SYNERGY BETWEEN JAW AND LIPS/TONGUE MOVEMENTS : CONSEQUENCES IN ARTICULATORY MODELLING

Gérard Bailly, Pierre Badin & Anne Vilain

Institut de la Communication Parlée — INPG & Université Stendhal
46, avenue Félix Viallet, 38031 Grenoble Cedex 1, France

Abstract

Linear component articulatory models [9, 10, 5] 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

Articulatory models shape the vocal tract (VT) with a minimal number of parameters [3]. In most of them, the controlled shape is the mid-sagittal contour. We may distinguish three model types: : (a) Geometric models [11] draw the contour with elementary geometric shapes or functions (straight lines, arcs, sinusoids . . .). Certain points or caracteristic angles are used as controlled parameters; (b) biomechanical models [12, 15, 13] consider the musculo-skelettal structure. Controlled parameters are then the levels of muscular activation ; (c) Statistical models [6, 9, 5] characterize each VT contour by a constant number of points and then perform a statistical analysis. The original points are obtained thanks an intersection grid [10] or the uniform sampling of a curvilinear abscissa [7].

The independence and the possible neuroanatomic interpretation of the controlled parameters are desirable properties of the models if we want them speaking. Among them the jaw is certainly the articulator whose degrees-of-freedom are best known and the easiest to measure. It is also the basis of language developpement. Mandibular oscillation explains the preferential associations between consonants and vowels of the first words : the emergence of the first consonants in the "pure frame" hypothesis [4] is produced by a jaw closure with a "inert" tongue corresponding to various vocalic configurations. A simulation using P1X model (see below) may be found in [2].

It is therefore crucial to have a realistic coupling between the jaw and organs that it carries (lips, tongue . . .). The "mechanist" coupling used in geometric models (see for example [8]) is described in section 3. Statistical models generally consider the jaw contribution as the statistical explanation of VT deformations by a few points directly measured on the jaw (for example, coordinates of the lower incise). In this paper we compare these two assumptions using radiofilm data from three subjects recorded at the Schiltigheim Hospital in Strasbourg. These data are of comparable size and have been all used to develop articulatory models: one female (B1X [10]) and two males (P1X [1] and J1X [14]).

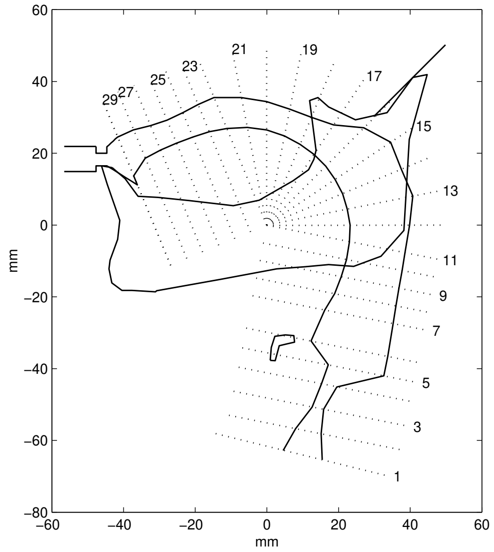


Figure 1: Grid used by the analysis/synthesis of P1X data. It is superposed with the "neutral" synthetic contour. Grid lines from 7 to 22 are fixed for each speaker. Grid lines 1..6 are attached to the larynx and 23..29 to the tongue tip.

2. Linear prediction of tongue shapes

The grid is defined in [5]. This grid has both fixed parts - attached to the upper teeth and the palate - and mobile parts - attached to the larynx, the tongue tip and the lips (see

