The potential role of speakers' vocal tract morphology onto speech perception: preliminary results

Susanne Fuchs[1], Jana Brunner[1,3], Pascal Perrier[2], Christian Geng[1], Bernd Pompino-Marschall[1,3]

1: Zentrum für Allgemeine Sprachwissenschaft (ZAS) Berlin, Germany
2: Institut de la Communication Parlée (ICP), INPG & CNRS, Grenoble, France
3: Humboldt University Berlin, Germany
fuchs@zas.gwz-berlin.de

Abstract

In order to investigate the influence of vocal tract morphology, i.e. different palate shapes, on the acoustic output and its perceptual consequences, we simulated /kVk/-sequences by means of a 2D biomechanical tongue model. Three palate shapes (normal, flat and dome shaped) differing in the coronal and mid-sagittal plane were implemented in the model. We assumed that small articulatory changes have a large impact on the area function and hence on the acoustics for subjects with a flat palate shape in comparison to normal or dome shaped palates. The results for the variability of the estimated area functions at the minimal constriction location and for the formant patterns clearly show vowel specific effects with respect to the three different palate shapes. KVk-sequences with the high front vowel /i/ were influenced to a greater extent than /kak/ and /kuk/-sequences.

1. Introduction

By means of an experiment investigating tongue movements and a perception test Perkell et al. (2004) confirmed for different vowel pairs the hypothesis that speakers with relatively sensitive perceptual capabilities produce sounds more distinctively. Their results exemplify how speech production and speech perception may interact and how they contribute to shape each other. In the same vein, our work addresses the issue whether morphological characteristics of the vocal tract can facilitate the emergence of speaker-specific articulatory strategies of speech production, which can in turn influence perceptual patterns and their discrimination. In a previous work on token-to-token variability during vowel production in German [5], we found speaker dependent differences, which were associated with the individual morphology of the palate shape. For one subject having a flat coronal palate shape, tongue sensors located in the mid-sagittal plane exhibited very small, nearly circular dispersion ellipses, whereas two speakers with dome shaped palates showed distinctively larger variability patterns. In accordance with Perkell (1999) we interpreted these speaker dependent results with respect to perceptual requirements: Small articulatory changes have a large impact on the area function and hence on the acoustics for subjects with a flat palate shape. Therefore these speakers should control their tongue positioning more precisely in order to avoid confusion with neighboring vowels. For speakers with a dome shaped palate a similar precision is not required. The possible role of the palate in the emergence of articulatory strategies has already been suggested by Johnson et al. (1993) who hypothesized on the basis of a vowel production study that palatal doming has an effect on articulatory organization, i.e. for speakers with a deeply vaulted palate it is necessary to move the tongue to a greater extent in order to reach the appropriate acoustic goal in comparison to speakers with a shallower palate shape. Additionally, they suggest that speakers with a flat palate move tongue and jaw more synchronously in comparison to speakers with a deeply vaulted palate who move their tongue more independently.

The relation between vocal tract morphology, its corresponding production and perceptual consequences could also play a role during speech acquisition. Mackenzie Beck (1999) reported how palatal dimensions change for males and females from the age of 6 to adulthood. During ontogenesis palates change their dimensions, especially with respect to the growing of the vault. Infants up to the age of 10 have a flat palate which becomes more and more vaulted above this age. The implications of this for characteristics of child speech and the development of speech motor control are not yet clear.

The current study aims at assessing the relation between vocal tract morphology, i.e. different palate shapes as they occur speaker dependently, and their consequences for the variability in production and perception. In order to do so, the following steps were carried out:

1. By means of a 2D biomechanical tongue model we simulated /kVk/-sequences with 3 different palate shapes (normal, flat, dome shaped).
2. We estimated for each tongue position and for three different palates the corresponding vocal tract area functions at the middle of the vowel from 3D geometries (for calculation, see [8]).
3. From the area functions we calculated formant values.
4. Using a synthesizer we transformed the area function into audio signals.
5. We will carry out a perception experiment to test the hypothesis that vowels synthesized from different palate shapes are perceptually distinguishable.

