

ZASPiL Nr. 42 – December 2005

**Papers in Phonetics and Phonology**

Editors: Christian Geng, Jana Brunner  
and Daniel Pape

## CONTENTS

**Silke Hamann and Hristo Velkov**

Airflow in stop-vowel sequences of German.....1

**Victoria Medina and Willy Serniclaes**

Late development of the categorical perception of speech sounds in pre-adolescent children.....13

**Silke Hamann and Anke Sennema**

Acoustic differences between German and Dutch labiodentals.....33

**Jana Brunner, Susanne Fuchs and Pascal Perrier**

The influence of the palate shape on articulatory token-to-token variability....43

**Marzena Zygis**

(Non)Retroflexivity of Slavic Affricates and Its Motivation. Evidence from Polish and Czech <č> .....69

**Katrin Dohlus**

Phonetics or Phonology: Asymmetries in Loanword Adaptations - French and German Mid Front Rounded Vowels in Japanese.....117

**Susanne Fuchs and Pascal Perrier**

On the complex nature of speech kinematics .....137

**Mariam Hartinger and Christine Mooshammer**

Articulatory variability of clutterers.....167

**Sabine Koppetsch**

Die motorische Funktionsprüfung bei oralen Tumoren.....181

**Katalin Mády, Krisztián Z. Tronka and Uwe D. Reichel**

Syllable cut and energy contour in vowels: a comparative study on German and Hungarian.....197

**Christian Geng and Phil Hoole**

Some comments on the reliability of three index factor analysis models in speech research.....219

# Airflow in stop-vowel sequences of German

**Silke Hamann**

*Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany*

**Hristo Velkov**

*Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany*

---

This study reports on the results of an airflow experiment that measured the duration of airflow and the amount of air from release of a stop to the beginning of a following vowel in stop vowel-sequences of German. The sequences involved coronal, labial and velar voiced and voiceless stops followed by the vocoids /j, i:, ɪ, ε, ʊ, a/. The experiment tested the influence of the three factors voicing of stop, place of stop articulation, and the following vocoid context on the duration and amount of air as possible explanation for assibilation processes. The results show that the voiceless stops are related to a longer duration and more air in the release phase than voiced ones. For the influence of the vocoids, a significant difference could be established between /j/ and all other vocoids for the duration of the release phase. This difference could not be found for the amount of air over this duration. The place of articulation had only restricted influence. Velars resulted in significantly longer duration of the release phase compared to non-velars. A significant difference in amount of air between the places of articulation could not be found.

---

## 1 Introduction

The present article investigates the difference in the amount of airflow between voiced and voiceless stops followed by the vocoids /j, ɪ, ɪ, ε, ʊ, a/ in German. Background for this investigation are phonological assibilation processes whereby stops are turned into affricates or fricatives before high vocoids, e.g. /ti/ surfaces as [s] in Finnish (Kiparsky 1973). In a typological study of assibilations in more than 30 typologically diverse languages, Hall & Hamann (to appear) postulated the following two implications:

- (1) a) Assibilation cannot be triggered by /i/ unless it is also triggered by /j/.
- b) Voiced stops cannot undergo assibilations unless voiceless ones do.

Following a study by Kim (2001) on assibilation in Korean, Hall, Hamann & Zygis (2004) give acoustic evidence for these implications in Polish and German. They measured the duration from stop burst until the beginning of the following vocoid /j, i/<sup>1</sup> (comprising burst frication, friction noise at the supralaryngeal place of articulation and aspiration), termed there and in the present article as ‘friction’ phase. This friction phase was significantly longer for /t/ than for /d/. Furthermore, for both voiced and voiceless stops, a following /j/ caused longer friction than a following /i/. Both observations are summarised in the following hierarchy of friction duration, where ‘>’ stands for ‘has longer friction duration than’:

$$(2) \quad /tj/ > /ti/ > /dj/ > /di/$$

The friction noise present in these sequences can be reinterpreted by listeners as lexically specified, i.e. as underlying fricative or as affricate, as Hall & Hamann (to appear) argue. Thus a longer friction phase is more likely to be reinterpreted as fricative than a shorter friction phase, which yields an acoustic motivation for the cross-linguistic implications in (1). The symbol ‘>’ in the hierarchy in (2) can therefore also be read as ‘is more likely to assibilate than’.

Hall et al. (2004) propose an aerodynamic explanation for the differences in friction length between voiced and voiceless stops: due to the open vocal folds, air can flow unimpeded for the voiceless stop, and more pressure builds up behind the constriction at closure, which results in longer (and stronger) friction at release. The difference between high vowel and glide is explained by referring to articulation and aerodynamics. The palatal glide might be articulated with a higher and more fronted tongue position than the high front vowel, and thus have a narrower constriction, which causes more air to built up behind the glide, again resulting in longer (and more forceful) friction. For earlier explanations along the same line, see Jäger (1978) and Ohala (1983).

The aim of the present study is to test the validity of the aerodynamic explanations by airflow measurements. If Hall et al.’s predictions are correct, then voiceless stops should not only show a longer duration of unimpeded airflow from the release of the stop until the beginning of the following vowel, see prediction (3a) below, but also a larger amount of air should be produced during this time interval, see prediction (3b). Furthermore, /j/ should cause a longer duration of airflow from burst until the onset of the following vowel and a larger amount of air over this time interval than /i/, cf. predictions (3c) and (d).

---

<sup>1</sup> The tense high front vowel is short in Polish and long in German. This difference is ignored in the present discussion of Hall et al.

(3) Four predictions:

- a) the voiceless stops show a longer duration of airflow in the friction phase than the voiced stops,
- b) the voiceless stops have a larger amount of air than the voiced stops over this time interval,
- c) the palatal glide causes a longer duration of airflow in the friction phase than the vowel /i/,
- d) the palatal glide causes a larger amount of air over this time interval than the vowel /i/.

Whereas Hall et al.'s investigation was restricted to coronal stops, the present study includes velar and bilabial stops, and in addition to the context of the palatal glide and the high front vowel /i/, the influence of a following /ɪ/, /ɛ/, /ʊ/ and /a/ on coronal stops is tested.

The predictions (3a) and (b) on the influence of stop voicing on the duration of airflow and amount of air lead to the following partial assibilation hierarchies in (4). These hierarchies have to be interpreted as /p/ has a longer duration of airflow in the friction phase and more air over this duration than /b/, and is thus more likely to assibilate than /b/, and so forth.

- (4)    p > b  
       t > d  
       k > g

The predictions on the influence of the following vowel or glide in (3c) and (d) can be extended to include further vowel contexts on the basis of the following principle. For a smaller area of constriction, i.e. a higher vowel, we expected a longer duration of airflow and larger amount of air. This results in the assibilation hierarchy in (5), where the vowels /ɪ/ and /ʊ/ are not ranked with respect to each other because they share the same vowel height.

- (5)    j > i: > {ɪ, ʊ} > e > a

The study by Hall et al. does not look at the influence of the place of articulation on the friction duration of stops and thus their likelihood to assibilate. We hypothesize that velars show a longer duration of airflow and amount of air in the friction phase than coronals. This is due to the shorter supralaryngeal cavity (looking downstream towards the glottis) in velars which results in more air pressure to built up behind the constriction, and which then yields a longer friction phase at the release and/or more air during the friction phase. For the same reason, coronals are expected to have a longer duration of airflow and

amount of air than labials. These expectations are formalised in the following hierarchy, which again predicts the highest likelihood for the item on the left to assibilate, and lowest for the item on the right:

(6) velar > coronal > labial

The following section describes the experimental setup to test the three (partial) hierarchies (4) – (6). In section 3, results of this experiment are presented. Section 4 concludes.

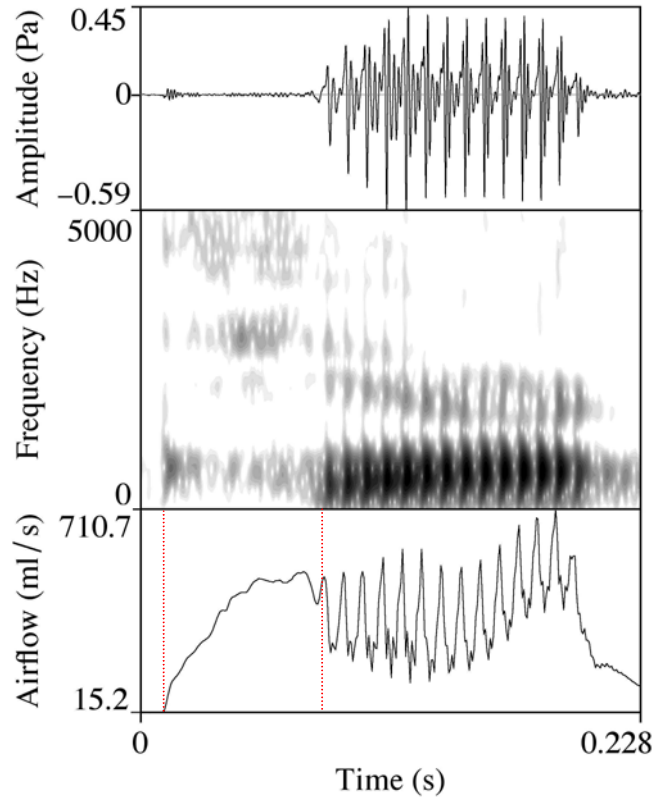
## 2 Method

Our subjects were four native German speakers (two male and two female). Each subject was asked to repeat the items in Table 1 five times in the carrier sentence “habe ... gesagt” ‘said ...’. This item set includes coronal, labial, and velar stops, both voiced and voiceless, followed by /i:/, /ɪ/ and /ja/. For the coronals, we furthermore used the following vowels /e/, /ʊ/, and /a/. Though all of these items are phonotactically well-formed in German, the sequences with stop plus glide have a very restricted occurrence and are mainly the result of an optional gliding process (e.g. *Opiat* [op.’ja:t] ‘opiate’, *Median* [me.’dja:n] ‘median’), see Hamann (2003) and Hall (to appear). For this reason we chose nonsense words.

**Table 1:** Test items (nonsense words).

tiek	tick	tjack	diek	dick	djack
teck	tuck	tack	deck	duck	dack
piek	pick	pjack	biek	bick	bjack
kiek	kick	kjack	giek	gick	gjack

We measured the oral airflow with the PCquirer hardware from Scicon, and carried out the data analysis with PRAAT (Boersma & Weenink 2005). For every test item, we measured the duration from release of the stop until the onset of the following vowel (i.e. the friction phase). The onset of the following vowel was determined by the beginning of the second vowel formant (in ambiguous cases, we took the beginning of higher formants and of periodicity as additional criteria). An example audio waveform, spectrogram and waveform of the airflow is given in Figure 1 for the word *tjack*. This figure shows the points of measurement in the waveform of the airflow with dotted lines. In addition to the duration of the friction phase, we calculated the sum of the amount of air that was produced over this time interval (i.e. the integral).



**Figure 1:** Waveform of the acoustic signal, spectrogram, and waveform of the airflow for *tjack*. The dotted lines in the airflow indicate the beginning and the end point of the measurements.

