Contribution de la mâchoire à la géométrie de la langue dans les modèles articulatoires statistiques

Gérard Bailly, Pierre Badin & Anne Vilain

Institut de la Communication Parlée — INPG & Université Stendhal
46, avenue Félix Viallet, 38031 Grenoble Cedex 1, France
Tél. : ++33 04 76 57 47 11 - Fax : ++33 04 76 57 47 10
e-mail : {bailly,badin,vilain}@icp.inpg.fr - http ://www.icp.inpg.fr/

RÉSUMÉ

Linear component articulatory models [Mae79, Mae88, Gab94] are built using an iterative substraction of linear predictors of the vocal tract geometry. In this paper we consider the contribution of jaw displacement to tongue and lips movements using sets of cineradiographic data from three different speakers. We show that linear prediction overestimates this contribution by capturing not only the intrinsic mechanical jaw-tongue coupling but also the synergetic control observed in the corpus. We then propose a substraction of the jaw contribution which do not affect the performance of the model in terms of data prediction.

1. INTRODUCTION

Les modèles articulatoires permettent de contrôler la forme du conduit vocal à l'aide d'un nombre minimal de paramètres [BMP94]. Dans la plupart des modèles, la forme contrôlée directement est la fonction sagittale. On peut distinguer trois types de modèle : (a) les modèles géométriques [Mer73] considèrent la forme du conduit vocal comme un ensemble de figures géométriques simples (droites, arcs). Certains points ou angles caractéristiques sont utilisés comme paramètres de contrôle ; (b) les modèles biomécaniques [Per69, WT95, SLO97] considèrent la structure musculo-skelettale. Les paramètres de contrôle sont alors les niveaux d'activation musculaire ; (c) les modèles statistiques [HLG77, Mae79, Gab94] cherchent à prédire les formes du conduit vocal d'un locuteur prononçant un corpus limité. Les formes sont obtenues par imagerie médicale dynamique (cinéradiographies, ultrasons ...) ou statique (images IRM ...). Les paramètres de contrôle sont ici les prédicteurs (en général linéaires) de ces formes caractérisées par un nombre constant de points. Ceci est rendu possible par l'usage d'une grille de lecture [Mae88] ou l'échantillonage d'une abscisse curviligne [KH96].

Faire parler ces modèles suppose de pouvoir accéder à des paramètres de contrôle les plus indépendants possibles et auxquels une interprétation neuroanatomique puisse être donnée. L'un des articulateurs dont les degrés de liberté sont les mieux connus, les plus facile à mesurer et qui sont à la base du développement du langage, est certainement la mâchoire. L'oscillation mandibulaire permet d'expliquer les associations préférentielles de l'ontogénèse : l'hypothèse "pure frame" [DM95] suppose l'émergence des premières consonnes par une élévation de la mâchoire avec une langue inerte correspondant à diverses configurations vocaliques (voir simulations dans [BBVB95]).

Il est donc crucial de disposer d'un couplage réaliste entre la mâchoire et les organes qu'elle porte. Si ce couplage est réalisé classiquement dans les modèles géométriques par l'hypothèse "mécaniste" (voir par exemple le modèle de Lindblom et Sundberg [LS71]) décrite à la section 3, les modèles statistiques laissent généralement ce soin à une prédiction linéaire. Nous allons soumettre ces modèles à des données cinéradiographiques de trois sujets enregistrées à l'hopital Schiltigheim de Strasbourg : B1X [Mae88], P1X [BGB+95] et J1X [Vil97].