In this paper the first three steps are described. The results of the perception experiment will be presented at the conference.

2. Simulations by means of a 2D biomechanical tongue model

2.1. The biomechanical model of the tongue

The biomechanical tongue model describes elastic properties of tongue tissues as well as static rigid structures such as the jaw and the hyoid bone in the mid-sagittal plane. Elastic properties of the tongue are accounted for by finite element (FE)

modeling. Muscles are modeled as force generators that (1) act on anatomically specified sets of nodes of the FE structure, and (2) modify the stiffness of specific elements of the model to account for muscle contractions within tongue tissues. Seven tongue muscles are included in the model: the Genioglossus posterior (Ggp), the Genioglossus anterior (Gga), the Styloglossus (Sg), the Verticalis (V), the Hyoglossus (Hg), the Inferior longitudinalis (Il), and the Superior longitudinalis (Sl). Limits of the vocal tracts in the mid-sagittal plane such as the palate and the pharynx are added based on one speaker's x-ray data.
The model is controlled according to the λ model [2] which specifies for each muscle a threshold length λ. If the muscle length is larger than λ, muscle force increases exponentially with the difference between the two lengths. Otherwise there is no active force. Hence, muscle forces are typically non linear functions of muscle lengths. The control space is called the λ space.

2.2. Palate shapes

From the 2D sagittal sections provided by the 2D biomechanical model of the tongue, area functions are estimated in the 3D space and hence, palate shapes were varied not only in the sagittal plane, but also in the coronal plane. We first changed palate shapes in the coronal plane by means of parameter C variations [8]. The greater C, the flatter the palate shape. The smaller C, the steeper the palatal vault (see Figure 1a,b).

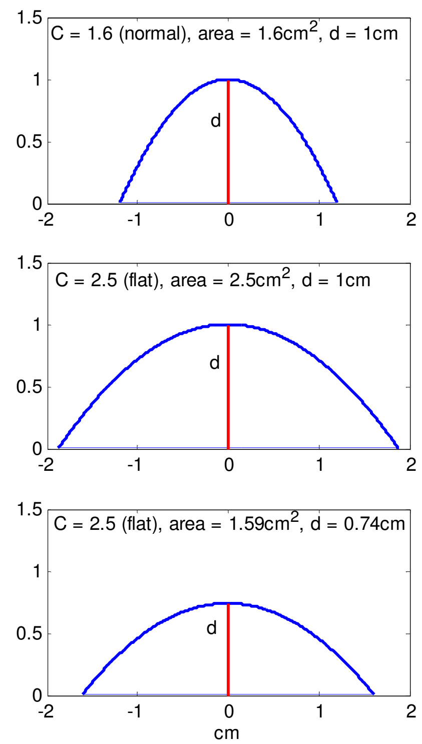

Figure 1:1a-c = from top to bottom; 1a: coronal view on the normal palate(C=1.6), 1b: flat palate (C=2.5), 1c: flat palate with an adapted palatal height (d), for explanation see text

Three different palate shapes were considered with C values of 1.3, corresponding to a realistic dome shaped palate, with C = 1.6, corresponding to the default value of the model (called hereafter normal palate shape), and a C value of 2.5, corresponding to a realistic flat palate. First, we defined one reference vowel for each /i/, /a/, and /u/ and calculated its area function (for further details see section 3). This resulted for instance, for the same sagittal distance of 10mm, in an area function difference of 0.9cm^2 between the normal and the flat palate (see Figure 1a: normal palate with an area function of 1.6cm^2; 1b: area function of 2.5cm^2 for the flat palate). These differences had already a large impact on the formant patterns of the reference vowels. Therefore, we also adapted the palatal shape in the sagittal plane (corresponding to d = palatal height in Figure 1) in such a way that the area functions for all the three palate shapes became more or less similar (Figure 1c for an adaptation of the area function of the flat palate to the normal palate). Resulting changes in the sagittal plane, i.e. a lower palate position for the flat and a higher palate position for the dome shaped palate, were further implemented in the model. Since these changes of the palatal shape in the mid-sagittal plane caused a step at the transition towards the velar region, we interpolated the values at the end of the palate with the default values for the velar region in order to get a smooth curvature.

