The influence of the palate shape on articulatory token-to-token variability

Jana Brunner

Humboldt-Universität zu Berlin & Zentrum für Allgemeine Sprach-wissenschaft, Berlin

Susanne Fuchs

Zentrum für Allgemeine Sprachwissenschaft, Berlin & Institut de la Communication Parlée, Grenoble

Pascal Perrier

Institut de la Communication Parlée, CNRS, INPG & Univ. Stendhal, Grenoble

Articulatory token-to-token variability not only depends on linguistic aspects like the phoneme inventory of a given language but also on speaker specific morphological and motor constraints. As has been noted previously (Perkell (1997), Mooshammer et al. (2004)) , speakers with coronally high "domeshaped" palates exhibit more articulatory variability than speakers with coronally low "flat" palates. One explanation for that is based on perception oriented control by the speaker. The influence of articulatory variation on the cross sectional area and consequently on the acoustics should be greater for flat palates than for domeshaped ones. This should force speakers with flat palates to place their tongue very precisely whereas speakers with domeshaped palates might tolerate a greater variability. A second explanation could be a greater amount of lateral linguo-palatal contact for flat palates holding the tongue in position. In this study both hypotheses were tested.

In order to investigate the influence of the palate shape on the variability of the acoustic output a modelling study was carried out. Parallely, an EPG experiment was conducted in order to investigate the relationship between palate shape, articulatory variability and linguo-palatal contact.

Results from the modelling study suggest that the acoustic variability resulting from a certain amount of articulatory variability is higher for flat palates than for domeshaped ones. Results from the EPG experiment with 20 speakers show that (1.) speakers with a flat palate exhibit a very low articulatory variability whereas speakers with a domeshaped palate vary, (2.) there is less articulatory variability if there is lots of linguo-palatal contact and (3.) there is no relationship between the amount of lateral linguo-palatal contact and palate shape. The results suggest that there is a relationship between token-to-token variability and palate shape, however, it is not

that the two parameters correlate, but that speakers with a flat palate always have a low variability because of constraints of the variability range of the acoustic output whereas speakers with a domeshaped palate may choose the degree of variability. Since linguo-palatal contact and variability correlate it is assumed that linguo-palatal contact is a means for reducing the articulatory variability.

1 Introduction

Intraspeaker variability is an inherent part of natural speech. In speech technology it has often been encountered as a problem, e.g. in automatic speech recognition (e.g. Ainsworth (1997)) or in speaker identification or verification (Nolan (1997)). Looking at it from a listener's point of view, however, this is different. A certain degree of intraspeaker variability is accepted and often not even noticed. Small token-to-token variability for example in formant values of a vowel due to a difference in tongue position is perfectly acceptable. However, the acceptability of variability is restricted by communciational needs.

One basis for these restrictions could be the phoneme boundaries of a language. It has been shown that there is a relation between the size of the phoneme inventory and the variability of the sounds. For some Australian languages with very small vowel inventories, for example, it has been found that the allophonic variation is huge (Dixon (1980): 130). This would mean that not not the same token-to-token variability is accepted in every language. A language in which /s/ and /ʃ/ are two different phonemes requires the speaker to produce less variability for each sound than a language where the contrast is only allophonic. If a speaker of a language where the two sounds form only one phoneme learns a language where both sounds are phonemes he or she has to reduce his or her token-to-token variability in order to distinguish between the two phonemes.

On the other hand, the relation between allophonic variation and size of the phoneme inventory has already been questioned. Tabain & Butcher (1999) have found he same degree of coarticulation, which can be seen as a kind of variability, for stops in two Australien languages with seven and six places of articulation as for English, which has only three places of articulation.

This leads to the conclusion that apart from the size of the phoneme inventory there should be other restrictions for the acceptability of token-to-token variability. Some of them are related to the relation between articulation and acoustics. As has been noticed by Stevens (1989), this relation is not linear.

A small change in articulation might cause a tremendous change in acoustics in one vocal tract configuration but not in another one. For example, at the place of a constriction, raising the tongue just a little can change an approximant into a fricative. If there is no constriction the same change in tongue height might cause a change in the acoustics which is not even noticed by the listener, for example when the quality of a rather open vowel like /ɛ/ is slightly changed. This non-linearity between articulation and acoustics should not only exist with respect to differences in vocal tract configurations, but also with respect to anatomical differences of the vocal tract. Same as there are vocal tract configurations which are very "sensitive" to small articulatory changes in that the acoustics change tremendously, there should also be vocal tracts which are more sensitive than others. In fact, evidence for such more or less "sensitive" vocal tract shapes has been found previously. Even if speakers are probably comparable in the token-to-token variability of their acoustic outputs, they seem to differ in their articulatory variability. Perkell (1997) compared six speakers with different palatal vaults who produced /i/, /ɪ/ and /ɛ/. He found that the speaker with the shallowest vault used the smallest differences in height of the tongue between the three vowels. The result has been supported by Mooshammer et al. (2004), who compared four speakers, three of them with a dome-shaped palate and one with a flat palate. They found that the speaker with a flat palate had a lower articulatory variability as compared to the other speakers.

Both results allow for the assumption that speakers with a flat palate have a more sensitive relation between articulation and acoustics than speakers with a more vaulted palate. One way to explain the sensitivity differences is to look at the different cross-sectional areas of the vocal tract which are most important for the acoustic output. These areas are not the same in the palatal region for speakers with different palates. For speakers with a shallow or "flat" palate the cross-sectional area resembles a quadrilateral, with the palate being the upper border, the tongue being the lower border and teeth and cheeks at the two sides. Speakers with a vaulted or "domeshaped" palate, on the other hand, have a cross- sectional area which can be schematized as a triangle with the tongue being the basis and the palate being the two legs of the triangle. The cross sectional area can be calculated for the quadrilateral (the flat palate) as

$$A_{orig} = ad \qquad (1)$$

with a being the width of the tongue and d the distance between tongue and

palate, and for the triangle as

$$A_{orig} = \frac{ad}{2}. \tag{2}$$

If there is some articulatory variability, for example if the tongue is raised by x, the distance d between tongue and palate changes to $d - x$ in both cases. When the area changes correspondingly, it becomes

$$A_{var} = a(d - x) \tag{3}$$

for the flat palate and

$$A_{var} = \frac{a(d - x)}{2} \tag{4}$$

for the domeshaped palate. The difference between A_{orig} and A_{var} is for the quadrilateral (the flat palate)

$$A_{diff} = A_{orig} - A_{var} \tag{5}$$

$$A_{diff} = ax \tag{6}$$

and for the triangle (the domeshaped palate):

$$A_{diff} = A_{orig} - A_{var} \tag{7}$$

$$A_{diff} = \frac{ax}{2} \tag{8}$$

This means that given the same articulatory variation the cross sectional area will change twice as much for the flat palate as compared to the domeshaped palate. Consequently, the acoustic signal will change more for the vocal tract with the flat palate than for the one with the domeshaped palate. This means that for a given amount of articulatory variation the acoustic output of a vocal tract with a flat palate responds in fact more sensitively to articulatory variation than a vocal tract with a domeshaped palate. Consequently, speakers with flat palates should articulate more precisely than speakers with domeshaped palates in order to limit the range of the acoustic variability and to facilitate perception. This explanation is based on perception oriented speaker's control and will be referred to as the *speaker's control hypothesis*. It is in line with Lindblom's Adaptive Variability Theory: Speech is an adaptive process,

and articulatory variability depends on the speaker's judgement of the communicative demands (Lindblom (1990)).

