Do Speakers' Vocal Tract Geometries Shape their Articulatory Vowel Space?

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

1: Center for General Linguistics (ZAS) Berlin, 2: Humboldt University (HU) Berlin, 3: ICP - GIPSA-lab, CNRS, Grenoble Institut National Polytechnique, Grenoble
E-mail: `fuchs@zas.gwz-berlin.de`

Abstract

This study investigates the relation between parameters describing differences between speaker-specific vocal tract geometries and articulatory distances between the corner vowels based on MRI data of 9 French speakers. Results provide evidence that speaker with a longer pharynx produce larger displacements between low back and high front vowels. Preliminary modeling results are also presented with the aim to study the relation between motor commands, articulation and acoustics.

1 Introduction

Speaker-specific articulatory behaviour seems to be an inherent characteristic of speech production. We are particularly interested in the question where inter-speaker variability stems from and draw our attention to the potential inter-speaker differences in vocal tract morphology. One of the difficulties in such an undertaking is to disentangle between the potential factors making one speaker distinguishable from another. Particularly, with regard to articulatory control and its variability, it is quite challenging to tease apart whether two speakers differ, since this is the intended goal (sociolinguistic and dialectal factors may play a role) or whether the differ due to differences in their vocal tract morphology.

The approach we favour to deal with such an issue is to gather experimental data for a large set of speakers and compare them with simulations carried out with realistic physical models of their vocal tract. Modelling has the advantage that different factors underlying the produced output can be controlled separately. However, it has to be considered carefully since it is based on hypotheses, estimations, theories how reality works, but it is not a copy of the reality.

2 Hypothesis and assumptions

The hypothesis and assumption we will make, are based on previous studies, which are briefly outlined below:

Honda and colleagues [2] analysed vocal tract geometries on the basis of x-ray data for 10 Japanese and 10 Caucasian American English speakers. They found a reciprocal relationship between pharyngeal distance (horizontal structure) and lower facial height (vertical structure). Speakers with longer horizontal structures tend to have smaller vertical structures and vice versa. Honda et al. also found some dependencies of vowel production on vocal tract morphology, but not on consonant production.

As a starting point we hypothesise that the shape of the vocal tract partly determines the organization of the articulatory vowel space. More specifically, speakers with a long pharynx (large vertical dimension) and a short palate (short horizontal dimension) use a larger articulatory distance between high and low vowels and a smaller articulatory distance between front and back vowels than speakers with a short pharynx. This hypothesis is based on the assumption that speakers primarily aim towards auditory goals (although the representation of speech is in general multi-modal, see [3]), i.e. they adapt their articulation to the respective vocal tract boundary in order to reach the intended goal.

3 Methods

Nine French speakers were recorded by means of Magnetic Resonance Imaging (MRI) with a Philips

Gyroscan T10-NT Powertrack 1000 scanner generating a static longitudinal magnetic field of 1.0 Tesla. An anterior neck coil was used. The repetition time was 1660 ms and the echo delay time was 9 ms. The image matrix was composed of 256x256 pixels, each with a spatial resolution of 1mm in the y-direction and 1.4mm in the x-direction. Data were originally collected for 10 isolated vowels /i e ɛ a y ø œ u o ɔ/ to study inter-speaker acoustic and articulatory variability [1]. For each vowel three 18 slice series of 3.6mm thick parallel sections with a 0.4mm distance between slices were collected. The acquisition lasted 48s for each sound. The air way was segmented from the surrounding tissues in a manual procedure. We used the itk-SNAP (version 1.4) software for segmentation. So far the corner vowels /i,a,u/ and schwa have been analysed in the 3D space.

3.1 Quantifying differences in vocal tract shapes

Quantifying the differences between individual vocal tract geometries turned out to be more complicated than we expected. The literature is full of parameters describing vocal tract growing curves from newborns to adults. However, it is sparse with respect to a parameterisation of inter-speaker differences during adulthood, going beyond the general statement that males have a longer pharynx than females.

We decided to run a Principal Component Analysis (hereafter PCA) on the dataset corresponding to the mid-sagittal outer vocal tract contour. Only the data for the neutral vowel schwa were included in order to avoid effects known for vocal tract length variation in different vowel productions. Using absolute coordinates in a PCA requires a joint spatial reference system of the speakers vocal tracts. The most consistent feature for all speakers' vocal tracts was the posterior pharyngeal wall, which could be approximated by a straight line. The origin of the new reference coordinate system was the lowest identifiable point of the pharyngeal wall. A second point was selected which occurred as high as possible along the pharyngeal wall. These two points were necessary for the translation and rotation of the 2D vocal tract contours so that the posterior pharyngeal wall always corresponded to the y-axis.

Hereafter a PCA was performed. The first factor explained 64 percent of the variance in the data and accounted for differences in pharyngeal length (see Fig.1, left). The second factor explained 29 percent of the variance in the data. It is less clear how to explain it, but we interpret it with respect to differences in palatal length and height (Fig. 1, right).

