Speaker-specific biomechanical models: From acoustic variability via articulatory variability to the variability of motor commands in selected tongue mucles

Ralf Winkler, Susanne Fuchs, Pascal Perrier, Mark Tiede

Speaker-specific biomechanical models: From acoustic variability via articulatory variability to the variability of motor commands in selected tongue muscles

Ralf Winkler[1*], Susanne Fuchs[1], Pascal Perrier[2], Mark Tiede[3]

[1]ZAS – Centre for General Linguistics, Schützenstrasse 18
10117 Berlin, Germany

[2]GIPSA-lab/ICP, Domaine Universitaire BP 46,
38402 Saint Martin d'Hères, France

[3]Haskins Laboratories, 300 George Street,
New Haven Connecticut 06511, U.S.A.

winkler@zas.gwz-berlin.de

Abstract. *In this work we have constructed biomechanical tongue models derived from MRI data in order to investigate the effects of differing locations of vocal tract bending on variability in motor command space and overall articulatory variability for vowel targets. Acoustic models predict negligible effects of the bend of the vocal tract if its length is held constant. However, the location of this bend crucially determines the relation between vertical and horizontal dimensions of the tract and thus the relative freedom of tongue movement within these dimensions. We predict that articulatory variability will be greater along those dimensions with more degrees of freedom as determined by vocal tract configuration imposed by bend location, and present simulation results that in general support this position.*

1. Theoretical background

Although native speakers of a given language may use the same phonemic inventory, speakers vary from one to another with respect to their acoustics, articulation, and probably also their respective motor commands. However, the latter is more speculative. The variability among speakers may be driven by anatomical, emotional, social and communicative factors. In a previous study (Brunner et al., 2009) we have focused on palatal morphology in constraining individual speech motor control. Here we concentrate on individual differences in the location of the vocal tract bend and its impact on speaker specific variability at three levels: acoustics, articulation and motor commands.

Such an investigation is not possible using human subjects for two reasons: first, one cannot control or disentangle the different factors described above; second, so far it is impossible to observe motor commands for human tongue movements. Instead, we

[*]Supported by a grant from the German Research Council to the SPRECHart project.

build speaker-specific biomechanical tongue models on the basis of human imaging data that allow us to link motor commands, articulation and acoustics. To our knowledge, such an investigation has not been carried out before. Unlike speaker-specific geometrical models which offer only superficial interpretations of underlying control, biomechanical tongue models have the advantage of controlling muscular length and forces as the results of the combined influences of central commands and feedback signals, and therefore they provide insights into the underlying muscle activations and dynamics which drive the tongue's movement.

2. Modeling

In this section the necessary steps to construct speaker-specific models are presented. After a short description of the labeling procedure, the steps towards building speaker-specific biomechanical models, running simulations and estimating the acoustics are presented.

2.1. Labeling human imaging data

Speaker-specific models were constructed based on Magnetic Resonance Imaging (MRI) data of two speakers (Figure 1, top row). Imaging data were originally collected for 10 isolated vowels to study inter-speaker acoustic and articulatory variability (Apostol, 2001). Details regarding the image acquisition procedure as well as the image resolution are specified in Apostol (2001).

Within each image slice the airway was segmented from the surrounding tissues manually by using the itk-SNAP software (Yushkevich et al., 2006). The biomechanical tongue model is implemented with standard teeth and standard lips. For that reason segmentation of the air channel terminates at the incisors. The lip region was ignored during segmentation because the front tube used during area function calculation was kept invariant. Furthermore, the epiglottis and the uvula were consistently excluded because the tongue contour is handled separately in the articulatory model and the uvula is not part of the model.

