REV LARYNGOL OTOL RHINOL.
2007;128,5:289-295.

PHONIATRIE

ARTICLE ORIGINAL

Têtes parlantes audiovisuelles virtuelles : données et modèles articulatoires - applications [1]

Virtual audiovisual talking heads: Articulatory data and models - applications

Badin P.
Elisei F.
Bailly G.
Savariaux C.
Serrurier A.
Tarabalka Y.
(Grenoble) [2]

Résumé

Dans le cadre de la phonétique expérimentale, notre approche de l'étude de la production de parole se base sur la mesure, l'analyse, et la modélisation des organes oro-faciaux tels que la mâchoire, le visage et les lèvres, la langue ou le voile du palais. Ainsi, nous présentons dans cet article des techniques expérimentales qui permettent de caractériser la forme et le mouvement des articulateurs de la parole (IRM statique et dynamique, tomodensitographie, articulographie électromagnétique, enregistrements vidéo). Nous décrivons ensuite les modèles linéaires des différents organes dérivés des données articulatoires obtenues sur un locuteur. Nous montrons que ces modèles, qui présentent une bonne résolution géométrique, peuvent être contrôlés à partir de données articulatoires de bonne résolution temporelle et peuvent ainsi permettre de reconstruire des animations d'articulateurs de haute qualité. Ces modèles, que nous avons intégrés dans une tête parlante virtuelle, peuvent servir à produire de la parole audiovisuelle augmentée. Dans ce cadre, nous avons évalué les capacités naturelles de lecture linguale de sujets humains à l'aide de tests perceptifs audiovisuels. Nous concluons par la proposition d'un certain nombre d'autres applications des têtes parlantes.

Mots-clés : Production de parole, tête parlante audiovisuelle, parole augmentée, modélisation articulatoire, mesure articulatoire.

Summary

In the framework of experimental phonetics, our approach to the study of speech production is based on the measurement, the analysis and the modeling of orofacial articulators such as the jaw, the face and the lips, the tongue or the velum. Therefore, we present in this article experimental techniques that allow characterising the shape and movement of speech articulators (static and dynamic MRI, computed tomodensitometry, electromagnetic articulography, video recording). We then describe the linear models of the various organs that we can elaborate from speaker-specific articulatory data. We show that these models, that exhibit a good geometrical resolution, can be controlled from articulatory data with a good temporal resolution and can thus permit the reconstruction of high quality animation of the articulators. These models, that we have integrated in a virtual talking head, can produce augmented audiovisual speech. In this framework, we have assessed the natural tongue reading capabilities of human subjects by means of audiovisual perception tests. We conclude by suggesting a number of other applications of talking heads.

Key-words: Speech production, audiovisual talking head, augmented speech, articulatory modeling, articulatory measurement.

INTRODUCTION

La production de la parole fait appel, entre autres, aux organes oro-faciaux tels que les lèvres, la mâchoire, la langue ou le voile du palais. Il est donc naturel que l'une des approches majeures de la recherche sur les phénomènes de production de parole se base, dans le cadre de la phonétique expérimentale, sur l'observation, la mesure, l'analyse et la modélisation de ces organes.

Cette approche a retrouvé depuis une décennie un souffle nouveau grâce au développement et à la relative disponibilité de méthodes d'imagerie médicale de plus en plus précises et rapides, en particulier l'Imagerie par Résonance Magnétique (IRM). Il est ainsi possible aujourd'hui de développer de véritables têtes parlantes audiovisuelles à partir de mesures faites sur des sujets humains.

Notre ambition dans cet article est de présenter un certain nombre des techniques expérimentales utilisées au Département Parole & Cognition de GIPSA-lab à Grenoble dans le domaine des études sur la production de la parole pour mesurer et caractériser la forme et le mouvement des articulateurs. Nous présenterons aussi des méthodes qui peuvent être utilisées avec ces données pour modéliser les différents articulateurs qui servent de

1. Communication présentée au LXIII Congrès de la Société Française de Phoniatrie et des Pathologies de la Communication, Paris, 16/10/07.
2. GIPSA-lab, ENSIEG, Domaine Universitaire, BP46, 38402 Saint Martin d'Hères, France.
Email: Pierre.Badin@gipsa-lab.inpg.fr

Article reçu : 20/12/07 accepté : 31/01/08

base aux têtes parlantes audiovisuelles, ainsi que des méthodes permettant le contrôle de ces têtes parlantes. L'article s'achèvera avec la description d'un récent travail d'exploration des possibilités de lecture linguale et la proposition de pistes pour des applications.

4.1. Méthodes de mesure articulatoire

La première étape dans notre approche de modélisation consiste à acquérir des données articulatoires les plus complètes possibles pour chaque locuteur étudié. Notre objectif étant l'obtention de représentations qui soient précises à la fois temporellement et spatialement, nous mettons en œuvre un ensemble de dispositifs de mesure dont certains fournissent de très bonnes données tridimensionnelles pour des articulations statiques, tandis que d'autres fournissent des données dynamiques avec une très bonne résolution temporelle sur quelques points particuliers des articulateurs. La section ci-dessous présente ces diverses méthodes.

