

# The relation between differences in vocal tract geometry and articulatory control strategies in the production of French vowels: Evidence from MRI and modeling

Ralf Winkler<sup>1,2\*</sup>, Susanne Fuchs<sup>1\*</sup>, Pascal Perrier<sup>3\*\*</sup>

<sup>1</sup>Zentrum für Allgemeine Sprachwissenschaft, Jägerstr. 10-11, 10117 Berlin, Germany

<sup>2</sup>KGW, Technische Universität, Ernst-Reuther-Platz 7, 10587 Berlin, Germany

<sup>3</sup>Institut de la Communication Parlée, INPG UMR 509, & Université Stendhal, Grenoble 46 Avenue Félix Viallet, 38031 Grenoble, Cedex 1, France

[winkler@kgw.tu-berlin.de](mailto:winkler@kgw.tu-berlin.de), [fuchs@zas.gwz-berlin.de](mailto:fuchs@zas.gwz-berlin.de), [perrier@icp.inpg.fr](mailto:perrier@icp.inpg.fr)

*This study investigates the relations between different vocal tract shapes and articulatory control strategies in vowel production by means of experimental data from MRI and simulations with a model. We particularly focused on the pharynx. Nine French native speakers were recorded with MRI. We found that speakers with a wider pharynx always show a larger maximal sagittal length of the area slices, but not necessarily a larger maximal axial length. In addition, we assumed that subjects with different vocal tract proportions in the vertical and horizontal direction also need to take these proportions into account in their vowel articulation. Our results confirm this assumption. They are discussed with respect to individual motor control strategies.*

## 1. Introduction

The overall aim of our work is to investigate the relations between individual morphological differences of the vocal tract and the corresponding articulatory control strategies. We assume that the representations of speech production are multi-modal (articulatory and acoustic), but that a hierarchy exists among them. The acoustic modality has the highest priority (Perrier 2005). Thus, under normal and non-perturbed conditions, speech is planned towards acoustic goals. Our previous work focused on individual palate shapes in the coronal plane, i.e. on dome shaped versus flat palates (Brunner et al. 2005). We hypothesized that subjects with a flat palate need to control their tongue position more precisely since small articulatory changes have a larger impact on the area function and on the acoustics in comparison to speakers with a dome-shaped palate. We tested this hypothesis by means of experimental data and simulations with a biomechanical tongue model. The experimental data provided evidence that speakers with a flat palate exhibit a small token-to-token variability,

---

\*Supported by the German Research Council DFG, GWZ 4/8-1 P1.

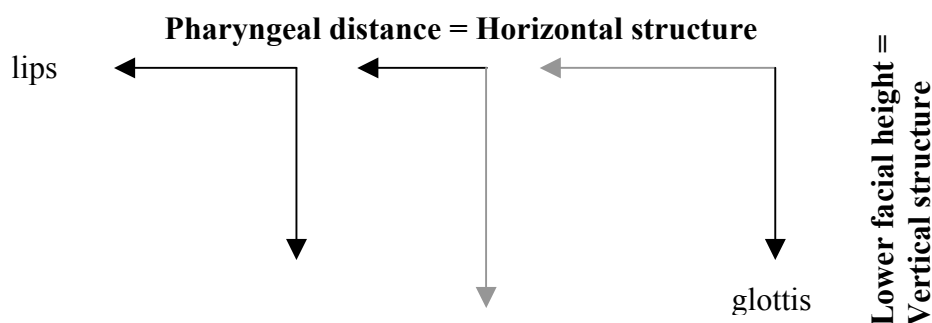
\*\*Supported by CNRS UMR 5009 & POPAART within the P2R framework.

whereas speakers with a dome-shaped palate varied in their articulatory precision. Some of them had a larger articulatory variability and others a smaller one. Based on these observations we proposed that speakers take into account their palate shape to set up their articulatory strategies, in particular the speakers with the flat palates. We tested the actual effects of the coronal palate shape on the acoustic variability by means of a model. Different palates were implemented in the model and acoustic variability was studied, holding articulatory variability constant for all the palates. The simulations confirmed our hypothesis. Results for simulations with a flat coronal palate shape in place showed larger variations in the area functions and in the first formants than the ones with dome shaped palates.