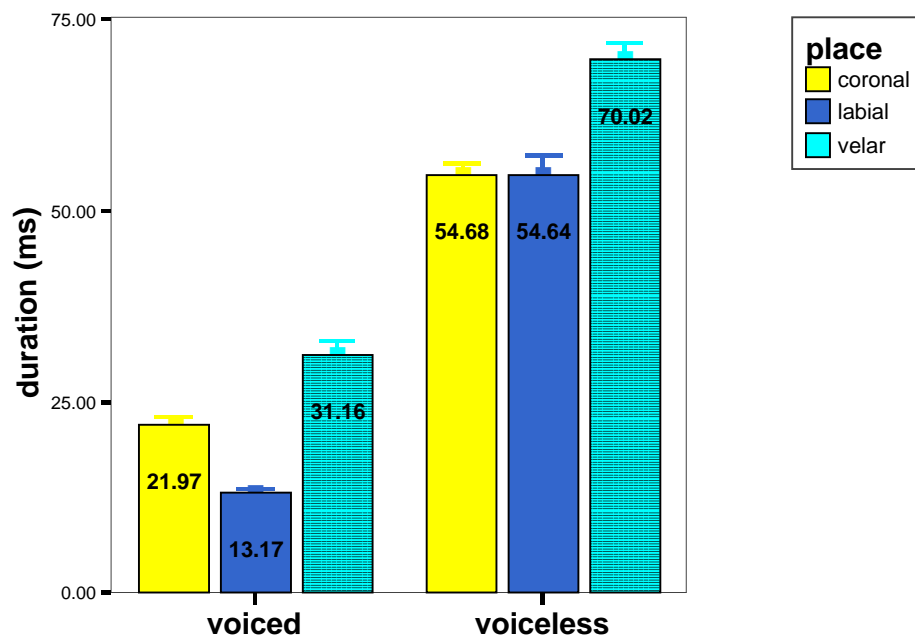
### 3 Results

The results are presented in the following order. In the first subsection (3.1), the influence of voicing of the consonant is given. In subsection 3.2, the influence of the following vocoid is presented, and in the last subsection (3.3), the influence of place of articulation is discussed. For each parameter, we give both the duration of friction and the amount of airflow produced over this duration. Due to the small number of repetitions, the following statistical analyses are all averaged over speakers. An interaction between speaker and duration could not be found, and an interaction between speaker and amount of air was observable only in half of the cases.

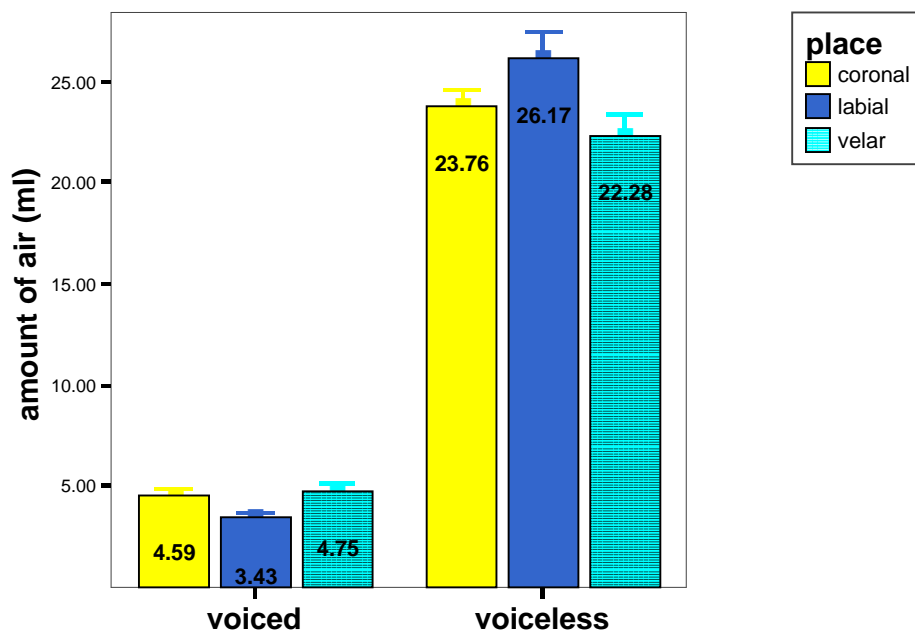
#### 3.1 Influence of voicing

Figures 2 and 3 show the average duration (in ms) from burst until onset of the following vocoid and the average amount of air (in ml) over this duration for all

stops and all four speakers. The vertical axes show the stops split by voicing, the different shading indicates the place of articulation (see the legends to the right).



**Figure 2:** The average duration from stop release to the start of the following vowel (in ms) split according to voicing of the stops for all four speakers. Error bars indicate standard error.



**Figure 3:** The average amount of air from stop release to the start of the following vowel (in ml) split according to voicing of the stops for all four speakers. Error bars indicate standard error.



A one-factorial ANOVA<sup>2</sup> with voicing as independent variable and the duration as dependent variable showed that the voicing had a significant influence, both for all places calculated together and for each place of articulation calculated separately (for all places of articulation together  $F(1, 480) = 577.409$ ,  $p < 0.001$ ; for coronals  $F(1, 239) = 269.4$ ,  $p < 0.001$ ; for labials  $F(1, 119) = 207.682$ ,  $p < 0.001$ ; for velars  $F(1, 120) = 206.626$   $p < 0.001$ ). The analysis of the results in Figure 2 thus supports prediction (3a).

Similarly, the analysis of the results presented in Figure 3 supports prediction (3b), because the voiceless stops all result in a larger amount of airflow over the friction duration than the voiced stops, both calculated for all three places of articulation together and separately (for all together  $F(1, 480) = 1018.969$ ; for coronals  $F(1, 239) = 511.068$ ,  $p < 0.001$ ; for labials  $F(1, 119) = 333.705$   $p < 0.001$ ; for velars  $F(1, 120) = 200.181$   $p < 0.001$ ).

These two results taken together give evidence in support of the assibilation hierarchy in (4).

### **3.2 *Influence of the following vocoid***

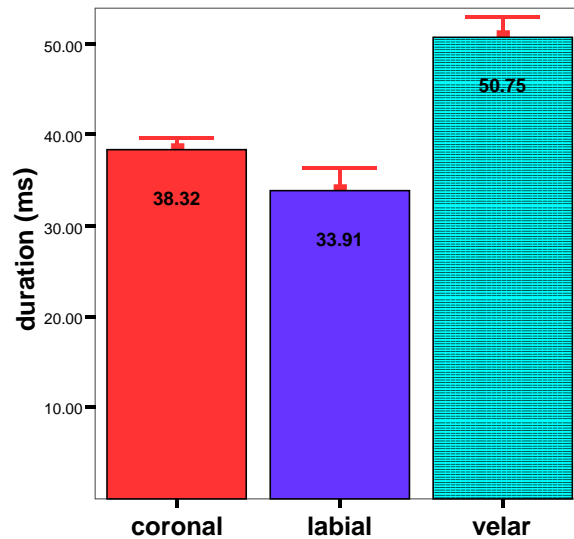
The influence of the following vowels /i, ɪ, ε, ʊ, a/ and the glide /j/ averaged over all four speakers are shown in the following two figures. The friction duration (in ms) is given in Figure 4 and the amount of air (in ml) over this duration in Figure 5. The vertical axes give the stops split according to the vocoid context.

A post-hoc Scheffé test showed that only the influence of the following glide on the duration of friction (as represented in Figure 4) is significantly different from the influence of all other contexts. The difference between the vowel /i:/ and /ʊ/ is almost significant ( $p < 0.007$ ). The analysis of the amount of air split according to the vowel context, as represented in Figure 5, did not yield any statistically significant results.

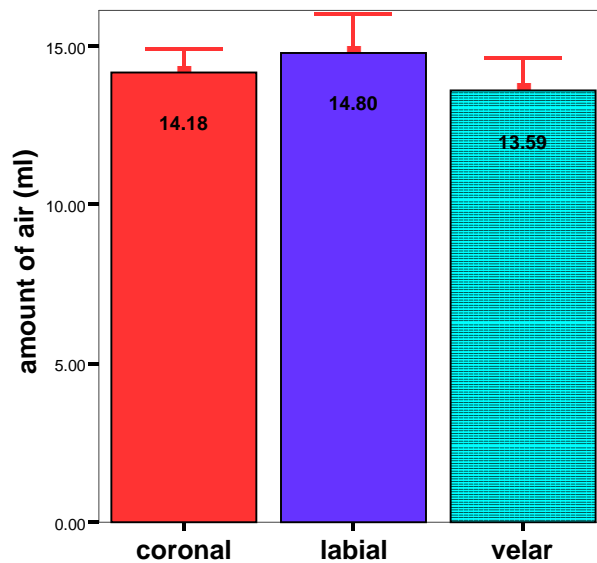
It has to be pointed out that the investigation of the influence of the vowels /a, e, ʊ/ was restricted to coronals (cf. the item set in Table 1).

---

<sup>2</sup> All statistical calculations were made in SPSS 11.5.1.



**Figure 4:** The duration from stop release to the start of the following vowel (in ms) split according to voicing of the stops for all four speakers. Error bars indicate standard error.

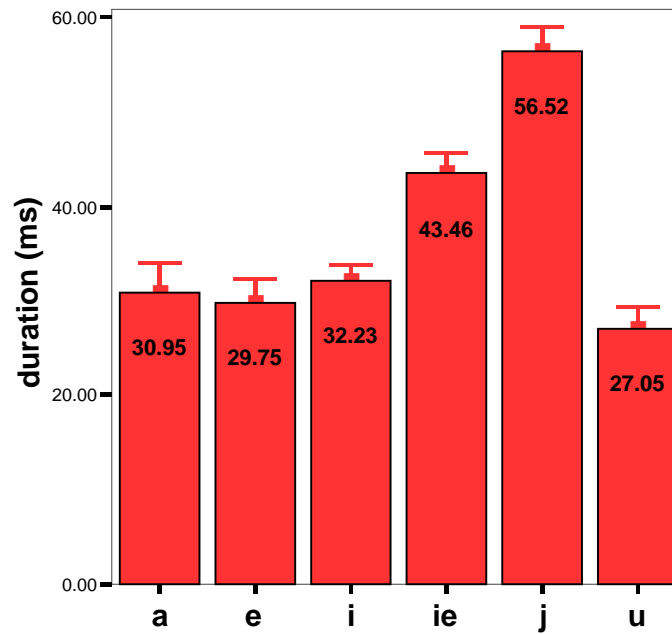


**Figure 5:** The average amount of air from stop release to the start of the following vowel (in ml) split according to voicing of the stops for all four speakers. Error bars indicate standard error.