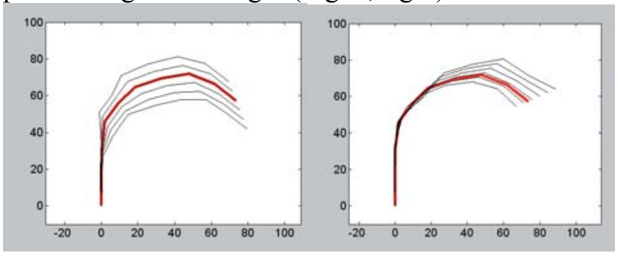

Figure 1. *Mean outer vocal tract contour (red line) varied solely by the full range of values for factor 1 (left) and solely by the full range of values for factor 2 (right).*

3.2 Quantifying the articulatory vowel space

To quantify the articulatory vowel space between the corner vowels, we adapted and modified a method proposed by [5] who defined flesh points (hereafter coils) on tongue surface configurations of MRI data (see Fig. 2).

The first coil (coil 1) was placed exactly at the tongue tip and all of the following 5 coils (coil 2-6) were placed in equal distances of 1cm (large dots in Fig. 3). Coils 4 to 6 correspond approximately to the tongue dorsum and tongue back sensors in EMA recordings.

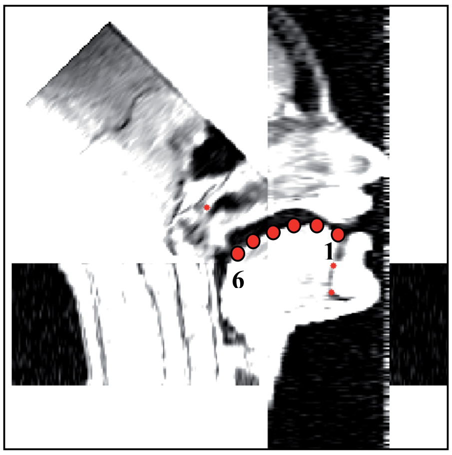

Figure 2. *Schematic view of flesh point markers (coils) defined on the tongue contour in the mid-sagittal plane (schwa, speaker fs).*

They are the most relevant in vowel production. An example of the coils' placement for the three corner vowels is given in Fig. 2.

Next, we drew triangles (corresponding to the vowel space of the corner vowels) for each coil respectively (an example is given in Fig. 3 for coil 3, black triangle in bold). Euclidian distances were calculated for the /ai/, /au/ and /ui/ distances of each coil. In order to normalise for the global differences in vocal tract size while taking into account the relation between high versus low and front versus back vowel distances, we additionally calculated the ratio (hereafter ratio) between the /au/ and /ui/ vowel distances. This ratio was in general >1, since distances between high and low vowels were larger than between front and back vowels. The higher the ratio value the larger the high-low distance relative to the front-back distance.

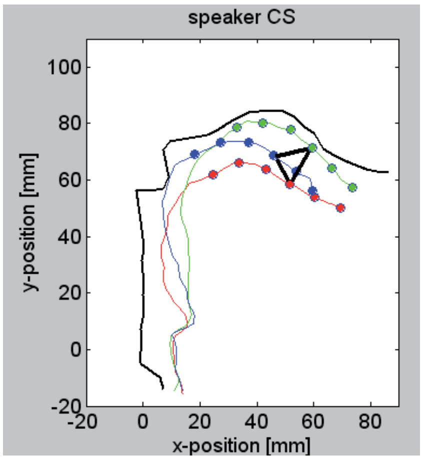

Figure 3. *Example of coil placement: green/highest tongue contour = /i/ configuration, blue/middle tongue contour = /u/ configuration, red/lowest tongue contour = /a/ configuration, in black mid-outer vocal tract contour*

4 Results: Relation between geometrical vocal tract parameters and articulatory vowel space

The following table lists only significant results for the correlations between vocal tract parameters (factor f1 and f2) and the articulatory measures (distances: ai, au, ui; ratio). The number of samples for all correlation is always 9 (= 9 speakers).

Factor 1 showed a significant correlation with the /ai/-distance, i.e. the longer the pharynx, the larger the Euclidian distance for high versus low vowels. A similar finding could not be found for the /au/ distance, which may on the one hand be explained with respect to the potential motor equivalence strategies occurring in /u/ (lip protrusion, tongue retraction or laryngeal lowering may equally well contribute to the low second formant values of /u/) but on the other hand it may also be related to the overall shape of the velo-pharyngeal part (the bend from the hard palate to the pharynx).

Table 1. *Significant correlations between factor scores (f1 and f2) and articulatory distances*

Significant correlation	Coil	R	P-value
f1 – ai	5	-0.716	0.027
	6	-0.719	0.026
f2 – au	6	-0.714	0.028
f1 – ratio	2	-0.703	0.032
	3	-0.684	0.041
f2 – ratio	4	-0.837	0.003
	5	-0.846	0.002
	6	-0.803	0.007

Factor 2 shows a correlation with the /au/-distance, i.e. the shorter the palate, the longer the distance between the low back and the back high vowels. The correlation of the factors with the ratio of the articulatory vowel distances shows in general a good agreement. Significant correlations of f1 with the ratio are found for the more anterior coils at the tongue (2, 3). Factor 2 correlates with the ratio for coils 4 to 6, the ones which are important for vowel production.