Now, we are further interested in the potential role of the vocal tract shape with particular focus on the pharynx. Differences in vocal tract length and size have been reported by Fitch and Giedd (1999) on the basis of MRI-data of 129 participants between 2 and 25 years of age. A greater increase in pharynx length in comparison to other parts of the vocal tract has been found for males between puberty and adulthood. This increased pharyngeal length is the most pronounced difference in comparison to female adults. In general, the authors also report a strong positive correlation between vocal tract length and body size, regardless of whether body height or weight was taken into account for the latter.

In order to characterize different vocal tract shapes, Honda et al. (1996) stamped the terms ‘pharyngeal distance’ and ‘lower facial height’ corresponding to landmarks of the horizontal and vertical structures of the vocal tract. A more simplified version with 2 parameters is displayed in Figure 1. Based on x-ray data of 10 Japanese and 10 Caucasian American English speakers Honda et al. found a reciprocal relationship between pharyngeal distance (horizontal structure) and lower facial height (vertical structure). Speakers with longer horizontal structures tended to have smaller vertical structures and vice versa. They also found some dependencies of vocal tract morphology on vowel production. Speakers with a short pharyngeal distance, a characteristic of males (e.g. figure 1 middle), showed a larger amplitude in their vertical tongue movements.

If the tongue position in the vocal tracts schematically depicted in figure 1 were kept constant, acoustics would differ, since the relations between horizontal structures and vertical structures affect the length and the proportion of the front and back cavities.



**Figure 1:** Simplified schematic view of different relations in lower facial height (vertical structures) and pharyngeal distance (horizontal structures).

In opposition, if the acoustics were kept constant, we suppose more vertical tongue displacement in speakers with a small pharyngeal distance and more horizontal tongue displacement in speakers with a large pharyngeal distance. Thus, to realize a similar acoustic output, speakers with different vocal tracts should show differences in tongue positioning. In addition, it is likely that they use different articulatory control strategies and muscle groups to realize the intended acoustic vowel targets. In this context, Honda et al. (1996) pointed to the two roles of the form of the speech organs, one in shaping the acoustic output and another, in the development of neural functions for speech production. We are particularly interested in the pharynx since its importance for vowel production has been recognized for a long time (Whalen et al. 1999). However, experimental evidence is often limited due to the difficulty of obtaining data from this region. Therefore, Whalen et al. (1999) tested data of 2 speakers by means of MRI to find out whether midsagittal fleshpoint data received from techniques as EMA could be sufficient to predict midsagittal pharynx shapes. They found a high predictability except from the extreme ends (larynx and lips) and the uvula. Whalen et al.'s study was however limited to the midsagittal plane and area functions (determining the acoustics) take the axial plane into account.

The aims of this study are: (1) to investigate the relation between vocal tract morphology, articulatory control strategies and acoustics in general, with particular focus on the pharynx, (2) to describe the relations between the maximal sagittal and axial length and their corresponding areas in the pharynx, (3) to find the region of greatest inter-speaker-variability, (4) to implement the most extreme vocal tract and tongue contours from the experimental data in a 2 D biomechanical tongue model, and to study the potentially underlying control strategies of the tongue assuming the same acoustic goals.

## **2. Methods**

### **2.1. Data acquisition**

The MR images were acquired using 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 256 x 256, and the spatial definition of each image is 1 mm in the y-direction and 1.4 mm in the x-direction. Data were originally collected for 10 isolated vowels /i e ε a y ø œ u o ɔ/ to study inter-speaker acoustic and articulatory variability (Apostol 2001). For each vowel three 18 slice series of 3.6 mm thick parallel sections were gathered. The interval between slices in each series was 4 mm. In the first series, starting from the front of the lips to the soft palate, the orientation of the section was vertical (coronal sections). In the second series, starting from the soft palate and ending at the top of the pharynx, the orientation of the sections was 45 degrees from horizontal. In the third series, starting from the top of the pharynx and ending below the glottis, the orientation of the sections was horizontal (axial sections). The three series overlapped in order to insure a scan of the whole vocal tract. The amount of time required to image a whole vocal tract configuration was 43s. Since sustaining the phonation during such a long time interval could be impossible for an untrained speaker, participants had the choice either to sustain the phonation or to start with phonation and maintaining the articulatory configuration. Here we will focus on the /œ/ since it has been associated

with a neutral vowel in French, similarly to schwa in other languages, and the corner vowels /a, i, u/. Nine subjects have been recorded for this study. Their gender, age weight and height at the time of the recording is displayed in table 1.