4.1.1. Cinéradiographie et IRM dynamique

La cinéradiographie a été pendant des décennies la méthode de référence pour la visualisation et la mesure de l'articulation en parole [1-3]. Grâce à une illumination latérale par rayons X du complexe orofacial, elle fournit une image 2D des articulateurs 3D. Ses principaux avantages sont une résolution temporelle de 50 images par seconde ou plus associée à la production d'une image complète de l'ensemble des articulateurs (fig. 1). Badin et al (1995) [4] ont en outre couplé à ce dispositif un système d'enregistrement vidéo synchrone qui permet de capturer ainsi les informations articulatoires au niveau labial (fig. 1). Outre la relative difficulté d'interprétation de ces images, liée au fait qu'elles résultent de la projection de l'ensemble des formes des articulateurs dans un plan sagittal, le principal inconvénient de cette méthode est sa nocivité qui découle du rayonnement ionisant employé. Cette méthode, qui n'est donc plus utilisée aujourd'hui dans le cadre de la recherche sur la parole avec des sujets sains, a laissé place à l'IRM dynamique [5]. Les images obtenues, avec des taux d'échantillonnage qui peuvent aller jusqu'à plusieurs dizaines de Hertz, sont plus faciles à interpréter puisqu'elles constituent de véritables coupes virtuelles (fig. 2), mais au prix, pour une bonne qualité d'image, de la nécessité pour le locuteur de répéter un grand nombre de fois la séquence de parole à analyser (jusqu'à plus de cent fois). Les progrès constants des imageurs IRM nous permettent cependant d'espérer réduire ou éliminer cette contrainte de répétions dans un futur proche.

Fig. 1 : Exemple d'images cinéradiographique et vidéo de face synchrones.

Fig. 2 : Exemple d'image obtenue par IRM dynamique (enregistrement effectué au Takenohara hospital avec l'aide de l'Advanced Telecomunication Institute de Kyoto au Japon).

4.1.2. Articulographie électromagnétique

La résolution temporelle de la cinéradiographie ou de l'IRM dynamique n'est pas suffisante pour suivre dans le détail la dynamique des articulateurs. L'articulographe électromagnétique constitue donc un dispositif complémentaire très intéressant qui permet de suivre les coordonnées dans le plan médiosagittal d'un certain nombre (une dizaine pour le modèle dont nous disposons) de petites bobines électromagnétiques réceptrices fixées sur les organes du sujet (pour une description détaillée du principe, Perkell et al [6]). Les bobines peuvent être fixées sur la mâchoire, la langue, le voile du palais, ou encore les lèvres (fig. 3). Les avantages de cette méthode sont d'une part sa bonne résolution temporelle (typiquement plusieurs centaines de Hertz), et d'autre part sa capacité à suivre des points de chair et non pas des contours. L'inconvénient majeur réside dans sa mauvaise résolution spatiale, qui ne permet qu'une vue très partielle des articulateurs ; nous verrons dans la suite que cet inconvénient peut être dépassé si l'on fait appel à des modèles des organes que l'on désire suivre. Le caractère partiellement invasif, dû à la présence des bobines et des fils à l'intérieur de la bouche du sujet constitue par contre un inconvénient qui n'est pas tout à fait négligeable.

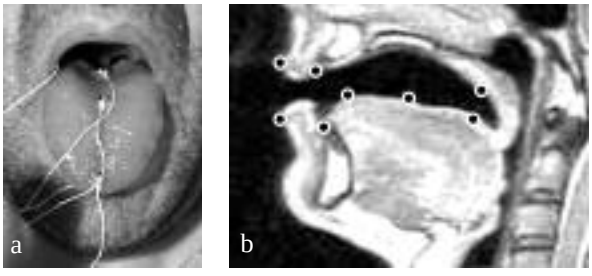


Fig. 3: (a) Photo de bobines réceptrices d'un articulographe électromagnétique fixées sur la langue d'un sujet ; (b) illustration du positionnement possible de huit bobines dans le plan médiosagittal.

4.1.3. Enregistrements vidéo

Pour les organes externes, facilement visibles, l'utilisation de caméras vidéo permet un enregistrement non invasif. De long corpus, avec un bon compromis entre résolution spatiale (de l'ordre de 2 pixels/mm) et résolution temporelle (typiquement 50 Hz), peuvent être enregistrés et stockés facilement au laboratoire.

4.1.4. IRM et tomodensitométrie volumiques

A présent, la technique la mieux adaptée à la mesure tridimensionnelle des articulateurs internes est l'IRM (Imagerie par Résonance Magnétique) volumique qui

peut fournir des séries d'images sagittales parallèles couvrant l'ensemble de la région des articulateurs. Ce type d'acquisition est relativement long, puisque le locuteur doit maintenir de manière artificielle chaque articulation pendant approximativement vingt-cinq secondes. Le protocole que nous utilisons actuellement produit 25 images sagittales de 4 mm d'épaisseur sans recouvrement avec une résolution d'image d'environ de 1 pixel/mm [7-10].

