998). Development of the Transmission Line Matrix method in acoustics. Application to higher modes in the vocal tract and other complex ducts. *International Journal of Numerical Modelling,* **11**, 133-151.

Elisei, F., Odisio, M., Bailly, G. & Badin, P. (2001). Creating and controlling video-realistic talking heads. In *Proceedings of the Auditory-Visual Speech Processing Workshop, AVSP 2001* (D.W. Massaro, J. Light & K. Geraci, editors), pp. 90-97. Aalborg, Denmark.

Engwall, O. (2000a). A 3D tongue model based on MRI data. In *Proceedings of the 6th International Conference on Spoken Language Processing* (B. Yuan, T. Huang & X. Tang, editors), vol. III, pp. 901-904. Beijing, China.

Engwall, O. (2000b). Replicating three-dimensional tongue shapes synthetically. *Tal Musik Hörsel - Quaterly Progress Status Report - Stockholm,* **2-3/2000**, 53-64.

Engwall, O. & Badin, P. (1999). Collecting and analysing two- and three-dimensional MRI data for Swedish. *Tal Musik Hörsel - Quarterly Progress Status Report - Stockholm,* **3-4/1999**, 11-38.

Engwall, O. & Badin, P. (2000). An MRI study of Swedish fricatives: coarticulatory effects. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 297-300. Kloster Seeon, Germany.

Fant, G. (1953). Speech communication research. *Ingeniör Vetenskap Akademie (IVA),* **24**(8), 331-337.

Fant, G. (1958). Modern instruments and methods for acoustic studies of speech. *Acta Polytechnica Scandinavica,* **1**, 1-81.

Fant, G. (1960). *Acoustic theory of speech production*. The Hague: Mouton & co.

Fant, G., Nord, L. & Branderud, P. (1976). A note on the vocal tract wall impedance. *STL-QPSR,* **4/1976**, 13-20.

Fant, G. & Pauli, S. (1974). Spatial characteristics of vocal tract resonance modes. In *Proceedings of the Speech Communication Seminar*, pp. Stockholm.

Feng, G. & Castelli, E. (1996). Some acoustic features of nasal and nasalized vowels: a target for vowel nasalisation. *Journal of the Acoustical Society of America,* **99**(6), 3694-3706.

Flanagan, J.L. (1955). A Difference Limen for vowel formant frequency. *Journal of the Acoustical Society of America,* **27**, 288-291.

Flanagan, J.L. (1972a). *Speech Analysis, Synthesis and Perception*. Berlin, Heidelberg, New-York: Springer Verlag.

Flanagan, J.L. (1972b). Voices of men and machines. *Journal of the Acoustical Society of America,* **51**, 1375-1387.

Flanagan, J.L. & Ishizaka, K. (1976). Automatic generation of voiceless excitation in a vocal cord-vocal tract speech synthesizer. *IEEE Transactions on Acoustics, Speech, and Signal Processing,* **24**, 163-170.

Flanagan, J.L., Ishizaka, K. & Shipley, K.L. (1980). Signal models for low bit-rate coding of speech. *Journal of the Acoustical Society of America,* **68**, 780-791.

Fowler, C.A. & Saltzman, E.L. (1993). Coordination and coarticulation in speech production. *Language and Speech,* **36**, 171-195.

Fujimura, O. & Lindqvist, J. (1971). Sweep-Tone measurements of vocal-tract characteristics. *Journal of the Acoustical Society of America,* **49**, 541-558.

Hadamard, J. (1923). *Lectures on the Cauchy problem in linear partial differential equations*. New Haven: Yale University Press.

Harshman, R.A. & Lundy, M.E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In *Research methods for multimode data analysis* (H.G. Law, C.W. Snyder, J.A. Hattie & R.P. MacDonald, editors), pp. 122-215. New-York: Praeger.

Heinz, J.M. & Stevens, K.N. (1965). On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech. In *Proceedings of the 5th International Conference on Acoustics*, pp. A44.

Hermansky, H. & Pavel, M. (1995). Psychophysics of speech engineering systems. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors), vol. 3, pp. 42-49. Stockholm, Sweden, Arne Strömbergs Grafiska Press.

Hixon, T.J. (1966). Turbulent noise sources for speech. *Folia Phoniatrica,* **18**, 168-182.

Hixon, T.J., Minifie, F.D. & Tait, C.A. (1967). Correlates of turbulence noise production for speech. *Journal of Speech and Hearing Research,* **10**, 133-140.

Hoole, P. (1999). On the lingual organization of the German vowel system. *Journal of the Acoustical Society of America,* **106**(2), 1020-1032.

Hoole, P. & Kroos, C. (1998). Control of larynx height in vowel production. In *Proceedings of the 5th International Conference on Spoken Language Processing* (R.H. Mannell & J. Robert-Ribes, editors), vol. 2, pp. 531-534. Sydney, Australia, Australian Speech Science and Technology Association Inc.

Jordan, M.I. (1990). Motor Learning and the degrees of freedom problem. In *Attention and Performance* (M. Jeannerod & N.J. Hillsdale, editors),: Lawrence Erlbaum.

Kaburagi, T. & Honda, M. (1994). Determination of sagittal tongue shape from the positions of points on the tongue surface. *Journal of the Acoustical Society of America,* **96**(3), 1356-1366.

Kaburagi, T. & Honda, M. (1998). Determination of the vocal tract spectrum from the articulatory movements based on the search of an articulatory-acoustic database. In *Proceedings of the 5th International Conference on Spoken Language Processing* (R.H. Mannell & J. Robert-Ribes, editors), vol. 2, pp. 433-436. Sydney, Australia, Australian Speech Science and Technology Association Inc.

Kaburagi, T. & Honda, M. (2001). Dynamic articulatory model based on multidimensional invariant-feature task representation. *Journal of the Acoustical Society of America,* **110**(1), 441-452.

Kelso, J.A.S., Saltzman, E.L. & Tuller, B. (1986). The dynamical theory of speech production: Data and theory. *Journal of Phonetics,* **14**, 29-60.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America,* **67**, 971-995.

Kröger, B.J., Schröder, G. & Opgen-Rhein, C. (1995). A gesture-based dynamic model describing articulatory movement data. *Journal of the Acoustical Society of America,* **98**, 1878-1889.

Laboissière, R., Ostry, D. & Feldman, A.G. (1996). Control of multi-muscle systems: Human jaw and hyoid movements. *Biological Cybernetics,* **74**(3), 373-384.

Laboissière, R. & Pelorson, X. (1995). Stability and bifurcations of the two-mass model oscillation: analysis of fluid mechanics effects and acoustical loading. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors), vol. 3, pp. 190-193. Stockholm, Sweden, Arne Strömbergs Grafiska Press.

Laboissière, R., Schwartz, J.-L. & Bailly, G. (1990). Motor control for speech skills: A connectionist approach. In *Connectionist models. Proceedings of the 1990 Summer School* (D.S.Touretzki, J.L. Elman, T.L. Seinowski & G.E. Hinton, editors).

Ladefoged, P. & Bladon, R.A.W. (1982). Attempts by human speakers to reproduce Fant's nomograms. *Speech Communication,* **1**, 185-198.

Lallouache, M.T. (1991). *Un poste "Visage-Parole" couleur. Acquisition et traitement automatique des contours des lèvres.* Unpublished Thèse doctorale : Signal, Image, Parole, Institut National Polytechnique, Grenoble.

Le Bel, J.G. (1990). *Traité de correction phonétique ponctuelle: essai systémique d'application*. Québec, Canada: CIRAL, Université Laval.

Lebart, L., Morineau, A. & Piron, M. (1995). *Statistique exploratoire multidimensionnelle.* Paris, France: Dunod.

Liberman, A.M., Cooper, F.S., Harris, K.S. & MacNeilage, P.F. (1962). A Motor Theory of speech perception. In *Proceedings of the Speech Seminar Stockholm*, pp. Stockholm.

Liberman, A.M. & Mattingly, I.G. (1985). The motor theory of speech perception revised. *Cognition,* **21**, 1-36.

Liénard, J.S. (1968). La machine parlante de Kempelen. *Bull. du G.A.M,* **34**, 1-15.

Liénard, J.S. (1977). *Les processus de la communication parlée. Introduction à l'analyse et à la synthèse de la parole.* Paris: Masson.

Liljencrants, J. (1971). Fourier series description of the tongue profile. *STL-QPSR,* **4/1971**, 9-18.

Lin, Q.G. (1990). *Speech production theory and articulatory speech synthesis.* Unpublished Ph Thesis, Kungliga Tekniska Högskolan, Stockholm.

Lindblom, B., Lubker, J. & Gay, T. (1979). Formant frequencies of some fixed-mandible vowels and a model of speech-motor programming by predictive simulation. *Journal of Phonetics,* **7**, 141-161.

Lindblom, B. & Sundberg, J. (1971). Acoustical consequences of lip, tongue, jaw and larynx movement. *Journal of the Acoustical Society of America,* **50**, 1166-1179.

Löfqvist, A., Koenig, L.L. & McGowan, R.S. (1995). Vocal tract aerodynamics in /aCa/ utterances: Measurements. *Speech Communication,* **16**, 49-66.

Maeda, S. (1979a). An articulatory model of the tongue based on a statistical analysis, *97th Meet. Acoust. Soc. Am.*

Maeda, S. (1979b). Un modèle articulatoire de la langue avec des composantes linéaires. In *Actes des 10èmes Journées d'Etude sur la Parole*, pp. 152-162.

Maeda, S. (1985). Une source d'excitation cohérente dans les occlusives. In *Actes des 14èmes Journées d'Etude sur la Parole*, pp. 43-46. Paris, France.

Maeda, S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In *Speech Production and Modelling* (W.J. Hardcastle & A. Marchal, editors), pp. 131-149. Kluwer: Academic Publishers.

Maeda, S. (1991). On articulatory and acoustic variabilities. *Journal of Phonetics,* **19**, 321-331.

Maeda, S. (1996). Phonemes as concatenable units: VCV synthesis using a vocal tract synthesizer. *Arbeitsberichte des Institut fŸr Phonetik und digitale Spachverarbeitung der UniversitŠt Kiel,* **31**, 145-164.

Maeda, S. & Honda, K. (1995). Articulatory co-ordination and its biological aspects: an essay. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors), vol. 2, pp. 76-83. Stockholm, Sweden, Arne Strömbergs Grafiska Press.

Markel, J.D. (1972). The Sift algorithm for fundamental frequency estimation. *IEEE Transactions on Audio ???,* **20**, 367-377.

Markel, J.D. & Gray, A.H. (1976). *Linear Prediction of Speech.* Berlin, Heidelberg, New-York: Springer Verlag.

Mawass, K. (1997). *Synthèse articulatoire des consonnes fricatives.* Unpublished Thèse doctorale : Signal, Image, Parole, Télécoms, Institut National Polytechnique, Grenoble.

Mawass, K., Badin, P. & Bailly, G. (2000). Synthesis of French fricatives by audio-video to articulatory inversion. *Acta Acustica,* **86**(1), 136-146.

Mawass, K., Badin, P., Vescovi, C. & Beautemps, D. (1996). Evaluation d'un modèle de source de friction pour la synthèse articulatoire des consonnes fricatives. In *Actes des 21èmes Journées d'Etude sur la Parole*, pp. 367-370. Avignon, France.

McGowan, R.S. (1994a). Knowledge from speech production used in speech technology: articulatory synthesis. *Haskins Laboratories Status Report on Speech Research,* **117-118**, 25-29.