**Table 1:** Participants of this study

<b>Speaker</b>	<b>Gender</b>	<b>Age</b>	<b>Height in m</b>	<b>Weight in kg</b>
AV	Female	23	1.65	52
CB	Female	31	1.70	60
FS	Female	28	1.79	60
GG	Female	24	1.65	53
<b>Mean</b>		<b>26.5</b>	<b>1.7</b>	<b>56.3</b>
CS	Male	30	1.82	75
DB	Male	32	1.73	75
JLS	Male	40	1.80	65
LR	Male	25	1.85	65
TG	Male	21	1.75	70
<b>Mean</b>		<b>29.6</b>	<b>1.79</b>	<b>70</b>

## 2.2. Airway segmentation

The airway was segmented from the surrounding tissues by a manual procedure regardless of the position and orientation of the single slices. We used the itk-SNAP (version 1.4) software for segmentation. During segmentation of the slices, three already known problems were faced consistently. Teeth are invisible in the MRI volumes, the end of the airway tube at the lips still lacks a proper definition, and the impact of the epiglottis as well as the piriform sinuses is still unclear.

The first two problems do not affect the vocal tract's region of interest because this study deals with pharyngeal morphology. For this reason, our reconstructed vocal tracts do not account for the teeth at the moment. To define the end of the airway tube at the lips, we simply segmented until the most front slice, where the lips delimited a closed contour.

The presence of the epiglottis in the airway was taken into account as follows: if the epiglottis is not unambiguously separable from the tongue body, and hence there is a closed contour between the epiglottis and the pharynx, this contour determines the cross-sectional area. In contrast, if the epiglottis is located in the middle of the airway, the tissue of the epiglottis was ignored and the contour of the tongue body and the pharyngeal walls determine the cross-sectional area. Because the separation between epiglottis and tongue body varies individually and in different vowel contexts, some reconstructed vocal tracts include the epiglottis, others do not account for that. The separation was decided on a case by case basis for every speaker and every vowel articulation, so that every reconstructed tract consequently either includes the whole epiglottis or not.

## 2.3. Computation of the area values

The contour coordinates obtained from the airway segmentation were then processed in a Matlab environment along the lines of Apostol (2001). The single contours were combined to form a vocal tract in the following way: in a first step every contour was subsampled with 100 equally spaced points, then the three series of 18 slices were

correctly positioned according to their original spatial orientation, and finally 100 longitudinal fibers were defined to connect the contours from the glottis to the lips by joining the points having the same index on each contour.

Five cross-sectional areas per speaker were calculated by intersecting each 3D volume with planes defined by a simple grid (see Figure 3, right). Two points, one on the upper half of the posterior pharyngeal wall (P1 in Figure 3) and one in the lower half (P2) define the line (hereafter P1-P2 line) to approximate the orientation of the posterior pharyngeal wall. A third marker P3 is placed below P2, at the point where the sagittal width the pharynx starts to decrease. The line intersecting P3 is perpendicular to the P1-P2 line and shows the lowest cutting plane in our data. The line passing P4 (placed below the lowest part of the velum) and perpendicular to the P1-P2 line is the highest cutting plane. Additionally, we defined 3 intermediate cutting planes, parallel and equally spaced between the lowest and the highest cutting planes. With this procedure, we also normalize for individual differences in pharyngeal size. All planes of the grid and their orientations were defined to ensure their best perpendicularity to the vocal tract midline.

### 3. Results

In this paper, we will firstly present the MRI-data from all subject's /œ/-production, and secondly from 2 speaker's /a, i, u/ realization.