Dans la mesure où l'IRM ne permet pas la visualisation des structures osseuses, un scanner tomodensitométrique complet de la tête du sujet a été obtenu afin d'obtenir de bonnes images de référence de ces différentes structures (mâchoire et os hyoïde pour les structures typiquement mobiles ; palais dur, sinus maxillaires et sphénoïdal, etc, pour les structures fixes de référence qui servent également à recaler dans un même repère commun les images acquises pour les différentes articulations).

4.1.5. Détermination des surfaces déformables

Nous présentons ici les méthodes qui permettent d'obtenir des représentations surfaciques des contours des divers articulateurs impliqués dans la production de parole audiovisuelle.

4.1.5.1. Visage et lèvres. La mesure des articulateurs visibles, essentiellement la surface du visage et celle des lèvres, peut se faire à partir d'enregistrements vidéo. Environ 250 marqueurs colorés sont fixés sur le visage du sujet [7, 9, 11]. L'utilisation de caméras multiples, synchrones et calibrées, permet de déterminer les coordonnées de ces marqueurs avec une précision meilleure que le millimètre. La surface du visage de la tête parlante est représentée par un maillage d'environ 450 triangles dont les sommets s'appuient sur ces points (fig. 4). Par ailleurs, les lèvres sont définies à l'aide d'un maillage générique. Celui-ci, indépendant du locuteur, est déformé puis ajusté aux images par un expert, en utilisant les vues multiples pour capturer le contour externe des lèvres et modéliser leur surface visible.

Fig. 4 : Exemple de vues de la tête d'un sujet produites par trois caméras synchrones et calibrées, avec superposition des maillages de visage et de lèvres.

4.1.5.2. Organes internes. La détermination de la surface des articulateurs internes se déroule en trois étapes [10, 12]. La première consiste à tracer manuellement, à l'aide d'un éditeur de courbes, les contours des organes sur chacune des images où ils apparaissent. Notons que nous utilisons en outre, pour aider à l'interprétation qui est parfois ardue, des images transverses reconstruites à partir des piles d'images sagittales, les plans utiles étant ceux correspondants à une grille semi-polaire médiosagittale conçue pour fournir des images toujours approximativement orthogonales au conduit vocal. Pour compléter l'aide et assurer la cohérence des tracés entre différentes images, les contours des structures osseuses et les contours déformables déjà tracés sont superposés sur les images IRM (fig. 5).

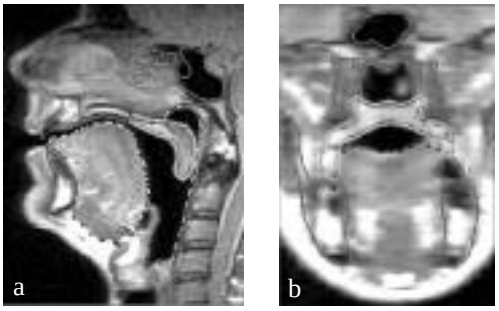


Fig. 5 : Exemple d'images IRM sagittale (a) ou transverse reconstruite (b). Les tracés de diverses structures rigides et déformables apparaissent en superposition.

La seconde étape consiste à recaler l'ensemble des contours plans obtenus pour chacune des images dans le même espace tridimensionnel. Finalement, un maillage générique de l'organe est transformé par un algorithme de déformation élastique pour s'ajuster au mieux à l'ensemble des contours et définir ainsi la forme de l'organe pour l'articulation étudiée.

On obtient finalement un maillage pour chaque structure rigide, et des collections de maillages de chaque organe déformable, qu'il soit interne ou externe, pour chaque articulation spécifique produite par le locuteur. Ces données sont cruciales pour le développement de modèles articulatoires.

MODÈLES ARTICULATOIRES LINÉAIRES BASÉS SUR LES DONNÉES

L'appareil de production de parole est constitué d'un grand nombre de composantes neuromusculaires et possède donc une dimensionnalité élevée. Ces composantes sont cependant fonctionnellement couplées afin de produire des gestes relativement simples [13] ; elles peuvent être assimilées à des articulateurs élémentaires que l'on peut encore appeler degrés de liberté. La parole peut donc être considérée comme le produit du recrutement soigneusement coordonné de tels articulateurs, chacun contrôlé par un paramètre. Notre approche de la modélisation articulatoire consiste en fin de compte à extraire ces composantes, par analyse en composantes linéaires non corrélées, de corpus de données représentatifs de l'ensemble des articulations pour un locuteur et une langue donnés. Les modèles articulatoires tridimensionnels que nous avons développés sont ainsi basés sur un corpus de 46 articulations artificiellement maintenues par un locuteur français : les quatorze voyelles