Even if the speaker's control hypothesis seems resonable one could still ask the question whether the differences are big enough to let speakers care about it. One could also argue that it is just a biomechanical matter. If the palate is flat the tongue has a greater area for linguo-palatal contact. During the production of consonants and high vowels the tongue could thereby be held in position in order to reduce the articulatory variability. This explanation will be called the *biomechanical hypothesis*. The aim of this study is to test both hypotheses.

The first hypothesis (based on speaker's control) could be supported experimentally by a correlation between palate shape and articulatory variability. For domeshaped palates a high articulatory variability should be found and for flat palates a lower one. There should be no correlation between the amount of palatal contact and variability nor between the amount of linguo-palatal contact and palatal shape.

In case the hypothesis based on speaker's control can be supported by exerimental data, another question worth investigating arises, namely whether given the same articulatory variability, the acoustic output is in fact more variable for a vocal tract with a flat palate than for one with a domeshaped palate. The problem with this kind of question is that it cannot be investigated experimentally because one would need subjects who can "switch off" their speaker's control and produce a certain degree of articulatory variability without paying attention to the acceptability of the output. Therefore, it has been decided to investigate this question by means of a tongue model.

If one would give preference to the second, biomechanical hypothesis, there should be a correlation between palate shape and the amount of lateral linguo-palatal contact for consonants and high vowels. Speakers with domeshaped palates should have less lateral contact whereas speakers with flat palates should have more lateral contact. Furthermore, the amount of lateral contact should correlate negatively with the articulatory variability. If there is lots of contact the articulatory variability should be low, if, however, there is hardly any contact the variability should be high.

To summarise the aims of this study: The relationships between palatal shape, articulatory variability and lateral linguo-palatal contact will be investigated. A correlation between palatal shape and variability without a relation of either of them with lateral contact should support the explanation based on speaker's control. The correlation betwen palatal shape and lateral contact on

the one hand, and between lateral contact and variability on the other hand should support the biomechanical hypothesis. These questions will be studied experimentally. The influence of the palate shape on the relation between articulatory and acoustic variability will be studied using a tongue model. If it turns out that the acoustic variability is higher for flat palates than for domeshaped palates given the same articulatory variability this would support the speaker's control hypothesis.

The following two sections describe the simulations carried out and their results. Sections 4 and 5 will describe the EPG experiment and its results respectively. In section 6 the results will be discussed.

2 Methods I: Simulations

The two dimensional biomechanical tongue model by Payan & Perrier (1997) has been used. Later versions of the model are sketched in Perrier et al. (1998), Perrier et al. (2003) and Perrier et al. (2004). The model will not be described here because it has been described extensively in the literature cited. Basically the model allows to simulate tongue movements due to the specification of recruitement thresholds muscle lengths that determine how muscle forces vary with muscle length. For a given set of recruitment thresholds of the tongue muscles, the target position of the tongue is reached for muscles lengths where a mechanical equilibrium is achieved. From a simulated tongue position an area function can be calculated (Perrier et al. (1992)) from which a sound can be synthesized.

The aim of the simulations was to investigate acoustic variability as a function of the palate shape while articulatory variability was held constant. Basically, three vowels, /a/, /i/ and /u/, were simulated with five palates differing in curvature. The tongue position of the vowels was changed slightly several times. For all the tongue positions corresponding sounds were synthesised and the formants of these sounds were calculated. This will now be described in more detail.

2.1 Building different palates

The width of the different palates was specified by α (Perrier et al. (1992)), a coefficient which gives information about the curvature of the palate. It has originally been designed for the $\alpha\beta$-model (Heinz & Stevens (1965)) which calculates the cross sectional area of a vocal tract as $A = \alpha * d^{\beta}$ with d being

the sagittal distance between the tongue and the palate (for β see Perrier et al. (1992)). As can be estimated from the formula, the lower α the more curved the palate is: For low α-values the area becomes smaller, which, given that the sagittal distance stays the same, has to be a result of a palate which is more curved. For the present purpose the following α values have been used: 1.3, 1.5, 2.0, 2.5 and 3.0. The corresponding curvatures of the palates can be seen in figure 1.

Figure 1: Palates with different curvatures used for the simulations (coronal perspective). The corresponding α values are 1.3 (very domeshaped palate), 1.5, 2.0, 2.5 and 3.0 (very flat palate). The values on the abscissa correspond to the coronal width, the ones on the ordinate to the height of the palate. A comparison with the alpha values calculated for the palates of human subjects shows that these palate shapes are realistic (cf. section 5).

Given the same distance between tongue and palate the cross sectional areas in figure 1 differ very much for the different palates. If one would carry out a synthesis for these cross sectional areas one would get very different sounds. This means that the premises for an investigation of the influence of a certain articulatory variation on the acoustics are not yet fulfilled since one needs as a starting point sounds which are comparable in terms of formant structure for all the palates. Therefore, the palates were raised or lowered until all the cross-sectional areas were the same (cf. figure 2).
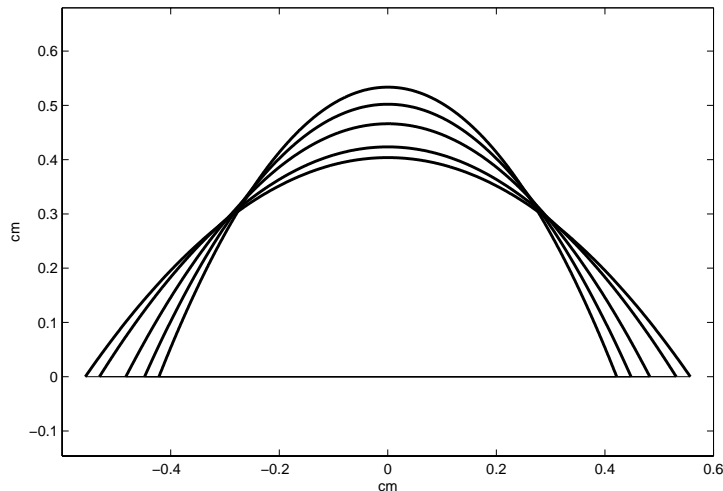
Figure 2: Palates with different curvatures from the coronal perspective. The values on the abscissa correspond to the coronal width, the ones on the ordinate to the height of the palate. In order to keep the cross sectional area constant, the palates have been moved up or down.

2.2 Simulations of "original" and "deviated" vowel positions

After the five palates had been built, the three vowels were simulated. As a result of this step we had three simulations with five different palates. The five vocal tract shapes corresponding to one vowel had approximately the same cross sectional area (figure 5, left side) and sounded about the same.

Now we simulated articulatory variation by moving the tongue position. The tongue position of /i/ was moved towards the one for /a/, the position of /a/ was moved towards /u/ and the position of /u/ was moved towards /i/ (cf. figure 3). The movement was carried out in six steps. The tongue positions will be called "deviated positions" in order to set them apart from the original tongue position. The important difference between the original tongue position and the deviated ones is that the cross-sectional areas of the original tongue positions are the same for all the palates whereas, following from the reasoning behind the speaker's control hypothesis stated in the introduction, for the deviated positions it is expected to differ for the five palates.

For each tongue position the area function was calculated. Afterwards, the sounds were synthesized and the formants were calculated (Badin & Fant (1984)).