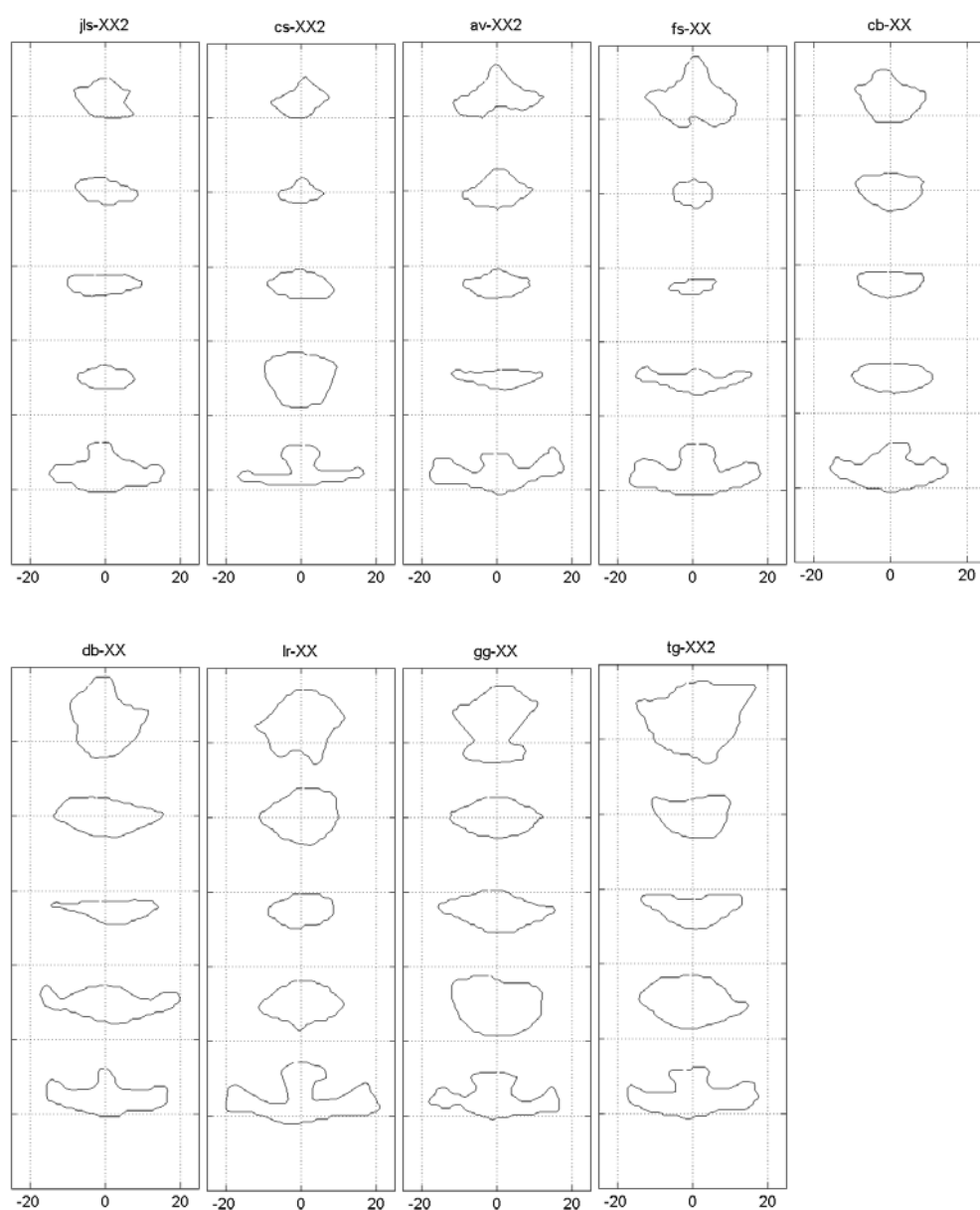
#### 3.1. Results for speaker's /œ/ production

In order to describe the relations between the maximal sagittal and axial length of the slices and their corresponding area function in the pharynx, we calculated means, standard deviations and the variability coefficient ( $\text{std} \cdot 100 / \text{mean}$ ; which normalizes the standard deviation with respect to the mean) for the 5 cutting planes. Results are ordered according to the size of the area in Table 2.