McGowan, R.S. (1994b). Recovering articulatory movement from formant frequency trajectories using task dynamics and a genetic algorithm: preliminary model tests. *Speech Communication,* **14**, 19-48.

McGowan, R.S., Koenig, L.L. & Löfqvist, A. (1995). Vocal tract aerodynamics in /aCa/ utterances: Simulations. *Speech Communication,* **16**, 67-88.

McGowan, R.S. & Saltzman, E.L. (1995). Incorporating aerodynamic and laryngeal components into tasks dynamics. *Journal of Phonetics,* **23**, 255-269.

Mermelstein, P. (1973). Articulatory model for study of speech production. *Journal of the Acoustical Society of America,* **53**, 1070-1082.

Morvan, Y. (2000). *Modélisation articulatoire d'une tête parlante virtuelle tenant compte de l'opposition Neutre / Sourire par analyse de données vidéo.* Unpublished Rapport de DEA Signal Image Parole Télécom, Institut National Polytechnique de Grenoble, Grenoble.

Motoki, K., Badin, P. & Miki, N. (1994). Measurement of acoustic impedance density distribution in the near field of the labial horn. In *Proceedings of the 3rd International Conference on Spoken Language Processing*, vol. 2, pp. 607-610. Yokohama, Japan.

Mrayati, M. & Carré, R. (1976). Relations entre la forme du conduit vocal et les caractéristiques acoustiques des voyelles françaises. *Phonetica,* **33**, 285-306.

Nishikawa, K., Asama, K., Hayashi, K., Takanobu, H. & Takanishi, A. (2000). Development of a talking robot. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 345-348. Kloster Seeon, Germany.

Nix, D.A., Papcun, G., Hogden, J. & Zlokarnik, I. (1996). Two cross-linguistic factors underlying tongue shapes for vowels. *Journal of the Acoustical Society of America,* **99**(6), 3707-3717.

Ostry, D.J., Vatikiotis-Bateson, E. & Gribble, P.L. (1997). An examination of the degrees of freedom of human jaw motion in speech and mastication. *Journal of Speech, Language, and Hearing Research,* **40**, 1341-1351.

Parke, F.I. (1982). Parametrized models for facial animation. *IEEE Computer Graphics and Applications,* **2**, 61-68.

Payan, Y. (1996). *Modèles biomécaniques et contrôle de la langue lors de la production de la parole.* Unpublished Thèse doctorale : Signal, Image, Parole, Télécoms, Institut National Polytechnique, Grenoble.

Payan, Y. & Perrier, P. (1997). Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication,* **22**, 185-205.

Payan, Y., Perrier, P. & Laboissière, R. (1995). Simulation of tongue shape variations in the sagittal plane based on a control by the Equilibrium-Point Hypothesis. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors), vol. 2, pp. 474-477. Stockholm, Sweden, Arne Strömbergs Grafiska Press.

Pelorson, X. (2000). L'instrument conduit vocal. Aspects acoustiques et aéroacoustiques. In *La parole, des modèles cognitifs aux machines communicantes, Actes de l'Ecole des techniques avancées, Signal, Image, Parole* (P. Escudier & J.-L. Schwartz, editors),: Hermes.

Pelorson, X., Badin, P., Motoki, K., Miki, N. & Plicque, M. (1995). On the radiation of sound at the lips during speech. Effects of lip geometry and of higher acoustical modes. In *Proceedings of the 15th International Conference on Acoustics*, vol. IV, pp. 497-500. Trondheim, Norway.

Pelorson, X., Hirschberg, A., Van Hassel, R.R., Wijnands, A.P.J. & Auregan, Y. (1994). Theoretical and experimental study of quasi-steady flow separation within the glottis during phonation. Application to a modified two-mass model. *Journal of the Acoustical Society of America,* **96**, 3416-3431.

Pelorson, X., Hirschberg, A., Wijnands, A.P.J. & Bailliet, H.M.A. (1995). Description of the flow through the vocal cords during phonation. *Acta Acustica,* **3**, 191-202.

Perkell, J.S. (1974). *A physiological-oriented model of the tongue activity during speech production.* Unpublished Ph. Diss., MIT, Cambridge.

Perkell, J.S. (1991). Models, theory and data in speech production. In *Proceedings of the XIIth International Congress of Phonetic Sciences*, vol. 1, pp. 182-191.

Perkell, J.S., Cohen, M.M., Svirsky, M.A., Matthies, M.L., Garabieta, I. & Jackson, M.T.T. (1992). Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America,* **92**, 3078-3096.

Perrier, P., Boë, L.-J., Majid, S.R. & Guérin, B. (1985). Modélisation articulatoire du conduit vocal. Exploration et exploitation. In *Actes des 14èmes Journées d'Etude sur la Parole*, pp. Paris.

Perrier, P., Boë, L.-J. & Sock, R. (1992). Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: modeling the transition with two sets of coefficients. *Journal of Speech and Hearing Research,* **35**, 53-67.

Perrier, P., Ostry, D. & Laboissière, R. (1996). The Equilibrium Point Hypothesis and its application to speech motor control. *Journal of Speech and Hearing Research,* **39**, 365-378.

Perrier, P., Payan, Y., Perkell, J.S., Jolly, F., Zandipour, M. & Matthies, M. (1998). On loops and articulatory biomechanics. In *Proceedings of the 5th International Conference on Spoken Language Processing* (R.H. Mannell & J. Robert-Ribes, editors), pp. Sydney, Australia, Australian Speech Science and Technology Association Inc.

Pham Thi Ngoc, Y. (1995). *Caractérisation acoustique du conduit vocal: fonctions de transfert acoustiques et sources de bruit. Étude des voyelles chuchotées et des consonnes fricatives non voisées.* Unpublished Thèse doctorale : Signal, Image, Parole, Institut National Polytechnique, Grenoble.

Pham Thi Ngoc, Y. & Badin, P. (1994). Vocal tract acoustic transfer function measurements: further developments and applications. In *Journal de Physique IV, Colloque C5, Supplément au Journal de Physique III. Proceedings of the 3rd French Congress of Acoustics*, vol. 4, pp. 549-552. Toulouse, France.

Portnoff, M.R. (1973). *A quasi one-dimensional digital simulation for the time-varying vocal tract.* Unpublished M.S. Thesis, MIT.

Rahim, M.G., Goodyear, C.C., Klejin, W.B., Schroeter, J. & Sonhi, M.M. (1993). On the use of neural networks in articulatory speech synthesis. *Journal of the Acoustical Society of America,* **93**, 1109-1121.

Reed, C.M., Rabinowitz, W.M., Durlach, N.I., Braida, L.D., Conway-Fithian, S. & Schultz, M.C. (1985). Research on the Tadoma method of speech communication. *Journal of the Acoustical Society of America,* **77**(1), 247-257.

Revéret, L. (1999). *Conception et évaluation d'un système de suivi automatique des gestes labiaux en parole.* Unpublished Thèse doctorale : Micro-électronique, Institut National Polytechnique, Grenoble.

Revéret, L., Bailly, G. & Badin, P. (2000a). MOTHER: a new generation of talking heads providing a flexible articulatory control for video-realistic speech animation. In *Proceedings of the 6th International Conference on Spoken Language Processing* (B. Yuan, T. Huang & X. Tang, editors), vol. II, pp. 755-758. Beijing, China.

Revéret, L., Bailly, G., Borel, P. & Badin, P. (2000b). Analyse par la synthèse d'un visage 3D parlant : inversion optico-articulatoire. In *Actes des 23èmes Journées d'Etude de la Parole*, pp. 125-128. Aussois, France.

Revéret, L. & Benoît, C. (1998). A new 3D lip model for analysis and synthesis of lip motion in speech production. In *Proceedings of the International Conference on Auditory-Visual Speech Processing / Second ESCA ETRW on Auditory-Visual Speech* (D. Burnham, J. Robert-Ribes & E. Vatikiotis-Bateson, editors), pp. 207-212. Terrigal-Sydney, Australia.

Riordan, C.J. (1977). Control of vocal-tract length in speech. *Journal of the Acoustical Society of America,* **62**(4), 998-1002.

Rokkaku, M., Hashimoto, K., Imaizumi, S., Niimi, S. & Kiritani, S. (1986). Measurement of the three-dimensional shape of the vocal tract based on the Magnetic Resonance Imaging technique. *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics,* **20**, 47-54.

Rossato, S. (2000). *Du son au geste, inversion de la parole: le cas des voyelles nasales.* Unpublished Thèse doctorale : Signal, Image, Parole, Telecoms, Institut National Polytechnique, Grenoble.

Rossato, S., Badin, P. & Feng, G. (2000). Estimation des mouvements du voile du palais à partir du signal de parole pour les voyelles nasales du Français. In *Actes des 23èmes Journées d'Etude de la Parole*, pp. 137-140. Aussois, France.

Rubin, P.E., Baer, T. & Mermelstein, P. (1981). An articulatory synthesizer for perceptual research. *Journal of the Acoustical Society of America,* **70**, 312-328.

Rubin, P.E., Saltzman, E.L., Goldstein, L., McGowan, R.S., Tiede, M.K. & Browman, C.P. (1996). CASY and extensions to the task-dynamic model. In *Proceedings of the 4th Speech Production Seminar - 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics*, pp. 125-128. Autrans, France.

Russell, O.G. (1928). *The vowel. Its physiological mechanism as shown by X-Ray.* Maryland: McGrath Publishing Company.

Sanguineti, V., Laboissière, R. & Ostry, D. (1998). A dynamic biomechanical model for neural control of speech production. *Journal of the Acoustical Society of America,* **103**(3), 1615-1627.

Savariaux, C., Perrier, P. & Orliaguet, J.-P. (1995). Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *Journal of the Acoustical Society of America,* **98**(5, Pt.1), 2428-2442.

Schönle, P.W. (1992). The developmental genealogy of Electromagnetic Articulography (EMA). *Forschungsberichte des Instituts für Phonetik und Spachliche Kommunikation der Universität München, FIPKM,* **30**, 83-90.

Schroeter, J. & Sondhi, M.M. (1986). Articulatory synthesizer for research in low bit-rate coding of speech. In *Proceedings of the 12th International Conference on Acoustics*, pp. A4-4.

Schwartz, J.-L., Boë, L.-J., Vallée, N. & Abry, C. (1997). The dispersion-focalization theory of vowel systems. *Journal of Phonetics,* **25**, 255-286.

Scully, C. (1986). Speech production simulated with a functional model of the larynx and the vocal tract. *Journal of Phonetics,* **14**, 407-413.

Scully, C. (1991). The representation in models of what speakers know. In *Proceedings of the XIIth International Congress of Phonetic Sciences*, vol. 1, pp. 192-197.

Scully, C. & Allwood, E. (1985). Production and perception of an articulatory continuum for fricatives of English. *Speech Communication,* **4**, 237-245.

Shadle, C.H. (1986). Models of fricative consonants involving sound generation along the wall of a tube. In *Proceedings of the 12th International Conference on Acoustics*, pp. A3-4.

Shadle, C.H. (1990). Articulatory-acoustic relationships in fricative consonants. In *Speech Production and Speech Modelling* (W.J. Hardcastle & A. Marchal, editors), pp. 187-209.: Kluwer Academic Publisher.

Shirai, K. & Honda, M. (1976). An articulatory model and the estimation of articulatory parameters by nonlinear regression model. *Transactions of the IECE Japan,* **59-A**(8), 668-674.