In some cases quite unusual formant patterns were found. For example, the first formant of /u/ for the first palate falls during the movement towards /i/, as expected, but in two steps (cf. figure 4): gradually until the third tongue position, then it jumps to the fourth position, and moves gradually again to the
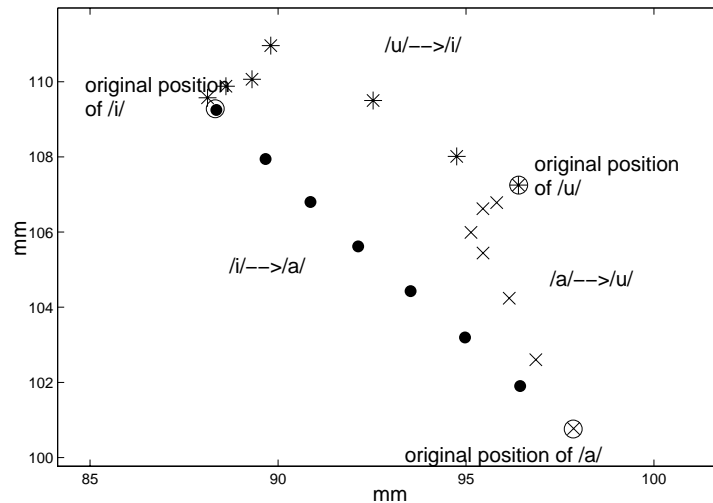
Figure 3: Original and deviated tongue positions in the sagittal perspective. The markers refer to a specific segment of the tongue, namely the highest point of the tongue back at the original simulations. A low value on the abscissa means that the tongue is near the teeth, a high value means that it is near the pharynx. A low value on the ordinate means that the vowel is very open, a high value means that it is very close. The tongue position of /i/ was moved towards /a/, /a/ was moved towards /u/, and /u/ was moved towards /i/. The original positions are circled.

seventh.

The reason for this jump in the middle could be that the place of constriction changes. Until the third step it is more in the back (tube 24-25 in the model), then it moves to the front (tube 27, and later tube 31). In order to restrict the study to local area changes, the analysis was limited to the steps for which the acoustic variability was within a reasonable range. These were for /a/ the third to seventh simulation and for /i/ and /u/ the fifth to seventh simulation.

3   Results I: Simulations

As can be seen in the area functions (figure 5) the cross-sectional areas of the deviated positions (right side in the figure) differ more for the five palates than the ones of the original position (left side) where the area functions are more or less the same). This means that the same articulatory variation caused by moving the tongue in the same way for all the palates results in different area functions.

Figure 6 shows the 95% confidential interval of the first and the second formant. The difference between the highest and the lowest value of these intervalls is nearly always greatest for the flattest palate and decreases for the more
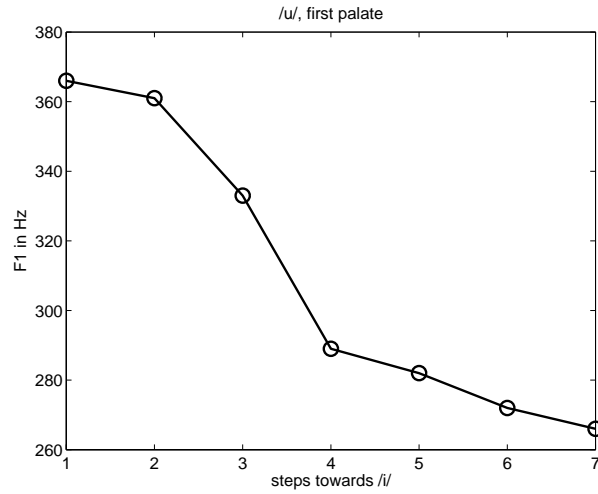
Figure 4: First formant values of /u/ for the first palate. The x-axis shows the seven steps in which the tongue position of /u/ was changed towards the one of /i/. The formant values do not decrease consistently, but there is a sudden jump at the fourth step. This jump can be explained by the changing position of the constriction.
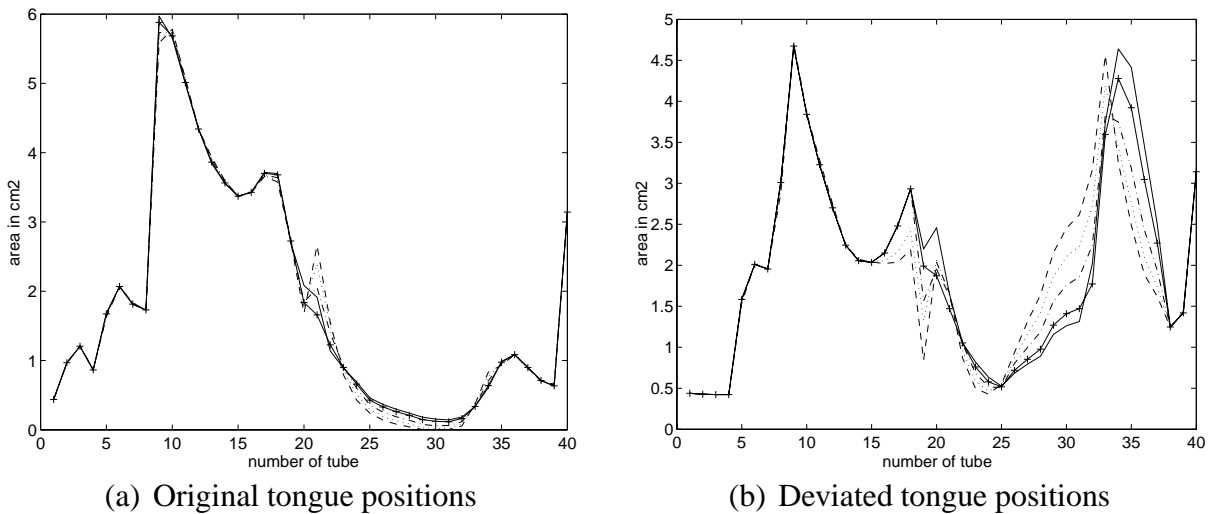

(a) Original tongue positions

(b) Deviated tongue positions

Figure 5: Area functions for the five different palates (marked by different line styles) for the original tongue positions of /i/ and the deviated ones, where the tongue has moved towards the configuration of /a/

domeshaped palates. This supports the reasoning behind the speaker's control hypothesis discussed in the introduction. For /a/ and /u/ the differences are greater for F1, for /i/ they are greater for F2.

4   Methods II: Experiment

The results of the simulations confirmed the basis for the speaker's control hypothesis: Vocal tracts with flat palates respond more sensitively to articulatory variability than vocal tracts with domeshaped palates. Now it has to be tested whether speakers really react differently to differences in vocal tract shape. Therefore an EPG experiment with 20 speakers has been carried out.

4.1   EPG-Recordings

Electropalatography (EPG) allows to investigate linguo-palatal contact. The subject wears an artificial palate made of acryl with 62 electrodes on it (EPG 3.0, Reading system). Each electrode is connected to the system over a small wire. During the recording the subject is holding a further electrode in his or her hand which is also connected to the system. Each time the tongue is touching an electrode at the palate the circuit is closed, which is registerd by the system. A parallel acoustic recording was carried out with a DAT recorder.

Since the question investigated here is not bound to a certain language but to human speech production in general, speakers of different languages have been recorded:

- two speakers of Bulgarian

- three speakers of Polish

- five speakers of English (two English, two Scottish and one Australian)

- ten speakers of German.