**Table 2:** Means, standard deviations & variability coefficient of the mean of areas, maximal sagittal and axial length; ordered by subject & area

Subject	Means			Standard deviation (Variability coefficient in %)		
	Area in cm <sup>2</sup>	Max. sagittal length in mm	Max. axial length in mm	Area in cm <sup>2</sup>	Max. sagittal length in mm	Max. axial length in mm
jls	1.163	8.59	19.79	0.566 (49)	2.84 (33)	5.76 (29)
cs	1.222	10.27	19.69	0.578 (47)	2.80 (27)	7.44 (38)
av	1.354	10.23	24.09	0.535 (40)	3.23 (32)	6.41 (27)
fs	1.471	10.50	22.70	0.979 (67)	5.28 (50)	9.71 (43)
cb	1.496	10.50	21.14	0.454 (30)	2.79 (27)	4.96 (23)
db	2.097	12.40	29.64	0.552 (26)	4.97 (40)	5.30 (18)
lr	2.269	14.93	25.27	0.686 (30)	3.45 (23)	8.21 (32)
gg	2.457	14.29	27.52	0.480 (20)	3.64 (25)	4.57 (17)
tg	▼ 2.631	▼ 14.10	28.81	0.966 (37)	4.35 (31)	4.75 (16)

As can be seen in Table 2 speakers with a larger area do consistently show larger means of the sagittal length, but not necessarily of the axial length (e.g. av, db). This means that if a sagittal view of the pharynx is given, we are not always able to predict the axial length. A partial correlation with area and sagittal length (axial length as a covariate) reached significance ( $R=0.886$ ,  $P=0.003$ ) whereas a partial correlation with area and axial length (sagittal length as a covariate) did not ( $R=0.7$ ,  $P=0.053$ ). Another interesting finding is that speakers with a larger mean area show a trend towards a smaller variability coefficient, i.e. the 5 areas are more similar than for subjects producing the vowel on average with a smaller pharyngeal area. All speaker's areas are depicted in Figure 2.

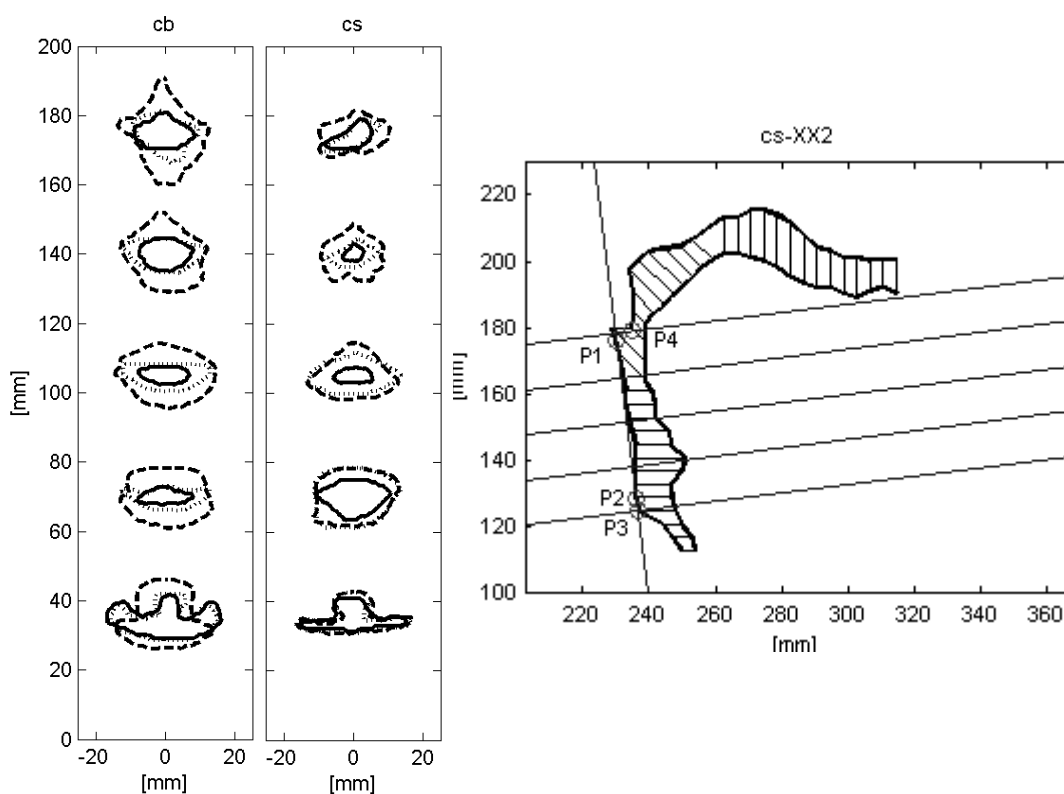


**Figure 2:** 5 cutting planes of the pharynx during /œ/ production from the lowest point of the pharynx (1st and lowest plane) to the highest (5<sup>th</sup> and highest plane) ordered by area; x = axial view, y = sagittal view