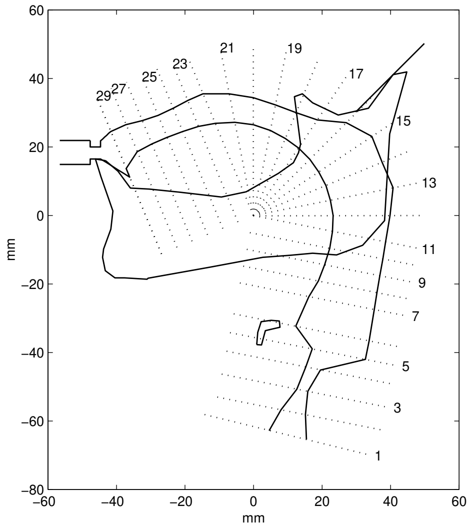
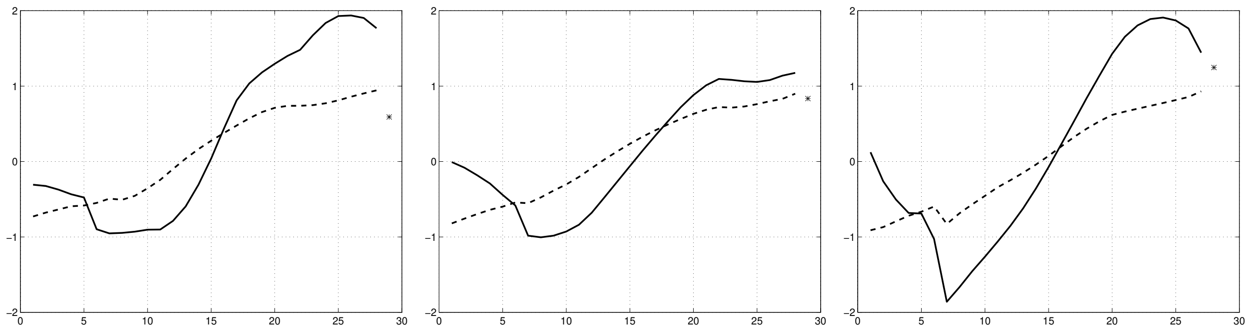


FIG. 1: Grille utilisée pour l'analyse/synthèse de P1X superposée au contour synthétique "neutre". Les droites de 7 à 22 de la grille sont fixes pour chaque locuteur. Les lignes 1..6 sont positionnées par rapport à la hauteur du larynx et les lignes 23..29 par rapport à la pointe de la langue.

2. PRÉDICTION LINÉAIRE DES FORMES DE LANGUES

La grille est définie dans [Gab94]. Cette grille possède des parties fixes, liées aux dents et à la région vélaire, et des parties mobiles liées au larynx, au bout de la langue et aux



(a) B1X (b) P1X (c) J1X

FIG. 2: Pente de la régression linéaire reliant le déplacement de la mâchoire aux points d'intersection de deux langues avec la grille pour les trois corpus. En traits pleins, la langue réelle, en traits pointillés, la langue moyenne supposée rigide. L'étoile figure le coefficient de régression pour l'écartement labial.

lèvres. Un exemple de grille est donné figure 1. Le point 15 de la grille correspond grossièrement à l'extrémité du vélum.

La figure 2 montre, pour les trois sujets, les relations existant entre le déplacement de l'incisive inférieure et les déplacements des points d'intersection du contour sagittal de la langue avec la grille. On constate tout d'abord que ces "fonctions de couplage" ont une grande cohérence : quand la mâchoire se ferme, (a) la partie avant de la langue (points de grille supérieurs à 18) monte, (b) la partie arrière de la langue (points de grille entre 7 et 15) s'avance et (c) le bas pharynx (points de grille entre 1 et 6) s'élargit. Les amplitudes des déplacements de la partie avant et arrière de la langue croissent de manière symétrique par rapport au point 15 de la grille.

On remarque cependant que cette régression linéaire opérée sur les données brutes met en évidence des déplacements dans la partie avant de la langue plus importants que les déplacements de la mâchoire elle-même alors que la grille est bien positionnée suivant l'axe des mouvements et que les points de la langue sont bien circonscrits à l'intérieur du lieu décrit par une rotation/translation simple de la mâchoire. Le rapport entre déplacements atteint 1,95 pour B1X et 1,9 pour J1X. Pour la partie arrière de la langue de J1X, la prédiction linéaire prédit de même des rapports atteignant 1,85. Seul, P1X montre une courbe qui évolue pratiquement entre les bornes physiquement réaliste [-1,+1]. On observe en outre que la pente de régression du déplacement des lèvres est voisine de l'unité. C'est d'ailleurs cette valeur qui est utilisée dans les modèles initiaux issus de ces données et cités en introduction.

3. HYPOTHÈSE "MÉCANISTE"

Ces rapports moyens ne peuvent s'expliquer physiquement qu'en faisant l'hypothèse "irréaliste" que les déplacements de la mâchoire et ceux de la langue sont couplés par un système dynamique largement sur-amorti. Dans l'interprétation quasi-statique des mouvements faite dans les modèles statistiques, ils donnent une grande importance aux mouvements de la mâchoire dans l'exécution des mouvements avant-arrière de la langue (voir figure 3(a)) et, de manière symétrique, un faible découplage langue-mâchoire dans les procédures d'inversion acousti-articulatoire. Nous avons ci-après caractérisé les rapports de déplacements induits par un couplage mécanique de la rotation/translation mandibulaire avec une langue moyenne supposée rigide.