We are aware that the size of the phoneme inventory possibly has an influence on the articulatory variability. This fact will be discussed later.

The sounds to be investigated were the consonants /s/, /ʃ/, /ɕ/ and /j/, and the vowels /i/, /e/, and /u/, with their lax counterpart /ɪ/, /ɛ/ and /ʊ/. The sounds were choosen because the vocal tract is rather narrow during their production and consequently an influence of the palate shape can be expected. In order to make the different languages comparable, nonsense words were used rather than real words because so it was possible to use the same items for all
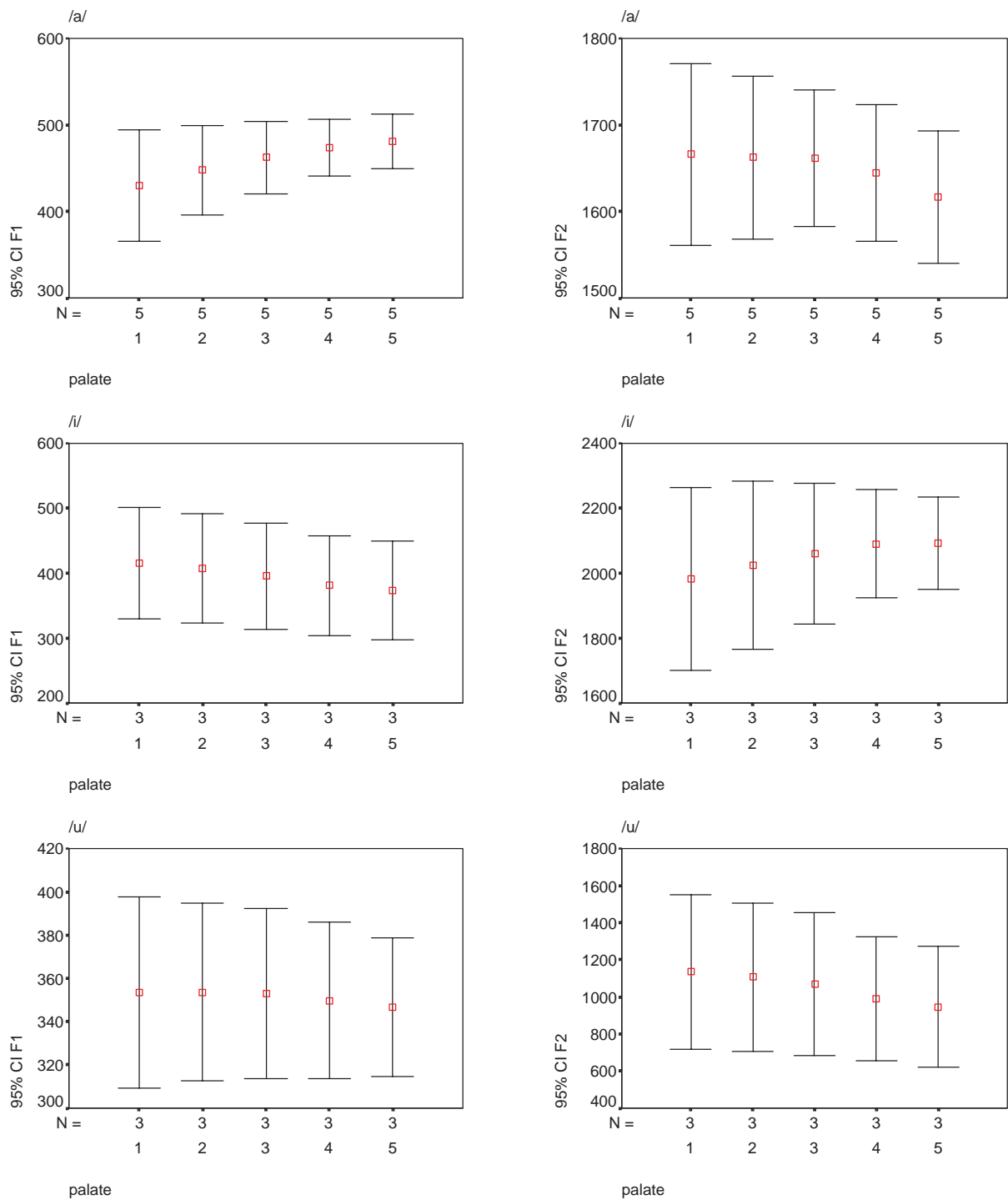
Figure 6: 95% confidential intervals for F1 (left) and F2 (right) for the five different palates. The first line gives the results for /a/, the second the ones for /i/ and the third the ones for /u/. In most cases the dispersion is greatest for the very flat palate (palate 1) and decreases with increasing palate height. An exception is the first formant of /i/, where the confidential interval stays about the same.

the languages. Since some of the sounds do not exist as phonemes in all the recorded languages not all the speakers were recorded speaking all items. For Bulgarian and Polish speakers no lax vowels were recorded, for English there was no palatal voiceless fricative nor the vowel /e/.

The items in which the sounds were embedded were: /'sasa/ (for German /'sasa/) , /'ʃaʃa/, /'mɪɕi/ (for the Polish speakers /ma'ɕina/), /'jaja/, /'tiːti/, /'tɪti/, /'tuːtu/, /'tʊtu/, /'teːtə/ and /'tɛtə/ (for the English speakers /'tɛtər/ or /'tɛtə/. The carrier phrases differed from language to language:

- for Bulgarian: *Kazah ... na teb.* (I have said ... to you.)

- for Polish: *Powiedzialem ... do ciebie.* (I said ... to you.)

- for German: *Habe ... gesagt.* ((I) have said ...)

- for English: *Say ... please.*

Each sentence was repeated 30 times in randomized order.

4.2   Calculation of the variability

Beginning and end of each segment of interest was labelled in the acoustic signal using PRAAT 4.2.17 (Boersma & Weenink (1992–2004)). The following points in time were labelled:

- friction onset and offset for the fricatives /s/, /ʃ/ and /ɕ/

- sonorant onset and offset for the sonorant /j/ as the points in the middle of the formant transitions between the surrounding vowels and the sonorant.

- onset and offset of the second formant for the vowels.

Afterwards, the percent of contact (poc) for each EPG frame within the measured interval has been calculated as

$$poc = \frac{noc * 100}{62} \tag{9}$$

with $noc =$ number of contacts and 62 as the maximal number of contacts.