In Figure 2, it can also be seen that the sagittal and axial length in the first cutting plane do not differ to a large extent between subjects. Individuals differ particularly from the 2<sup>nd</sup> to the 5<sup>th</sup> plane and may be grouped in those with a relatively large sagittal length (CS, DB, LR, GG, TG) and those with a small sagittal length (JLS, AV, FS, CB) focusing on the 2<sup>nd</sup> and 3<sup>rd</sup> cutting plane.

### 3.2. Results for CB's and CS's /a, i, u/ production

Although we have already labelled the data for all subjects, we will here concentrate on 2 speakers, CS and CB who show differences in the proportions of their vocal tracts. CS displays a structure that vertically longer rather than horizontally, whereas for CB the opposite is the case (see also Figure 1 for a schema). So far we haven't found a reliable way to quantify these relations since individuals also have differences in the curvature of the vocal tract. Hence, this description is based on visual inspection.



**Figure 3:** 5 cutting planes for /a, i, u/-production for CB and CS (left); /a/ = solid line, /i/ = dashed line, /u/ = dotted line; right figure = grid showing the landmarks and calculation of the cross-sectional area in the pharyngeal region; long lines show = cutting planes of the grid, short lines connecting inner and outer contour show the original orientation of the image slices.

In CB's data (slice 2, 3, 4), /a/ (solid line) and /u/ (dotted line) are more similar in their pharyngeal area with some differences in the axial dimension. Data for these two vowels clearly differ from /i/ data (dashed line). For CS /i/ and /u/ data are more similar in comparison to /a/ (in this speaker the epiglottis is included in /a/ and /u/). For all

speakers the pharyngeal area is comparable in /i/. The different results are interpreted with respect to the extreme vocal tract proportions of these subjects. Since CB has a vocal tract with a longer horizontal structure than a vertical one, more horizontal tongue movements were assumed which also affect the pharynx. Results in Figure 3 show more differences in the area between the front and the back vowels, but fewer between high and low vowels. Hence, this subject can potentially show more variability in tongue movement in the horizontal direction, but is limited in its degrees of freedom for vertical movements. For CS the opposite can be assumed. CS has a vocal tract with vertical structures that are larger than the horizontal ones, and more tongue movement affecting the pharynx shape were assumed in the vertical direction. Results support this assumption, since the pharynx shape differs more between high and low vowels than between front and back vowels. Hence, is limited in its degrees of freedom for horizontal movements.

In a next step we will run simulations with a biomechanical tongue model to provide further evidence of the potential underlying control strategies used for the different vocal tracts. We have already implemented vocal tract shapes in the model and have adapted the different tongue contours. In a next step, we will use additional x-ray data available for some of the speakers to implement the jaw. Results of the simulations will be presented at the conference.

#### **4. Acknowledgements**

We like to thank Lian Apostol for making available the MRI data that he collected during his PhD thesis. This work is dedicated to Dieter Fuchs.

#### **References**

- 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.
- Brunner, J., Fuchs, S. & Perrier, P. The influence of palate shape on token-to-token variability. *ZASPiL* 42: 43-67, 2005.
- Fitch, W.T., Giedd, J. Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *JASA* 106(3): 1511-1522, 1999.
- 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.
- Magen, H.S., Kang, A.M., Tiede, M.K. & Whalen, D.H. Posterior pharyngeal wall position in the production of speech. *JSLHR* 46: 241-251, 2003.
- Perrier, P. Control and representations in speech production. *ZASPiL*, 40:109-133, 2005.
- 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. *Journal of Speech, Language, and Hearing Research* 42: 592-603, 1999.