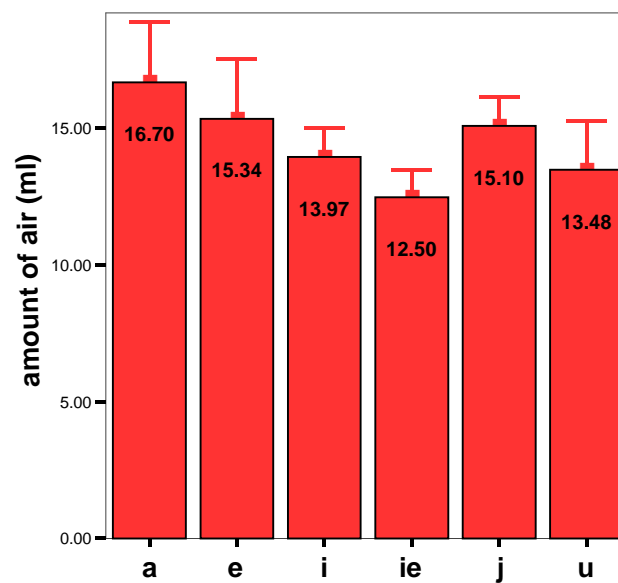
### 3.3 Influence of the place of articulation

Figures 6 and 7 on the next page show the average duration (in ms) for the friction phase and the average amount of air (in ml) over this duration,

respectively, for all places of articulation and all four speakers. The vertical axes show the stops split according to their place of articulation.



**Figure 6:** The duration from stop release to the start of the following vowel (in ms) split according to voicing of the stops for all four speakers. Error bars indicate standard error.



**Figure 7:** The average amount of air from stop release to the start of the following vowel (in ml) split according to voicing of the stops for all four speakers. Error bars indicate standard error.

The difference in duration (Figure 6) between coronal and velar place of articulation is statistically significant ( $p < 0.001$ ) and so is the difference between labial and velar place of articulation ( $p < 0.001$ ). The difference between coronals and labials is not significant (all results obtained by a post-hoc Scheffé test). None of the differences in amount of air (Figure 7) are significant.

#### **4 Summary and discussion**

The present study experimentally tested the influence of stop voicing, the following vowel and the place of articulation on the possible assibilation of a stop. We measured both the duration of airflow and the amount of air from release of a stop to the beginning of the following vowel in stop vowel-sequences of German.

Both in the duration measurement and the measurement of the amount of airflow a statistically significant difference was found between voiced and voiceless segments. The present study thus reproduced the findings by Hall et al. (2003), where the coronal voiceless segments showed a longer duration of the release phase than their voiced counterparts. In addition, the difference in duration and amount of air for the voicing condition could be established for the labial and velar places of articulation. This gives evidence for the partial hierarchies established in (4), repeated here in (7):

- (7)  $p > b$   
       $t > d$   
       $k > g$

The findings on the difference in release duration between voiced and voiceless stops are in accordance with the literature, see e.g. Isshiki & Ringel (1964), Klatt et al. (1968), and Warren (1996). It can be accounted for with the fact that the vocal fold vibration impedes the flow of air and consequently the duration and amount of air in the friction phase, see the discussion of Hall et al. (2003) in section 1.

For the influence of the following vocoid, only one context was significantly different from all others, namely that of the following glide /j/. This finding holds only for the duration of friction, the amount of air did not significantly differ between any of the vocoids. The assibilation hierarchy on the vocoid influence in (5) has to be changed accordingly to the one in (8).

- (8)  $j > \{i:, \text{ɪ}, \text{ʊ}, \text{e}, \text{a}\}$

This hierarchy is again in accordance with the findings by Hall et al. (2003). The fact that the quality of the other following vowels does not matter is not expected, see, however, similar results in Klatt et al. (1968: 45). The small number of tokens with the vowels /a, e, u/ might be responsible for these findings.

The influence of the place of articulation was mainly not significant, only the duration measurement showed a significant difference between coronals or labials and velars. The assibilation hierarchy for place of articulation in (6) therefore has to be modified in the following way:

(6) velar > {coronal, labial}

Klatt (1975) and Keating, et al. (1980) found a difference in voice onset time (VOT) for voiceless plosives that is similar to the present durational hierarchy. The measure of friction duration employed in the present study is identical to VOT, but only for voiceless stops. According to Keating et al. the duration in VOT is “somewhat larger for alveolars than for labials and substantially larger for velars than for either” (p.93). Thus our present durational findings confirm those of previous studies. The hierarchy in (6) could not be attested with the measurements on the amount of air.

Summing up, there is durational evidence for the assibilation hierarchies established in Hall & Hamann (to appear), namely the difference in influence between voiced and voiceless stops and the difference in influence of the following glide and high front vowel on the likelihood of assibilation for the stop. We could not only confirm a difference for coronals (as in Hall et al.’s measurements) but also for velars and labials. And in addition to the durational differences, we found statistical differences in the amount of air depending on the place of articulation. For the special status of the glide /j/ in assibilation processes, our durational measurements attested this (again supporting Hall et al.’s study). The difference in the amount of air for glide versus non-glide context did not prove to be significant.

In general, the present study showed that the amount of air seems not to be a reliable predictor for assibilation processes, although this assumption has to be further tested with studies that involve larger samples.

## **Acknowledgements**

We would like to thank Jana Brunner, Christian Geng and Daniel Pape for helpful comments, and Jörg Dreyer for technical support. We gratefully

acknowledge funding by the German Science Foundation (DFG) grant GWZ 4/8-1-P2 to Silke Hamann.

## References

- Boersma, P. & Weenink, D. (2005). Praat: doing phonetics by computer (version 4.3.27) [computer programme], Retrieved from <http://www.praat.org/>.
- Hall, T. A. (to appear) Derived Environment Blocking Effects in Optimality Theory. *Natural Language and Linguistic Theory*.
- Hall, T. A. & Hamann, S. (to appear) Towards a typology of stop assibilation. *Linguistics*.
- Hall, T. A., Hamann, S. & Zygis, M. (2004). The phonetic motivation for phonological stop assibilation. *ZAS Working Papers in Linguistics*, 37: 187-219.
- Hamann, S. (2003). German glide formation functionally viewed. *ZAS Working Papers in Linguistics*, 32: 137-154.
- Isshiki, N. & Ringel, R. (1964). Airflow during the production of selected consonants. *Journal of Speech and Hearing Research*, 7: 233-244.
- Jäger, J. (1978). Speech aerodynamics and phonological universals. In: *Proceedings of the 4th Annual Meeting of the Berkeley Linguistic Society*, Berkeley: 311-329.
- Keating, P. A., Westbury, J. R. & Stevens, K. N. (1980). Mechanisms of stop-consonant release for different places of articulation. *Journal of the Acoustical Society of America*, 67: 93.
- Kim, H. (2001). A phonetically based account of phonological stop assibilation. *Phonology*, 18: 81-108.
- Kiparsky, P. (1973). Abstractness, opacity and global rules. In: O. Fujimura (ed.) *Three Dimensions of Linguistic Theory*. Tokio: Taikusha; 57-86.
- Klatt, D. H. (1975). Voice onset time, frication and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research*, 18: 686-706.
- Klatt, D. H., Stevens, K. N. & Mead, J. (1968). Studies of articulatory activity and airflow during speech. *Annals of the New York Academy of Sciences*, 155: 42-54.
- Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In: P. F. MacNeilage (ed.) *The production of speech*. New York: Springer; 189-216.
- Warren, D. W. (1996). Regulation of Speech Aerodynamics. In: N. J. Lass (ed.) *Principles of Experimental Phonetics*. St Louis: Mosby; 46-92.

# **Late development of the categorical perception of speech sounds in pre-adolescent children**

**Victoria Medina**

*UFR- Linguistique, Université Denis Diderot (Paris 7) France*

*LPE, UMR8581 CNRS & Université René Descartes (Paris 5) France*

**Willy Serniclaes**

*LPE, UMR8581 CNRS & Université René Descartes (Paris 5) France*

---

While the perilinguistic child is endowed with predispositions for the categorical perception of phonetic features, their adaptation to the native language results from a long evolution from the end of the first year of age up to the adolescence. This evolution entails both a better discrimination between phonological categories, a concomitant reduction of the discrimination between within-category variants, and a higher precision of perceptual boundaries between categories. The first objective of the present study was to assess the relative importance of these modifications by comparing the perceptual performances of a group of 11 children, aged from 8 to 11 years, with those of their mothers. Our second objective was to explore the functional implications of categorical perception by comparing the performances of a group of 8 deaf children, equipped with a cochlear implant, with normal-hearing chronological age controls. The results showed that the categorical boundary was slightly more precise and that categorical perception was consistently larger in adults vs. normal-hearing children. Those among the deaf children who were able to discriminate minimal distinctions between syllables displayed categorical perception performances equivalent to those of normal-hearing controls. In conclusion, the late effect of age on the categorical perception of speech seems to be anchored in a fairly mature phonological system, as evidenced the fairly high precision of categorical boundaries in pre-adolescents. These late developments have functional implications for speech perception in difficult conditions as suggested by the relationship between categorical perception and speech intelligibility in cochlear implant children.

---

## **1 Introduction**

Categorical perception (CP) is the phenomenon by which differences between stimuli are not perceptible except if they belong to different categories (Liberman, Harris, Hoffman & Griffith, 1957). The functional interest of CP in speech perception is to filter out irrelevant information for the recognition of lexical units. CP takes part at two levels of speech treatment, a phonetic level and phonological one (Serniclaes, 2000). The CP of phonetic features makes it possible to neutralize acoustic differences in the realization of the same phonetic category in different contexts (for example, the difference in the acoustic salience of the burst vs. formant transitions in /ki/ vs. /ka/). The CP of phonological features allows to neutralize phonetic differences in the realization of the same phonological category in different contexts (for example, the difference between velar vs. palatal place of articulation in /ku/ vs. in /ki/).

The perception of phonetic features is largely innate. Prelinguistic children are able to perceive all the phonetic contrasts of the world's languages, even those which do not exist in their linguistic environment (for a review: Vihman, 1996). Categorical perception was first found for a voicing continuum in babies between 1 month and 4 months of age (Eimas, Siqueland, Jusczyk & Vigorito, 1971), the children reacting to a 20 ms VOT difference when accompanied by a phonetic difference (a change from /ba/ to /pa/). However, the children hardly did react to the same acoustic difference when it was not accompanied by a phonetic difference.

The linguistic environment has a crucial influence on categorical perception development. Categorical perception changes in the first year of life and adapts itself to the phonemic oppositions of native language (Werker & Tees, 1984a; Werker & Logan, 1985; Werker, 2003). The native language contrasts become more categorical than foreign ones. Other discrimination data show that categorical perception evolves between 2 and 6 years (Burnham, Earnshaw & Clark, 1991). Finally identification data suggest that categorical performances still evolve between 6 and 12 years (Hazan & Barret, 2000). However, the performances assessed in identification experiments do not pertain to categorical perception (CP) but on the precision of the categorical boundary or "Boundary Precision" (Serniclaes, submitted).

The progressive evolution of categorical perception from childhood to adolescence probably has beneficial consequences for spoken communication. The increase in CP makes that within-category differences become less discriminated, thereby preventing non-relevant information to reach the mental lexicon. This should facilitate word recognition, especially under difficult listening conditions. The effect of age on categorical perception should therefore enhance communication.