orales et nasales, et les consonnes dans les contextes vocaliques symétriques [a i ou], une articulation préphonatoire et une position de repos. Notre méthode d'identification des degrés de liberté, déjà décrite en détail [7], peut être résumée à une analyse en composantes linéaires guidée : les composantes sont imposées à partir de mesures particulières (comme par exemple le paramètre de hauteur de mâchoire qui pilote la contribution de la mâchoire au mouvement de la langue) ou au contraire extraites par analyse en composantes principales d'un ensemble de variables (par exemple les coordonnées des points de la langue qui se trouvent dans le plan médiosagittal). Ces choix visent à trouver un compromis optimal entre d'une part une explication maximale de la variance des données observées et d'autre part l'extraction de composantes qui soient à la fois en nombre minimal et interprétables en termes biomécaniques ou phonétiques. Ces choix visent également à minimiser les biais du corpus, en particulier la sous représentation de gestes importants et/ou faibles énergétiquement. Le fait d'analyser un sujet unique permet d'éviter de brouiller la lisibilité des stratégies d'articulation qui sont souvent différentes d'un sujet à l'autre.

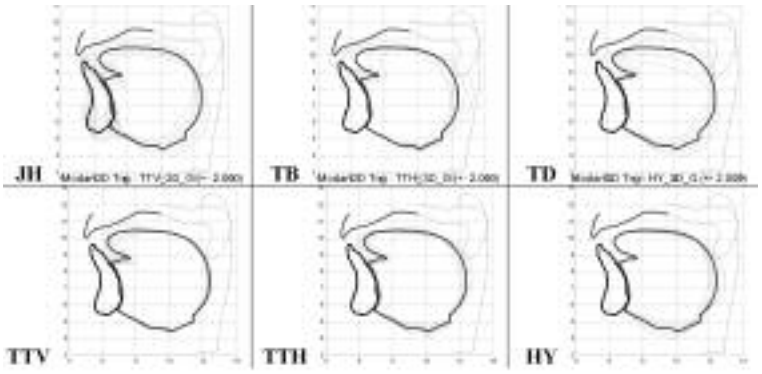


Fig. 6 : Coupes dans le plan médiosagittal des nomogrammes tridimensionnels du modèle de langue (variation entre -2 et +2 par pas de 1 des divers paramètres articulatoires).

Nous avons donc développé un modèle articulatoire mâchoire-lèvres-visage, un modèle articulatoire mâchoire - langue, et un modèle de voile du palais.

Le modèle mâchoire - langue [12] possède six composantes. La première, liée à l'ouverture de la mâchoire JH, correspond à une rotation globale de la masse de la langue autour d'un point situé à l'arrière (fig. 6). Les composantes corps de langue TB, et dos de langue TD correspondent respectivement à des mouvements avant – arrière et aplatissement – bombement, tandis que les composantes hauteur d'apex TTV et avancée d'apex TTH correspondent respectivement principalement à des mouvements verticaux et horizontaux de la pointe de la langue (fig. 6). Le dernier paramètre, HY, lié à la hauteur de l'os hyoïde, contrôle une remontée de la racine de la langue, liée à un mouvement de recul. L'ensemble de ces six composantes permet de représenter près de 90 % de la variance totale des données tridimensionnelles de la langue.

Le modèle mâchoire-lèvres-visage possède six composantes pour la parole neutre (fig. 7) ([14] ou [9] pour une extension à de la parole plus expressive). La première, JH, est liée au mouvement d'ouverture/fermeture de la mâchoire ; la deuxième, JA, correspond à un mouvement d'avancée/recul de la mâchoire utile dans les articulations labio-dentales /f/ par exemple. Trois autres composantes sont essentielles pour les lèvres : le mouvement de protrusion/étirement commun aux deux lèvres (LP) qui caractérise la différence /i/ vs. /u/ ; le mouvement d'élévation de la lèvre supérieure (LU), utile pour la réalisation de la consonne labio-dentale /f/ par exemple ; le mouvement d'abaissement de la lèvre inférieure (LI) que l'on retrouve dans le /cheu/ pour lequel les deux lèvres sont ouvertes au maximum, alors que la mâchoire est en position haute. Notons enfin une composante correspondant à l'abaissement du larynx (LX). Au total, plus de 90 % de la variance des données est prise en compte par ces composantes.