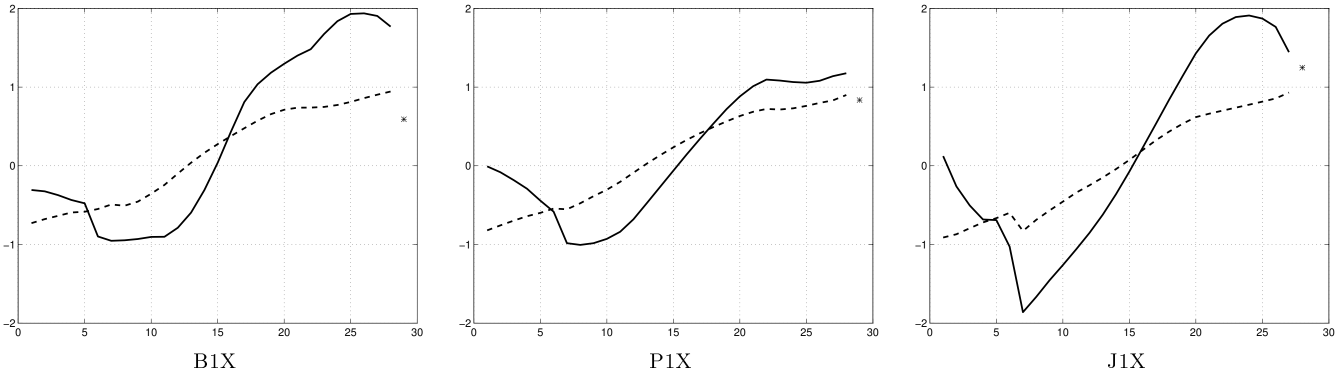


Figure 2: Comparing slopes of the linear regression between jaw height and displacements of the intersection points of the real (-) and rigid (- -) tongue with the grid lines. The star figures the predicted displacement for the labial aperture.

figure 1). The line 15 coincides roughly with the end of the velum.

The figure 2 shows, for the three subjects, the relations between the displacement of the lower incisive and the displacements of the intersections of the mid-sagittal contours with the grid lines. These "coupling functions" have a strong coherence : when jaw closes, (a) the front part of the tongue (18 upwards) raises, (b) the back part (between 7 and 15) advances and (c) the low pharynx (between 1 and 6) gets larger. The amplitudes of these displacements grow symmetrically according to the distance to 15.

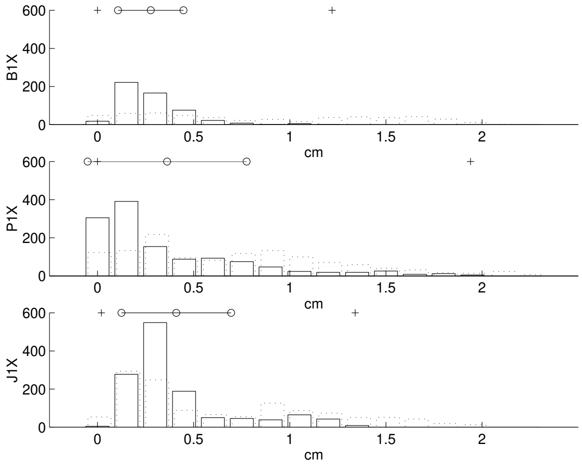


Figure 3: For each subject: statitics of vertical displacements (cm) of the lower incisor referenced to the upper one (-) and of lip aperture (–). The average, standard deviation, minimal and maximal values of the jaw amplitude are given on top of each caption.

The linear regression overestimates the displacements in the front region of the tongue compared to the amplitude of the vertical mouvements of the lower incisor, although the grid lines are approximatively parallel to these mouvements and tongue points are within the space predicted by a simple translation/rottation of the jaw. The slopes reach 1.95 for B1X and 1.9 for J1X. For the posterior of the J1X tongue, the slopes reach -1.85. P1X is the only speaker that exhibits a curve that evolves within the realistic $[-1, +1]$ interval. We observe moreover that the slope for the lips is near the unity[1]. These observations should be compared to the comparative statistics on jaw displacements shown Fig. 3 : J1X has also the larger jaw variance.

These oversestimated values are due to the regression analysis that captures in a single coefficient both jaw and TB contributions that have similar and often synergetic actions on VT shapes but from different muscular origins[2].