5 Building speaker specific biomechanical tongue models

In a next step we implemented the individual vocal tract configurations and the individual tongue surface contours for the neutral vowel (corresponding to the rest position in the model) in the 2D biomechanical tongue model described in [4] to build individual models. The implementation involved the following procedures:

1. A rough division of the individual outer vocal tract contour into different anatomical sections in the mid-sagittal plane (e.g. pharynx, velum, palate etc.)
2. A definition of the beginning and end of the tongue surface contour for the neutral vowel configuration and an adaptation of the associated finite element mesh
3. A calculation of the λ commands at rest (according to [4] recruitment thresholds λ; values equal the length of the muscle fibers at rest, no force

generation.) Forces are generated as soon as the tongue shape changes slightly.

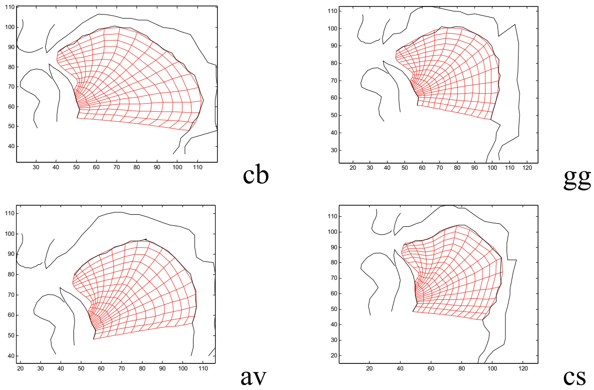
cb
gg
av
cs

Figure 4. *Individual 2D biomechanical tongue models based on the MRI data (gg, cs = long pharynx, cb, av = short pharynx)*

Four examples are shown in Figure 4. These individual models will allow us to discover the nonlinear relations between motor commands, kinematic output, and acoustics, and to account for the potential effects emerging from the supine position of the subject during the MRI recording. From the 3D MRI volume data during schwa production we also computed the acoustics (first four formants) in order to check the actual realization of the vowel corresponding to the rest position in the model.

So far we run simulations for the model cb (short pharynx) and gg (long pharynx). To do so, we used the λ commands for the rest position and the corner vowels /a, i, u/ provided in [4] as a reference and calculated the differences between λ at rest and λ at one of the corner vowels. The differences in λ rather than absolute λ commands were applied to the cb and gg model. Starting from the rest position 50 simulations were run for each of the corner vowels. λ commands were randomly chosen among all the possible combinations in the specified range of the target command.

Figure 5 displays the results of our simulations showing /a/ variations around the target position.

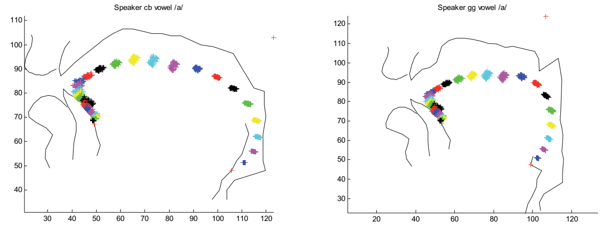

Figure 5. *Simulations of /a/ variability using a similar set of λ variation: left for cb and right for gg*

Tongue variability for /a/ looks similar for both speakers, but for cb the pharynx gets nearly closed due to the downward and backward movement of the tongue whereas for gg it gets more constricted. This is in agreement with our hypothesis that speakers with a short pharynx and a long palate have to control their vertical movements precisely since their vocal tract constraints them to move the tongue in the vertical direction. However, there is probably a trade-off between tongue and jaw movements. With an open jaw the tongue can move downwards without a pharyngeal constriction. To what extent more or less vigorous jaw movement may be related to the individual vocal tract geometry seems to be an interesting question for future studies.

Acknowledgements

We like to thank Lian Apostol for providing the MRI data and Mark Tiede for the idea running a PCA on the vocal tract data in order to parameterize the differences. This work was primarily supported by the Christian Benoit Association by means of an award dedicated to the first author. It is dedicated to Christian Benoit and Dieter Fuchs.

References

[1] Apostol, L. *Étude et simulation des caractéristiques individuelles des locuteurs par modélisation du processus de production de la parole*. Unpubl. PhD thesis at INP Grenoble, 2001.

[2] Honda, K., Maeda, S., Hashi, M., Dembowski, J.S., & Westbury, J.R. Human palate and related structures: their articulatory consequences. *Proc. ICSLP* 2: 784-787, 1996.

[3] Perrier, P. Control and representations in speech production. *ZASPiL*, 40:109-133, 2005.

[4] Perrier, P., Payan, Y., Zandipour, M., & Perkell, J. Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *JASA*, 114, 1582-1599, 2003.

[5] Whalen, D.H., Kang, A.M., Magen, H.S., Fulbright, R.K. & Gore, J.C. Predicting midsagittal pharynx shape from tongue position during vowel production. *JHSLR* 42: 592-603, 1999.