Figures 7 and 8 illustrate this method. Figure 7 shows the EPG frames of a production of /u/ of a certain speaker. In the beginning one can still see contact in the anterior region which is a remnant of the /t/ preceeding the sound. Gradually the /t/ disappears and the /u/ shows up with contact only in the
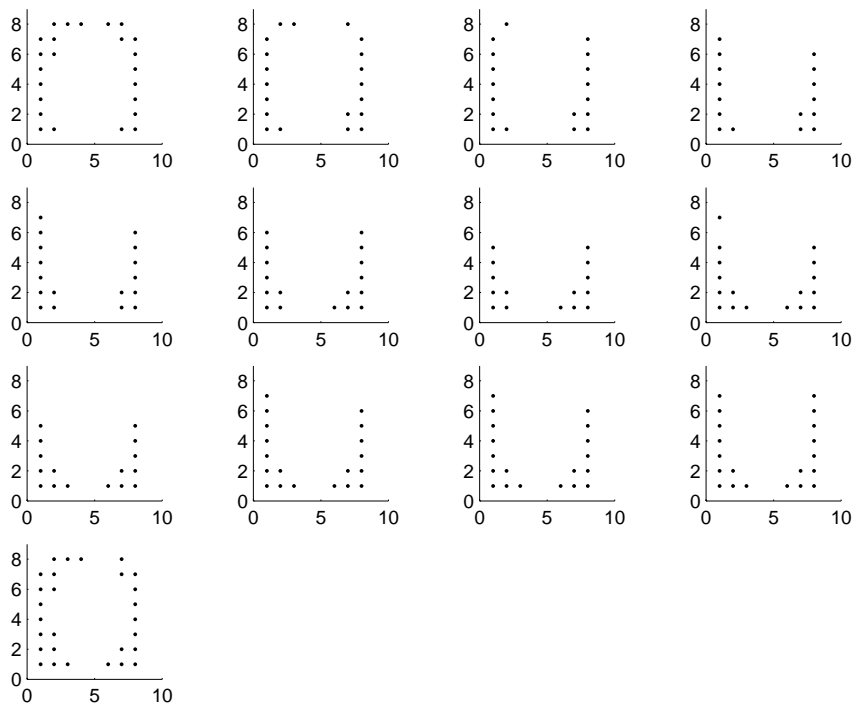
Figure 7: EPG frames for /u/ surrounded by /t/

posterior region. Towards the end one can again see contact in the anterior region which marks the /t/ following the /u/.

Figure 8 shows the *poc* for each of the frames in figure 7. As one can see there is much contact in the beginning (when the /t/ is still present), then there is less contact in the middle of the segment (during the /u/), and towards the end the *poc* rises again because of the second /t/.

These calculations were carried out for the 30 repetitions of each item. A problem for further calculations was that the length of the segments differed depending on the the velocity of speaking. Therefore, the segments had to be made the same length. This was done by a spline interpolation on 20 points (cf. figures 9 and 10).

Now a mean value and the standard deviation of the 30 repetitions was calculated for each of the 20 sample points (cf. figure 11). The mean of all the standard deviations was calculated. Because the standard deviation is highly dependent on the mean value it was normalised at the mean value due to the variability coefficient:

$$v = \frac{s}{\bar{x}} \tag{10}$$

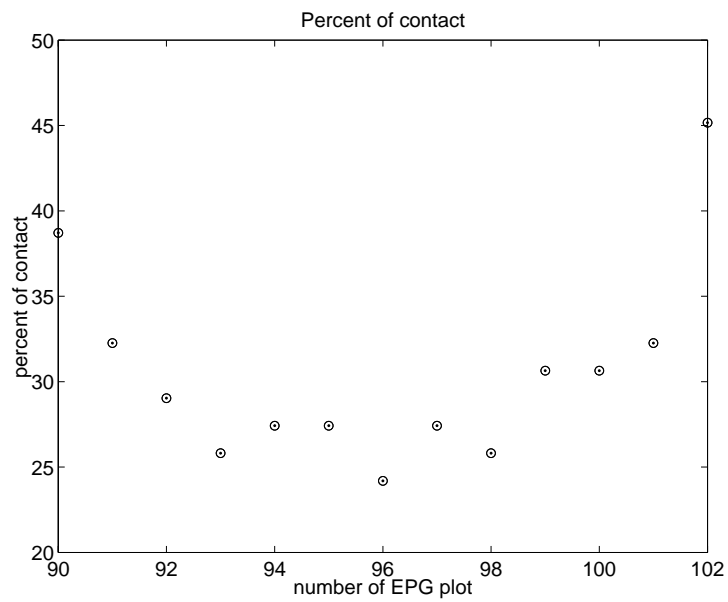This resulting number was treated as the variability of a segment uttered by a certain speaker.

Figure 8: Percent of contact for the EPG frames of /u/ shown in figure 7 (with spline interpolation for easier understanding)
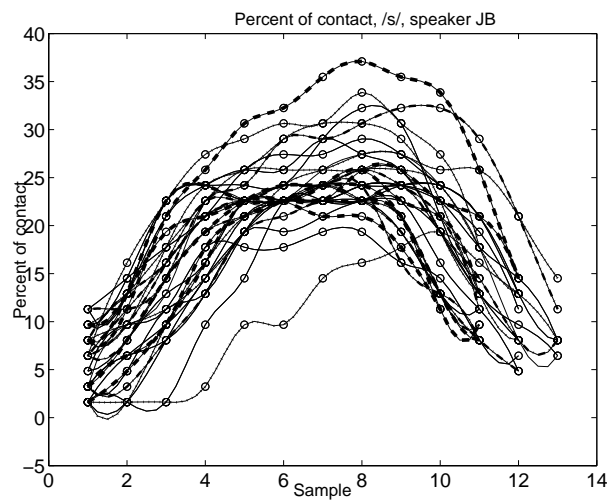

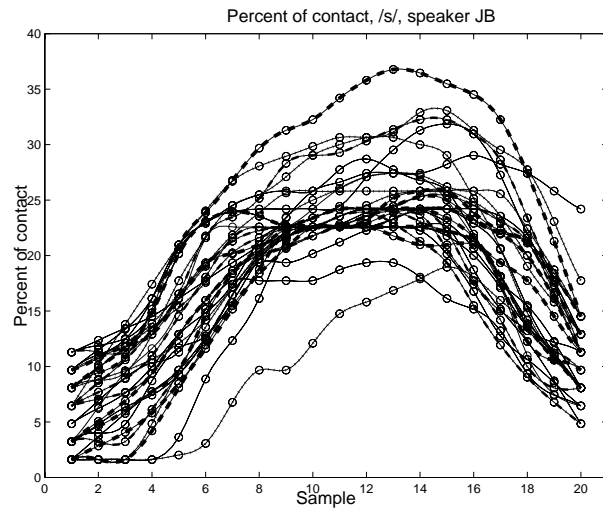**Figure 9:** Percent of contact for /s/ for 30 repetitions.

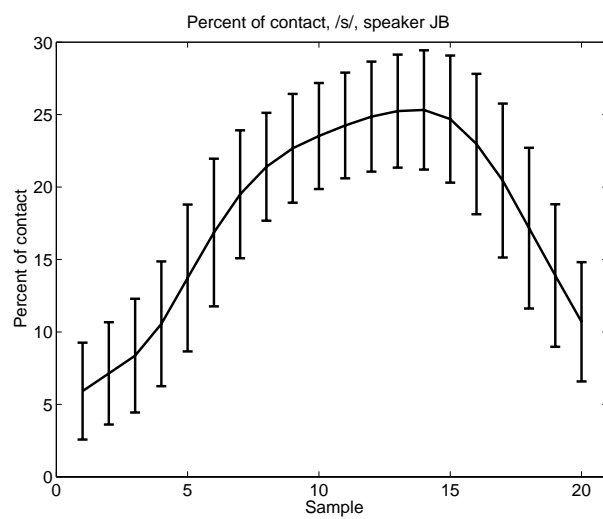**Figure 10:** Percent of contact for /s/ for 30 repetitions after the interpolation on 20 points.