3. "Mecanist" hypothesis

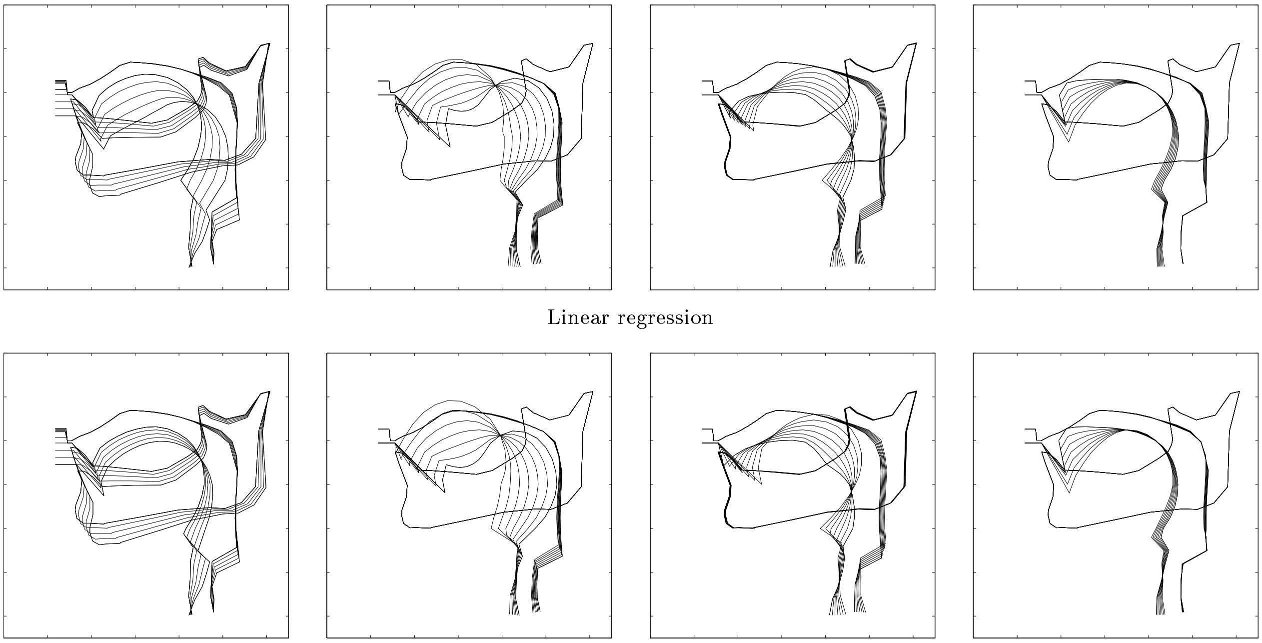
These average slopes can be physically explained only by the unrealistic hypothesis that jaw and tongue are coupled by an overdamped dynamical system. In the quasi-static approximation of movements made in statistical models, they give an over-estimated importance of the jaw in the front/back movements of the tongue (see fig. 4(a)) and, in a symmetric way, a weak decoupling of the two articulators in acoustico-articulatory inversion procedures. We have thereafter computed the slopes produced by the mecanichal coupling of the jaw rotation/translation with a rigid tongue.

3.1. A rigid tongue attached to the jaw

Fig. 2 gives also, for the three subjects, the relations between displacements of the lower incisive and the rigid tongue. The rotation/translation of the jaw was determined by the movement of two fixed points on the jaw. We took the two points of the reference jaw contour adjusted by experts to fit with the tracings of the lower incisive. For each frame, the average tongue is then rotated/translated and the intersection with the grid computed. The displacement of the rigid tongue is finally computed.

[1]Note that this is this value that is used in the models derived from these data and cited in the introduction.

[2]A statistical analysis using muscular activations in a biomechanical model [13] have however the same difficulties in separating the contributions.



Linear regression

Composite model

Figure 4: Articulatory nomograms for J1X using the two mlethods for extracting jaw component. From left to right, the movements are produced by jaw, TB, TD and TT.

Fig. 2 (dotted lines) shows that the slopes have a similar shape as those commented in the previous section 2. As expected, these curves are however in 2:1 ratio and the maximal values do not go over 1,0. For the lower pharynx, this model predicts an increasing influence of the jaw from the epiglottis to the larynx whereas the previous regression predicts the opposite.

3.2. Composite statistical model

We propose a composite model that minimizes the influence of the jaw on the displacements of the carried articulators : we choose for each grid point the coefficient of minimal absolute value. This fusion keeps the initial model for the lower pharynx but opt for the "mechanist" model for the lingual region. The resulting model is remarkably continuous in the transition zone.

We have verified that the guided Principal Component Analysis will still be able to substract intellegible articulatory components. After the jaw, the extracted linear components are in decreasing order of tongue variance: the tongue body (TB) and dorsum (TD), the tongue tip raising (TT) and advance (TA). This last parameter accounts for the residual variance between grid lines 23 and 29. Fig 5 compares, for each speaker, the variance explained by an increasing number of linear components. B1X has a very small amplitude of jaw movements : the incidence of the model is thus reduced. For the two other speakers, we note: (a) as expected, the composite model only explains half of the variance explained by the initial model, (b) this deficit is compensated by the two next parameters : TB and TD, (c) TB does most of the job except for the central region which is the main domain of action for TD.

