Comparative articulatory modelling of the tongue in speech and feeding

Antoine Serrurier[1]*, Anna Barney[1], Pierre Badin[2], Louis-Jean Boë[2] & Christophe Savariaux[2]

[1] Institute of Sound and Vibration Research, University of Southampton, UK
[2] GIPSA-lab (Département Parole & Cognition / ICP), CNRS – Universités de Grenoble, France
* E-mail: A.Serrurier@soton.ac.uk

Abstract

Two of the major functions of the human vocal tract are feeding and speaking. Ontogenetically and phylogenetically feeding tasks precede speaking tasks and it has been hypothesised that speaking movements constitute a subset of feeding movements. From ElectroMagnetic Articulography data we have extracted two estimates of the degrees of freedom of the tongue, one from a speech task and one from a feeding task. These have been used to build two corresponding articulatory models. For both tasks, about 95% of the tongue variance can be explained by 4 parameters. Contrary to the accepted behaviour for speech, for feeding tasks, the tip, mid and back regions of the tongue appear to be controlled independently. The reconstruction errors of speech and feeding articulations by the two models show slightly better efficiency for the feeding model, suggesting that the speech movements might realistically be considered a subset of the feeding movements.

1 Introduction

In the framework of speech evolution, the study of the point at which speech movements emerged in the vocal tract raises a number of questions. A current hypothesis is that during evolution, humans developed articulatory movements for speech by borrowing skills related to other tasks such as chewing or swallowing [3, 4]. Three main functions can be ascribed to the vocal tract: breathing, feeding and speaking. Whilst breathing does not require complex vocal tract articulation, we know that both ontogenetically and phylogenetically feeding tasks precede speaking tasks. Based on this observation and relying on the frame-content theory [4], [2, 3] have hypothesized that the geometrical and articulatory spaces covered by the articulators while speaking might be a subset of the spaces covered while feeding. In the present study, we have considered this hypothesis in terms of degrees of freedom of the jaw and tongue. Our method involves recording articulatory measurements during speaking and feeding tasks using an ElectroMagnetic Articulograph (EMA) and building the associated articulatory models [1]. Although some comparative articulatory studies between speech and feeding have already been conducted [3, 5], our study aims to provide the first articulatory model of the vocal tract for feeding tasks. The linear modelling relies on Principal Component Analysis (PCA) and linear regression to extract the degrees of freedom of the jaw and the tongue from the measurements, as described in [1].

2 Data

2.1 Subject and corpora

To model the feeding mechanism using EMA, a French subject already used in another articulatory modelling study [1] was recorded.

The speech corpus consisted of (1) a set of artificially sustained articulations representative of the range of French articulations (44 phonemes including oral and nasal vowels, and consonants in three contexts [a i u], see [1]), and (2) a set of 9

This work formed part of the HandToMouth project funded by the European Commission under the NEST initiative.

French continuous, phonetically balanced sentences. The feeding corpus was designed according to the clinical protocol for swallowing disorder assessment. The food was divided into three categories covering a wide range of food textures: liquids (saliva, water, single cream, custard), solid food (Angel Delight, Weetabix softened in milk) and hard food (Rice Crispies in single cream, shortbread biscuits). The subject fed himself by means of a plastic teaspoon from a container in front of him, performing between 3 and 6 swallows for each type of food. The water was drunk from a glass by 3 different methods: (1) using a plastic teaspoon, (2) continuously using a plastic straw and (3) continuously directly from the glass.