Sinder, D.J., Krane, M.H. & Flanagan, J.L. (1998). Synthesis of fricative sounds using an aeroacoustic noise generation model. In *Proceedings of the 16th International Conference on Acoustics and 135th Meeting of the Acoustical Society of America* (P.K. Kuhl & L.A. Crum, editors), vol. I, pp. 249-250. Seattle, USA.

Stark, J., Lindblom, B. & Sundberg, J. (1996). APEX an articulatory synthesis model for experimental and computational studies of speech production. *Tal Musik Hörsel - Quaterly Progress Status Report - Stockholm,* **2/1996**, 45-48.

Stevens, K.N. (1971). Airflow and turbulence noise for fricative and stop consonants: static considerations. *Journal of the Acoustical Society of America,* **50**, 1180-1192.

Stevens, K.N. (1993). Models for the production and acoustics of stop consonants. *Speech Communication,* **13**, 367-375.

Stevens, K.N., Blumstein, S.E., Glicksman, L., Burton, M. & Kurowski, K. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. *Journal of the Acoustical Society of America,* **91**, 2979-3000.

Stevens, K.N. & House, A.S. (1955). Development of a quantitative description of vowel articulation. *Journal of the Acoustical Society of America,* **27**, 484-493.

Stevens, K.N., Kasowski, S. & Fant, G. (1953). An electrical analog of the vocal tract. *Journal of the Acoustical Society of America,* **25**, 734-742.

Stewart, J.Q. (1922). An electrical analogue of the vocal organs. *Nature,* **110**(2757), 311-312.

Vallée, N., Schwartz, J.-L. & Escudier, P. (1999). Phase Spaces of vowel systems. A typology in the light of the Dispersion-Focalization Theory (DFT). In *Proceedings of the 14th International Congress of Phonetic Sciences* (J.J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A.C. Bailey, editors), vol. 1, pp. 333-336. San Francisco, USA, Congress organizers at the Linguistics Department, University of California, Berkeley.

Vatikiotis-Bateson, E., Kuratate, T., Kamachi, M. & Yehia, H. (1999). Facial deformation parameters for audiovisual synthesis. In *Proceedings of AVSP'99 (Auditory-Visual Speech Processing)* (D.W. Massaro, editor, pp. 118-122. University of California, Santa Cruz, California, USA.

Vescovi, C., Castelli, E. & Pelorson, X. (1995). Adaptation of a two-mass model of the vocal cords to a particular speaker. In *Proceedings of the 4th EuroSpeech Conference* (J.M. Pardo, E. Enríquez, J. Ortega, J. Ferreiros, J. Macías & F.J. Valverde, editors), vol. 3, pp. 1933-1936. Madrid, Spain, GrÀficas Brens.

Vilain, A. (2000). *Apports de la modélisation des degrés de liberté articulatoires à l'étude de la coarticulation et du développement de la parole.* Unpublished Thèse doctorale : Sciences du Langage, Université Stendhal, Grenoble.

Vilain, A., Abry, C. & Badin, P. (1998a). Coarticulation and degrees of freedom in the elaboration of a new articulatory plant: Gentiane. In *Proceedings of the 5th International Conference on Spoken Language Processing* (R.H. Mannell & J. Robert-Ribes, editors), vol. 7, pp. 3147-3150. Sydney, Australia, Australian Speech Science and Technology Association Inc.

Vilain, A., Abry, C. & Badin, P. (1998b). A propos des degrés de liberté dans la coarticulation d'un locuteur français filmé 50 fois par seconde. In *Actes des 22èmes Journées d'Etude sur la Parole*, pp. 311-314. Martigny, Suisse.

Vilain, A., Abry, C. & Badin, P. (1999). Motor equivalence evidenced by articulatory modelling. In *Proceedings of the 6th EuroSpeech Conference*, vol. 1, pp. 169-172. Budapest, Hungary.

Vilain, A., Abry, C. & Badin, P. (2000). Coproduction strategies in French VCVs: confronting Öhman's model with adult and developmental articulatory data. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 81-84. Kloster Seeon, Germany.

von Kempelen, W.R. (1791). *Mechanismus der menschlichen Sprache nebst der Beschreibung einer sprechende Maschine.* Wien: Degen, J.B.

Wilhelms-Tricarico, R. (1995). Physiological modeling of speech production: Methods for modeling soft-tissue articulators. *Journal of the Acoustical Society of America,* **97**(5, Pt. 1), 3085-3098.

Wilhelms-Tricarico, R. & Perkell, J.S. (1995). Towards a physiological model of speech production. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, editors), vol. 2, pp. 68-75. Stockholm, Sweden, Arne Strömbergs Grafiska Press.

Wu, H.Y., Badin, P., Cheng, Y.M. & Guérin, B. (1987a). Continuous variations of the vocal tract length in a Kelly-Lochbaum type speech production model. In *Proceedings of the 11th International Congress of Phonetic Sciences*, vol. 2, pp. 340-343. Tallinn, Estonia.

Wu, H.Y., Badin, P., Cheng, Y.M. & Guérin, B. (1987b). Simulation du conduit vocal: réalisation de la variation continue de longueur dans un modèle de Kelly-Lochbaum - Effets de l'échantillonnage spatial de la fonction d'aire. *Bulletin du Laboratoire de la Communication Parlée*, **1**, 1-27.

Wu, H.Y., Badin, P., Cheng, Y.M. & Guérin, B. (1987c). Vocal tract simulation: implementation of continuous variations of the length in a Kelly-Lochbaum model, Effects of area function spatial sampling. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 9-12.

Yu, Z. & Zeng, S. (2000). Articulatory synthesis using a vocal-tract model of variable length. In *Proceedings of the 6th International Conference on Spoken Language Processing* (B. Yuan, T. Huang & X. Tang, editors), vol. I, pp. 693-696. Beijing, China.

Deuxième partie : Participation à la vie scientifique

Cette deuxième partie présente de façon plus administrative les diverses activités auxquelles j'ai été amené à participer en tant que chercheur.

V. ENCADREMENT DE CHERCHEURS

Les thèses de doctorat de l'INPG que j'ai dirigées ou co-dirigées illustrent les divers aspects de mon propre travail de recherche: mesures in-vitro, établissement de modèles, et enfin à partir de ces acquis, récupération des gestes articulatoires et synthèse articulatoire. J'ai co-encadré la thèse d'Eric Castelli avec Bernard Guérin, et Pascal Perrier a assuré l'encadrement de la dernière année, lors de laquelle j'étais à Stockholm. La thèse de Denis Beautemps comportait deux aspects complémentaires liés à l'inversion en parole : j'ai encadré le travail sur l'inversion des consonnes fricatives lié à la production de la parole tandis que Jean-Luc Schwartz s'est occupé de la partie liée à la perception. J'ai assuré entièrement l'encadrement des thèses de Yen Pham Thi Ngoc et de Khaled Mawass. J'ai enfin co-encadré la thèse de Solange Rossato avec Gang Feng et Rafael Laboissière et celle d'Anne Vilain avec Christian Abry, pour les aspects liés aux données et modèles articulatoires.

Outre ces (co-)encadrements doctoraux, j'ai guidé les pas d'un certain nombre d'étudiants en DEA SIPT, SL et SC, sans compter des ingénieurs et d'autres étudiants. La liste détaillée de ces encadrements figure ci-dessous.

A. Thèses (7)

- Vilain, Anne : Apports de la modélisation des degrés de liberté articulatoires à l'étude de la coarticulation et du développement de la parole (Thèse de Doctorat de l'Université Stendhal, SL, septembre 1997 - décembre 2000 ; co-encadrement à 50 % avec Ch. Abry)
- Rossato, Solange : Du son au geste, inversion de la parole: le cas des voyelles nasales (Thèse de Doctorat de l'INPG, SIPT, septembre 1997 - décembre 2000 ; co-encadrement à 30% avec G. Feng et R. Laboissière)
- Borel, Pascal : Modélisation articulatoire et acoustique des nasales et des latérales (Thèse de Doctorat de l'INPG, SIPT, septembre 1999 - juillet 2000, abandon)
- Mawass, Khaled : Synthèse articulatoire des consonnes fricatives (Thèse de Doctorat de l'INPG, SIP, septembre 1994 - septembre 1997)
- Pham Thi Ngoc, Yen : Caractérisation acoustique du conduit vocal: fonctions de transfert acoustiques et sources de bruit. Étude des voyelles chuchotées et des consonnes fricatives non voisées (Thèse de Doctorat de l'INPG, SIP, octobre 1991 - janvier 1995).
- Beautemps, Denis : Récupération des gestes de la parole à partir de trajectoires formantiques: Identification de cibles vocaliques non-atteintes et modèles pour les profils sagittaux des consonnes fricatives (Thèse de Doctorat de l'INPG, SIP, septembre 1990 - février 1993; co-encadrement à 50 % avec J.L. Schwartz)
- Castelli, Eric : Caractérisation acoustique des voyelles nasales du français. Mesures, modélisation et simulation temporelle (Thèse de Doctorat de l'INPG, SIP, septembre 1986 - juin 1989 ; co-encadrement à 30 % avec B. Guérin et P. Perrier)

B. DEA (11)

- Trompat, Julien : Perception audiovisuelle de la parole: effet McGurk et réalité augmentée (DEA *Sciences Cognitives*, UJF-INPG, septembre 2000 - juin 2001, co-encadrement avec Marie-Agnès Cathiard).
- Morvan, Yan : Modélisation articulatoire d'une tête parlante virtuelle tenant compte de l'opposition Neutre / Sourire par analyse de données vidéo (Ingénieur ENSERG + DEA SIPT, septembre 1999 - septembre 2000)
- Borel, Pascal : Modélisation articulatoire linéraire d'un visage incluant des lèvres (Ingénieur ENSERG + DEA SIPT, septembre 1998 - septembre 1999)

- Vilain, Anne : Un nouveau modèle articulatoire pour la synthèse et le contrôle robotique de la parole (DEA SL, septembre 1996 – Juin 1997, co-encadrement avec Christian Abry).
- Houcine, Kermiche : Développement d'un modèle de passage de la fonction sagittale à la fonction d'aire à l'aide de techniques d'optimisation sous contrainte. (DEA SIP, septembre 1994 - juin 1995, co-encadrement avec Denis Beautemps)
- El-Masri, Samir : Application de la méthode TLM aux ondes acoustiques. (DEA OOM, Co-encadrement avec Xavier Pelorson et Pierre Saguet, septembre 1993 - juin 1994)
- Karami-Mollaei, Mohamad : Modèle de passage de la fonction sagittale à la fonction d'aire pour l'inversion du conduit vocal à l'aide de techniques d'optimisation sous contrainte. (DEA SIP, septembre 1993 - juin 1994)
- Issa, Imad : Etude de la parole en atmosphère Hélium / Oxygène (DEA SIP, septembre 1991 - juin 1992)
- El-Jakl, Jalil : Modélisation de la source de bruit pour les consonnes fricatives (DEA SIP, septembre 1990 - juin 1991)
- Kamlé,Tawfik : Implantation d'une source de bruit fricatif à spectre commandable dans un modèle de conduit vocal (ligne à réflexion) (DEA SIP, septembre 1989 - septembre 1990)
- Udo, Ogemdi : Etude et simulation des fonctions de sensibilité du conduit vocal appliquées aux configurations nasales (DEA SIP, septembre 1989 - juin 1990)