The abilities to communicate in deaf children might depend on their CP performances. The cochlear implant (CI) improves hearing, but communication abilities also depend on different other factors, mainly on deafness duration without implant and implantation age (Miyamoto, Osberger, Todd, Robbins, Stroer, Zimmerman-Phillips & Carney, 1995). These factors seem to act on the development of the phonological level before implantation, as suggested by the effect of the size of the phonemic repertory before implantation on the rate of development after implantation (Serniclaes, Ligny, Schepers, Renglet, & Mansbach, 2002). A shorter amount of hearing deprivation has the virtue of preserving phonetic predispositions and their developmental potential. However, the precise link between phonological development and speech communication performances remains unknown. One possibility is that phonological development merely depends on the amount of exposure to speech sounds. This would be the case if the innate potential for categorical perception remained intact during deprivation. However, another possibility is that innate capacities need to be activated during some sensitive period in order to preserve their categorical properties. Hearing deprivation would then affect categorical perception to a degree which depends on the duration of deprivation and on the moment at which it occurs during language development. The question then is whether speech communication performances in deaf children with cochlear implants depend or not on their degree of categorical perception of speech sounds.

The first purpose of the present study was to confirm the effect of age on the development of both categorical perception and boundary precision by comparing identification and discrimination performances in a group of pre-adolescents, aged from 8 to 11 years, with those of their mothers (experiment 1). Our second purpose was to explore the functional implications of CP by comparing the performances of a group of 8 deaf children, equipped with a cochlear implant, with normal-hearing chronological age controls (experiment 2).

## **2 Experiment 1**

### **2.1 Method**

#### **2.1.1 Participants**

Two groups of native French speakers took part to this study. One group included 10 normal-hearing children (7 boys and 3 girls) aged from 8 to 11 years (average age: 9.3 years, SD: 0.8) and were attending normal schools (classes from the second to the fourth grade). The second group included 10 normal-

hearing adults who were the mothers of these children. The mothers group was aged from 35 to 50 years (average age: 43.2 years, SD: 5.5). All the participants had a normal audition level, as indicated by an audiometric test.

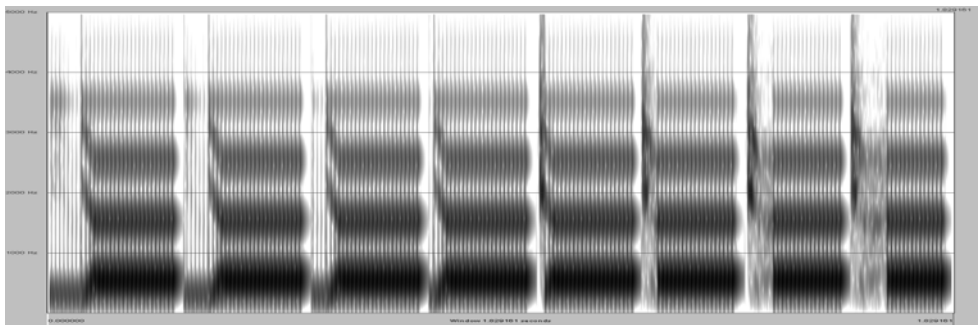
### 2.1.2 Stimuli

#### **Minimal Pairs of Perception and Speech Production Evaluation Protocol with Standardized Stimuli (PEPS).**

The protocol was a modified version of the “Evaluation Test of Speech Production and Perception” (Test d’Evaluation de la Production et de la Perception de la Parole, TEPPP: Vieu, Mondain, Sillon, Piron & Uziel, 1999). We used the 2 simplified lists of CV syllables recorded of a French speaker. These lists assessed several features per pair (e.g. labiality and frontness, voicing and nasality) and each list had 3 different pairs (e.g. /fo/-/fa/ or /sa/-/ka/) and 3 similar pairs (e.g. /fa/-/fa/ or /sa/-/sa/) for the vowels and the consonants. The vowels were presented in /f/ context and the consonants in /a, u, i/ contexts.

#### **Voicing continuum for the categorical perception test**

CP tests were based on a /də/-/tə/ voicing continuum, composed of 8 stimuli differing in VOT, from -70 ms to +70 ms, by 20 ms steps (figure 1). The stimuli were generated by modulated sinewave synthesis using software implemented by R. Carré (CNRS, France). The starting frequencies of F1, F2 and F3 transitions were of 200, 2100 and 3100 Hz, respectively. The end values of the transitions were fixed at 500, 1500 and 2500 Hz respectively for F1, F2, and F3. The F0 was fixed to 120 Hz, transition duration was 24 ms and the stable vocalic duration was 180 ms.



**Figure 1:** Voicing continuum from /də/ to /tə/. The synthetic stimuli varied along a VOT continuum from -70 to +70 ms in 20 ms steps.

### **2.1.3 Procedure**

The procedure comprised three successive stages: minimal pairs test, discrimination continua endpoints after training, CP tests. The stimuli were delivered through headsets.

The purpose of the training phase was to check correct discrimination of the endpoints of the /də/- /tə/ continuum. Participants had to collect at least 75% correct responses in order to be admitted to the CP test. The latter included 2 tasks, an identification task and a discrimination task, both run with a computer program ("Percep", realized by R. Carré, CNRS). The identification task was run first. Each stimulus was presented 10 times in a pseudo-random order and participants had to identify the stimuli as either /də/ or /tə/. The identification responses were given by pressing one among two different colored keys on the computer keyboard. In the discrimination task, the stimuli were presented by pairs (AX format), comprising either different stimuli (in two different orders: e.g. -70 ms VOT followed by -50 ms VOT, or -50 ms VOT followed by -70 ms VOT) or the same stimulus presented twice (e.g. two times -70 ms VOT, or two times -50 ms VOT). There were 14 different pairs (7 stimulus combinations x 2 orders) and 8 same pairs. Each pair was presented 5 times in a pseudo-random order and participants had to respond by answering either "same" or "different". The answers were delivered by pressing either the "M" key for same (i.e. "même" in French) or the "D" key for different on the keyboard. The total procedure took approximately 60 minutes.

Categorical Perception (CP) was assessed by comparing the observed discrimination scores with those expected from the labeling data, the latter being computed with elementary probability formulas (adapted from Pollack & Pisoni, 1971). The degree of CP is inversely related to the size of the difference between the observed and expected discrimination scores. Boundary precision (BP) tests were based on the slope of the labeling curve, a shallower slope indicating lesser precision. The slope was assessed separately for each participant using Logistic Regression (Mc Cullagh & Nelder, 1983). Individual assessments of slopes were then used for testing the difference between groups with ANOVA.

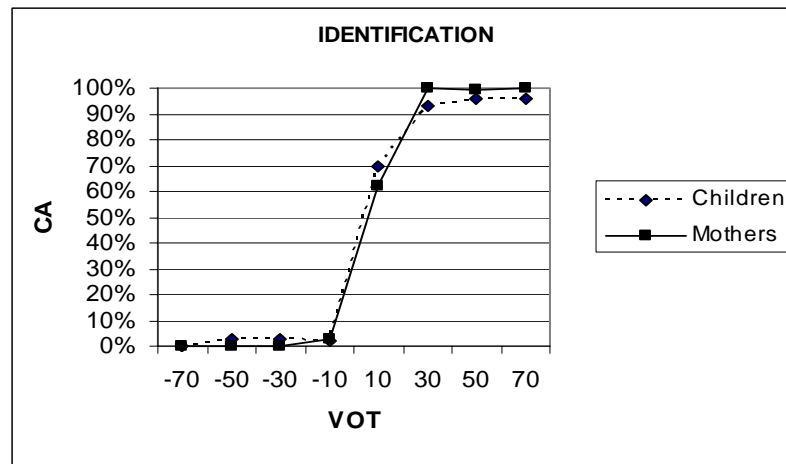
## **2.2 Results**

Minimal pairs were perfectly discriminated by all participants. The endpoints of the /də/-/tə/ voicing continuum were correctly discriminated with scores above 75% by all participants. Consequently, we decided to include all the participants in the CP and BP assessment tests.

### 2.2.1 Categorical Perception test (/də-/ /tə/ voicing continuum)

#### Identification:

Examination of the average identification curves of the mothers and children (figure 2) indicates that slope is slightly steeper for the mothers. Comparison between the mean slopes, separately assessed for each participant, showed a significant difference between groups ( $F(1.9) = 7.42$ ,  $p < 0.05$ ). The VOT boundary was close to 5 ms VOT for both groups ( $F < 1$ ).

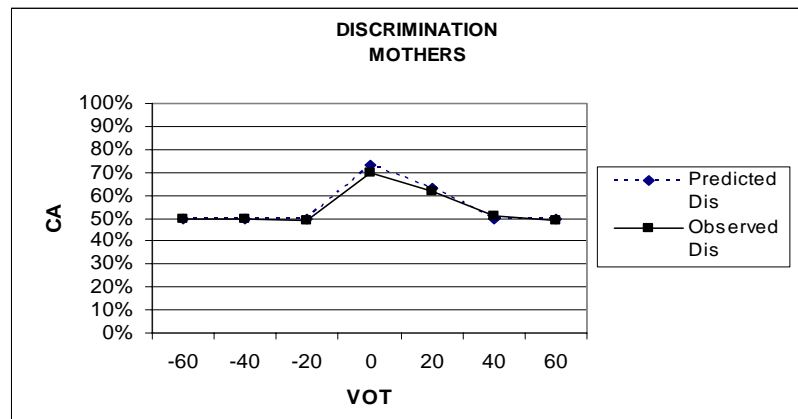


**Figure 2:** Identification of /də-tə/ continuum for the mothers and children.

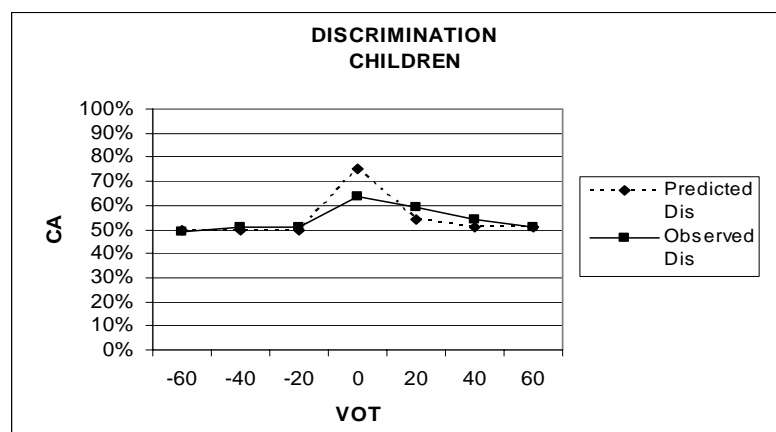
#### Discrimination:

Figures 3 and 4 show the predicted discrimination and observed discrimination functions for the mothers and children, respectively. The predicted and observed functions were quite similar for the mothers. Discrimination scores were at chance level (50%), except for the 0 and 20 ms pairs. These two pairs surround the phonemic boundary (at 5 ms). Predicted and observed functions were different for the children. The predicted score was larger than the observed one for the 0 ms pair while the observed score was larger than the predicted one for the 20 ms pair. Arcsine transforms of the scores were tested in a repeated measures ANOVA with age (mother vs. child), score type (observed vs. predicted) and VOT as factors. As Mauchly's sphericity test was significant ( $p < .001$ ), Greenhouse-Geisser corrected F values were used. The main effect of VOT was significant ( $F(1.6, 14.7) = 16.8$ ,  $p < .001$ ), while those of score type and age were not significant (both  $F < 1$ ). The score type x VOT interaction was marginally significant ( $F(1.6, 14.0) = 3.58$ ,  $p = .065$ ), while the age x score type

and age x VOT were not significant (both  $F < 1$ ). The age x score type X VOT interaction was also marginally significant ( $F(1.6, 14.7) = 3.58$ ,  $p = .065$ ). The examination of contrasts for this latter interaction indicated that when the differences in observed vs. expected discrimination between 0 ms vs. 20 ms VOT in the children's results (figure 4) are compared to the absence of such differences in their mothers' results (figure 3), the difference was significant ( $F(1, 9) = 6.39$ ,  $p < .05$ ).