2.2 Articulatory data

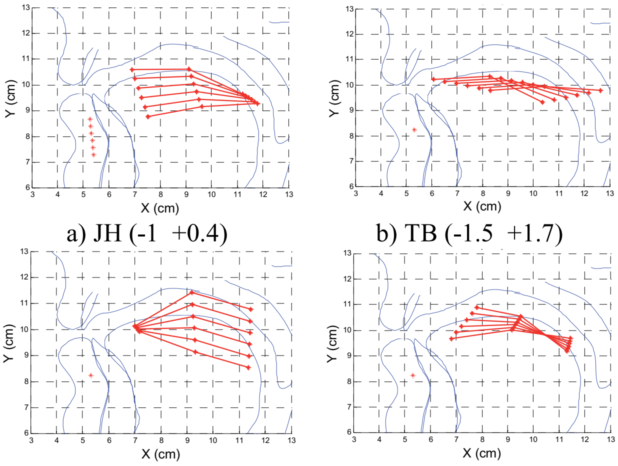
Eight EMA sensors were glued on the subject's articulators in the midsagittal plane: upper incisors and top of the nose (used as references for the head), lower incisors, upper and lower lips, and tip, mid and back regions of the tongue (about 1 cm, 4 cm and 6 cm from the tip point respectively). The EMA data used in this study consist of the coordinates of the jaw and tongue sensors centres.

The middle instant of each sustained phoneme was manually labelled and chosen as representative of the phoneme articulation. For the sentences, all the samples between the beginning and the end of each utterance have been used. For the feeding data, each sequence represents the entire process of swallowing, from the opening of the mouth for placing the food through to the final swallowing of the bolus. Note however that for the water drunk from the glass or from the straw, the sequence is considered to run from the beginning of the first swallow to the end of the last swallow, considering the movement as continuous rather than as succession of elementary swallows. In summary, with repetitions during the recordings, we obtained 82 sequences for the sustained phonemes, 12 for the sentences and 34 for the entire food corpus.

3 Articulatory model of the tongue in speech

The speech data used for the articulatory modelling of the tongue consist of 82 articulations of 8 variables (3 locations on the tongue + 1 on the jaw each with 2 coordinates).

It is generally agreed that the jaw constitutes the primary articulator that impacts on the tongue position. Around 95% of the variance of the jaw position can be explained by a single articulatory parameter, *Jaw Height* (JH), obtained by PCA on the lower teeth sensor coordinates. A linear regression of the 82×6 tongue coordinate measurements on the JH parameter allows to determine the contribution of JH to the tongue (see Fig 1a) and to remove its contribution from the data. A PCA is then applied to the 82×4 residue coordinates of the mid and back tongue sensors. The first two parameters correspond to the articulatory parameters *Tongue Body* (TB) and *Tongue Dorsum* (TD). Their contribution to the tongue is visible in Fig 1b and 1c. Finally, the parameter *Tongue Tip* (TT) is extracted by PCA on the residue of the tongue tip sensor coordinates (see Fig 1d). The explained variance and the root mean square (RMS) reconstruction error for each articulatory parameter are summarised in Table 1: four parameters only explain 94% of the variance. We observe that the degrees of freedom extracted in this study are in good general agreement with those extracted for the same subject by the same method from similar corpuses but different data sets [1].



a) JH (-1 +0.4) b) TB (-1.5 +1.7)

c) TD (-1.4 +1.8) d) TT (-0.5 +1.1)

Figure 1: *Nomograms of the tongue corresponding to the 4 speech articulatory parameters, superimposed on a fix contour of the vocal tract at rest. The ranges of variation correspond to those found in the data.*

Table 1. *Explained variance, cumulative explained variance and cumulative RMS reconstruction error for the tongue by each of the articulatory parameters*

	Var.	Cum.Var.	RMS
Speech model parameters			
JH	19 %	19 %	0.55 cm
TB	48 %	67 %	0.35 cm
TD	20 %	87 %	0.22 cm
TT	7 %	94 %	0.15 cm
Feeding model parameters			
JH	65 %	65 %	0.46 cm
Tbck	11 %	76 %	0.35 cm
Tmid	8 %	84 %	0.28 cm
Ttip	13 %	97 %	0.13 cm
TtipV	2 %	99 %	0.08 cm

4 Articulatory model of the tongue in feeding

The model presented in this section constitutes a first attempt to extract the degrees of freedom for feeding tasks. Recall that unlike speech data, the feeding sequences used here consist of EMA trajectories, each sequence being a succession of positions of the sensors.