2.2. Speaker-specific biomechanical tongue models

The speaker-specific models are based on the well established 2D biomechanical tongue model of Perrier et al. (2003), which consists of a deformable Finite Element Mesh (FEM) embedded in rigid vocal tract walls in the mid-sagittal plane (Figure 1, 2nd row). The geometry of the mesh has been specifically designed to facilitate anatomical implementation of the muscles within the tongue. The model is controlled according to the λ-model (Feldman, 1986), which generates muscle force as a function of the difference between a centrally specified threshold length λ and the actual muscle length. Details regarding the assumptions underlying the design of the speaker-specific models are given in Winkler et al. (submitted). Anatomical landmarks necessary for constructing speaker-specific models include the following: the mid-sagittal tongue contour at rest position, the mid-sagittal palate contour, the velar contour, the posterior pharyngeal wall, the lower and upper limits of the tongue insertions on the mental spine (P1 and P2), and the styloid process (P3).

These mid-sagittal landmarks were extracted from labeled imaging data for each of the two subjects with the tongue in rest position.

The three anatomical landmarks (P1, P2 and P3) were measured from a high-definition mid-sagittal view of the speaker's MRI data. P1 and P2 were determined on the basis of grey level changes in the mental spine region. It is not possible to determine the exact location of the styloid process in the mid-sagittal plane directly. As an approximation in the mid-sagittal plane, P3 was placed on the internal contour of the sphenoid bone at 1/3 of its length.

Based on these anatomical landmarks the original generic biomechanical model was adapted to a speaker-specific anatomy (for details of the matching procedure see Winkler et al. (submitted)). The matching procedure fully determines the geometry of the new mesh and consequently the muscle arrangement within the new speaker-specific tongue model. It preserves the original topology of the mesh while accounting for the speaker-specific morphology.

In order to simulate tongue movements with the two speaker-specific biomechanical tongue models, motor commands and their activation duration have to be specified for successive targets. In order to compare speaker-specific biomechanical tongue models with each other, not absolute but relative motor command values were chosen. The motor commands were defined relative to their rest position ($\Delta\lambda = \lambda_{target} - \lambda_{rest}$).

2.3. Acoustic modeling

In order to determine acoustics from articulation, the 2D distance function resulting from the biomechanical tongue model has to be converted to its corresponding area function. There are basically two different methods. On the one hand, in absence of the 3D airway contours a set of physiologically realistic α-values from the literature (for instance in Perrier et al. (1992)) can be applied to estimate the areas from distance values. On the other hand, the coefficients can be determined speaker-specifically if 3D data are available.

In our approach this is accomplished by determining the distance function based on a pre-defined grid (Perrier et al., 2003), and subsequent reconstruction of the area function proceeds according to the speaker-specific coefficients for an adapted version of the α-β-model (Perrier et al., 1992) associated with the respective grid lines.

Estimating acoustics (i.e. formants) from area functions does not involve any speaker-specific adaptation. In our experiments formants are computed by coupling an acoustic analog of the vocal tract with the reconstructed area functions.