**Figure 3:** Predicted and observed discrimination of /də-tə/ continuum for mothers group. The VOT represents the average VOT of the stimuli in the pair (for example, -60 corresponds the -70 vs -50 ms VOT pair).



**Figure 4:** Predicted and observed discrimination of /də-tə/ continuum for children group.

## 2.3 Discussion

The results of this experiment confirm the effect of age on the precision of the phoneme boundary, previously reported by Hazan and Barret (2000). The results also evidence the presence of a late effect of age on categorical perception. For the adults in this study, observed discrimination scores were closely similar to the expected ones, reflecting almost perfect categorical perception. Children exhibited weaker than expected discrimination for a VOT pair centered on 0 ms VOT, the stimulus pair closest to the phoneme boundary in this experiment (5 ms VOT). Furthermore, higher than expected discrimination performance was observed for another stimulus pair centered on 20 ms VOT.

## 3 Experiment 2

### 3.1 Method

#### 3.1.1 Participants

Two groups of native French speakers took part to this study. One group of 11 normal-hearing children (8 boys and 3 girls) aged from 6 to 11 years (average age: 9 years, SD: 1.3) at a normal school (classes from the second to the fourth grade). This group included the 10 children who already participated in experiment 1.

The second group included 8 deaf children with cochlear implant (CI), among which 3 were under observation at Trousseau Hospital and 5 were under observation at Robert Debré Hospital, in Paris. These children (1 boy and 7 girls) were aged from 5 years 9 months to 11 years (average age: 7.6, SD: 1.6) with a minimum of 3 years of implantation (average duration of implantation use: 4.3, SD: 0.9). We did not use exclusion criteria in relation with either the origin of deafness or the cochlear implant type.

#### 3.1.2 Stimuli

#### **Minimal Pairs of Perception and Speech Production Evaluation Protocol with Standardized Stimuli (PEPS)**

We used the same simplified minimal pair test as in the experiment 1. The complete version of the minimal pair test was also used in the present experiment. The test included 2 completed lists which were based on CV syllables of French. The vowels list had 10 different pairs (e.g. /fo/-/fa/) and 10 similar pairs (e.g. /fa/-/fa/), and it assessed 4 features: aperture, frontness, labiality and nasality. The consonants list included 16 French consonants in /a,

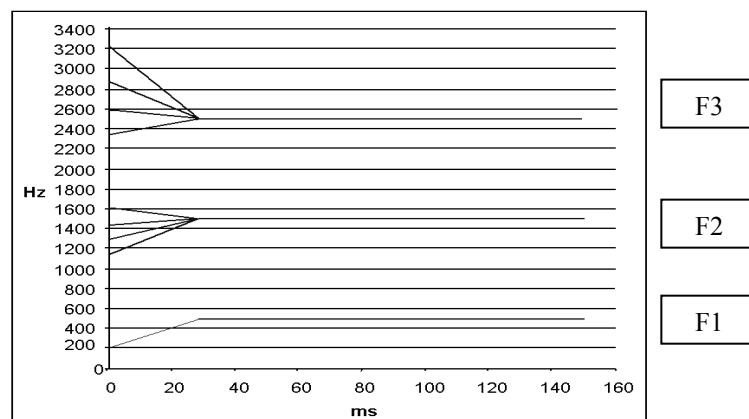
u, i/ context, 16 different pairs and 16 similar pairs. This list assessed 4 features: manner, place, voicing and nasality.

### **Voicing continuum from /də/ to /tə/ for the categorical perception test**

We used the same voicing continuum from /də/ to /tə/ as in the first experiment.

### **Place continuum for the categorical perception test**

In this experiment we also used a /bə/-/də/ place of articulation continuum, composed of 4 stimuli generated by modification of F2 and F3 transitions (figure 5). The stimuli were generated by modulated sinewave synthesis using software realized by R. Carré (CNRS, France). The F2 and F3 transition onset frequencies varied from 1168 Hz to 1604 Hz and from 2330 Hz to 3211 Hz, respectively. The F1 transition onset was fixed at 200 Hz. The offset transition values were fixed at 500, 1500 and 2500 Hz for F1, F2 and F3, respectively. A 10 ms friction noise, with pole frequencies equal to formant onset frequencies, preceded the onset of formant transitions. The F0 was stable at 100 Hz, the durations of negative VOT, formant transitions and vocalic stable segment were of 90, 27 and 154 ms, respectively.



**Figure 5:** Place continuum from /bə/ to /də/. Stimuli generated by sinewave synthesis, with F2 and F3 modifications.

### **3.1.3 Procedure**

The same three-stage procedure as in experiment 1 was used: minimal pairs test, discrimination continua endpoints after training, CP tests.

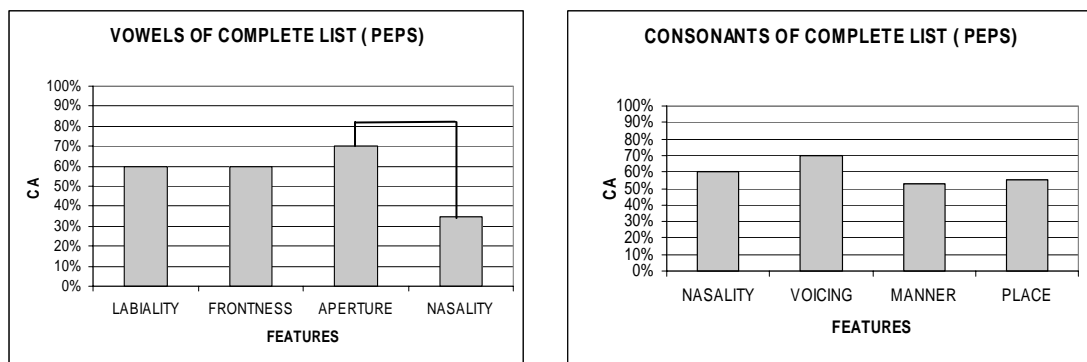
### 3.2 Results

All the control children passed the selection stages with 100% correct responses for the simplified minimal pair discrimination tests and with above 75% correct discrimination of the endpoints of the /də/-/tə/ voicing continuum.

However, of 8 CI children, 4 did not pass the simplified minimal pair test and 2 did not pass the continuum endpoint discrimination test. These 6 CI children were given the complete minimal pair test. The 2 CI children who passed both preliminary discrimination tests were given the categorical identification and discrimination tests instead.

#### Minimal Pairs (PEPS)

*Simplified lists.* The success average for the control children group was 100 % CA. The average of CI group was 75 % CA. The difference between 2 groups was significant (Mann-Whitney  $U=5.5$ ;  $p=0.01$ ).



**Figure 6:** Vocalic features (on the left) and consonant features (on the right) of the complete minimal pairs test (PEPS).

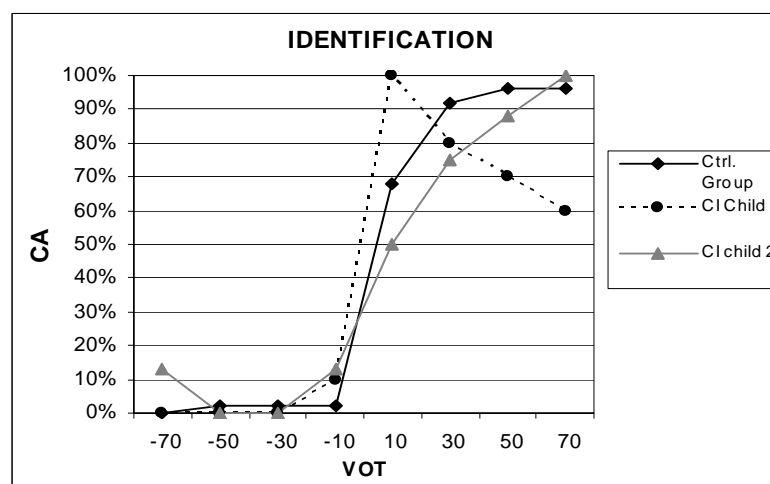
*Complete Lists.* Six CI children received the complete lists of the PEPS. For the vowels, aperture collected the highest score with 70% correct discrimination (figure 6). Vowel labiality and frontness obtained 60% correct discrimination each. The feature most difficult to discriminate was nasality with only 35% correct responses. Although discrimination differences between vowel features were not significant ( $F(3,6) = 3.12$ ;  $p=0.10$ ), the examination of individual contrasts indicated a significant difference between aperture and nasality ( $F(3,6) = 3.12$ ;  $p<0.05$ ). The results for consonants showed that the best discriminated features were voicing and nasality with 70% and 60% scores respectively. Manner and place were less well discriminated with scores of 53%



and 55% respectively. However, discrimination differences between consonant features were not significant ( $F < 1$ ).

### Voicing continuum - Identification:

Examination of the average identification curves of the control children and the curves of the two who were given the CP tests (figure 7) indicates that the slope was shallower for the CI children vs. the mean of the controls. The VOT boundary of both CI children was similar to that of the control group (about 5 ms VOT). However, voiceless answers of CI child 1 decreased at longer positive VOTs. Table 1 presents the slopes of the two CI children in regard with the distribution of the controls. The distribution of the slopes in usual logit units was positively skewed, logarithmic transforms were used. As can be seen, the slopes of the two CI children fall inside the distribution of the controls (less than 2 SD differences).



**Figure 7:** Identification of voicing continuum from /də/ to /tə/ for the control group (11 children) and two CI children.

### Voicing continuum - Discrimination:

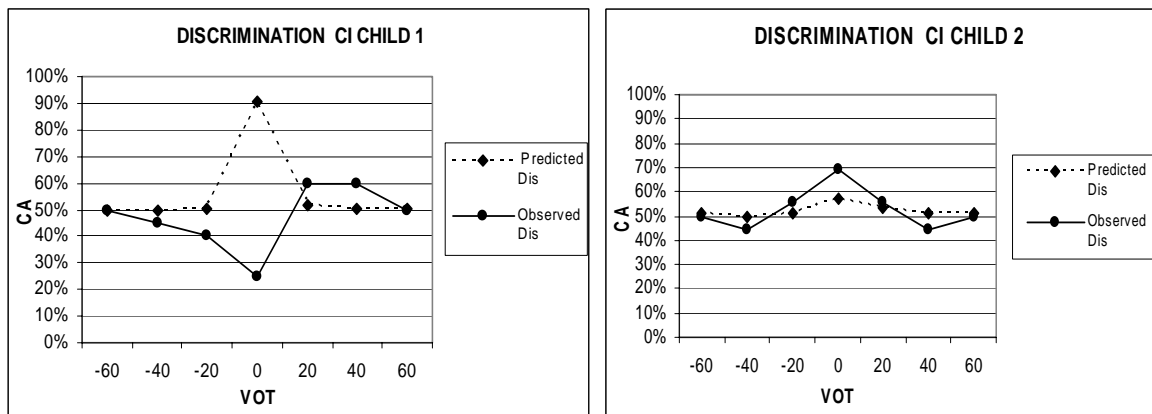
Predicted and observed functions along the VOT continuum for the two CI children are presented in figure 8. The functions of the 11 control children in this experiment were quite similar to those of the subgroup of 10 children already included in experiment 1 for the mother-child comparison (figure 2). CI child 1 presented a drop of observed discrimination instead of an increase around 0 ms. Examination of responses to similar and different pairs revealed that this was due to the conjunction of a bias towards “similar” responding for pairs of stimuli with negative VOT with a bias towards “different” responding

for pairs of stimuli with positive VOT. CI child 2 presented a better observed discrimination than the predicted discrimination.