2.3. Simulations

Simulations were run for /kVk/-sequences. The surrounding plosive context was chosen in order to model more realistic sequences than pure single vowels. The velar targets for /kak/ and /kuk/-sequences are similar whereas a fronted k-target was chosen for /kik/-sequences which is in general agreement with production studies. Lambda values for the reference vowels (ref.) were selected following Perrier et al. (2003). Changes in the λ commands (var.) for /a, i, u/ were chosen which are likely to give token-to-token variability typical of normal speech. Table 1 shows the relevant commands. Only those muscles are included in the table where lambda commands were varied.

Table 1: Lambda values for reference vowels (ref.) and vowel variations used in the simulations

	Lambda	Lamda	Lambda	n
/a/ ref. /a/ var.	GGA: 38 35-41	HG:38 35-41	GGP:63 60-63	 196
/i/ ref. /i/ var.	SG:88 85-91	IL:60 60-63	GGP:44 43-46	 112
/u/ ref. /u/ var.	SG:74 80:84	HG:66 63-69	GGP:50 47-53	 245
			for n*3 palates ∑ 1659	

3. Estimation of area functions

3.1. Estimating area functions from a 2D model

For the generation of area functions from mid-sagittal data (see Figure 2) we adopted Perrier et al.'s (1992) procedure. First, the sagittal distances (d's) were computed along the segments of a semi-polar grid (see Figure 2). On the basis of geometrical measurements of the vocal tract extracted from a cast of a cadaver vocal tract and from x-ray CT-Scan data on a human speaker, the authors suggested to divide the vocal tract into 5 different regions from the glottis to the lips. According to the vocal tract morphology in each of these regions, they suggested to use two sets of 5 parameters Ci,j, one for small sagittal dimensions and the other for large sagittal dimensions, and to compute the cross-sectional area A(x) at a distance x from the glottis by applying the following equation:

$$A = C_{i,j} * d^{1.5} \quad (1)$$

In the current study, still in the vein of Perrier et al.'s (1992) suggestion the palatal shapes and the tongue surface were modelled as parabolic curves (see Figure 1 for different parabolic C values of the palatal shape). In order to compare the area functions of the different palate shapes we changed C values in the interval between the alveolars and the end of the hard palate with everything else being equal.

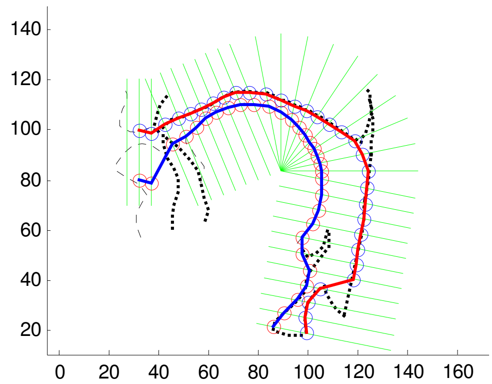

Figure 2: Example of the mid-sagittal grid in /kik/ with a dome shaped palate

3.2. Results

Table 2 exhibits the means of area functions for the relevant vowels at the appropriate constriction section with their corresponding standard deviation in brackets. For the low vowel /a/ differences in palate shape revealed no significant effects on the area functions. Figure 3a-c shows the variation of the area functions for /kik/-simulations with a flat palate (3a), a normal palate (3b) and a dome shaped palate (3c). The relevant area for the /i/-constriction is located between 10-15cm from the glottis. The area functions for /i/ in /kik/-sequences clearly exhibit the predicted influences of vocal tract morphology on the area with a greater variability for the flat palate in comparison to the dome shaped and the normal palate. The latter two are quite similar.
Comparing the standard deviations of the area functions at the constriction location in the /kuk/-sequences for the different palate shapes did not show an effect.