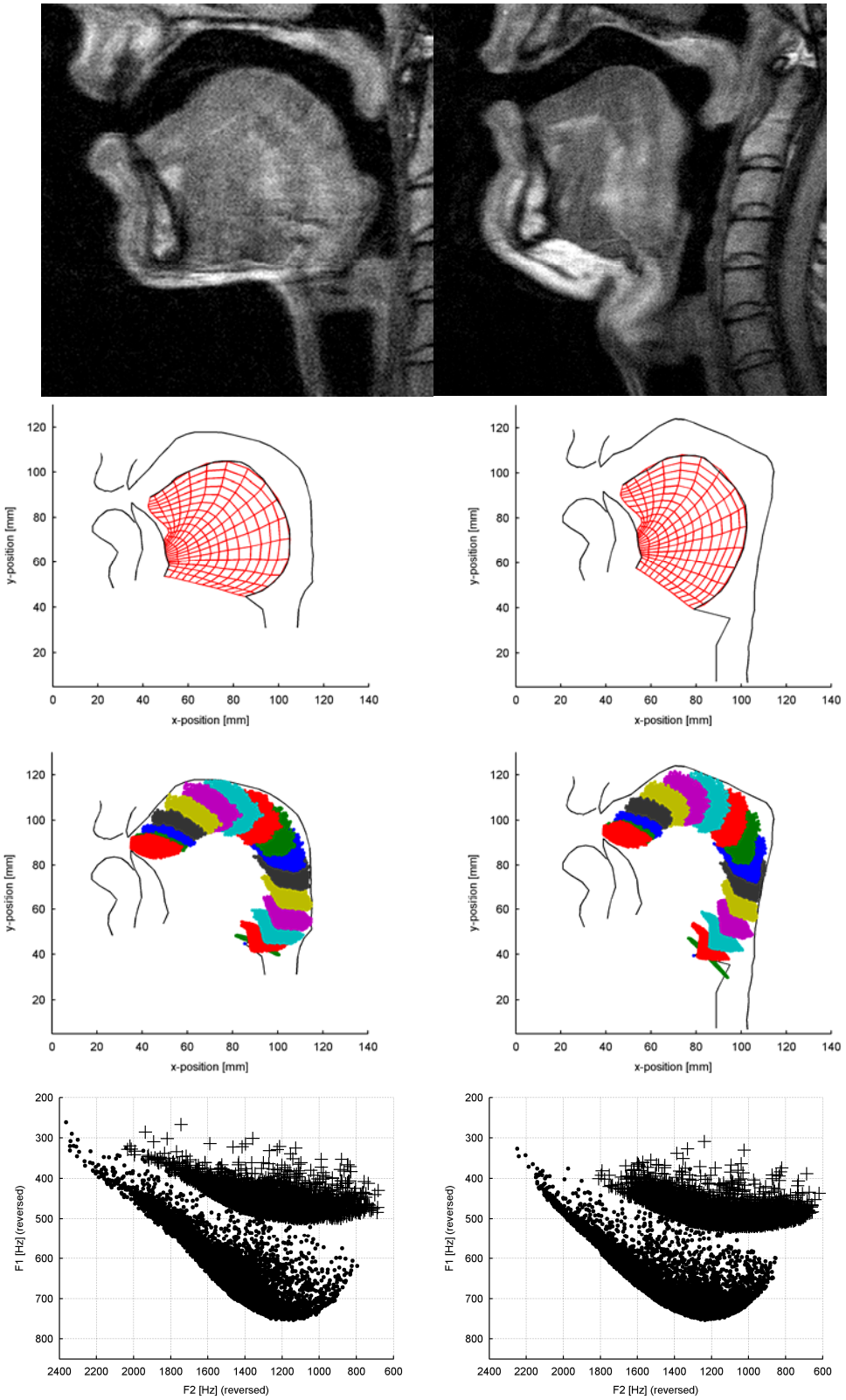

Figure 1: 1st row: Mid-sagittal MRI data (left AV, right CS), 2nd row: biomechanical tongue models with FE mesh for both speakers, 3rd row: articulatory simulations, 4th row: corresponding vowel spaces

3. Experimentation

Acoustic models of speech production assume that the length of the vocal tract, and the length and location of the constriction defining the resonance cavities are crucial parameters for the description of the spectral properties of vowels. The location of the vocal tract bend seems to play a negligible role with respect to the acoustics, assuming the length of the tract is kept constant (Sondhi, 1986). However, the location of the bend within the vocal tract affects the relative vertical and horizontal dimensions of the tract and we suppose that it is likely to influence the degrees of freedom of articulatory motion in the respective directions. The two speakers we selected to build speaker-specific models show two different locations of vocal tract bending. Speaker CS has a relatively long vertical dimension in comparison to the horizontal dimension whereas speaker AV shows about equal proportions between the two. In other words, for CS the location of the vocal tract bend is more anterior than for AV (assuming the same vocal tract length). These relations may also be representative for differences between males (CS with a long pharynx) and females (AV with a shorter pharynx).