**Table 1:** Slopes of labeling functions for the /də-tə/ voicing continuum.

	Slope in log (logits)
Control Group Average and SD	1.01 (1.04)
CIC 1 and difference from control's mean in SD units	-0.24 (-1.2)
CIC 2 and difference from control's mean in SD units	0.11 (-0.9)

Table 2 presents the discrimination results of the two IC children in comparison with the distribution of the controls. Data are summarized in terms of the “Phoneme Boundary Effect” (PBE: Wood, 1976), i.e. the difference in observed discrimination between the across-boundary pair and the mean of within-category pairs. The PBE of Child IC 1 falls outside the distribution of controls, while the one of child IC 2 is inside the distribution.



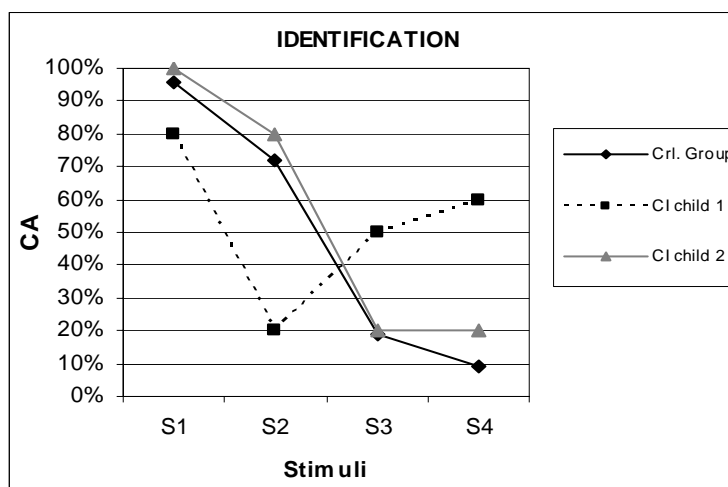
**Figure 8:** Predicted and observed discrimination on the /də-tə/ continuum for the CI children 1 and 2.

**Table 2:** Phoneme Boundary Effect for the /də-tə/ voicing continuum.

Pairs	PBE	Between-category	Within-category
Control Group Average and SD	17% (9.6)	69% (10.4)	51% (2.6)
CIC 1 and difference from control's mean in SD units	-7% (-2.5)	25% (-4.2)	51% (0)
CIC 2 and difference from control's mean in SD units	14% (0.3)	69% (0)	50% (-0.4)

### Place continuum - Identification:

As can be seen in figure 9, the labeling function of CI child 2 is similar to the one of the controls, while the function of CI child 1 is quite irregular and different from the one of the controls. However, the slopes of both CI children fall inside the distribution of the controls (table 3).



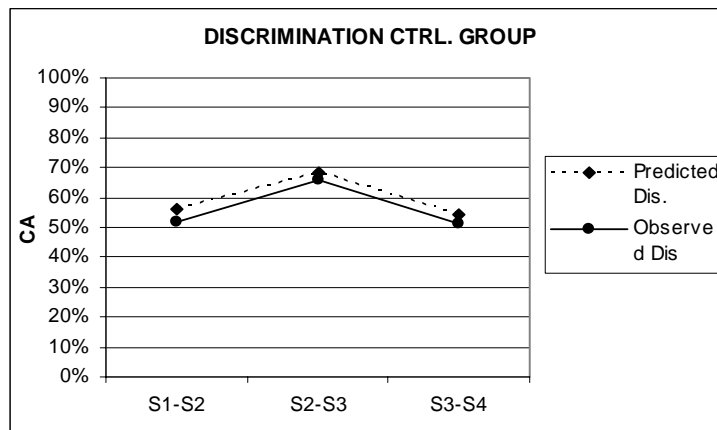
**Figure 9:** Identification of place continuum from /bə/ to /də/ for the control group (11 children) and two CI children.

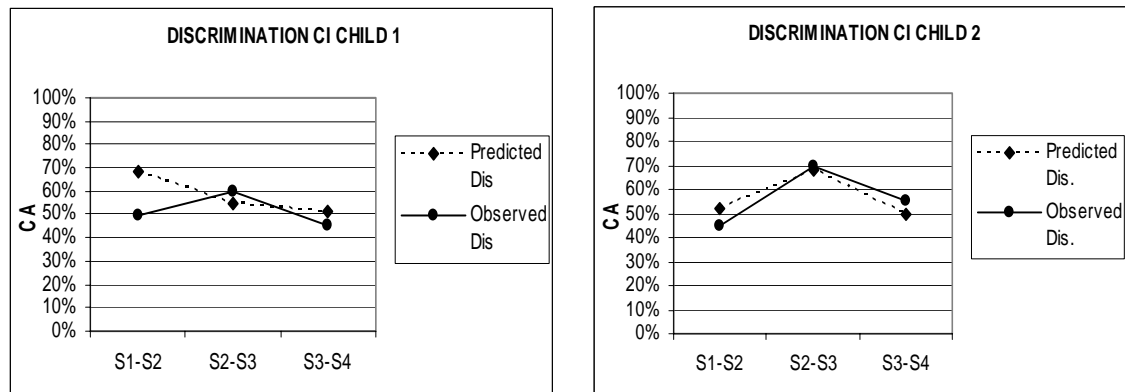
**Table 3:** Slopes of labeling functions for the /bə- də/ place continuum.

	Slope in logits
Control Group Average and SD	-10.4 (9.7)
CIC 1 and difference from control's mean in SD units	-0.12 (-1.1)
CIC 2 and difference from control's mean in SD units	1.79 (-0.9)

Place continuum - Discrimination:

For the control group, there was an increase in observed discrimination around the boundary identification (figure 10, pair S2-S3), and the predicted and the observed discrimination functions were fairly similar. The predicted and the observed discrimination functions of CI child 1 were quite different (figure 11). The predicted discrimination varied considerably compared to predicted discrimination of the control group, though the observed discrimination was similar to that of control group. The predicted and observed discrimination functions of CI child 2 were similar. As shown in table 4, for both CI children the PBE was inside the distribution of the controls.

**Figure 10:** Predicted and observed discrimination of /bə-də/ continuum for the control group.



**Figure 11:** Predicted and observed discrimination of /bə-də/ continuum for the CI children 1 and 2.

**Table 4:** Phoneme Boundary Effect for the /bə-də/ place continuum.

Pairs	PBE	Between-category	Within-category
Control Group Average CA and SD	17% (12.1)	68% (12.5)	51% (1.7)
CA CIC 1 and difference from control's mean in SD units	13% (-0.3)	60% (-0.6)	52% (0.6)
CA CIC 2 and difference from control's mean in SD units	20% (0.3)	70% (0.2)	57% (3.5)

### 3.3 Discussion

As we have mentioned in an earlier paper (Medina, Loundon, Busquet, Petroff & Serniclaes, 2004), the CI children in this study have a development of speech perception performances in relation to the duration of cochlear implant use and the chronological age. The 4 children who failed minimal pairs discrimination test had an average of 3.8 years of CI use vs. 4.8 years for the 4 successful children. For the discrimination of the VOT and place continua endpoints, mean durations of CI use were 4.6 years for the 2 children who failed vs. 4.10 years for the 2 successful children. The two children who passed the VOT and place CP tests had almost the same duration of CI use (5.1 years and 4.9 years), although the most categorical one (CI child 2) was 1 year older than the less categorical child (CI child 1). Notice however that there was a significant correlation between chronological age and age of implantation ( $r=0.85$ ,  $p<0.001$ ).

### 3.3.1 *Minimal Pairs*

Compared to normal-hearing controls, CI children show difficulties in minimal pair discrimination. For CI children the vocalic feature most difficult to discriminate is nasality and the easiest one is aperture. This does not seem to be specific to the speech perception with CI. The examination of vowel perception by normal-hearing subjects suggests that aperture is more perceptible than other vocalic distinctions (frontness, labiality and nasality; Papçun, 1980). Differences in perceptibility also exist between consonant features, voicing being more solid than other consonant features (Miller and Nicely, 1955). Furthermore, for deaf subjects without CI, voicing is the most robust feature (Vickers et al., 2001) and the same trends appear in the present experiment with CI children, although not in a significant way. To increase statistical power, it would be interesting to continue this study with a greater number of observations by feature.

### 3.3.2 *Categorical Perception*

The control children are characterized by fairly similar predicted and observed discrimination performance, both for the voicing and place continua used in this experiment. This is also found for one of the two CI children (CI 2) who successfully passed the minimal pair test. The discrimination functions of the other CI child (CI 1) are somewhat different from those of the control group. However, these differences arise from factors not directly related to CP. For the voicing continuum, CI 1 adopted a strategy consisting in answering “similar” to the pairs with negative VOT and to answer “different” those with positive VOT. These two strategies cross near to 0 ms with a consequence of a fall to correct answers below 50%. This strategy is probably due to the fact that the negative VOT differences are less audible than positive VOT differences for this child. If the child did not perceive the differences between negative VOT stimuli, but well those between positive VOT stimuli, she might have adopted a strategy consisting in giving “similar” responses to the negative VOT pairs and “different” responses to positive VOT pairs, perhaps to balance the total number of similar and different responses. If this is correct, the absence of a discrimination peak on the voicing continuum for this subject would not be due to a lack of categorical perception, but to the reduced audibility of fairly small (20 ms) VOT differences, especially those involving negative VOT. For the place of articulation continuum, this subject (CI 1) presents an observed discrimination peak similar to the one of the control group. This suggests that this subject is endowed with CP, although the observed discrimination function is somewhat different from the expected one. However, the discrepancy between

observed and expected discrimination scores is due to a somewhat erratic labeling function.

### *3.3.3 Boundary Precision*

For one of the two CI children who successfully passed the minimal pair test, the /də-tə/ and /bə-də/ labeling functions are not very different from those of the control group, suggesting similar boundary precision. The other CI child with successful minimal pair discrimination had fairly inconsistent labeling functions, though boundary precision was inside the distribution of control values.

### *3.3.4 Minimal pairs and Categorical Perception*

In this study, only the children who were successful in the minimal pairs test were given the categorical perception tests. The results suggest that children who can discriminate minimal pairs above some threshold (75% correct discrimination here) have categorical perception, although this has to be confirmed with a larger sample. While one can reasonably suppose that failure to discriminate minimal pairs means lack of categorical perception, this should also be verified. It is indeed possible to be endowed with categorical discrimination even with low labeling performances.