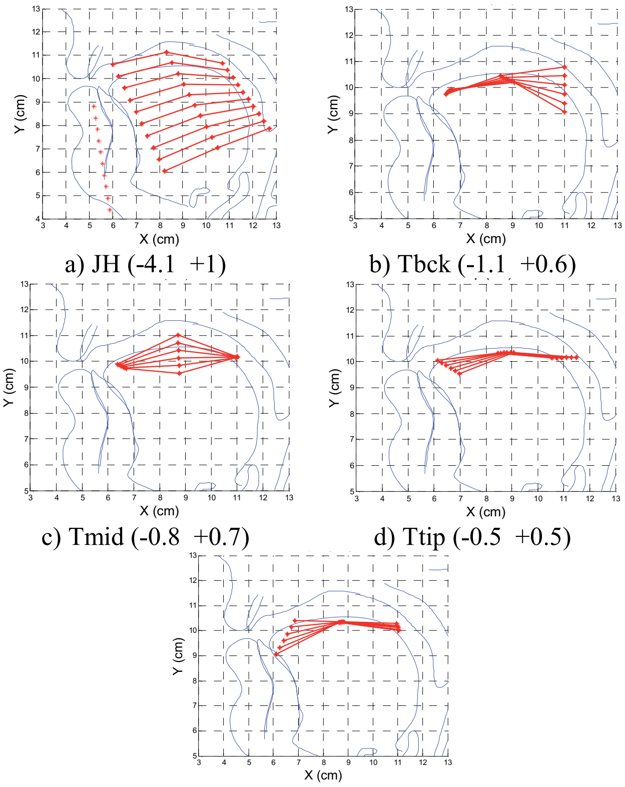
Again, the jaw is the principal influence on tongue position. The parameter JH is thus extracted from the jaw sensor coordinates on the full feeding corpus (see it contribution to tongue in Fig 2a). The range of variation of the jaw and thus of the tongue is greater for feeding than for speech. The movement of the tongue differs from that during speech: the back of the tongue follows a downward-backward movement where the movement is simply backward for speaking.

We make the assumption that the EMA sequence for water drunk from a glass is typical of swallow patterns for feeding. This sequence shows a cyclic, continuous movement of the tongue, from the front of the mouth to collect the liquid to the back of the mouth to swallow it, without any discontinuity or break. The model described next has thus been extracted from this sequence only.

A study of the cross-correlations of the residues of the tongue variables on this sequence has shown very little correlation suggesting that, unlike speech, the three points of the tongue act quite independently. Three articulatory parameters are thus extracted corresponding to the three tongue points, starting with the back point and finishing with the front, in order of decreasing variance. First *Tongue Back* (Tbck), was obtained by PCA on the back sensor coordinates and its contribution to the 6 tongue variables assessed and removed by linear regression. The same procedure was then applied to the mid sensor coordinates to extract *Tongue Mid* (Tmid) and finally to the front sensor coordinates to extract *Tongue Tip* (Ttip). The nomograms of these parameters are shown in Fig 2b to 2d. The contribution of each articulatory parameter to the full feeding corpus is summarized in Table 1.

We observe that Tbck corresponds mainly to a vertical movement of the back of the tongue acting as an opening-closing connection with the pharyngeal cavity. Tmid corresponds mainly to a vertical movement of the mid of the tongue, the two extremities remaining fixed. This movement appears related to the creation of a puddle in the middle of the mouth to gather the liquid before lifting the tongue up to the hard palate to pass the liquid to the pharynx. Finally Ttip corresponds to a global front-back movement of the tongue associated with an up-down movement of the front of the tongue, so as to collect the liquid arriving in the mouth.