Table 2: Means of area function values in cm^2 at the minimal constriction section, standard deviations in brackets

V.	Area Flat Pal.	Area Normal Pal.	Area Dome Pal.
/a/	0,927 (0,011)	0,943 (0,011)	0,943 (0,011)
/i/	0,363 (0,015)	0,355 (0,012)	0,35 (0,01)
/u/	0,422 (0,06)	0,423 (0,06)	0,418 (0,06)

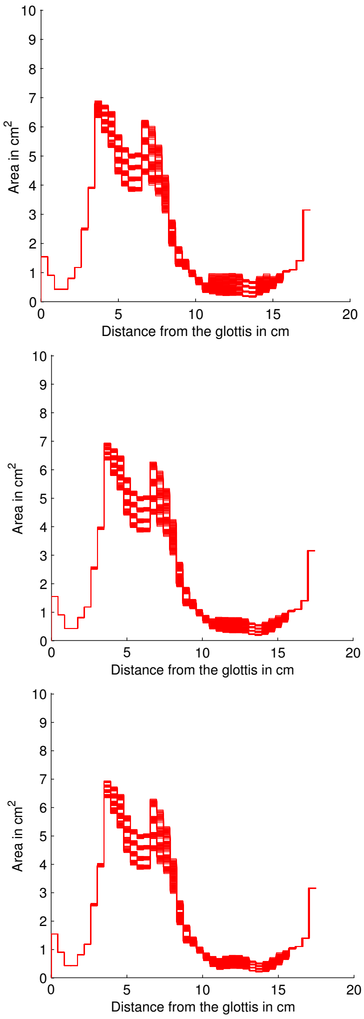

Figure 3: Area functions for /kik/;3a (top): flat palate, 3b (middle): normal palate,3c (bottom): dome shaped palate; relevant area for /i/ constriction around 10-15cm from the glottis

4. Formants

The coupling of the tongue model to a harmonic model of the vocal tract enables formant frequencies to be obtained from the articulatory commands. To do so, once the area function is calculated, the corresponding transfer function can be estimated with a harmonic acoustic model of vowel production

(Badin and Fant, 1984), taking into account losses by wall vibration, heat and viscosity, and a low frequency lip radiation model. The pole frequencies of the transfer function provide formant frequencies.

Table 3: Means of formant values with standard deviations in brackets

V.	Formants Flat Pal.	Formants Normal Pal.	Formants Dome Pal.
/a/			
F1	582 (9)	574 (9)	560 (10)
F2	1397 (36)	1321 (29)	1253 (23)
F3	2637 (32)	2682 (28)	2699 (19)
/i/			
F1	334 (34)	331 (28)	331 (25)
F2	2242 (129)	2258 (95)	2255 (86)
F3	2718 (124)	2734 (93)	2716 (73)
/u/			
F1	327 (6)	232 (5)	322 (5)
F2	1267 (71)	1306 (67)	1312 (65)
F3	2529 (28)	2524 (25)	2521 (24)