3.1. Method

We assume that for vowel production, the goals of speech production are primarily auditory (Perrier, 2005). Based on this assumption, in the region of maximal overlap of the formant spaces of model AV and CS three formant ellipses were defined in the F1/F2 plane for the three corner vowels. Prior to the definition of the formant ellipses, the area functions were manipulated to match the acoustic vowel space of the two speakers. In a first step the length of the epilaryngeal tube of speaker CS was shortened to match that of speaker AV. Secondly, the original area values in the laryngeal region were fixed to a physiologically realistic value of 0.4 cm^2. In the region of the epiglottis the area values were decreased by 50% for both speakers to avoid an artificially large cavity resulting from the excised epiglottis in the tongue model. Finally, the area functions were re-scaled to a vocal tract length of 17 cm, since we wanted to disentangle the impact of vocal tract length and the location of vocal tract bending. For each acoustic vowel target 36 simulations per speaker were randomly selected whose respective first two formant values were equally distributed over the corresponding ellipse area.

Assuming the same acoustic target regions of the corner vowels for both models, we now look at the articulatory correspondences in the two models and their respective motor commands. Of particular interest is the amount of variability at the articulatory and motor command levels.

3.2. Results

3.2.1. Articulatory results

In order to obtain the articulatory results for the simulations of each model which correspond to the acoustic dispersion ellipses, we selected three nodes along the surface of the Finite Elemente Mesh for /i, a/ and two for /u/. These nodes are located at the constriction location of the relevant vowel, i.e. node locations differ among the vowels.

For each node the standard deviation in the vertical (y) and horizontal (x) direction was calculated. We then compared the standard deviation of the two models by subtracting the respective x- or y-standard deviations of CS from AV. If the difference in standard deviation is close to zero, both models behave in a similar fashion. Negative values correspond to a larger standard deviation for CS than AV and positive values to a larger standard deviation for AV than CS. We expect that the CS model shows generally more variability in the vertical direction, since it has a longer pharynx than the AV-model. For AV we expect that more variability is allowed in the horizontal direction, since it has a relatively long horizontal dimension in comparison to the CS-model.

Figure 2 displays the results, which are split by vowel. The red triangles show the results in the vertical direction and the black dots the results in horizontal direction.

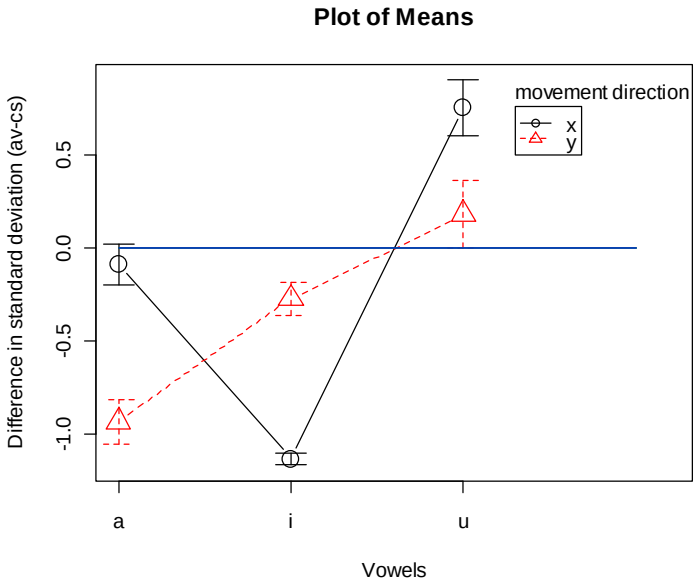

Figure 2: Differences in variability (standard deviation) of the articulatory data in vertical (y) and horizontal (x) direction between model AV and CS; if this difference is negative then $variability_{CS}$>$variabillity_{AV}$; for positive values $variability_{CS}$<$variability_{AV}$. Data are split by vowel.