C. Ingénieurs (14)

- De Penguern, Guillaume (Ingénieur ENSERG, Mars 2001 – Juin 2001) Modélisation articulatoire 2D / 3D et acoustique des nasales pour une tête parlante audiovisuelle.
- Morvan, Yan : Modélisation articulatoire d'une tête parlante virtuelle tenant compte de l'opposition Neutre / Sourire par analyse de données vidéo (Ingénieur ENSPG, septembre 1999 - juin 2000)
- Pouchoy, Laurent : Modèle articulatoire 3D de conduit vocal (Ingénieur Institut Polytechnique de Sévenans, février 1999 - juillet 1999).
- Pouchoy, Laurent : Logiciel de traitement d'images IRM (Ingénieur Institut Polytechnique de Sévenans, septembre 1997 - février 1998).
- López, GermÀn : Inversion d'un modèle articulatoire de langue à partir de données articulographique et audiovisuelles (Ingénieur Institut Polytechnique de Madrid, octobre 1997 - mars 1998).
- Baricchi, Enrico : Inversion d'un modèle articulatoire de langue à partir de données articulographiques (Ingénieur Institut Polytechnique de Parme, octobre 1996 - mars 1997).
- Chavagnat, Sylvain : Logiciel de traitement d'images cinéradiographiques du conduit vocal (Ingénieur ENSERG, septembre 1995 - juin 1996, co-encadrement avec P.Y. Coulon et T. Lallouache).
- Embling, Clare: Logiciel de traitement d'images cinéradiographiques du conduit vocal (Auditrice libre ENSERG, Ingénieur University College, London, septembre 1995 - juin 1996, Co-Encadrement avec P.Y. Coulon et T. Lallouache).
- Bouffartigue, Christophe : Modélisation mécanique et aérodynamique des lèvres pour la production de consonnes plosives bilabiales. Acquisition de données labiométriques à la caméra vidéo rapide (Ingénieur ENSERG, co-encadrement avec X. Pelorson, septembre 1993 - juin 1994).
- Tourret, Stéphane : Modélisation mécanique et aérodynamique des lèvres pour la production de consonnes plosives bilabiales. Acquisition de données labiométriques à la caméra vidéo rapide. (Ingénieur ENSERG, co-encadrement avec X. Pelorson, septembre 1993 - Juin 1994)
- Orlando, Paolo : Modellizzazione della glottide come sorgente di segnali vocali (Stagiaire Erasmus, Université de Bologne, Italie, mars - juin 1993)
- Ritterhaus, Diane: Midsagittal distance and area functions for a set of fricatives and vowels (Stagiaire TU Dresden, janvier - avril 1992)
- Dridi, Ryad : Mise en place du logiciel de mesure de fonction de transfert du CV sur PC (Ingénieur, ICPI, Lyon, 1990)

- Hastreiter, Peter : Système de mesure de la fonction de transfert acoustique du CV en temps réel (Ingénieur, Université de Münich , Allemagne, 1990)

D. Divers (7)

- Engwall, Olle : Modèle articulatoire 3D du Suédois à partir de données IRM (Doctorant de KTH, Juin-Juillet et septembre 1999)
- Beautemps, Denis (Post-Doc, septembre 1994 - juillet 1996, co-encadrement avec G. Bailly])
- Bouvier, David : Modélisation acoustique du conduit vocal dans le domaine fréquentiel (IUT Informatique de Grenoble, avril - juin 1996)
- Chosson, Cédric : Modélisation acoustique du conduit vocal dans le domaine fréquentiel (IUT Informatique de Grenoble, avril - juin 1996)
- Wellner, Torsten (Stagiaire Erasmus, octobre 1993 - avril 1994)
- Barraclough, Lorna (Assistante de recherche, Contrat SCIENCE "Mesure, caractérisation et modélisation des sons fricatifs", septembre 1989 - août 1990).
- Tzavali, Efthasia (Assistante de recherche, Contrat SCIENCE "Mesure, caractérisation et modélisation des sons fricatifs", septembre 1990 - août 1991)

E. Participation à des jurys

Pour terminer cette section, je dois mentionner que j'ai été membre du jury de thèse de Denis Beautemps, Yen Pham Thi Ngoc, Khaled Mawass, Solange Rossato, Anne Vilain, et que j'ai été invité en tant qu'*opposant* à la défense de thèse de licence (*licenciate thesis,* Towards an articulatory speech synthesizer: Model development and simulations) de Mats Båvegård, KTH, Stockholm (16 février 1996).

VI. PARTICIPATION A COLLOQUES ET CONGRES

La liste exhaustive de mes publications dans des colloques internationaux est donnée à la section XI, et je me limite donc ici à mentionner les manifestations auxquelles j'ai participé en tant que présentateur.

2000

- *Fifth Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, Münich, Allemagne
- *23èmes Journées d'Etudes de la Parole*, Aussois

1999

- *Satellite workshop on "Speech Technology applications in CALL", EUROCALL'99*, Besançon (invité)

1998

- *ESCA Tutorial and Research Workshop on Speech Technology in Language Learning*, Stockholm, Suède
- *Hokkaido Workshop on Speech Production*, Kutchan, Japon

1997

- *Fifth EuroSpeech Conference,* Rhodes, Grèce

1996

- *Fourth Speech Production Seminar - First ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics*, Autrans (contribution invitée avec Christian Abry)
- *International Conference on Spoken Language Processing* Philadelphie, USA, (invité)

1995

- *13th International Congress of Phonetic Sciences*, Stockholm, Suède

1994

- *Third International Conference on Spoken Language Processing*, Yokohama, Japon (invité)

1993

- *Third Seminar on Speech Production, Data and Models*, New Haven, CT, USA
- *Acoustical Society of America*, Ottawa, Canada

1991

- *12th International Congress of Phonetic Sciences,* Aix-en-Provence

1990

- *Second Seminar on Speech Production, Data and Models*, Leeds, UK

1989

- *First EuroSpeech Conference*, Paris

1988

- *First Seminar on Speech Production, Data and Models*, Grenoble

1987

- *11th International Congress of Phonetic Sciences*, Tallinn, URSS

1985

- *French-Swedish Seminar on Speech*, Grenoble

1982

- *IEEE International Conference on Acoustics, Speech and Signal Processing*, Paris

D'autre part, au cours de visites de laboratoires à l'étranger, j'ai été amené à donner des conférences sur divers aspects de la production de la parole, en particulier à l'Université de Tokyo et à l'Université du Hokkaido au Japon, et à l'Université *Johns Hopkins* de Baltimore et aux *Laboratoires Haskins* à New Haven aux Etats-Unis.

VII. PARTICIPATION A LA VIE DU LABORATOIRE

Cette section présente les différents aspects de ma participation à la vie du laboratoire.

Animation de l' équipe Acoustique (1990- présent). Un aspect important est mon rôle d'animateur de l'équipe *Acoustique*, depuis 1990. Cette équipe a compté jusqu'à une dizaine de personnes (3-4 permanents, 3-4 thésards et 4-5 stagiaires divers), même si elle est plus réduite aujourd'hui. Dans ce cadre, j'ai été régulièrement amené à participer à l'encadrement scientifique de chercheurs encadrés par d'autres membres de l'équipe, et à coordonner les différents travaux de recherche afin de les rendre plus cohérents, et de faire avancer notre projet, à savoir la compréhension et la modélisation des phénomènes acoustiques et aérodynamiques qui président à la production des sons de parole.

Activité d'éditeur (1990- présent). Je participe également au rayonnement international du laboratoire à travers l'édition des *Cahiers de l'ICP*, créés en 1990 avec C. Abry, L.J. Boë, P. Perrier et J.L. Schwartz, et dont 12 numéros sont sortis à ce jour. Les *Cahiers de l'ICP* regroupent le *Bulletin de la Communication Parlée*, revue francophone sur la parole, avec comité de lecture, les *Rapports de Recherche de l'ICP*, et des *Monographies*. Ces cahiers font suite au *Bulletin du Laboratoire de la Communication Parlée*, créé avec P. Perrier et J.L. Schwartz, dont quatre numéros sont parus entre 1987 et 1990, lorsque nous n'avions encore pas encore réalisé une fusion complète avec le *Bulletin de l'Institut de Phonétique de Grenoble.*

Gestion de la bibliothèque de l'ICP-gare (1986-présent). Depuis 1986, je gère la bibliothèque de l'ICP, site Viallet, à l'aide d'une bibliothécaire à temps partiel, en liaison avec la bibliothèque de l'ENSERG.

Conseil de laboratoire (1990- présent). Je suis membre élu du Conseil de Laboratoire de l'ICP depuis Janvier 1994, après avoir été invité permanent de ce même conseil, au titre de responsable de l'équipe *Acoustique*, depuis 1990.

Plan informatique de l'ICP (1985). En équipe avec J.L. Schwartz, j'ai assuré, en 1985, la coordination pour la remise à niveau du parc informatique du laboratoire (définition des besoins, contacts avec les constructeurs, programmes d'essai, rapports auprès de la Commission Informatique). Ce travail avait permis au laboratoire de combler un certain retard en équipement et en organisation informatique.

VIII. ADMINISTRATION DE LA RECHERCHE

A. Contrats et projets

Mesure, caractérisation et modélisation des consonnes fricatives (SCIENCE, 1989-1990). De 1989 à 1991, j'ai été co-responsable, avec Bernard Guérin, de l'Action de la CEE SCIENCE *SC1*0147 C (EDB)*: *Mesure, caractérisation et modélisation des consonnes fricatives*, en collaboration avec l'Université de Leeds et l'Université de Southampton. Ce projet m'a en particulier permis d'embaucher une assistante de recherche pendant deux ans, et a donné lieu à plusieurs publications et contributions à des conférences.

Base de données articulatoires et acoustiques pour la parole (Réseau Européen, 1991). En 1991, j'ai joué le rôle de Coordinateur du Réseau Européen de Laboratoires *Base de données articulatoires et acoustiques pour la parole*, financé par le Ministère de la Recherche et de la Technologie, avec l'ENST à Paris, et les Universités de Stockholm, Leeds et Southampton.

Speech Maps (ESPRIT/BR, 1992-1995). Le réseau *Base de données articulatoires et acoustiques pour la parole* a joué un rôle important dans l'élaboration du projet européen ESPRIT/BR N°6975 *Speech Maps*, que j'ai co-dirigé avec Christian Abry de 1992 à 1995. Ce projet, d'un budget total de 10 MF, a rassemblé 14 laboratoires européens travaillant dans le domaine de la production de la parole et de la robotique anthropomorphique, pour essayer de construire un robot parlant capable de reproduire des sons de parole par apprentissage. Ce projet, sur lequel j'ai quasiment travaillé à temps plein pendant quatre ans, a permis de manière générale de formaliser le concept de *robotique de la parole*, de faire progresser considérablement nos modèles de production et de perception de la parole, de développer et tester des méthodes d'inversion, et a finalement abouti à de nombreuses publications et présentations à des conférences internationales, ainsi qu'à un film audiovisuel de démonstration qui a été présenté à de nombreuses occasions depuis la fin du projet en décembre 1995.

Modélisation et simulation acoustique du conduit vocal (projet ELESA, 1995-1996). J'ai joué le rôle de correspondant à l'ICP de ce projet intégré à l'axe « modélisation et simulation des phénomènes acoustiques, électroniques ou électromagnétiques » de la fédération *ELESA* visant à appliquer la méthode TLM à l'acoustique du conduit vocal.