We observe that the four parameters JH, Tbck, Tmid and Ttip explain 97 % of the full feeding corpus variance, although three of them have been extracted on a restricted version of this corpus. However some tongue variability may have been missed by this approach, as only jaw movements and liquid swallowing have been considered for the model. A more precise study of the residue of the feeding data unexplained by the model shows a large variability for the front point of the tongue. PCA on the residue of the front point coordinates allows a *Tongue Tip Vertical* parameter (TtipV) to emerge, whose contribution to the tongue obtained by linear regression is displayed in Fig 2e. TtipV corresponds to an upward-backward movement of the front of the tongue which might be ascribed to a residue of mastication or a complementary action to move food from the front to the middle of the mouth. Although TtipV has a relatively low impact on the variance explanation, it corresponds to a plausible movement and it reduces significantly the RMS reconstruction errors found articulation by articulation.



a) JH (-4.1 +1)
b) Tbck (-1.1 +0.6)
c) Tmid (-0.8 +0.7)
d) Ttip (-0.5 +0.5)
e) TtipV (-0.9 +0.7)

Figure 2: *Nomograms of the tongue corresponding to the 5 feeding articulatory parameters. The ranges of variation correspond to those found in the data.*

5 Comparison of the models

We observe first that the same number of articulatory parameters for the two models leads to a comparable accuracy of reconstruction (see Table 1).

We attempted to evaluate the discrepancy between the two models by trying to reconstruct both the speech and the feeding articulations from both the speech and the feeding models. The observation of the global RMS reconstruction errors of the tongue for the 12 sentences shows a similar accuracy for the two models, with an error approaching 0.1 cm for each sentence. The feeding model appears slightly more efficient, however, than the speech model. Equivalently, the global RMS reconstruction errors of the tongue for the 34 feeding sequences have been computed. For both models the error increases with increased viscosity of the food. Across all sequences, the feeding model shows reconstruction errors between 0.05 cm and 0.1 cm while the speech model shows between 0.1 cm and 0.2 cm.

6 Conclusion and Future Work

The performance of the feeding model compares favourably with current speech articulatory models. The number of degrees of freedom and their description are plausible and we note the prediction that the front, mid and back of the tongue can be controlled independently during feeding.

We conclude that the feeding model as presented allows more accurate reconstruction of the tongue positions for both speech and feeding samples. This suggests the speech movements might well constitute a subset of the feeding movements as hypothesised in the framework of speech evolutionary theory: a key point in the discussion of speech emergence. A more systematic assessment of the efficiency of the two models will be required to fully explore this hypothesis. In particular, a study of each tongue position reconstructed by the two models would highlight which articulations are present or absent, and thus are predictable or not, for each task. Such a study is currently under way.

Acknowledgements

We would like to thank Morwenna Collins (Univ. of Southampton, UK) for helpful discussions about the feeding corpus design and the feeding model.

References

[1] P. Badin & A. Serrurier, A. Three-dimensional linear modeling of tongue: Articulatory data and models. *ISSP7*, 395-402, Ubatuba, Brazil, 2006.

[2] K.M. Hiiemae, J.B. Palmer, S.W. Medicis, J. Hegener, B.S. Jackson & D.E. Lieberman. Hyoid and tongue surface movements in speaking and eating. *Arch Oral Biol,* 47(1):11–27, 2002.

[3] K.M. Hiiemae & J.B. Palmer. Tongue movements in feeding and speech. *Crit Rev Oral Biol Med,* 14(6):413–429, 2003.

[4] P.F. MacNeilage. The frame/content theory of evolution of speech production. *Behav Brain Sci*, 21:499–511, 1998.

[5] D.J. Ostry, E. Vatikiotis-Bateson & P.L. Gribble. An examination of the degrees of freedom of human jaw motion in speech and mastication. *JSLHR,* 40:1341-1351, 1997.