Figure 11: Mean values and standard deviation of the percent of contact of 30 repetitions of /s/

4.3 Calculation of lateral linguo-palatal contact

In order to investigate the relationship between linguo-palatal contact and variability which was suggested in the biomechanical hypothesis, the lateral index li of the segments was calculated. This was done similarly to the calculation of the percent of contact. For each EPG frame the percent of contact in the lateral region (the two very left and the two very right rows, cf. figure 12) was calculated as

$$li = \frac{noc * 100}{30} \tag{11}$$

with noc = number of contacts and 30 the maximal number of contacts in the lateral region.
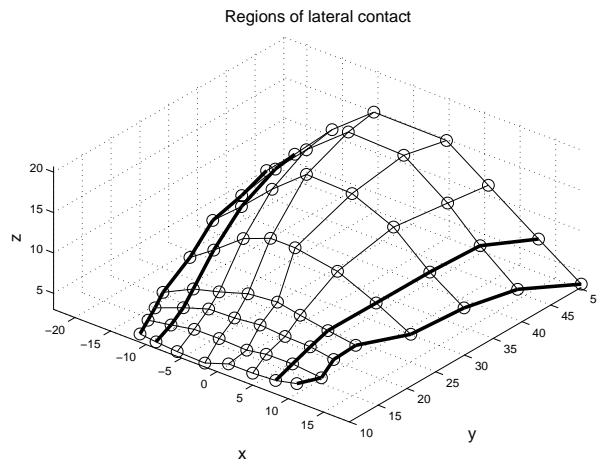

Figure 12: The lateral index is calculated as the percentage of contact in the marked lateral regions.

Same as for the percent of contact, an interpolation on 20 points was carried out in order to make all the 30 repetitions the same length. Afterwards the mean value of each of the 20 sample points was calculated. From these 20 numbers a mean was calculated as average lateral contact.

4.4 Determination of the palate shape

In order to investigate the relationship between palate shape and variability one needs to find a method to determine the curvature of the palate. Because an image of each palate existed in form of the EPG palates these were taken as a basis for this step.

At first the coordinates of each of the 62 electrodes were measured using a caliper. A result of this can be seen in figure 13. The sixth row (marked bold in the figure) was found to represent the palatal curvature best. That is why this

row was chosen for calculating the curvature coefficient α (cf. section 2).
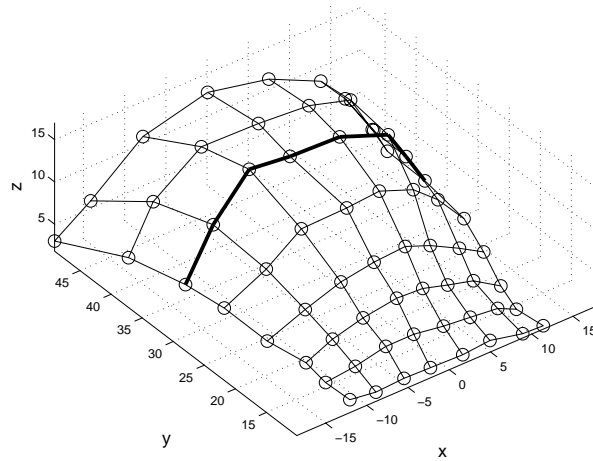

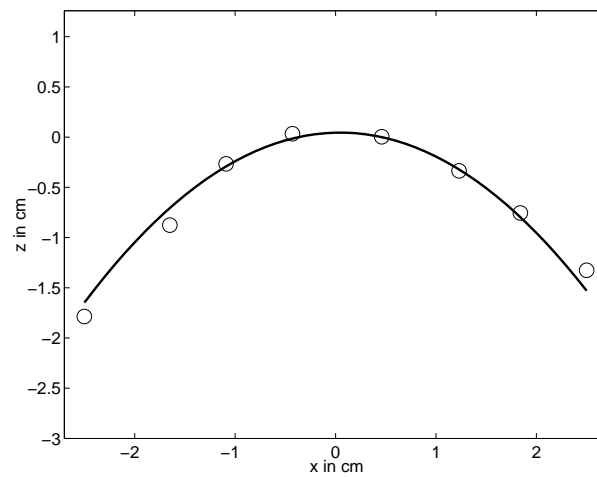**Figure 13:** The curvature of the palate was estimated for the sixth row.

Figure 14: Linear approximation of the measured points from the sixth row

As can be seen in figure 14, a linear approximation with two coefficients was carried out for the measured points of row six. The palatal shape could now be described by

$$y = ax^2 + b \tag{12}$$

α was calculated as (Perrier et al. (1992))

$$\alpha = \frac{\frac{4}{3}}{\sqrt{|a|}} \tag{13}$$

A high α value corresponds to a flat palate and a low value to a domeshaped palate.

4.5 Calculation of the correlation

The correlations (Pearson) between α and variability, variability and lateral contact and α and lateral contact were calculated using SPSS 11.5.1.

5 Results II: Experiment

Table 1 gives the correlation coefficients for the correlations between the three parameters. Even if most correlations are not significant one can still find some general tendencies. The strongest relationship seems to be the one between variability and lateral contact (cf. third column in the table). Five of the ten correlations are significant. This means that the more lateral contact there is the smaller the articulatory variability is. There seems to be no correlation between α and lateral linguo-palatal contact since some of the correlations are positive and some are negative. Sometimes a speaker with a domeshaped palate has lots of contact, sometimes he or she hasn't. The correlation between variability and α is always negative, which suggests that speakers with a flat palate have a low variability and speakers with a domeshaped palate have a high variability. However, the correlation is nearly never significant.

The results differ very much with respect to the item. They are best for palatal vowels and sonorants, not as good for the fricatives and worst for the back vowels.

Table 1: Correlation coefficients r of variability and α (second column), variability and lateral contact (third column) and lateral contact and α (fourth column), significance p is given in brackets.

item	variability - α	variability - lat. contact	lateral contact - α
/ɕ/	-.145 (.606)	-.472 (.076)	.266 (.337)
/s/	-.301 (.129)	-.569** (.009)	257 (.275)
/ʃ/	-.254 (.281)	-.429 (.059)	.259 (.269)
/j/	-.466* (.039)	-.137 (.564)	-.136 (.567)
/i/	-.294 (.208)	-.577** (.008)	.637** (.003)
/ɪ/	-.559* (.030)	-.342 (.212)	.487 (.066)
/u/	-.001 (.998)	-.679** (.001)	-.110 (.644)
/ʊ/	-.011 (.969)	-.589* (.021)	-.195 (.486)
/e/	-.382 (.160)	-.705** (.003)	.524* (.045)
/ɛ/	-.291 (.293)	.080 (.777)	.400 (.139)