Le modèle de voile du palais [8, 10] comporte deux composantes. La première composante VL correspond à un mouvement oblique haut – bas qui est le principal mouvement d'ouverture/fermeture du port vélopharyngé (fig. 8). Une seconde composante VS correspond à un mouvement horizontal couplé avec un allongement vertical qui complète la première composante pour contrôler

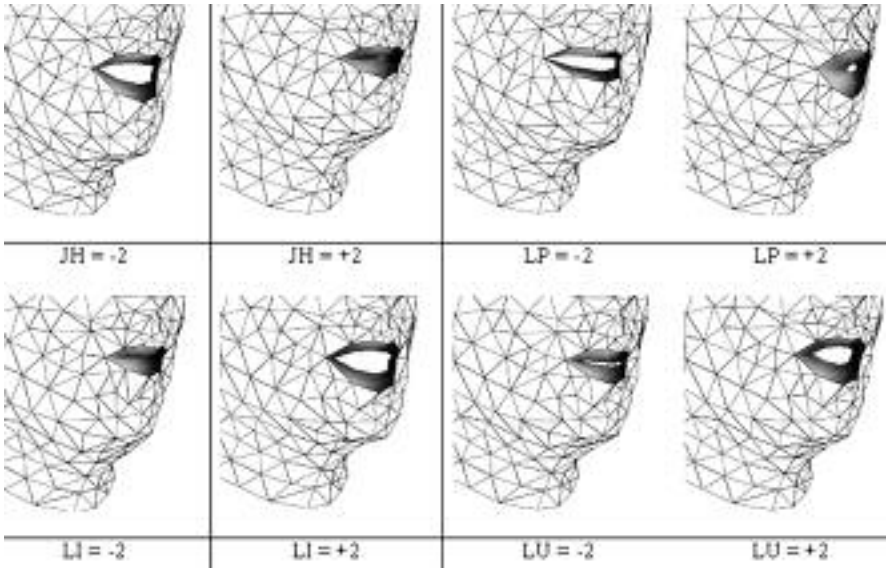


Fig. 7 : Présentation du modèle de mâchoire–lèvres–visage pour des valeurs -2 et +2 des quatre premiers paramètres articulatoires.

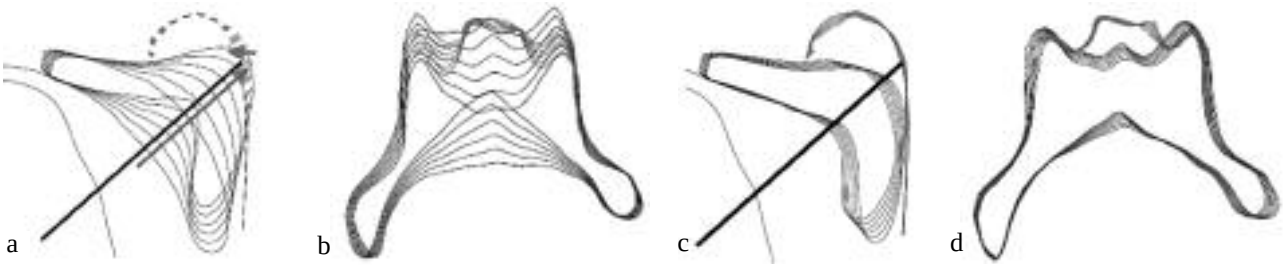


Fig. 8 : Coupes dans le plan médiosagittal (a-c) et dans un plan transverse (b-d indiqué dans le plan médiosagittal) des nomogrammes tridimensionnels du modèle de voile du palais et de paroi nasopharyngée (variation entre -3 et +3 par pas de 1 des paramètres articulatoires : VL (a-b), et VS (c-d).

l'aire du port vélopharyngé. Ces deux composantes, qui expliquent près de 90 % de la variance des données, sont associées à de petits mouvements de la paroi nasopharyngée dans des directions contraires, qui semblent assurer une fonction de sphincter comme l'illustre la figure 8.

En conclusion, tous ces modèles articulatoires constituent les éléments d'une tête parlante audiovisuelle virtuelle qui peut être contrôlée à partir d'un nombre réduit de paramètres articulatoires. Nous décrivons dans la section suivante différentes méthodes qui permettent de déterminer ces paramètres.

CONTRÔLE DE LA TÊTE PARLANTE

Trois types de méthodes peuvent être envisagés pour générer les paramètres de contrôle de la tête parlante : (1) la capture de mouvement, que nous allons détailler dans la suite, repose sur l'obtention des trajectoires temporelles de mesures articulatoires ; (2) l'inversion à partir du signal audiovisuel enregistré par le sujet, processus qui tente d'optimiser les paramètres de contrôle afin d'obtenir un son de synthèse et des formes de lèvres les plus proches possible du signal original [15] ; (3) la synthèse à partir du texte, qui demande une mise en œuvre complexe [16].