Given the same acoustic output for CS and AV, the articulatory variation differs. Model CS allows more variability (negative values for the difference in standard deviation) in the *vertical* direction for /a/ and /i/, but not for /u/. Findings for /u/ may differ from the other vowels since the constriction location is placed along the bending of the vocal tract and may be affected by it. Model AV allows more variability (positive values for the difference in standard deviation) in the *horizontal* direction for /u/ only. In /a/ the two models are quite similar (close to zero) and in /i/ the CS-model can vary more freely than the AV-model. This result goes against our original hypothesis. However, /i/ is to a large extent constrained by the shape of the palate which varies between the two models.

To summarize, the articulatory results provide evidence that the location of the vocal tract bending affects the articulatory variability for back and low vowels. Speakers with

long vertical vocal tract dimensions have larger degrees of freedom to move vertically than speakers with shorter vertical vocal tract dimensions. Similarly, speakers with long horizontal dimensions have the possibility to move more freely in this direction than speakers with shorter horizontal dimensions. The variability of front vowels may be more constrained by the shape and steepness of the palate than by vocal tract bending.

3.2.1. Results for the motor commands

In this section we will concentrate on the variability in lambda values (=motor commands) for /a/ and /u/ only. In principle, 6 different muscles can be activated, but only a few of them are necessary to realize the constriction of the relevant vowel.

To produce the velar constriction of /u/, a combined activation of the Styloglossus and Genioglossus Posterior are necessary. Both muscles are responsible for pulling the tongue up and back. We found slightly more variation in both muscles in the CS-model.

For /a/ a constriction is formed in the pharyngeal region. This constriction is produced by a combination of Hyoglossus (moving the tongue down and back) and Styloglossus (up and back) activation. Results from our simulations show much more variation for model CS in both muscles than in model AV.

4. Summary and conclusion

We have constructed speaker-specific biomechanical tongue models with the aim of investigating the effect of the location of vocal tract bending on articulatory variability and the variability in motor command space. Assuming the length of the vocal tract is kept constant, acoustic models predict only negligible effects of the bend of the vocal tract. However, the bending location within the vocal tract may be crucial for the relation between vertical and horizontal dimensions in the vocal tract and the respective degrees of freedom the tongue has within these dimensions.

Based on similar acoustic targets for the corner vowels we ran simulations using the speaker-specific models and analyzed the corresponding variability at the articulatory and motor command level.

Evidence for our assumptions was found for back and low vowels. The high front vowels may be more sensitive to palate shape than the location of the bend in the vocal tract. Further investigations are planned which will allow us to disentangle effects due to different vocal tract shapes and biomechanics.

References

Apostol, L., “Étude et simulation des caractéristiques individuelles des locuteurs par modélisation du processus de production de la parole”, Unp. PhD thesis, INP Grenoble, France, 2001.

Brunner, J., Fuchs, S. & Perrier, P. On the relationship between palate shape and articulatory behavior. JASA 125(6): 3936-3949, 2009.

Feldman, A.G., Once More on the Equilibrium-Point Hypothesis (Lambda-Model) for Motor Control", J Mot Behav., 18(1):17-54, 1986.

Perrier, P. Control and representations in speech. ZAS Papers in Linguistics. 40: 109-132, 2005.

Perrier, P., Payan, Y., Zandipour, M. and Perkell, J., Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study, JASA, 114(3):1582-1599, 2003.

Perrier, P., Boë, L.J., and Sock, R. Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: Modeling the transition with two sets of coefficients, JSLHR, 35:53–67, 1992.

Sondhi, M.M., Resonances of a bent vocal tract, JASA, 79(4):1113-1116, 1986.

Winkler, R., Fuchs, S., Perrier, P. and Tiede, M., Biomechanical tongue models: An approach to studying inter-speaker variability, Interspeech Florence, Italy, (submitted)

Yushkevich, P.A., Piven, J., Hazlett, H.C., Smith, R.G., Ho, S., Gee, J.C. and Gerig, G., "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability", Neuroimage, 31(3):1116-28, 2006.