4. Conclusions

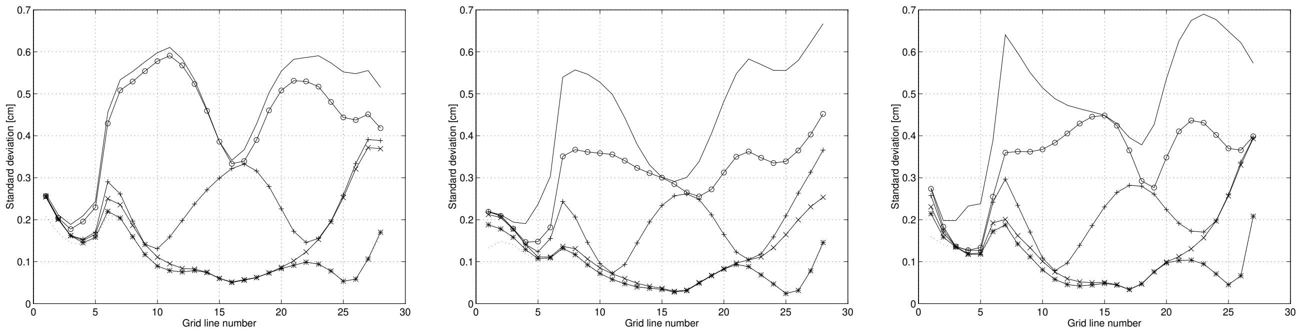
Using radiofilms of three subjects, we have proposed a correction of the first step of the guided Principal Component Analysis used in statistical articulatory models. The composite model does not generate a sub-optimal prediction of the original mid-sagittal contours. This correction will enable us to better understand the contribution of jaw to speech articulation and development.

Acknowlegments

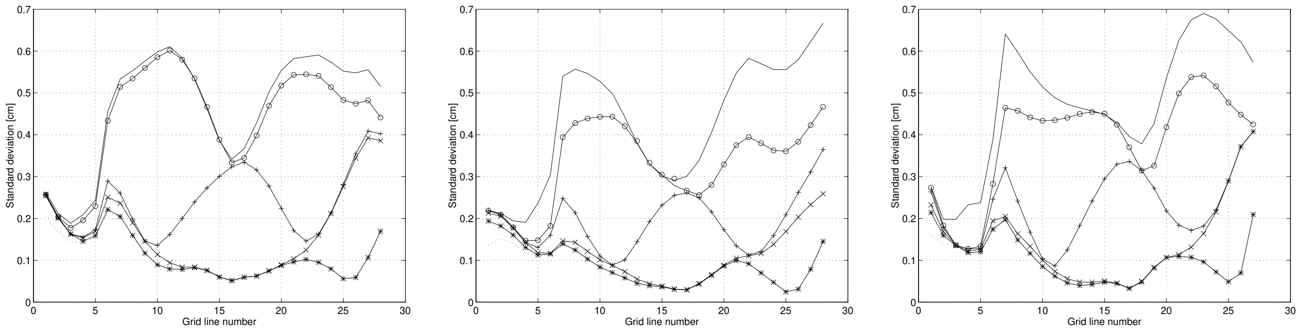
We thank Shinji Maeda and our colleagues of the Institut de Phonétique de Strasbourg, more specifically Gilbert Brock, for data recording and processing.

5. references

1. Badin, P., Gabioud, B., Beautemps, D., Lallouache, T., Bailly, G., Maeda, S., Zerling, J.P., and Brock, G. Cineradiography of VCV sequences: articulatory-acoustic data for a speech production model. In *International Congress on Acoustics*, pages 349–352, Trondheim - Norway, 1995.
2. Bailly, G., Boë, L.J., Vallée, N., and Badin, P. Articulatori-acoustic prototypes for speech production. In *Proceedings of the European Conference on Speech Communication and Technology*, volume 2, pages 1913–



Linear regression



Composite model

Figure 5: For two jaw models, accounted variance of each grid point as a function of the number of linear components. From top to bottom in each caption, JH (∘), TD(+), TD (x) and TT ($*$). From left to right : B1X, P1X and J1X.

1916, Madrid - Spain, 1995.

3. Boë, L.J., Maeda, S., and Perrier, P. 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.

4. Davis, B.L. and MacNeilage, P.F. The articulatory basis of babbling. *Journal of Speech and Hearing Research*, 38:1199–1211, 1995.

5. Gabioud, B. Articulatory models in speech synthesis. In Keller, E., editor, *Fundamentals of speech synthesis and speech recognition*, pages 215–230. John Wiley and Sons, Chichester, 1994.

6. Harshman, R., Ladefoged, P., and Goldstein, L. Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, 62:693–707, 1977.

7. Kaburagi, T. and Honda, M. 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.

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

9. Maeda, S. An articulatory model of the tongue based on a statistical analysis. *Journal of the Acoustical Society of America*, 65:S22, 1979.

10. Maeda, S. Improved articulatory model. *Journal of the Acoustical Society of America*, 81(S1):S146, 1988.

11. Mermelstein, P. An articulatory model for the study of speech production. *Journal of the Acoustical Society of America*, 53:1070–1082, 1973.

12. Perkell, J. *Physiology of Speech Production*. MIT Press, Cambridge, MA, 1969.

13. Sanguineti, V., Laboissière, R., and Ostry, D. 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.

14. Vilain, A. 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.

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