## **4 Conclusions**

This study provides some hints on the relationship between minimal pair discrimination, categorical perception and precision of the labeling boundary in both normal-hearing children and the deaf children with CI. The results of experiment 1 confirm the effect of age on boundary precision and show that age also contributes to improve categorical perception. The results of experiment 2 give the first insight on categorical perception in children with cochlear implants.

From these very preliminary results, it seems that CI children who can discriminate minimal phonological distinctions between syllables also possess categorical perception. This lends support to the idea of a functional link between speech intelligibility and categorical perception. However, this needs further confirmation with larger samples of CI children.

## Acknowledgments

We would like to thank Marie-Pierre Geron for her help and involvement in the experiment 1, and to Natalie Loundon, Denise Busquet and Nathalie Petroff for their participation in the experiment 2.

## References

- Burnham, D.K., Earnshaw, L.J. & Clark, J.E. (1991). Development of categorical identification of native and non-native bilabial stops: infants, children and adults. *Journal of Child Language*, 18, 231-260.
- Eimas, P.D., Siqueland, E.R., Jusczyk, P. & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303-306.
- Hazan, V. & Barrett, S. (2000). The development of phonemic categorization in children aged 6-12. *Journal of Phonetics*, 28, 377-396.
- Lieberman A.M., Harris, K.S. Hoffman, H.S. & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358-368.
- McCullagh, P. & Nelder, J.A. (1983). Generalized Linear Models. London: Chapman & Hall.
- Medina V., Loundon, N., Busquet, D., Petroff, N. & Serniclaes, W. (2004). Perception catégorielle des sons de parole chez des enfants avec Implant Cochléaire. *JEP-TALN-RECITAL*. Fès, Maroc. 19-22 April 2004.
- Miller, G.A. & Nicely, P.E. (1955). An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 27, 338-352.
- Miyamoto, R.T., Osberger, M.J., Todd, S.L., Robbins, A.M., Stroer, B.S. Zimmerman-Phillips, S. & Carney, A.E. (1995). Variables affecting implant performance in children. *Laryngoscope*, 104, 1120-1124.
- Papçun, G. (1980). Discriminate analyses of four imitation dialects. *UCLA WPP*, Feb. 80, 48.
- Pollack S. & Pisoni, D. (1971) On the comparison between identification and discrimination tests in speech perception. *Psychon.Science*, 24, 299-300
- Serniclaes, W. (submitted). "Allophonic perception in developmental dyslexia: origin, reliability and implications of the categorical perception deficit."
- Serniclaes, W. (2000). La perception de la parole. In P. Escudier, G. Feng, P. Perrier, J.-L. Schwartz (eds.) *La parole, des modèles cognitifs aux machines communicantes*. Paris: Hermès, 159-190.
- Serniclaes, W., Ligny, Ch., Schepers, F., Renglet, Th. & Mansbach, A.-L. (2002). *Evaluation du bénéfice thérapeutique de l'Implant Cochléaire à l'aide de mesures de production de la parole*. In W. Serniclaes (Ed.) *Méthodes d'évaluation des performances de l'Implant Cochléaire - Methods for the assessment of Cochlear Implant performances*. Brussels: Etudes et Travaux N° 5, ILVP –ULB, 31-48



- Vickers, D.A., Moore, B.C.J. & Baer T. (2001). Effects of low-pass filtering on the intelligibility of speech in quiet for people with and without dead regions at high frequencies. *Journal of Acoustical Society of America*, 110, 1164-1174.
- Vieu, A., Mondain, M., Sillon, M., Piron, J. P. & Uziel, A. (1999). Test d'Evaluation des Perceptions et Productions de la Parole (TEPPP). *Revue de Laryngologie, Otologie et Rhinologie*, 120, 219-225, 1999.
- Vihman, M.V. (1996) *Phonological development : The origins of language in the child*. Cambridge (MA): Blackwell.
- Werker, J.F. & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development*, 7, 49-63.
- Werker, J.F. & Logan, J.S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, 37, 35-44.
- Werker, J.F. (2003). Baby steps to learning language. *The Journal of Pediatrics*, 143, 62-69.
- Wood, C.C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America*, 60, 1381-1389.

# Acoustic differences between German and Dutch labiodentals

**Silke Hamann**

*Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany*

**Anke Sennema**

*Universität Potsdam, Potsdam, Germany*

---

The present article is a follow-up study of the investigation of labiodentals in German and Dutch by Hamann & Sennema (2005), where we looked at the perception of the Dutch labiodental three-way contrast by German listeners without any knowledge of Dutch and German learners of Dutch. The results of this previous study suggested that the German voiced labiodental fricative /v/ is perceptually closer to the Dutch approximant /ʋ/ than to the corresponding Dutch voiced labiodental fricative /v/. These perceptual indications are attested by the acoustic findings in the present study. German /v/ has a similar harmonicity median and a similar centre of gravity to Dutch /ʋ/, but differs from Dutch /v/ in these parameters. With respect to the acoustic parameter of duration, German /v/ lies closer to the Dutch /v/ than to the Dutch /ʋ/.

---

## 1 Introduction

Dutch has three labiodental segments, namely a voiceless fricative /f/, a voiced fricative /v/, and a voiced approximant /ʋ/ (Booij 1995, Gussenhoven 1999). Minimal triplets of the three sounds in word-initial position are given in (1).

- |     |      |      |      |                       |
|-----|------|------|------|-----------------------|
| (1) | /f/  | /v/  | /ʋ/  |                       |
|     | fee  | vee  | wee  | ‘fairy, cattle, ache’ |
|     | feil | vijl | wijl | ‘error, rasp, while’  |

Many speakers of Standard Dutch, apart from those from the Southern part of the Netherlands, neutralize the voiced and voiceless distinction for labiodental fricatives (as for all fricatives) word-initially, see Gussenhoven (1999, p.74).

German learners of Dutch usually have problems acquiring the three-way contrast since their native language differentiates only a voiced and a voiceless

labiodental fricative [v] and [f] (Kohler 1999, Wiese 1996), see the minimal pairs in (2).<sup>1</sup>

- (2)    /f/            /v/  
          fein        Wein        ‘fine, wine’  
          fort        Wort        ‘away, word’

In a perception experiment, Hamann & Sennema (2005) tested the categorisation of the Dutch labiodental contrast in a closed-set identification task by three groups of listeners, namely Dutch native listeners, German native listeners, and German learners of Dutch. Since the experiment tested not only the categorisation of the labiodentals, the German native listeners had all German consonants as response categories, and the Dutch listeners and the German learners of Dutch had all Dutch consonants as response categories.

**Table 1:** Mean identification scores (percent correct) of the three test groups in the perception experiment by Hamann & Sennema (2005), with stimuli in rows, and responses, sorted by language group, in columns. The numbers in each row per language group do not add up to 100 percent, because miscategorisations involving non-labiodental sounds are not included.

		German L1		Dutch L2			Dutch L1		
		/f/	/v/	/f/	/v/	/v/	/f/	/v/	/v/
stimulus	/f/	99.5%	0%	79.0%	17.7%	2.1%	94.8%	5.2%	0%
	/v/	16.7%	82.8%	5.2%	74.6%	18.5%	5.2%	94.8%	0%
	/v/	0%	99.5%	0.1%	6.1%	92.6%	0%	0%	99.5%

The results of the experiment (see Table 1) illustrate that the categorization of Dutch /f/ - /v/ - /v/ by German listeners departs from what would have been expected on the basis of the phonemic descriptions of these sounds. German listeners without knowledge of Dutch perceived the Dutch labiodental approximant as their voiced fricative in almost all of the cases, but the Dutch voiced fricative as their voiceless fricative only in 16.7 percent of the cases. Furthermore, German learners of Dutch appeared to have no problems perceiving the Dutch labiodental approximant correctly, even though they do not have such a category in their native language. At the same time, the German L2

<sup>1</sup> German has the grapheme <v>, which is used both for /v/ and for /f/, see e.g. *Vase* [va:zə] ‘vase’ and *Vieh* [fi:] ‘cattle’, respectively.

learners had problems perceiving the Dutch labiodental fricative, though they have the same category in their native language. These findings indicate that Germans acquiring Dutch set the Dutch approximant equal to their native voiced fricative.

To account for our findings in Hamann & Sennema (2005), we proposed that the mismatch in the perception of the Dutch labiodental fricative stems from German /v/ sharing more acoustic properties with the Dutch approximant /ʋ/ than with the corresponding Dutch labiodental /v/. In the following experiment we tested this hypothesis by comparing the acoustic characteristics of the German labiodentals /f, v/ with those of the Dutch labiodentals /f, v, ʋ/.

## **2 Experiment**

### **2.1 Subjects and material**

Subjects of our experiment were five female speakers of Dutch from Nijmegen, their age ranging from 20 to 34. For the German stimuli set, five female speakers of German from the Berlin area read the items. They ranged in age from 24 to 47. Speakers either volunteered for the experiment or they were paid for their participation.

The Dutch speakers read the monosyllabic nonsense words /pa, ba, ta, da, ka, xa, fa, va, ʋa, sa, za, ɕa/ ten times in randomized order in isolation. The German speakers read the sequences /pa, ba, ta, da, ka, ga, fa, va, sa, za, ɕa, ʃa/ ten times in randomized order in the carrier sentence “Sage ...”, ‘say ...’.<sup>2</sup> All words were presented orthographically (the Dutch sequence /Na/ was represented as <sja>, German /sa/ as <ssa> and /ɕa/ as <cha>). For both languages the whole set of obstruents was included in the reading material because speakers should not be aware of the contrast under investigation.

All speakers were recorded on a DAT recorder with an audio sampling frequency of 48 kHz, except for the recordings of two German speakers, which were digitised at a sampling rate of 22.05 kHz. Acoustic analyses of the recordings were performed with PRAAT (Boersma & Weenink 2005), statistical analyses were made with SPSS, version 12.

### **2.2 Acoustic parameters**

The duration of the labiodentals was computed by measuring the point of time of the consonant onset to the beginning of the continuous formants of the

---

<sup>2</sup> One Dutch speaker produced the sequences in the carrier sentence “Hoor je ...”, ‘Do you hear ...’, and one German speaker produced the sequences without a carrier sentence.

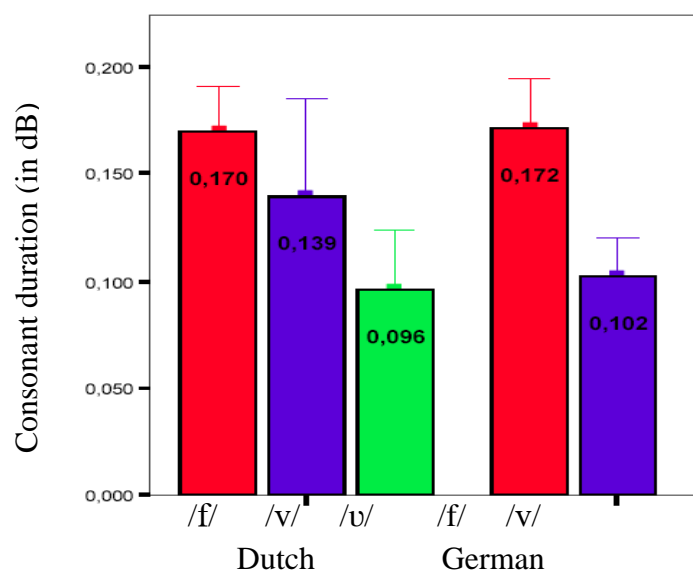
following vowel. To determine the beginning of the vowel after the Dutch approximant, the rise in amplitude was taken as additional criterion.