Nous avons vu plus haut que l'articulographe électromagnétique permet de suivre les coordonnées médiosagittales de petites bobines fixées en différents points de chair sur les articulateurs de la parole : si le nombre de ces points est suffisant, il est alors possible de déterminer par optimisation les paramètres de commande des modèles articulatoires de telle sorte à faire coïncider les points correspondants des modèles avec les points mesurés. Nous décrivons brièvement ici la mise en œuvre de cette méthode décrite plus en détail par ailleurs [17]. Une bobine unique permet de déterminer l'ouverture de mâchoire (paramètre commun au modèle de lèvres et à celui de langue) ; les deux bobines de lèvres fournissent quatre mesures (les coordonnées horizontales et verticales) ; nous disposons donc de cinq mesures pour déterminer cinq paramètres de contrôle du modèle de lèvres, ce qui permet de résoudre le problème. De manière similaire, les trois bobines de langue fournissent six mesures, qui, associées à l'ouverture de mâchoire, représentent sept variables qui nous permettent de déterminer les six paramètres de contrôle du modèle de langue. Enfin, nous avons montré [10] qu'il est possible de contrôler les deux paramètres qui contrôlent notre modèle de voile du palais à partir des deux coordonnées d'une bobine unique fixée sur cet articulateur.

Dans le cadre d'une étude que nous allons présenter à titre d'illustration d'applications de notre tête parlante [17], nous avons donc enregistré, pour le locuteur qui a servi au développement des modèles articulatoires présentés ci-dessus, un important corpus de séquences Voyelle – Consonne – Voyelle (VCV) comprenant toutes les combinaisons possibles symétriques pour le français, et nous avons déterminé les trajectoires temporelles des paramètres de contrôle correspondantes.

EXEMPLE D'APPLICATION : ÉVALUATION DE L'APPORT DE LA VISION DE LA LANGUE À L'INTELLIGIBILITÉ DE LA PAROLE

L'apport de la vision du visage et des lèvres à l'intelligibilité de la parole est bien documenté depuis longtemps [18, 19]. Il est établi que lorsque le rapport signal sur bruit du signal de parole diminue, l'utilisation de la composante visuelle du signal croît, afin de compenser le déficit auditif. Il est également établi que le taux de reconnaissance du signal audiovisuel est supérieur à la fois au taux de reconnaissance du signal audio seul et à celui du signal visuel seul, que le visage entier apporte plus d'information que les lèvres seules, et qu'un visage naturel apporte plus qu'un visage de synthèse. Ces propriétés de lecture labiale / faciale sont très vraisemblablement le résultat d'un apprentissage des relations entre vue des lèvres et sons de parole qui commence dès la naissance [20]. De manière plus générale, on pourrait faire l'hypothèse que les capacités de conscience articulatoire, c'est-à-dire de savoir/sentir comment sont positionnés nos articulateurs [21], pourraient permettre aux sujets d'utiliser la vision de la langue, comme ils utilisent la vision des lèvres dans le cas de la lecture labiale. Nous avons donc tenté d'évaluer la contribution potentielle de la vision directe et complète de la langue à l'intelligibilité de la parole [17].

Nous avons mis en œuvre la tête parlante décrite ci-dessus, qui permet d'afficher tous les articulateurs de la parole, y compris la langue, en mode «parole augmen-

tée», en la pilotant pour générer un ensemble de stimuli audiovisuels VCV à partir d'enregistrements articulographiques effectués sur le même sujet. Nous avons présenté ces stimuli à un ensemble de sujets dans une série de tests de perception audiovisuelle suivant diverses conditions de présentation (audio seul, audiovisuel avec écorché de profil avec ou sans langue, visage complet (fig. 9), et avec différent niveaux de rapport signal sur bruit. L'analyse des résultats montre un certain effet d'apprentissage implicite de la lecture linguale, une préférence pour le rendu plus écologique de la présentation du visage complet par rapport à la présentation en écorché, une prédominance de la lecture labiale sur la lecture linguale, sauf dans les cas où – le signal audio étant tellement dégradé (ou absent) – la lecture linguale prend le relais. Ces résultats préliminaires sont à compléter par des tests plus systématiques impliquant notamment des mesures d'attention visuelle, pour confirmer que nos capacités naturelles de lecture linguale sont faibles, ou qu'elles sont simplement dominées par celles en lecture labiale. Nous envisageons d'élaborer des protocoles pour montrer que l'apprentissage de la lecture linguale est rapide et facile.

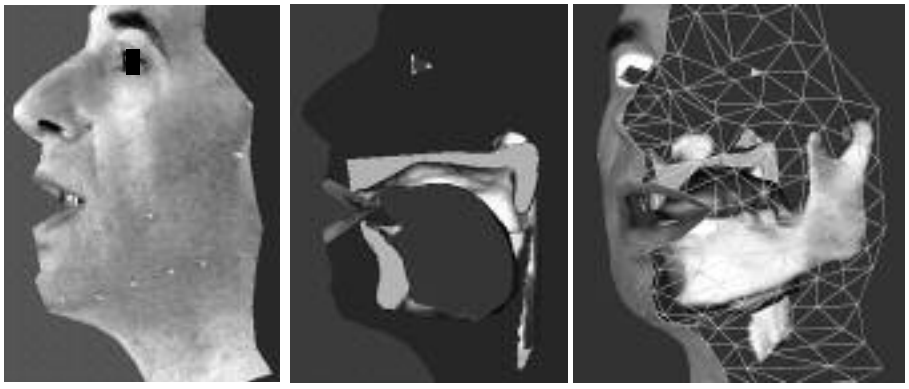


Fig. 9 : Exemples de présentation de la tête audiovisuelle: tête complète vue de profil avec texture de peau, écorché de profil, vue de trois quart à moitié texturée.