Une tête parlante virtuelle : données et modèles en production de parole (ARASSH, 1997-1999). De 1997 à 1999, j'ai coordonné avec Louis-Jean Boë, un projet financé dans le cadre du Programme de Recherche de l'ARASSH (Agence Rhône-Alpes pour les Sciences Sociales et Humaines) et intitulé *Une tête parlante virtuelle : données et modèles en production de parole*. Ce projet a permis de mettre en place une base de données articulatoires, développer des modèles pour les articulateurs et les intégrer dans la tête parlante virtuelle, d'étudier la variabilité et l'espace de contrôle de ces modèles, et d'induire des prototypes pour les phonèmes les plus fréquents des langues du monde.

En dehors de la coordination de ces projets, j'ai participé, ou je participe, directement à une série de projets coordonnés par d'autres collègues de l'ICP :

- ***Les robots parlent aux robots. Un générateur des structures sonores du langage (1996-1998)***. Projet du *GIS Sciences de la Cognition*, Cognition naturelle / Cognition artificielle ; Resp. Christian Abry.
- ***Valorisation de la banque de données cinéradiographiques de l'Institut de Phonétique de Strasbourg. Numérisation des données et élaboration d'une plate-forme multimédia pour leur analyse (1996-1999)***. Projet *Ingénierie des langues du SHS / CNRS* ; Resp. Pascal Perrier.
- ***Modèles 3D de visages parlants (1999-2000)***. Projet *France Telecom R&D* ; Resp. Gérard Bailly.
- ***TEMPO-VALSE, Terminal Expérimental MPEG4 POrtable de Visiophonie et Animation Labiale ScalablE (2000-2002)***. Projet du *RNRT* (Réseau National de Recherche en Télécommunications) ; Resp. à l'ICP, Gérard Bailly.
- **ABISPA, Apprentissage Bayésien Intersensoriel de Structures Phonologiques par un Androïde bébé (2001-2003)**. Projet *Cognitique, Action 2000*, Thème « Langage et Cognition » du MENRT ; Resp. Louis-Jean Boë.
- ***De 1'hyper-articulation au langage parlé complété: robustesse et versatilité de la communication langagière. Etude des stratégies adaptatives humaines et modélisation d'interfaces multimodales (2000-2002)***. Projet *jeune équipe CNRS* ; Resp. Denis Beautemps.
- ***Tête parlante audiovisuelle dotée de réalité augmentée, augmentée de Langage Parlé Complété pour la réhabilitation des déficients auditifs (2001-2003)***. Projet financé par le *Programme Cognitique du MENRT* ; Resp. Denis Beautemps.
- ***ARTUS, Animation Réaliste par Tatouage audiovisuel à 1'Usage des Sourds (2001- 2004)***. Projet du *RNRT* (Réseau National de Recherche en Télécommunications). Resp. Gérard Bailly.

B. Organisation de séminaires

Suite à un séjour de 20 mois à Stockholm, j'ai participé à l'organisation de rencontres franco-suédoises (*French-Swedish Seminar on Speech Production*, Grenoble, 1985; *French-Swedish Seminar on Speech Recognition and Perception* , Grenoble, 1987).

J'ai par ailleurs participé à l'organisation du *First Seminar on Speech Production* (Grenoble, 1988), et j'ai été membre du comité scientifique du *Fourth Speech Production Seminar, First ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics* (Autrans, 1996).

Dans le cadre du projet *Speech Maps*, j'ai été amené à organiser quatre rencontres internationales qui ont a chaque fois réuni une quarantaine de participants (Grenoble, 1992, 1993, 1995; Stockholm, 1994).

Enfin, plus récemment, j'ai co-organisé avec Gérard Bailly les XXIII^{èmes} Journées d'Etude sur la Parole à Aussois du 19-23 juin 2000, colloque qui a rassemblé plus de 130 participants francophones sur tous les thèmes qui touchent à la communication parlée.

C. Activités éditoriales

Je suis régulièrement relecteur pour divers journaux internationaux :

- *Journal of the Acoustical Society of America*
- *Journal of Phonetics*
- *Speech Communication*
- *IEEE Transactions on Speech and Audio Processing*

J'ai également été relecteur d'un ouvrage pour *Kluwer Academic Publishers.*

IX. SEJOURS A L'ETRANGER ET MISSIONS SUR LE TERRAIN

À de nombreuses occasions, j'ai pu effectuer des séjours scientifiques dans différents pays. Je présente ici deux séjours de longue durée à Stockholm, et un ensemble de campagnes de mesures dans plusieurs grands laboratoires internationaux.

A. Séjours de longue durée

KTH, Stockholm, Suède (Post-Doc INRIA, avril 1983 – mars 1984). Mon premier séjour de longue durée à l'étranger s'est déroulé au *Department of Speech Communication and Music Acoustics* du *Kungliga Tekniska Högskolan* (KTH), Stockholm. Dans le cadre de ce stage post-doctoral, financé par une bourse de recherche de l'INRIA, j'ai travaillé en collaboration avec le Professeur Gunnar Fant sur l'*acoustique du conduit vocal*, et développé un programme de simulation acoustique du conduit vocal.

KTH, Stockholm, Suède (Assistant de recherche, avril 1984 – septembre 1984). J'ai ensuite eu la chance de pouvoir continuer mon travail de *simulation acoustique du conduit vocal* grâce à un poste d'assistant de recherche au KTH.

KTH, Stockholm, Suède (mission CNRS, octobre 1984 – novembre 1984). J'ai achevé mon séjour à Stockholm, en tant qu'attaché de recherche au CNRS en mission, pour travailler à l'*amélioration d'un système de transcription orthographique-phonétique pour le français*.

KTH, Stockholm, Suède (année sabbatique, mars 1988 – août 1989). Un Programme International de Coopération Scientifique (PICS) établi entre le KTH et l'ICP m'a permis d'effectuer à Stockholm un séjour sabbatique au cours duquel j'ai en particulier effectué ma première série d'*enregistrements in vivo de données acoustiques et aérodynamiques*, et travaillé sur la *théorie acoustique des consonnes fricatives.*

B. Campagnes de mesures *in vivo* et *in vitro*

1. In vivo

Depuis mon deuxième séjour à Stockholm, j'ai commencé à développer une politique d'acquisition de données *in vivo*. Pour comprendre et modéliser les phénomènes complexes de la production de la parole, il est nécessaire de mesurer le maximum de paramètres aérodynamiques, acoustiques, géométriques ou articulatoires correspondant à la production de sons donnés. Un certain nombre de dispositifs expérimentaux ont été développés afin de mesurer certains paramètres, mais malheureusement ces dispositifs sont pour la plupart incompatibles entre eux. Dans le cadre de la politique du sujet de référence exposée dans l'autre partie du mémoire, j'ai servi de sujet en prononçant le même corpus de séquences parfaitement contrôlées à l'aide des différents dispositifs compatibles, afin de reconstituer une image la plus

complète possible du phénomène global. C'est ainsi que j'ai été amené à effectuer des séjours dans un certain nombre de laboratoires. Je peux également ajouter que mon premier contact avec la plupart des dispositifs expérimentaux que j'utilise à Grenoble a eu lieu lors de collaborations et de séjours à l'étranger.

Huddinge Hospital, Suède (juillet 1989). Enregistrement d'un film en vidéo normale de mes cordes vocales illuminées par stroboscopie, en collaboration avec le KTH.

Speech Laboratory, University of Leeds, UK (mars 1990). *Enregistrements acoustiques* de référence de corpus de fricatives statiques et dynamiques. *Enregistrements aérodynamiques* à l'aide d'un masque pneumotachographique.

Department of Electronics and Computer Sciences, University of Southampton, UK (mars 1990). *Enregistrements Hifi et laryngographiques* pour les mêmes corpus de fricatives statiques et dynamiques.

Département de Radiologie Faciale, CHRU, Grenoble (avril 1990). Enregistrement d'une série de *téléradiographies statiques* de voyelles et de fricatives.

Speech Laboratory, University of Leeds, UK (mars 1993). *Enregistrements électropalatographiques* des mêmes corpus de fricatives statiques et dynamiques.

Department of Electronics and Computer Sciences, University of Southampton, UK (mars 1993). *Enregistrements vidéo en lumière structurée pour la détermination de la forme des lèvres* des mêmes corpus de fricatives statiques et dynamiques.

Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, USA (mai 1993). Au cours d'une mission d'étude aux Etats-Unis en mai 1993, j'ai pu effectuer des *enregistrements de la forme de la langue par imagerie ultrasonique*, en relation avec l'hôpital Johns Hopkins.

Speech Laboratory, University of Leeds, UK (mars 1994). *Enregistrements acoustiques* de corpus complets de séquences Voyelle–Consonne–Voyelle. *Enregistrements aérodynamiques* à l'aide d'un masque pneumotachographique.

Department of Electronics and Computer Sciences, University of Southampton, UK (mars 1994). *Enregistrements Hifi, laryngographiques, et en lumière structurée* pour les mêmes corpus de séquences Voyelle–Consonne–Voyelle.

Service de cardiologie, Hôpital Schiltigheim, Strasbourg (juin 1994). Enregistrement d'un *film cinéradiographique synchrone avec un labiofilm vidéo*, en collaboration avec l'*Institut de Phonétique* de Strasbourg.

Service de cardiologie, Hôpital Schiltigheim, Strasbourg (juin 1995). Enregistrement d'un *film cinéradiographique synchrone avec un labiofilm vidéo du sujet J1X*, en collaboration avec l'*Institut de Phonétique* de Strasbourg.

General Hospital, Nara, Japon (février 1996). Grâce à une invitation de la JSPS (*Japan Society for the Promotion of Science*), j'ai pu enregistrer des séries d'*images bi- et tri-dimensionnelle du conduit vocal par IRM* au, en collaboration avec *ATR* (Advanced Telecommunication Research) à Kyoto, Japon.

Service de radiologie, CHRU, Grenoble (février 1998). Enregistrement d'*images IRM*, en collaboration avec l'*UM Université Joseph Fourier / INSERM U438, LRC CEA*. Plusieurs autres séances ont suivi.

Research Institute for Logopedics and Phoniatrics, University of Tokyo (août 1998). Enregistrement d'un court film à 4000 images / sec. de mes cordes vocales pour des séquences contenant des fricatives.

2. In vitro

Pour compléter les mesures *in vivo*, une longue collaboration avec le Japon m'a permis d'effectuer des mesures acoustiques sur diverses maquettes. Une partie de ce travail collaboratif a été financée dans le cadre d'un *Projet de Collaboration Internationale* par le Ministère de l'Education Japonais *Monbusho*.

Institute of Applied Electricity, Hokkaido University, Sapporo, Japon (décembre 1990 – janvier 1991). Mesures d'*impédance d'entrée de modèles de lèvres en plâtre*.

Institute of Applied Electricity, Hokkaido University, Sapporo, Japon (juillet 1993). Mesures de *pression acoustique dans des modèles simplifiés en plexiglas*.

Hokkai Gakuen University, Sapporo, Japon (août–septembre 1994). Mesures du *rayonnement acoustique d'un tuyau rectangulaire baffle*.

X. DIFFUSION DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE

L'un des résultats important du projet ESPRIT/BR *Speech Maps* était une démonstration de synthèse articulatoire audiovisuelle. J'ai donc coordonné et participé à la réalisation d'une bande audiovisuelle de démonstration qui a été présentée à de nombreuses occasions depuis la fin du projet en décembre 1995 (voir ci-dessous).

J'ai par ailleurs participé depuis 1995 à la présentation des travaux de l'ICP aux diverses éditions de la « Science en Fête » à Grenoble.