There are two possible reasons for the lack of significance of the correlation between α and variability: Either the sample is to small and one needs to

record more speakers. The other possibility is that there is a relation but it is
not a correlation. Even if the first possibility certainly plays a role, looking at
the data more closely shows that the second explanation is also reasonable. In
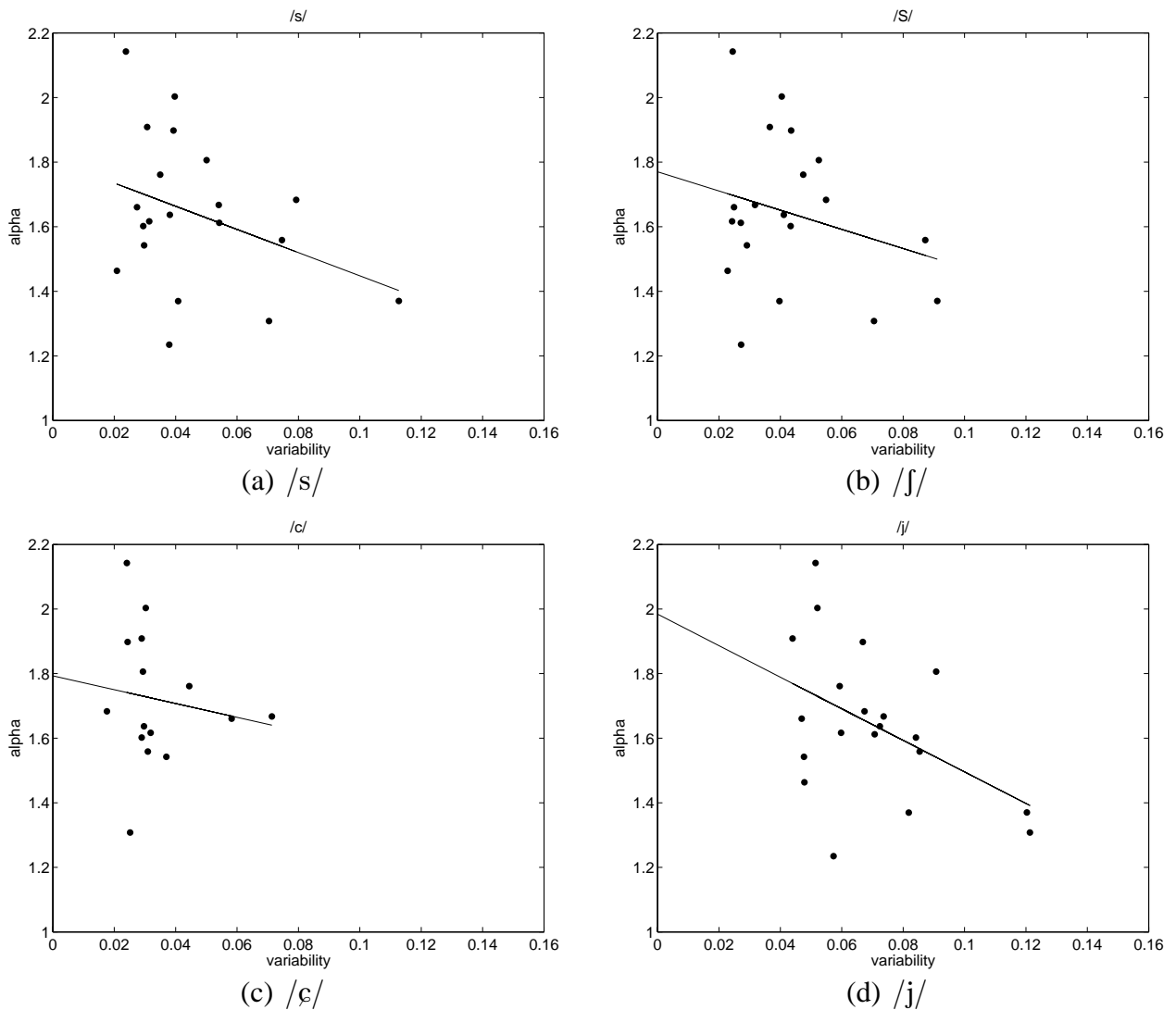figures 15 and 16 α is plotted against variability.

(a) /s/

(b) /ʃ/

(c) /ç/

(d) /j/

Figure 15: Correlation between variability (abscissa) and α (ordinate) for the consonants

What can be seen is that speakers with a flat palate (a high α) always have
a low variability. Speakers with a domeshaped palate, however, sometimes have
a high variability, sometimes they have a low variability. There are no speakers
with a flat palate who have a high variability.

(a) /i/

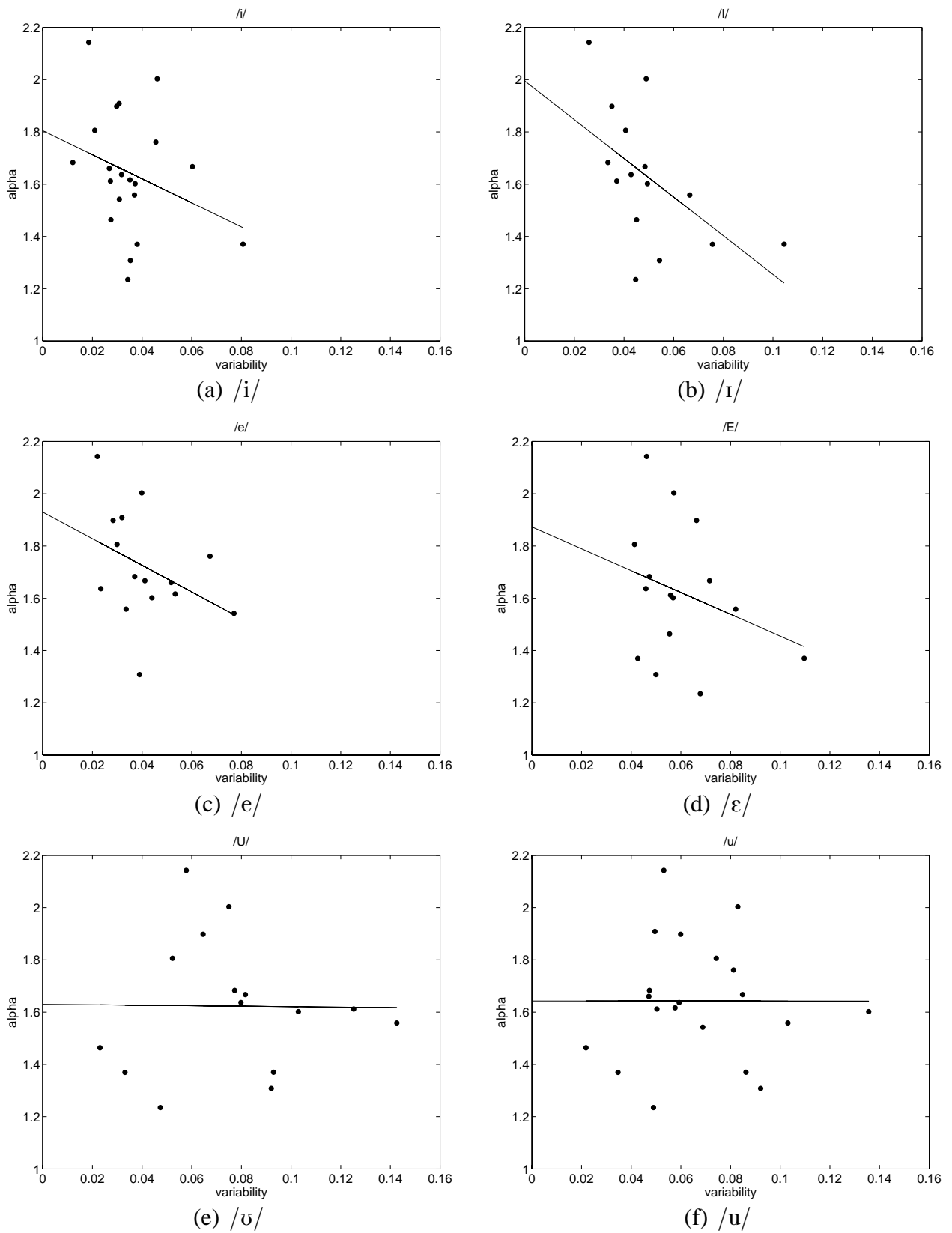(b) /ɪ/

(c) /e/

(d) /ɛ/

(e) /ʊ/

(f) /u/

Figure 16: Correlation between variability (abscissa) and α (ordinate) for the vowels

6 Discussion

The aim of the current study was to investigate the relation between palate shape and articulatory variability. Two possible explanations for the previously detected relation between the two parameters have been supposed.