CONCLUSIONS – PERSPECTIVES

Nous avons présenté un certain nombre de méthodes de mesure et d'analyse de l'articulation en parole, aussi bien au niveau statique tridimensionnel qu'au niveau de la dynamique. Nous avons décrit le corpus de données originales qui a pu être accumulées sur un sujet humain, et qui ont permis le développement d'une tête parlante audiovisuelle virtuelle, véritable clone du sujet réel. Ces données et ces modèles constituent une base précieuse pour l'étude des mécanismes de production de la parole au niveau périphérique et pour l'étude du contrôle de leur coordination.

Nous avons illustré les possibilités applicatives offertes par cet ensemble d'outils par une évaluation des facultés humaines naturelles en lecture linguale, en nous basant sur les capacités d'affichage en parole augmentée de notre tête parlante virtuelle. Dans cette première application, les stimuli utilisés étaient fixes, issus d'enregistrements sur un locuteur ; à terme, on peut imaginer un système capable de déterminer les formes et les mouvements des articulateurs internes à partir du simple signal audiovisuel externe, et de fournir ainsi au locuteur un retour perceptif augmenté pour sa langue. Les capacités de parole augmentée de notre tête parlante virtuelle ouvrent la voie à de nombreuses applications dans les domaines de (1) l'orthophonie pour les enfants atteints de troubles de parole, (2) la réhabilitation en perception et production pour les enfants handicapés auditifs, et (3) l'apprentissage de la prononciation et la correction phonétique pour des apprenants en langue seconde.

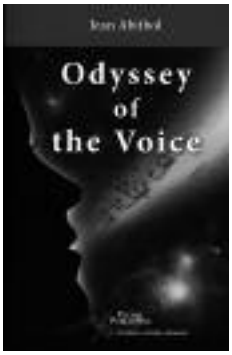
Références

1. HEINZ JM, STEVENS KN. On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech. *Proceedings of the 5th International Conference on Acoustics*, 1965: A44.
2. BOTHOREL A, SIMON P, WIOLAND F, ZERLING JP. Cinéradiographie des voyelles et consonnes du français. 1986:296 *ISBN : 2-902022-00-X*.
3. BEAUTEMPS D, BADIN P, BAILLY G. 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. 2001;109:2165-2180.
4. BADIN P, GABIOUD B, BEAUTEMPS D, LALLOUACHE TM, BAILLY G, MAEDA S, ZERLING JP, BROCK G. Cineradiography of VCV sequences: Articulatory-acoustic data for a speech production model. *Proceedings of the 15th International Conference on Acoustics, Trondheim, Norway*, 1995;IV:349-352.
5. MASAKI S, TIEDE MK, HONDA K, SHIMADA Y, FUJIMOTO I, NAKAMURA Y, NINOMIYA N. MRI-based speech production study using a synchronized sampling method. JOURNAL OF THE ACOUSTICAL SOCIETY OF JAPAN. 1999;20:375-379.
6. PERKELL JS, COHEN MM, SVIRSKY MA, MATTHIES ML, GARABIETA I, JACKSON MTT. Electromagnetic midsagittal articulometer systems for transducing speech articulatory movements. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA. 1992;92: 3078-3096.
7. BADIN P, BAILLY G, REVÉRET L, BACIU M, SEGEBARTH C, SAVARIAUX C. Three-dimensional linear articulatory modeling of tongue, lips and face based on MRI and video images. JOURNAL OF PHONETICS. 2002;30:533-553.
8. BADIN P, SERRURIER A. Three-dimensional modeling of speech organs: Articulatory data and models. *IEICE Technical Report, Kanazawa, Japan*, 2006;106(177),SP2006-26:29-34. .
9. BAILLY G, ELISEI F, BADIN P, SAVARIAUX C. Degrees of freedom of facial movements in face-to-face conversational speech. *Proceedings of the International Workshop on Multimodal Corpora, Genoa, Italy*. 2006.
10. SERRURIER A, BADIN P. A three-dimensional articulatory model of nasals based on MRI and CT data. JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, *in revision*. 2008;123,4 *in press*.
11. ODISIO M, ELISEI F, BAILLY G, BADIN P. Clones parlants 3D vidéo-réalistes: application à l'analyse de messages audio-visuels. *Actes des 7èmes Journées d'Études et d'Échange "Compression et représentation des signaux audivisels" (CORESA' 2001), Dijon, France, France Telecom R&D, IRSIA-Université de Bourgogne*. 2001;141-144.