Enfin tout récemment, j'interviens en tant que sujet de référence dans l'émission scientifique *Archimède* de la chaîne de télévision *Arte*.

XI. PUBLICATIONS CLASSEES

Je présente ici l'ensemble de mes 113 publications, répertoriées selon la classification traditionnelle.

1. Revues internationales avec comité (10)

Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C. & Savariaux, C. (In press). Three-dimensional articulatory modeling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*.

Beautemps, D., Badin, P. & Bailly, G. (2001). Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. *Journal of the Acoustical Society of America, 109*(5), 2165-2180.

Mawass, K., Badin, P. & Bailly, G. (2000). Synthesis of French fricatives by audio-video to articulatory inversion. *Acta Acustica, 86*(1), 136-146.

El Masri, S., Pelorson, X., Saguet, P. & Badin, P. (1998). Development of the Transmission Line Matrix method in acoustics. Application to higher modes in the vocal tract and other complex ducts. *International Journal of Numerical Modelling, 11*, 133-151.

Badin, P., Beautemps, D., Laboissière, R. & Schwartz, J.-L. (1995). Recovery of vocal tract geometry from speech signal for vowels and fricative consonants using a midsagittal-to-area function conversion model. *Journal of Phonetics, 23*, 221-229.

Beautemps, D., Badin, P. & Laboissière, R. (1995). Deriving vocal-tract area functions from midsagittal profiles and formant frequencies: A new model for vowels and fricative consonants based on experimental data. *Speech Communication, 16*, 27-47.

Badin, P., Motoki, K., Miki, N., Ritterhaus, D. & Lallouache, T.M. (1994). Some geometric and acoustic properties of the lip horn. *Journal of the Acoustical Society of Japan (English), 15*(4), 243-253.

Badin, P. (1991). Fricative consonants: acoustic and X-ray measurements. *Journal of Phonetics, 19*, 397-408.

Djéradi, A., Guérin, B., Badin, P. & Perrier, P. (1991). Measurement of the acoustic transfer function of the vocal tract: a fast and accurate method. *Journal of Phonetics, 19*, 387-395.

Badin, P., Perrier, P., Boë, L.-J. & Abry, C. (1990). Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *Journal of the Acoustical Society of America, 87*, 1290-1300.

2. Ouvrages ou chapitres dans un ouvrage (2)

Bailly, G., Vatikiotis-Bateson, E., Badin, P., Revéret, L. & Yehia, H. (In preparation). Visible characteristics of speech production. In *Audiovisual speech* (E. Vatikiotis-Bateson, G. Bailly & P. Perrier, Eds.). Cambridge, MA, USA: MIT Press.

Abry, C., Badin, P. & Scully, C. (1994). Sound-to-gesture inversion in speech: The *Speech Maps* approach. ESPRIT Research Report No. 6975. In *Advanced Speech Applications* (K. Varghese, S. Pfleger & J.P. Lefèvre, Eds.), pp. 182-196. Berlin: Springer Verlag.

3. Colloques internationaux avec comité (45)

Elisei, F., Odisio, M., Bailly, G. & Badin, P. (2001). Creating and controlling video-realistic talking heads. In *Proceedings of the Auditory-Visual Speech Processing Workshop, AVSP 2001* (D.W. Massaro, J. Light & K. Geraci, Eds.), pp. 90-97. Aalborg, Denmark, 2001.

Apostol, L., Perrier, P., Baciu, M., Segebarth, C. & Badin, P. (2000). Using the formant/cavity affiliation to study the inter-speaker variability: assessment from MRI data. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 213-216. Kloster Seeon, Germany, May 2000.

Arnal, A., Badin, P., Brock, G., Connan, P.Y., Florig, E., Perez, N., Perrier, P., Simon, P., Sock, R., Varin, L., Vaxelaire, B. & Zerling, J.P. (2000). An X-ray database for French. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 293-296. Kloster Seeon, Germany, May 2000.

Badin, P., Borel, P., Bailly, G., Revéret, L., Baciu, M. & Segebarth, C. (2000). Towards an audiovisual virtual talking head: 3D articulatory modeling of tongue, lips and face based on MRI and video images. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 261-264. Kloster Seeon, Germany, May 2000.

Engwall, O. & Badin, P. (2000). An MRI study of Swedish fricatives: coarticulatory effects. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 297-300. Kloster Seeon, Germany, May 2000.

Motoki, K., Badin, P., Pelorson, X. & Matsuzaki, H. (2000). A modal parametric method for computing acoustic characteristics of three-dimensional vocal tract models. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 325-328. Kloster Seeon, Germany, May 2000.

Motoki, K., Pelorson, X., Badin, P. & Matsuzaki, H. (2000). Computation of 3D vocal tract acoustics based on mode-matching technique. In *Proceedings of the 6th International Conference on Spoken Language Processing* (B. Yuan, T. Huang & X. Tang, Eds.), vol. I, pp. 461-464. Beijing, China, October 2000.

Revéret, L., Bailly, G. & Badin, P. (2000). MOTHER: a new generation of talking heads providing a flexible articulatory control for video-realistic speech animation. In *Proceedings of the 6th International Conference on Spoken Language Processing* (B. Yuan, T. Huang & X. Tang, Eds.), vol. II, pp. 755-758. Beijing, China, October 2000.

Vilain, A., Abry, C. & Badin, P. (2000). Coproduction strategies in French VCVs: confronting Öhman's model with adult and developmental articulatory data. In *Proceedings of the 5th Seminar on Speech Production: Models and Data & CREST Workshop on Models of Speech Production: Motor Planning and Articulatory Modelling*, pp. 81-84. Kloster Seeon, Germany, May 2000.

Vilain, A., Abry, C. & Badin, P. (1999). Motor equivalence evidenced by articulatory modelling. In *Proceedings of the 6th EuroSpeech Conference*, vol. 1, pp. 169-172. Budapest, Hungary, September 1999.

Vilain, A., Abry, C., Badin, P. & Brosda, S. (1999). From idiosyncratic pure frames to variegated babbling: Evidence from articulatory modelling. In *Proceedings of the 14th International Congress of Phonetic Sciences* (J.J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A.C. Bailey, Eds.), vol. 3, pp. 2497-2500. San Francisco, USA, August 1999. Congress organizers at the Linguistics Department, University of California, Berkeley.

Badin, P., Bailly, G. & Boë, L.-J. (1998). Towards the use of a Virtual Talking Head and of Speech Mapping tools for pronunciation training. In *Proceedings of the ESCA Tutorial and Research Workshop on Speech Technology in Language Learning*, pp. 167-170. Stockholm, Sweden, May 1998. ESCA and Dept. Speech, Music and Hearing, KTH, Stockholm.

Badin, P., Bailly, G., Raybaudi, M. & Segebarth, C. (1998). A three-dimensional linear articulatory model based on MRI data. In *Proceedings of the 5th International Conference on Spoken Language Processing* (R.H. Mannell & J. Robert-Ribes, Eds.), vol. 2, pp. 417-420. Sydney, Australia, December 1998. Australian Speech Science and Technology Association Inc.

Badin, P., Bailly, G., Raybaudi, M. & Segebarth, C. (1998). A three-dimensional linear articulatory model based on MRI data. In *Proceedings of the Third ESCA / COCOSDA International Workshop on Speech Synthesis*, pp. 249-254. Jenolan Caves, Australia, December 1998.

Bailly, G., Badin, P. & Vilain, A. (1998). Synergy between jaw and lips/tongue movements: Consequences in articulatory modelling. In *Proceedings of the 5th International Conference on Spoken Language Processing* (R.H. Mannell & J. Robert-Ribes, Eds.), vol. 5, pp. 1859-1862. Sydney, Australia, December 1998. Australian Speech Science and Technology Association Inc.

Vilain, A., Abry, C. & Badin, P. (1998). Coarticulation and degrees of freedom in the elaboration of a new articulatory plant: Gentiane. In *Proceedings of the 5th International Conference on Spoken Language Processing* (R.H. Mannell & J. Robert-Ribes, Eds.), vol. 7, pp. 3147-3150. Sydney, Australia, December 1998. Australian Speech Science and Technology Association Inc.

Badin, P., Baricchi, E. & Vilain, A. (1997). Determining tongue articulation: from discrete fleshpoints to continuous shadow. In *Proceedings of the 5th EuroSpeech Conference*, vol. 1, pp. 47-50. Rhodos, Greece, September 1997. University of Patras, Wire Communication Laboratory, Patras, Greece.

Mawass, K., Badin, P. & Bailly, G. (1997). Synthesis of fricative consonants by audiovisual-to-articulatory inversion. In *Proceedings of the 5th EuroSpeech Conference*, vol. 3, pp. 1359-1362. Rhodos, Greece, September 1997. University of Patras, Wire Communication Laboratory, Patras, Greece.

Abry, C. & Badin, P. (1996). Speech mapping as a framework for an integrated approach to the sensori-motor foundations of language. In *Proceedings of the 4th Speech Production Seminar - 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics*, pp. 175-178. Autrans, France, May 1996.

Badin, P. & Abry, C. (1996). Articulatory synthesis from X-rays and inversion for an adaptive speech robot. In *Proceedings of the 4th International Conference on Spoken Language Processing*, vol. 2, pp. 1125-1128. Philadelphia, PA, USA, October 1996. University of Delaware & Alfred I. du Pont Institute.

Badin, P., Mawass, K., Bailly, G., Vescovi, C., Beautemps, D. & Pelorson, X. (1996). Articulatory synthesis of fricative consonants : data and models. In *Proceedings of the Fourth Speech Production Seminar - First ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics*, pp. 221-224. Autrans, France, May 1996.

Beautemps, D., Badin, P., Bailly, G., GalvÀn, A. & Laboissière, R. (1996). Evaluation of an articulatory-acoustic model based on a reference subject. In *Proceedings of the 4th Speech Production Seminar - 1st ESCA Tutorial and Research Workshop on Speech Production Modeling: from Control Strategies to Acoustics*, pp. 45-48. Autrans, France, May 1996.

El Masri, S., Pelorson, X., Saguet, P. & Badin, P. (1996). Vocal tract acoustics using the transmission line matrix (TLM) method. In *Proceedings of the 4th International Conference on Spoken Language Processing*, vol. 2, pp. 953-956. Philadelphia, PA, USA, October 1996. University of Delaware & Alfred I. duPont Institute.

Badin, P., Gabioud, B., Beautemps, D., Lallouache, T.M., Bailly, G., Maeda, S., Zerling, J.P. & Brock, G. (1995). Cineradiography of VCV sequences: articulatory-acoustic data for a speech production model. In *Proceedings of the 15th International Conference on Acoustics*, vol. IV, pp. 349-352. Trondheim, Norway, June 1995.

Badin, P., Mawass, K. & Castelli, E. (1995). A model of frication noise source based on data from fricative consonants in vowel context. In *Proceedings of the 13th International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, Eds.), vol. 2, pp. 202-205. Stockholm, Sweden, 1995. Arne Strömbergs Grafiska Press.

Bailly, G., Boë, L.-J., Vallée, N. & Badin, P. (1995). Articulatory-acoustic vowel prototypes for speech production. In *Proceedings of the 4th EuroSpeech Conference* (J.M. Pardo, E. Enríquez, J. Ortega, J. Ferreiros, J. Macías & F.J. Valverde, Eds.), vol. 3, pp. 1913-1916. Madrid, Spain, September 1995. GrÀficas Brens.