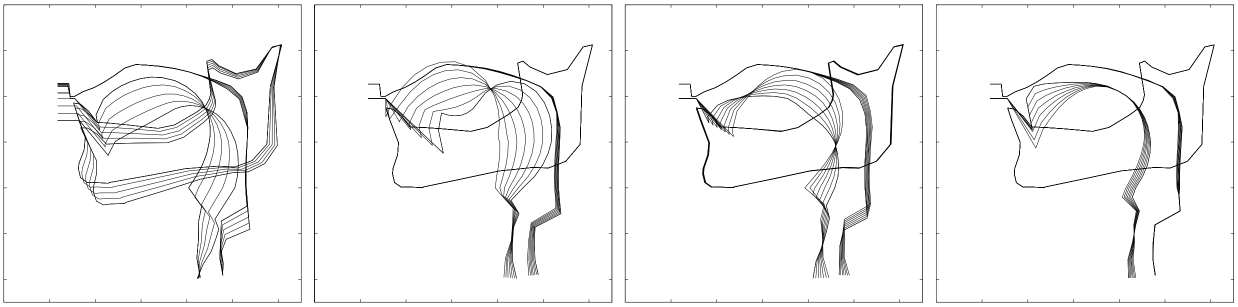
3.1. Langue rigide attachée à la mâchoire

La figure 2 montre également pour les trois sujets, les relations existant entre le déplacement de l'incisive inférieure et les déplacements de la langue moyenne. Pour ceci, on considère un ensemble mâchoire/langue moyenne rigide. Le mouvement de deux points sur la mâchoire, extraits des tracés cinéradiographiques, permet de calculer pour chaque contour la rotation/translation subie par cet ensemble. L'intersection avec la grille de la langue moyenne ainsi déplacée est calculée et le déplacement induit sur la langue moyenne ainsi déterminé. La figure 2 montre (courbes en pointillés) que, pour la partie linguale, les rapports entre le déplacement de l'incisive inférieure et les points de la langue moyenne présentent une allure similaire à celle des pentes de régression commentées à la section 2. Comme attendu, les deux courbes sont dans un rapport de 2 :1 environ et les valeurs maximales absolues dans notre cas ne dépassent pas 1,0. Pour la partie du bas-pharynx cependant, le modèle de couplage mécanique prévoit une influence grandissante de l'articulation mandibulaire de l'épiglotte au larynx alors que la régression laisse cette zone de plus en plus insensible aux mouvements de mâchoire.

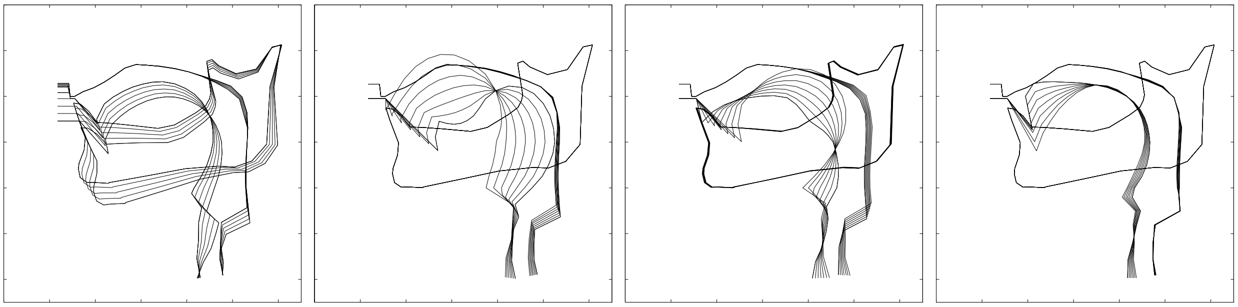
3.2. Modèle statistique composite

Ces données nous ont suggéré une composition des deux modèles visant à minimiser l'influence de la mâchoire sur les déplacements des organes portés : le modèle composite que nous proposons choisit pour chaque point du contour labial le coefficient d'amplitude minimale en valeur absolue. Cette fusion de modèles revient à conserver le modèle initial pour la partie bas-pharynx et à opter pour le modèle "mécanique" pour la partie linguale. Ces modèles sont remarquablement continus dans la zone de transition entre bas-pharynx et langue.