12. BADIN P, SERRURIER A. Three-dimensional linear modeling of tongue: Articulatory data and models. *Proceedings of the 7th International Seminar on Speech Production, ISSP7, Ubatuba, SP, Brazil (H.C. Yehia, D. Demolin & R. Laboissière, editors) UFMG, Belo Horizonte, Brazil.* 2006:395-402.
13. KELSO JAS, SALTZMAN EL, TULLER B. The dynamical theory of speech production: Data and theory. JOURNAL OF PHONETICS 1986;14:29-60.
14. BAILLY G, BÉRAR M, ELISEI F, ODISIO M. Audiovisual speech synthesis. INTERNATIONAL JOURNAL OF SPEECH TECHNOLOGY. 2003;6:331-346.
15. MAWASS K, BADIN P, BAILLY G. Synthesis of French fricatives by audio-video to articulatory inversion. ACTA ACUSTICA. 2000; 86:136-146.
16. GOVOKHINA O, BAILLY G, BRETON G. Learning optimal audiovisual phasing for a HMM-based control model for facial animation. *6th ISCA Workshop on Speech Synthesis, Bonn, Germany*. 2007.
17. TARABALKA Y, BADIN P, ELISEI F, BAILLY G. Can you "read tongue movements"? Evaluation of the contribution of tongue display to speech understanding. *1ère Conférence internationale sur l'accessibilité et les systèmes de suppléance aux personnes en situation de handicaps (ASSISTH'2007) (N. Vigouroux, P. Gorce & J.-L. Nespoulous, editors), Editions Cépaduès, Toulouse, France*. 2007:187-193.
18. ERBER NP. Auditory-visual perception of speech. JOURNAL OF SPEECH AND HEARING DISORDERS. 1975;XL:481-492.
19. BENOÎT C, LE GOFF B. Audio-visual speech synthesis from French text: Eight years of models, designs and evaluation at the ICP. SPEECH COMMUNICATION. 1998;26:117-129.
20. MILLS AE. The development of phonology in the blind child. *In: Hearing by eye. The psychology of lipreading. R. Campbell, Ed. London, Lawrence Erlbaum Associates.* 1987:145-161.
21. MONTGOMERY D. Do dyslexics have difficulty accessing articulatory information? PSYCHOLOGICAL RESEARCH. 1981;43.

PRESSE / PRESS

ODYSSEY OF THE VOICE
Jean Abitbol
2006 - color illustration - 489 pages - 154x230
Price £29,95 - $49,95
ISBN 1-59756-029-4
Plural Publishing Inc, 49 Bath street, Abingdon, Oxfordshire, OX14 1EA, United Kingdom. www.pluralpublishing.com
Email: noelmcpherson@pluralpublishing.com



Renowned French otolaryngologist Jean Abitbol, a lifetime student of the human voice, takes readers on an unforgettable odyssey spanning man's first use of voice through the acquisition of language to the use of voice as an expression of self. With great wit and charm, Dr Abitbol's narrative encompasses everything from the psychological to the physiological, from explaining the workings of the voice to celebrating the human voice's highest achievements.

He describes a fascinating history of the voice, its origins, its course since the Homo Sapiens' first sentence, its episodes of hoarseness, and its achievements, from the newborn cry to the coloratura soprano, from the impersonator to the ventriloquist. After exploring what is known about the voice, Dr Abitbol tells us what our voices are capable of. He examines what he describes as "the magic of the voice": The voice as a fingerprint, a reflection of our personality in expressing our sex and sexuality.

A great portion of this odyssey is devoted to singing and singers, both to the complexity of singing in general and to lyrical singing, the intricacies of which requires participation of the mechanical, emotional, and cerebral systems. The mysteries of the voice unfold as Dr Abitbol guides readers through the latest physiological and pathological research using examples of historical figures', patients', and celebrities' voices to explain how the ways in which the body moves affect the way the voice sounds and how vocal quality is unique to each human being.

A unique *tour de force* of the human vocal instrument, Odyssey of the Voice changes the way we thing about our voices.

Dr Abitbol is Ancien Chef de Clinique at the Faculty of Medicine of Paris. An otorhinolaryngologist, he specializes in phoniatrics and voice surgery. Based on his clinical experience and research, he has published extensively on voice medicine and is recognized internationally for his contributions to voice surgery and care of the professional voice.

A lifelong student of the intricacies of the human voice, His interest in the human voice led him to develop innovative diagnostic and therapeutic techniques, including vocal dynamic exploration, a method that allows physicians to look at the vocal folds speaking or singing and, more recently, the use of three-dimensional imaging of the larynx. His research has centered on the effects of hormones on the human voice.

Dr Abitbol has published over 300 articles in the medical literature and produced several medical literature and produced several medical movies for which he has won film prizes. He hosts an annual international seminar for physicians and voice professonals on laser voice surgery and voice care and gives conferences to teachers and singers on voice care. A popular international presenter at scientific and progessional meetings, he has received many honors, the most recent of which was the prestigious *Chevalier de la Légion d'Honneur* or which he was nominated by the Ministry of Health on behalf of the President of the French Republic.