The first explanation is based on perception oriented speaker's control. Speakers with a domeshaped palate can afford more articulatory variability because the cross sectional area of their vocal tracts is not affected to the same extent by small articulatory changes as compared to speakers with a flat palate. Consequently, the acoustic output can be kept constant more easily. Speakers with a flat palate, on the other hand, have to invest more effort to keep their acoustic output constant because even a small articulatory variation can change the acoustic output immensely.

In contrast to that, the biomechanical hypothesis suggests that speakers with a flat palate have more linguo-palatal contact which holds the tongue in position. Speakers with a domeshaped palate, on the other hand, have less contact. Therefore their articulation is more variable.

To repeat the hypotheses from the introduction, greater variation of the acoustic signal for a flat palate as compared to a domeshaped one, given the same articulatory variation, would lay a basis for the speaker's control hypothesis. A correlation between palate shape and variability would support this explanation. A correlation between palate shape and linguo-palatal contact and between linguo-palatal contact and variability would support the biomechanical hypothesis.

Looking at the results of this study shows that things are not as simple as suggested in either of the two explanations. There seems to be a relation between variability and palate shape, however, it is not a correlation, which would support the speaker's control hypothesis. Speakers with a flat palate have a very low articulatory variability. Speakers with a domeshaped palate, however, vary. Some of them have a high articulatory variability, some of them don't. Support for the speaker's control hypothesis is given by the simulations. The acoustic signal changes more easily if the palate is flat than if it is domeshaped. Linguo-palatal contact negatively correlates with variability. The more linguo-palatal contact there is the less articulatory variability. This would support the biomechanical explantion. However, the biomechanical hypothesis traces the amount of palatal contact in the palate shape. But there seems to be no relation between palate shape and linguo-palatal contact.

A way out could be a modified version of the speaker's control hypothe-

sis. The relation between variability and palate shape is actually quite easy to explain: Speakers with a flat palate have to have a low articulatory variability in order to keep their acoustic output constant. Speakers with a domeshaped palate, however, have the choice. They can afford to have more variability, but some of them don't. Reasons for that could be the situation in which they were during the experiment. Maybe they intended to speak very clearly, which would be in line with the H&H Theory: The degree of variability depends on situational demands (Lindblom (1990)). Another reason could be a difference in the ability to perceive subtle acoustic differences among the speakers with a domeshaped palate. As has been suggested by Perkell et al. (2003) speakers who produce speech with less variability also perceive smaller differences in acoustics. This does not seem to go against perception oriented speaker's control as such. What is essential for the speaker's control hypothesis is that speakers with a flat palate have a low articulatory variability.

There is still the question why variability and lateral contact correlate. But even here one can find an explanation. Since there is no relation between palate shape and lateral contact speakers seem to be free to choose to have more or less linguo-palatal contact. It seems likely that speakers use palatal contact as a means to reduce their articulatory variability. Speakers with a flat palate always do it, whereas speakers with a domeshaped palate can do it (then they have a low variability) but they do not have to because the acoustic output does not change as easily. In any case linguo-palatal contact does not seem to be a consequence of a certain palate shape.

The relation between palate shape and articulatory variability is stronger for some items than for others. There are several possible reasons for that. For the fricatives one can assume that there is so much linguo-palatal contact that the articulatory variability is generally very low so that differences between speakers are hard to find. For /u/ one can assume that at least some speakers will compensate for their articulatory variability via lip rounding.

As has been stated in the introduction, articulatory variability might depend on the size of the phoneme inventory of the language spoken, even if this has also been questioned. For example, if there is a phonemic difference between /s/ and /ʃ/ in a language, the sounds should be less variable when spoken by speakers of this language as compared to speakers of a language where the two sounds are only one phoneme. Up to now it is hard to find a relationship between the size of the phoneme inventory and the variability detected by the speakers of the language. The variability of the vowels seems to be greater for the Polish and Bulgarian speakers than for the others. This could be because

there is no phonemic difference in tenseness in these languages. However, up to now too few speakers have been recorded to be definite about that.

References

Ainsworth, W. A. (1997). Some approaches to automatic speech recognition. In: W. J. Hardcastle & J. Laver (eds.) *The Handbook of Phonetic Sciences*, 721–743. Oxford: Blackwell.

Badin, P. & Fant, G. (1984). Notes on vocal tract computation. *STL-QPSR*, 2–3:53–108.

Boersma, P. & Weenink, D. (1992–2004). Praat, a system for doing phonetics by computer. URL `www.PRAAT.org`.

Dixon, R. M. W. (1980). *The languages of Australia*. Cambridge: Cambridge University Press.

Heinz, J. & Stevens, K. (1965). On the relations between lateral cineradiographs, area functions, and acoustic spectra of speech. In: *Proceedings of the Fifth International Congress of Acoustic*, A44. Liège.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In: W. J. Hardcastle & A. Marchal (eds.) *Speech Production and Speech Modelling*, 403–439. Dordrecht: Kluwer Academic Publishers.

Mooshammer, C., Perrier, P., Geng, C., & Pape, D. (2004). An EMMA and EPG study on token-to-token variability. *AIPUK*, 36:47–63.

Nolan, F. (1997). Speaker recognition and forensic phonetics. In: W. J. Hardcastle & J. Laver (eds.) *The Handbook of Phonetic Sciences*, 744–767. Oxford: Blackwell.

Payan, Y. & Perrier, P. (1997). Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication*, 22:185–205.

Perkell, J., Matthies, M., Guenther, F., Tiede, M., Zandipour, M., Stockmann, E., & Marrone, N. (2003). Sensory goals for speech movements: Cross-subject relations among production, perception and the use of an articulatory saturation effect. In: *Proceedings of the 6th International Seminar on Speech Production*, 219–224.

Perkell, J. S. (1997). Articulatory processes. In: W. J. Hardcastle & J. Laver (eds.) *The Handbook of Phonetic Sciences*, 333–370. Oxford and Cambridge, Massachusetts: Blackwell.

Perrier, P., Boë, L. J., & Sock, R. (1992). Vocal tract area function estimation from midsagittal dimensions with CT scans and a vocal tract cast: modelling the transition with two sets of coefficients. *Journal of Speech and Hearing Research*, 35:53–67.

Perrier, P., Payan, Y., & Marret, R. (2004). Modéliser le physique pour comprendre le contrôle: le cas de l'anticipation en production de parole. In: R. Sock & B. Vaxelaire (eds.) *L'anticipation à l'horizon du présent*. Liège: Mardaga.

Perrier, P., Payan, Y., Perkell, J., Zandipour, M., & Matthies, M. (1998). On loops and articulatory biomechanics. *Proceedings of the 5th International Conference on Spoken Language Processing (Sydney)*, 2:421–424.

Perrier, P., Payan, Y., Zandipour, M., & Perkell, J. (2003). Influences that shape tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America*, 114:1582–1599.

Stevens, K. N. (1989). On the quantal nature of speech. *Journal of Phonetics*, 17:3–45.

Tabain, M. & Butcher, A. (1999). Stop consonants in Yanyuwa and Yindjibarndi: A locus equation perspective. *Journal of Phonetics*, 27:333–357.