Boë, L.-J., Badin, P. & Perrier, P. (1995). From sensitivity functions to macro-variations. In *Proceedings of the 13th International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, Eds.), vol. 2, pp. 234-237. Stockholm, Sweden, 1995. Arne Strömbergs Grafiska Press.

Miki, N., Badin, P., Takemura, K., Kuroda, M. & Ogawa, Y. (1995). Pitch dependency of vocal tract transfer functions. In *Proceedings of the 13th International Congress of Phonetic Sciences* (K. Elenius & P. Branderud, Eds.), vol. 4, pp. 444-447. Stockholm, Sweden, 1995. Arne Strömbergs Grafiska Press.

Pelorson, X., Badin, P., Motoki, K., Miki, N. & Plicque, M. (1995). On the radiation of sound at the lips during speech. Effects of lip geometry and of higher acoustical modes. In *Proceedings of the 15th International Conference on Acoustics*, vol. IV, pp. 497-500. Trondheim, Norway, June 1995.

Badin, P., Shadle, C.H., Pham Thi Ngoc, Y., Carter, J.N., Chiu, W., Scully, C. & Stromberg, K. (1994). Frication and aspiration noise sources: contribution of experimental data to articulatory synthesis. In *Proceedings of the 3rd International Conference on Spoken Language Processing*, vol. 1, pp. 163-166. Yokohama, Japan, September 1994.

Miki, N., Badin, P., Pham Thi Ngoc, Y. & Ogawa, Y. (1994). Vocal tract model and 3-dimensional effect of articulation. In *Proceedings of the 3rd International Conference on Spoken Language Processing*, vol. 1, pp. 167-170. Yokohama, Japan, September 1994.

Motoki, K., Badin, P. & Miki, N. (1994). Measurement of acoustic impedance density distribution in the near field of the labial horn. In *Proceedings of the 3rd International Conference on Spoken Language Processing*, vol. 2, pp. 607-610. Yokohama, Japan, September 1994.

Pelorson, X., Lallouache, T.M., Tourret, S., Bouffartigue, C. & Badin, P. (1994). Modeling the production of bilabial plosives: aerodynamical, geometrical and mechanical aspects. In *Proceedings of the 3rd International Conference on Spoken Language Processing*, vol. 2, pp. 599-602. Yokohama, Japan, September 1994.

Pham Thi Ngoc, Y. & Badin, P. (1994). Vocal tract acoustic transfer function measurements: further developments and applications. In *Journal de Physique IV, Colloque C5, Supplément au Journal de Physique III. Proceedings of the 3rd French Congress of Acoustics*, vol. 4, pp. 549-552. Toulouse, France, May 1994.

Stromberg, K., Scully, C., Badin, P. & Shadle, C.H. (1994). Aerodynamic patterns as indicators of articulation and acoustic sources for fricatives produced by different speakers. In *Proceedings of the Institute of Acoustics*, vol. 16(5), pp. 325-333, 1994.

Beautemps, D., Badin, P. & Laboissière, R. (1993). Recovery of vocal tract midsagittal and area function from speech signal for vowels and fricative consonants. In *Proceedings of the 3rd EuroSpeech Conference*, vol. 1, pp. 73-76. Berlin, Germany, September 1993.

Shadle, C.H., Badin, P. & Moulinier, A. (1991). Towards the spectral characteristics of fricative consonants. In *Proceedings of the 12th International Congress of Phonetic Sciences*, vol. 3, pp. 42-45. Aix-en-Provence, France, 1991.

Badin, P. & Fant, G. (1989). Fricative production modelling: aerodynamic and acoustic data. In *Proceedings of the 1st EuroSpeech Conference*, vol. 2, pp. 23-26. Paris, France, September 1989.

Castelli, E. & Badin, P. (1989). Nasopharyngeal tract transfer functions measurements with white noise excitation. In *Proceedings of the 13th International Conference on Acoustics*, vol. 2, pp. 511-514, 1989.

Castelli, E., Perrier, P. & Badin, P. (1989). Acoustic considerations upon the low nasal formant based on nasopharyngeal tract transfer function measurements. In *Proceedings of the 1st EuroSpeech Conference*, vol. 2, pp. 412-415. Paris, France, September 1989.

Castelli, E. & Badin, P. (1988). Vocal tract transfer functions measurements with white noise excitation. Application to the naso-pharyngeal tract. In *Proceedings of the 7th FASE Symposium*, pp. 415-422. Edinburgh, UK, 1988.

Badin, P. & Boë, L.-J. (1987). Vocalic nomograms: acoustic considerations. A crucial problem: formant convergence. In *Proceedings of the 11th International Congress of Phonetic Sciences*, vol. 2, pp. 352-355. Tallinn, Estonia, 1987.

Wu, H.Y., Badin, P., Cheng, Y.M. & Guérin, B. (1987). Continuous variations of the vocal tract length in a Kelly-Lochbaum type speech production model. In *Proceedings of the 11th International Congress of Phonetic Sciences*, vol. 2, pp. 340-343. Tallinn, Estonia, 1987.

Wu, H.Y., Badin, P., Cheng, Y.M. & Guérin, B. (1987). Vocal tract simulation: implementation of continuous variations of the length in a Kelly-Lochbaum model, Effects of area function spatial sampling. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 9-12, 1987.

Badin, P. & Murillo, G. (1983). An analysis method for high quality formant synthesis. In *Abstract of the 10th International Congress of Phonetic Sciences*, pp. 378. Utrecht, The Netherlands, August 1983.

Badin, P. & Degryse, D. (1982). Speech Communication Hardware. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 514-516. Paris, France, 1982.

4. Colloques internationaux sans comité (12)

Bailly, G., Revéret, L., Borel, P. & Badin, P. (2000). Hearing by eyes thanks to the "labiophone": exchanging speech movements. In *COST254 Workshop Friendly Exchanging Through The Net* (C. Germain, E. Grivel & O. Lavialle, Eds.), pp. 67-72. Bordeaux, France, March 2000. COST254 (Intelligent Processing and Facilities for Communication Terminals).

Badin, P., Bailly, G. & Boë, L.-J. (1999). Speech production models and Virtual Talking Heads: Useful aids for pronunciation training ? In *InStill*. Besançon, France, September 1999.

Badin, P., Bailly, G., Raybaudi, M. & Segebarth, C. (1998). A three-dimensional linear articulatory model based on MRI data. In *Proceedings of the Hokkaido Workshop on Speech Production*, pp. 1-7. Kutchan, Japan, August 1998. Hokkaido University.

Pelorson, X., Fahas, S. & Badin, P. (1997). On the meaning and the accuracy of the pressure-flow technique to determine constriction areas within the vocal tract. *Journal of the Acoustical Society of America, 102*(5, Pt. 2), 3167.

Badin, P., Beautemps, D. & Laboissière, R. (1993). Using a new model of vocal-tract midsagittal profile to area-function conversion to constrain an optimisation algorithm for the articulatory-acoustic inversion of fricative consonants. *Journal of the Acoustical Society of America, 93*(4, Pt. 2), 2416-2417.

Badin, P., Beautemps, D., Laboissière, R. & Schwartz, J.-L. (1993). Inversion of fricative consonants in vocalic context by optimization under constraints using a new model of vocal-tract midsagittal profile to area function conversion, *Invited contribution to the 3rd Seminar on Speech Production, New Haven. 11-13 May 1993.*

Scully, C., Georges, E. & Badin, P. (1991). Movement paths: different phonetic contexts and different speaking styles. *Papers from the symposium: Current Phonetic Research Paradigms: Implications for Speech Motor Control. Phonetic Experimental Research at the Institute of Linguistics University of Stockholm (PERILUS), XIV*(69-74), 13-16.

Fant, G., Lin, Q.G. & Badin, P. (1988). Speech production models: constraints and control strategies. *Journal of the Acoustical Society of America, 84*, S125.

Badin, P. & Fant, G. (1985). Vocal tract frequency domain calculation techniques. In *French-Swedish Seminar on Speech*. Grenoble, France, 1985.

Badin, P. & Murillo, G. (1984). An analysis method for high quality formant synthesis. In *Proceedings of the 10th International Congress of Phonetic Sciences* (M.P.R. VandenBrocke & A. Cohen, Eds.), pp. 221-224. Utrecht, The Netherlands, August 1984.

Badin, P. & Murillo, G. (1983). Méthode d'analyse des sons en vue d'une synthèse de très haute qualité et de la détection des indices acoustiques de la parole. In *Proceedings of the 11th International Conference on Acoustics*, pp. 85. Paris, France, 1983.

Badin, P. (1982). Strategy for high quality speech elements formant synthesis. *Journal of the Acoustical Society of America, 71*, S7.

Al-Ansari, A., Badin, P. & Guérin, B. (1981). The characteristics of a static and dynamic model of the vocal source. *Journal of the Acoustical Society of America, 69*, S65.

5. Colloques nationaux avec comité (8)

Arnal, A., Badin, P., Brock, G., Connan, P.Y., Florig, E., Perez, N., Perrier, P., Simon, P., Sock, R., Varin, L., Vaxelaire, B. & Zerling, J.P. (2000). Une base de données cinéradiographiques du français. In *Actes des 23èmes Journées d'Etude de la Parole*, pp. 425-428. Aussois, France, juin 2000.

Borel, P., Badin, P., Revéret, L. & Bailly, G. (2000). Modélisation articulatoire linéaire 3D d'un visage pour une tête parlante virtuelle. In *Actes des 23èmes Journées d'Etude de la Parole*, pp. 121-124. Aussois, France, juin 2000.

Revéret, L., Bailly, G., Borel, P. & Badin, P. (2000). Analyse par la synthèse d'un visage 3D parlant : inversion optico-articulatoire. In *Actes des 23èmes Journées d'Etude de la Parole*, pp. 125-128. Aussois, France, juin 2000.

Rossato, S., Badin, P. & Feng, G. (2000). Estimation des mouvements du voile du palais à partir du signal de parole pour les voyelles nasales du Français. In *Actes des 23èmes Journées d'Etude de la Parole*, pp. 137-140. Aussois, France, juin 2000.

Badin, P., Pouchoy, L., Bailly, G., Raybaudi, M., Segebarth, C., Lebas, J.-F., Tiede, M.K., Vatikiotis-Bateson, E. & Tohkura, Y.I. (1998). Un modèle articulatoire tridimensionnel du conduit vocal basé sur des données IRM. In *Actes des 22èmes Journées d'Etude sur la Parole*, pp. 283-286. Martigny, Suisse, juin 1998.

Bailly, G., Badin, P. & Vilain, A. (1998). Contribution de la mâchoire à la géométrie de la langue dans les modèles articulatoire statistiques. In *Actes des 22èmes Journées d'Etude sur la Parole*, pp. 287-290. Martigny, Suisse, juin 1998.

Vilain, A., Abry, C. & Badin, P. (1998). A propos des degrés de liberté dans la coarticulation d'un locuteur français filmé 50 fois par seconde. In *Actes des 22èmes Journées d'Etude sur la Parole*, pp. 311-314. Martigny, Suisse, juin 1998.

El Masri, S., Pelorson, X., Saguet, P. & Badin, P. (1996). Etude et analyse par la méthode TLM de la propagation acoustique dans le conduit vocal: effet des modes d'ordre supérieur. In *Actes des 21èmes Journées d'Etude sur la Parole*, pp. 243-246. Avignon, France, juin 1996.

Mawass, K., Badin, P., Vescovi, C. & Beautemps, D. (1996). Evaluation d'un modèle de source de friction pour la synthèse articulatoire des consonnes fricatives. In *Actes des 21èmes Journées d'Etude sur la Parole*, pp. 367-370. Avignon, France, juin 1996.