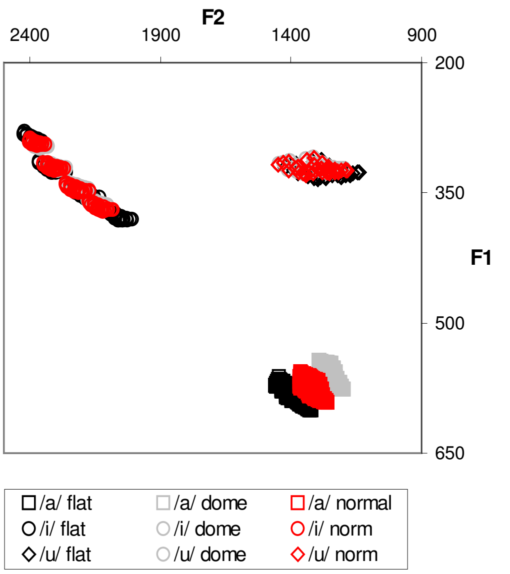

Figure 4: Formant space for the simulated vowels and different palatal shapes, x-axis: F2 in Hz, y-axis: F1 in Hz

In Table 3 and Figure 4 formant values are presented. Comparing the variance of the formant values with respect to the different palate shapes, the most pronounced results were obtained for /kik/-sequences, where simulations for the flat palate double the variance in F2 compared to normal (F2 was affected to a greater extent than F1). In /kak/ and /kuk/-sequences mean values shifted from a lower to a higher F2 in /kak/, and from a lower to a higher F1 in /kuk/-sequences comparing the normal with flat palate. However, the various palate shapes did not influence the variance of the formant patterns as much as the high front vowel.

5. Discussion

We investigated the relationship between differences in palate morphology and its consequences on the variability of the corresponding area functions and formant patterns by means of simulations of /kVk/-sequences with a 2D biomechanical tongue model. The results for the variability of the estimated area functions at the minimal constriction location and for the formant patterns clearly show vowel specific effects with respect to the three different palate shapes. KVk-sequences with the high front vowel /i/ were influenced to a greater extent than /kak/ and /kuk/-sequences, i.e. the variability of the area functions and the formant patterns was greater for simulations with a flat palate in comparison to the normal and the dome shaped palates. We will further assess the perceptual consequences of the calculated variability.

We suggest that during speech acquisition infants are not capable of the articulatory precision required for their vocal tract morphology (flat palate) and are therefore more variable in their speech production.

6. Acknowledgements

This work was supported by a grant from the German Research Council (DFG) GWZ 4/8-1, P.1 and is a cooperation of the French-German network, supported by the P2R-program.

7. References

[1] Badin, P. and Fant, G. "Notes on vocal tract computations" *STL QPSR, 2-3: 53-108, 1984.*

[2] Feldman, A.G., "Once More on the Equilibrium-Point Hypothesis (λ Model) for Motor Control", *J. of Mot. Behav., Vol. 18 (1): 17-54, 1986.*

[3] Johnson, K., Ladefoged, P., and Lindau, M., "Individual differences in vowel production", *J. Acoust. Soc. Amer., Vol. 94(2), 701-714, 1993.*

[4] Mackenzie Beck, J., "Organic variation of the vocal apparatus" In Hardcastle, W.J. and Laver, J. (eds.), *The Handbook of Phonetic Sciences*, Blackwell Publishers, Oxford: 257-297, 1999.

[5] Mooshammer, C., Perrier, P., Fuchs, S., Geng, C., and Pape, D., "An EMMA and EPG study on token-to-token variability", *AIPUK, Vol. 36: 46-63, 2004.*

[6] Perkell, J.S., "Articulatory processes" In Hardcastle, W.J. and Laver, J. (eds.), *The Handbook of Phonetic Sciences*, Blackwell Publishers, Oxford: 333-370.

[7] Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M., and Zandipour, M., "The distinctness of speakers' productions of vowel contrasts", *J. Acoust. Soc. Amer., Vol. 116(4): 2338-2344, 2004.*

[8] Perrier, P., Boe, 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", *JSHR Vol. 35: 53-67,1992.*

[9] Perrier, P., Payan, Y., Zandipour, M, and Perkell, J. S. Influences of tongue biomechanics on velar stop consonants: A modeling study. *J. Acoust. Soc. Amer., Vol. 114(3), 1582-1599, 2003.*