Nous avons vérifié que la poursuite de la procédure de



(a) régression linéaire



(b) modèle composite

FIG. 3: Nomogrammes articulatoires pour J1X obtenus par deux méthodes de soustraction de la composante mâchoire. Les mouvements à partir de la configuration moyenne sont induits par, de gauche à droite : la mâchoire, le corps, le dos et l'élévation de la pointe de la langue.

soustraction de composantes linéaires n'entraînait pas de pertes de prédiction des contours sagittaux. Rappelons que les composantes linéaires extraites sont dans l'ordre après la mâchoire (MA) : le dos (TD), le corps (TB), l'élévation (TT) et l'avancée (TA) de la pointe de la langue. Le paramètre TA permet de contrôler finement la partie non expliquée par les paramètres MA, TD et TB de la distance entre les droites 23 et 29 de la figure 1. La figure 4 compare, pour chaque locuteur, l'évolution de la variance expliquée en chaque point de la grille et en fonction du nombre de composantes linéaires considérées, ceci pour deux modèles : (a) le modèle classique de régression linéaire (RL) et (b) le modèle composite (C). Notons tout d'abord que B1X a une très faible dynamique mandibulaire : l'incidence du prédicteur de couplage langue/mâchoire est donc réduite. Sur les deux autres locuteurs, l'évolution comparée de la variance expliquée suggère quelques commentaires : (a) comme attendu, la composante mâchoire de C n'explique qu'environ la moitié de la variance expliquée par celle de RL, (b) cet écart de variance est rattrapé par les deux paramètres suivants : corps et dos de langue, (c) TB rattrape pratiquement toute la variance sauf une petite portion de la partie centrale, domaine d'action principal de TD.

4. CONCLUSIONS

En prenant appui sur les bases de données cinéradiographiques de trois locuteurs (deux hommes : P1X et J1X et une femme : B1X), nous avons proposé une correction de la première étape de soustraction de composantes linéaires dans les modèles articulatoires statistiques. Le modèle composite proposé ne produit pas une prédiction sous-optimale des contours sagittaux des corpus originaux. Cette correction devrait permettre de mieux caractériser la contribution de la mâchoire aux mouvements des effecteurs finaux de l'articulation et ainsi mieux comprendre et modéliser la négociation entre canaux vocaliques et consonantiques du langage articulé.

5. REMERCIEMENTS

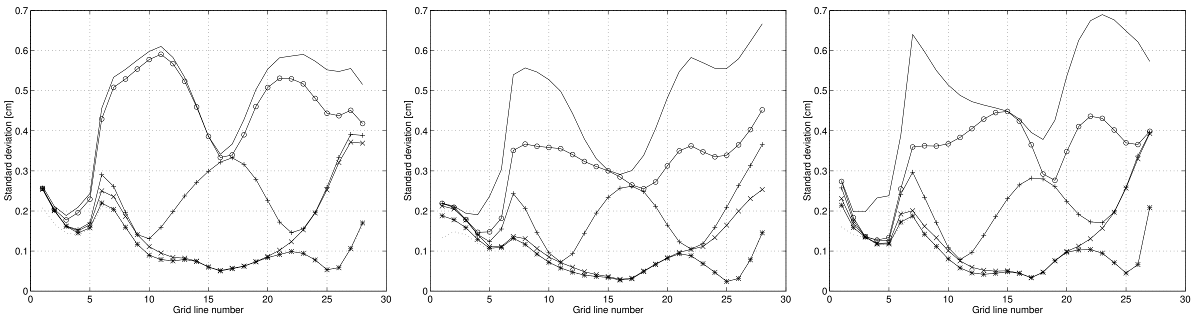
Nous tenons à remercier Shinji Maeda et nos collègues de l'Institut de Phonétique de Strasbourg, plus particulièrement Gilbert Brock pour leur collaboration dans l'obtention des données.