6. Colloques nationaux sans comité (4)

Djéradi, A., Badin, P. & Guérin, B. (1992). Effets de couplage subglottique: mesure et modélisation dans le domaine fréquentiel pour les fricatives d'arrière de l'arabe. In *Actes des 19èmes Journées d'Etude sur la Parole*, pp. 13-18. Bruxelles, Belgique, 1992. SFA.

Castelli, E. & Badin, P. (1988). Mesures de fonctions de transfert du conduit vocal - Application à la détermination des fonctions de transfert du conduit nasopharyngal. In *Actes des 17èmes Journées d'Etude sur la Parole*, pp. 189-193, 1988. SFA.

Perrier, P., Badin, P. & Boë, L.-J. (1987). Nomogrammes du conduit vocal par modélisation articulatoire. In *Actes des 16èmes Journées d'Etude sur la Parole*, pp. 124-127. Hammamet, Tunisie, 1987. SFA.

Murillo, G. & Badin, P. (1983). Méthodes et logiciels pour synthèse de parole de haute qualité. In *Séminaire GALF/GRECO: Analyse du signal de parole*. Paris, France, 1983.

7. Rapports d'activité (8)

Engwall, O. & Badin, P. (1999). Collecting and analysing two- and three-dimensional MRI data for Swedish. *Tal Musik Hörsel - Quarterly Progress Status Report - Stockholm, 3-4/1999*, 11-38.

Badin, P., Hertegård, S. & Karlsson, I. (1990). Notes on the Rothenberg mask. Speech Transmission Laboratory - Quarterly Progress Status Report - Stockholm, 1/1990, 1-7.

Castelli, E., Perrier, P. & Badin, P. (1990). Caractérisation acoustique de la nasalité. A propos du premier formant nasal : Quelques hypothèses (résonn...ables ?). *Bulletin du Laboratoire de la Communication Parlée, 3.*, 187-212.

Badin, P. (1989). Acoustics of voiceless fricatives: production theory and data. *Speech Transmission Laboratory - Quarterly Progress Status Report - Stockholm, 3/1989*, 33-55.

Badin, P., Boë, L.-J., Perrier, P. & Abry, C. (1988). Vocalic nomograms: acoustic considerations upon formant convergence. *Bulletin du Laboratoire de la Communication Parlée, 2*, 65-94.

Boë, L.-J., Abry, C., Perrier, P., Guérin, B. & Badin, P. (1988). From the linguistic system to the signal through articulatory and acoustic modeling. *Bulletin du Laboratoire de la Communication Parlée, 2*, 1-10.

Wu, H.Y., Badin, P., Cheng, Y.M. & Guérin, B. (1987). Simulation du conduit vocal: réalisation de la variation continue de longueur dans un modèle de Kelly-Lochbaum - Effets de l'échantillonnage spatial de la fonction d'aire. *Bulletin du Laboratoire de la Communication Parlée, 1*, 1-27.

Badin, P. & Fant, G. (1984). Notes on vocal tract computation. Speech Transmission Laboratory - Quarterly Progress Status Report - Stockholm, 2-3/1984, 53-108.

Badin, P. (1983). Les techniques d'analyse et de synthèse de la parole. *Bulletin de l'Institut Phonétique de Grenoble, 12*, 95-140.

8. Divers (7)

Badin, P., Brockhaus, W., Boë, L.-J. & Abry, C. (2000) Editorial. Current Trends in Phonology and Phonetics II : Relationship between phonetics and phonology. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée,* **5**, 3.

Boë, L.-J., Vallée, N., Badin, P., Schwartz, J.-L. & Abry, C. (2000) Tendencies in phonological structures : the influence of substance on form. Current Trends in Phonology and Phonetics II : Relationship between phonetics and phonology. *Les Cahiers de l'ICP, Bulletin de la Communication Parlée,* **5**, 35-55.

Badin, P., Bailly, G. & Boë, L.-J. (1999) Speech production models and Virtual Talking Heads: Useful aids for pronunciation training ? In *EUROCALL'99, Satellite workshop "Speech Technology applications in CALL"*, Besançon, France, EUROpean association for Computer Assisted Language Learning. Invited contribution.

Abry, C. & Badin, P. (1998) Foreword to comments on "The Equilibrium Point Hypothesis and its application to speech motor control". *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, **4**, 3-4.

Abry, C., Badin, P., Mawass, K. & Pelorson, X. (1998) The Equilibrium Point Hypothesis and control spaced for relaxation movements or "When is movement actually needed to control movement ?". *Les Cahiers de l'ICP, Bulletin de la Communication Parlée*, **4**, 27-33.

Badin, P. (1983). Analyse de la parole - synthèse à formants. Application à la synthèse des contoïdes constrictives voisées du Français. Discussion d'une méthode en vue de la détermination des indices acoustiques de la parole. Unpublished Thèse de Docteur Ingénieur, Institut National Polytechnique de Grenoble.

Badin, P. (1980). Simulation numérique en temps réel d'un synthétiseur à formants. Unpublished Rapport de D.E.A., ENSERG, Institut National Polytechnique de Grenoble, Grenoble.

9. Rapports de contrats (15)

Badin, P. & Boë, L.-J. (2000). Une Tête Parlante Virtuelle - Données et modèles en production de parole. In *Rapport final de projet ARASSH (Agence Rhône-Alpes pour les Sciences Sociales et Humaines)*.

Badin, P. & Boë, L.-J. (1999). Une Tête Parlante Virtuelle - Données et modèles en production de parole. In *Rapport intermédiaire de projet ARASSH (Agence Rhône-Alpes pour les Sciences Sociales et Humaines)*.

Abry, C., Badin, P., Vilain, A., Stefanuto, M. & Boë, L.-J. (1997). Annexe 5. Une théorie de l'ontogénèse à l'épreuve de trois modèles articulatoires anthropomorphiques. In *Rapport d'avancement des recherches du projet GIS Sciences de la Cognition, Cognition naturelle / cognition artificielle "Les robots parlent aux robots. Un générateur des Structures Sonores du Langage"* (C. Abry, Ed.), pp. 29-37. Grenoble: INPG/Stendhal/CNRS.

Abry, C. & Badin, P. (1995). Periodic Progress Report N°3. In Reports of European project ESPRIT/BR N° 6975 Speech Maps (Mapping of Action and Perception in Speech) (C. Abry & P. Badin, Eds.).

Abry, C. & Badin, P. (1994). Periodic Progress Report N°2. In Reports of European project ESPRIT/BR N° 6975 Speech Maps (Mapping of Action and Perception in Speech) (C. Abry & P. Badin, Eds.).

Badin, P. (1994). Voice and noise source modelling. In From speech signal to source, Periodic Progress Report N°2, European ESPRIT/BR N° 6975 Speech Maps project (C. Scully, Ed.).

Badin, P. (1994). 2D and 3D vocal tract geometry data. In From speech signal to vocal tract geometry, Periodic Progress Report N°2, European ESPRIT/BR N° 6975 Speech Maps project (S. Maeda, Ed.).

Badin, P. & Dolmazon, J.-M. (1994). ASTER : "Analyse Spectrale du Signal de Parole en Temps Réel". In Compte rendu scientifique de fin de projet. Programme d'Aide à l'Investissement de la Région Rhône-Alpes (N° M06 6 0 9).

Abry, C. & Badin, P. (1993). Periodic Progress Report N°1. In Reports of European project ESPRIT/BR N° 6975 Speech Maps (Mapping of Action and Perception in Speech) (C. Abry & P. Badin, Eds.).

Badin, P., Castelli, E. & Pham Thi Ngoc, Y. (1993). Acoustic transfer functions for vowels and consonants. In *From speech signal to vocal tract geometry, Periodic Progress Report N°1, European ESPRIT/BR N° 6975 Speech Maps project* (S. Maeda, Ed.).

Castelli, E. & Badin, P. (1993). Time and frequency domain acoustic models of the vocal tract. In From speech signal to vocal tract geometry, Periodic Progress Report N°1, European ESPRIT/BR N° 6975 Speech Maps project (S. Maeda, Ed.).

Guérin, B., Badin, P., Grabbe-Georges, E., Moulinier, A., Shadle, C.H., Scully, C. & Tzavali, E. (1991). Mesure, caractérisation et modélisation des sons fricatifs, *Quatrième rapport semestriel du contrat SCIENCE SC1*147-C.*

Guérin, B., Badin, P., Castelli, E., Grabbe-Georges, E., Moulinier, A., Shadle, C.H., Scully, C. & Tzavali, E. (1990). Mesure, caractérisation et modélisation des sons fricatifs, *Troisième rapport semestriel du contrat SCIENCE SC1*147-C.*

Guérin, B., Barraclough, L., Badin, P., Castelli, E., Moulinier, A., Shadle, C.H. & Scully, C. (1990). Mesure, caractérisation et modélisation des sons fricatifs, *Deuxième rapport semestriel du contrat SCIENCE SC1*147-C.*

Guérin, B., Barraclough, L., Badin, P., Castelli, E., Shadle, C.H. & Scully, C. (1989). Mesure, caractérisation et modélisation des sons fricatifs, *Premier rapport semestriel du contrat SCIENCE SC1*147-C.*

10. Vulgarisation (2)

Badin, P., Bailly, G., Beautemps, D., Guiard-Marigny, T., Laboissière, R. & Lallouache, T.M. (1995). Speech Maps Audio-Visual Articulatory Synthesis, *Video VHS, 26mn.*

Abry, C. & Badin, P. (1994). Speech Maps: Exploring new territory. *Elsnews, The Newsletter of the European Network in Language and Speech, 3*(3), 4-5.

Troisième partie : Articles annexés

J'ai sélectionné ici une courte liste des publications que je considère les plus représentatives de mes activités de recherche.

- **Badin, P., Perrier, P., Boë, L.-J. & Abry, C. (1990)**. Vocalic nomograms: Acoustic and articulatory considerations upon formant convergences. *Journal of the Acoustical Society of America*, **87**, 1290-1300.
- **Badin, P. (1991)**. Fricative consonants: acoustic and X-ray measurements. *Journal of Phonetics*, **19**, 397-408.
- **Badin, P., Motoki, K., Miki, N., Ritterhaus, D. & Lallouache, T.M. (1994)**. Some geometric and acoustic properties of the lip horn. *Journal of the Acoustical Society of Japan (English)*, **15**(4), 243-253.
- **Badin, P., Bailly, G. & Boë, L.-J. (1998)**. Towards the use of a Virtual Talking Head and of Speech Mapping tools for pronunciation training. In *Proceedings of the ESCA Tutorial and Research Workshop on Speech Technology in Language Learning*, pp. 167-170. Stockholm, Sweden, ESCA and Dept. Speech, Music and Hearing, KTH, Stockholm.
- **Mawass, K., Badin, P. & Bailly, G. (2000)**. Synthesis of French fricatives by audio-video to articulatory inversion. *Acta Acustica*, **86**(1), 136-146.
- **Beautemps, D., Badin, P. & Bailly, G. (2001)**. Linear degrees of freedom in speech production: Analysis of cineradio- and labio-film data and articulatory-acoustic modeling. *Journal of the Acoustical Society of America*, **109**(5), 2165-2180.
- **Badin, P., Bailly, G., Revéret, L., Baciu, M., Segebarth, C. & Savariaux, C. (In press)**. Three-dimensional articulatory modeling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*.