RÉFÉRENCES

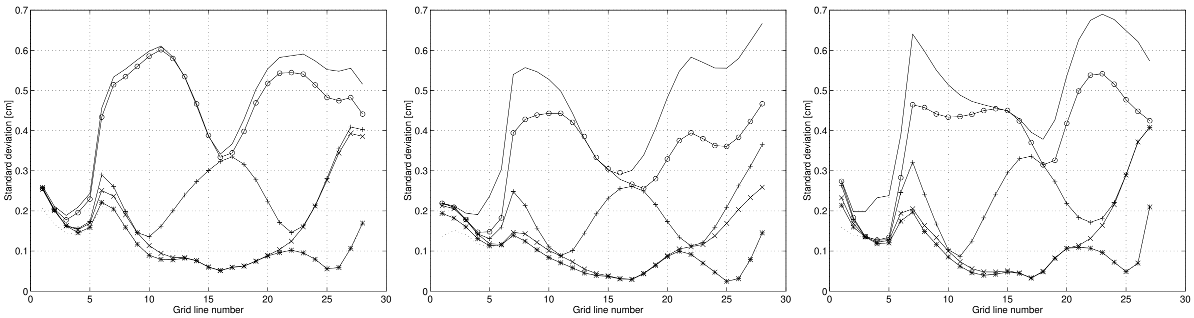
[BBVB95] Gérard Bailly, Louis-Jean Boë, Nathalie Vallée, and Pierre Badin. Articulatori-acoustic prototypes for speech production. In *Proceedings of the European Conference on Speech Communication and Technology*, volume 2, pages 1913–1916, Madrid - Spain, 1995.

[BGB+95] Pierre Badin, Bernard Gabioud, Denis Beautemps, Tahar Lallouache, Gérard Bailly, Shinji Maeda, Jean-Pierre Zerling, and Gilbert Brock. Cineradiography of VCV sequences : articulatory-acoustic data for a speech production model. In *International Congress on Acoustics*, pages 349–352, Trondheim - Norway, 1995.

[BMP94] Louis-Jean Boë, Shinji Maeda, and Pas-



(a) régression linéaire



(b) modèle composite

FIG. 4: Evolution comparée de la variance expliquée en chaque point de la grille et en fonction du nombre de composantes linéaires considérées pour le contour sagittal de la langue pour les deux méthodes d'estimation de la composante mâchoire. De haut en bas dans chaque dessin : la mâchoire ($\circ$), TB ($+$), TD (x) et TT ($*$). De gauche à droite : B1X, P1X et J1X.

cal Perrier. La modélisation articulatoire :un demi-siècle d'évolution entre fonctionnel, physique et biomécanique. In *Journées d'Etudes sur la Parole*, pages 41–54, Trégastel–France, 1994.

[DM95] Barbara L. Davis and Peter F. MacNeilage. The articulatory basis of babbling. *Journal of Speech and Hearing Research*, 38 :1199–1211, 1995.

[Gab94] Bernard Gabioud. Articulatory models in speech synthesis. In Eric Keller, editor, *Fundamentals of speech synthesis and speech recognition*, pages 215–230. John Wiley and Sons, Chichester, 1994.

[HLG77] Richard Harshman, Peter Ladefoged, and Louis Goldstein. Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, 62 :693–707, 1977.

[KH96] Tokihiko Kaburagi and Masaaki Honda. A model of articulator trajectory formation based on the motor tasks of vocal-tract shapes. *Journal of the Acoustical Society of America*, 99(5) :3154–3170, 1996.

[LS71] Bjorn Lindblom and J. Sundberg. Acoustic consequences of lip, tongue, jaw, and larynx movement. *Journal of the Acoustical Society of America*, 50 :1166–1179, 1971.

[Mae79] Shinji Maeda. An articulatory model of the tongue based on a statistical analysis. *Journal of the Acoustical Society of America*, 65 :S22, 1979.

[Mae88] Shinji Maeda. Improved articulatory model. *Journal of the Acoustical Society of America*, 81(S1) :S146, 1988.

[Mer73] Paul Mermelstein. An articulatory model for the study of speech production. *Journal of the Acoustical Society of America*, 53 :1070–1082, 1973.

[Per69] Joseph Perkell. *Physiology of Speech Production*. MIT Press, Cambridge, MA, 1969.

[SLO97] Vittorio Sanguineti, Rafaël Laboissière, and David Ostry. An integrated model of the biomechanics and neural control of the tongue, jaw, hyoid and larynx system. In *Proceedings of the European Conference on Speech Communication and Technology*, volume 4, pages 2023–2026, Rhodes - Greece, 1997.

[Vil97] Anne Vilain. Un nouveau modèle articulatoire pour la synthèse et le contrôle robotique de la parole : Gentiane. Rapport de DEA Sciences du Langage, Université Stendhal – Grenoble, 1997. sous la direction de Christian Abry et Pierre Badin.

[WT95] Reiner Wilhelms-Tricarico. Physiological modeling of speech production : Methods for modeling soft-tissue articulators. *Journal of the Acoustical Society of America*, 5 :3085–3098, 1995.