To compare the relation of voicing to friction, the degree of acoustic periodicity was determined by calculating the median of the harmonics-to-noise ratio for each labiodental with time steps of 0.01 s, a minimum pitch of 75 Hz, a silence threshold of 0.1 and 1 period per window. This measure is called “harmonicity median” in the following. A harmonicity median of 0 dB means that there is equal energy in the harmonics and in the noise of a signal, and a harmonicity median of 20 dB that there is almost 100% of the energy of the signal in the periodic part (Boersma 1993).

The spectral qualities of the labiodentals were compared by measuring the centre of gravity (see e.g. Jassem 1979 and Gordon, et al. 2002), which is the average of frequencies over the entire frequency domain weighted by the amplitude (with the power spectrum). Signals were high-pass filtered with a centre frequency of 500 Hz and a smoothing of 100 Hz to exclude the influence of the fundamental frequency in the voiced fricatives.

## 2.3 Results

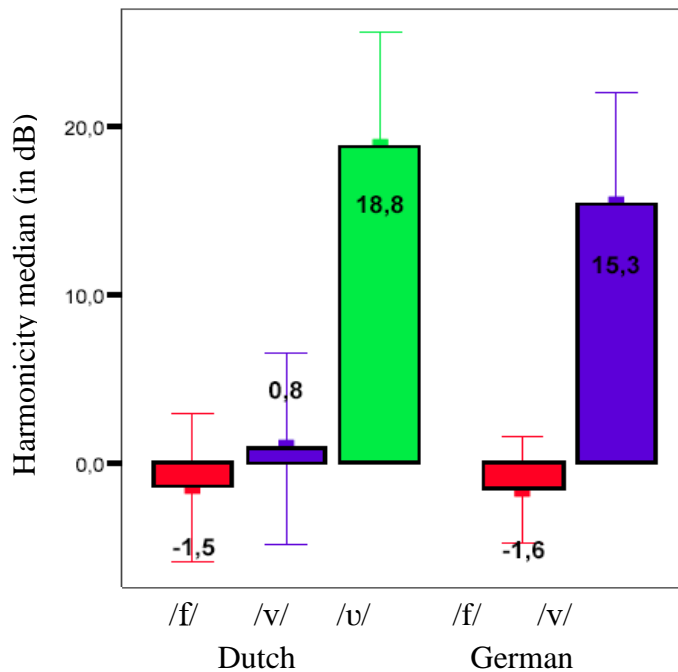
Mean values for the parameters consonant duration, harmonicity and gravity centres were computed across languages for subjects and items, and analyses of variance with language as fixed factor and acoustic parameters as dependent variable were carried out on these data. The mean duration (s) of the labiodental consonants are given in Figure 1.



**Figure 1:** Mean values for consonant duration (s). Error bars indicate standard deviation.

Results indicate that consonant duration was significantly different for both Dutch ( $F(2,12) = 14.24$ ,  $p < .001$ ) and German ( $F(1,8) = 49.81$ ,  $p < .0001$ ). Bonferroni-adjusted post-hoc tests for the Dutch data revealed significant differences between /f/ and /v/, and between /v/ and /ʋ/. A two-sided t-test showed no significant difference in duration between Dutch /f/ and German /f/, and between Dutch /ʋ/ and German /v/. The difference between Dutch /v/ and German /v/ failed to be significant ( $t(8) = -2.175$ ,  $p = .061$ ).

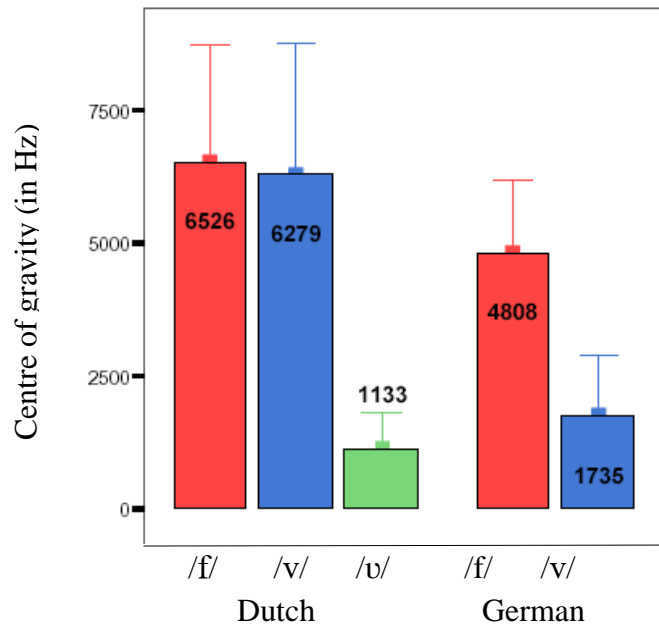
Average values of the harmonicity median (dB) for the three Dutch sounds and the two German sounds are given in Figure 2.



**Figure 2:** Mean values for the harmonicity median (in dB). Error bars indicate standard deviation.

Analyses of variance showed that harmonicity median within language was significantly different for both Dutch ( $F(2,12) = 48.46$ ,  $p < .0001$ ) and German ( $F(1,8) = 27.22$ ,  $p < .001$ ). Again, Bonferroni-adjusted post-hoc tests for the Dutch data revealed significant differences between /f/ and /v/, and between /v/ and /ʋ/. A two-sided t-test ( $t(8) = 4.468$ ,  $p = .002$ ) showed a significant difference in harmonicity median between Dutch /v/ and German /v/, yet no significant difference between Dutch /ʋ/ and German /v/ and between Dutch /f/ and German /f/.

For the comparison of the mean values for the centres of gravity, the two German speakers who were recorded at a lower sampling rate were excluded. This was done because labiodental fricatives show energy in the high frequency domain, which is not taken into account when calculating gravity centres from signals with a low sampling rate.



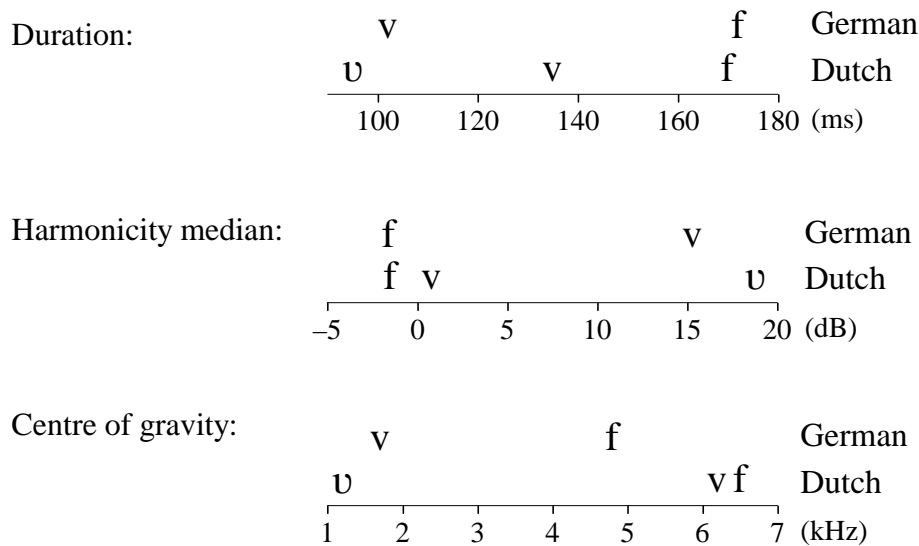
**Figure 3:** Mean values for centre of gravity (in Hz) of the filtered signals. Error bars indicate standard deviation.

Analyses of variance showed that centre of gravity within language was significantly different for both Dutch ( $F(2,12) = 15.138$ ,  $p < .001$ ) and German ( $F(1,4) = 23.534$ ,  $p < .008$ ). Post-hoc tests with Bonferroni-adjustments for the Dutch data revealed significant differences between /f/ and /v/ and between /v/ and /ʋ/. A two-sided t-test ( $t(6) = -3.346$ ,  $p = .015$ ) showed a significant difference in gravity centre values between Dutch /v/ and German /v/. The difference between Dutch /ʋ/ and German /v/ and that between Dutch /f/ and German /f/ were not significant.

### 3 Discussion and conclusions

The results of the present acoustic study show that the German speech sound traditionally described as labiodental fricative and referred to with the IPA symbol /v/ is, with regard to the parameters tested, acoustically closer to the Dutch labiodental approximant /ʋ/ than to the corresponding Dutch labiodental fricative /v/. Both German /v/ and Dutch /ʋ/ share a similar harmonicity median

and a similar centre of gravity. German /v/ is different from Dutch /v/ in these parameters. With respect to the acoustic parameter of duration, German /v/ lies in-between the two Dutch sounds /v/ and /ʋ/. A summary of these findings is given with the comparison of the three Dutch and the two German labiodentals along three scales in Figure 4.



**Figure 4:** Scales comparing the realisations of the two German and the three Dutch labiodentals with respect to the acoustic parameters of duration, harmonicity median, and centre of gravity.

These results can be interpreted as indication that the German voiced labiodental sound is more a glide than a fricative from a phonetical point of view.<sup>3</sup> In the phonetic literature on German, the voiced labiodental sound is usually described as a fricative (see e.g. Jessen 1998, Kohler 1999, Wängler 1974). However, some notable exceptions exist. Kohler (1995: 154), for instance, mentions that German /v/ can turn into an approximant, especially in initial position. This phrasing implies that the default pronunciation of German /v/ is nevertheless a fricative. A picture more in line with the present findings emerges from Scherer & Wollmann (1985) who write that German speakers produce little contact for labiodentals, which might cause an approximant-like articulation that does not exist in English (p.93).

<sup>3</sup> The present study does not consider phonological arguments in favour of a fricative status of the German voiced labiodental sound.



The present study thus provides acoustic evidence for the results of Hamann & Sennema's (2005) perception experiment, where German listeners without knowledge of Dutch classified the Dutch /ʊ/ as their /v/ in almost all of the cases, and German learners of Dutch had problems categorising the Dutch /v/ correctly. Furthermore, our findings illustrate the problems of equating the phonological categories of two languages that are described as identical but have different phonetic realizations. Due to the limited number of speakers, the present results cannot be more than tentative, and further investigations with more speakers need to confirm the present findings.

## Acknowledgements

We would like to thank Jana Brunner, Christian Geng and Daniel Pape for their careful reading of an earlier version of this article. We gratefully acknowledge funding by the German Science Foundation (DFG) grant GWZ 4/8-1-P2 for Silke Hamann and grant SFB 632-C4 for Anke Sennema.

## References

- Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *Proceedings of the Institute of Phonetic Sciences (University of Amsterdam)*, 17: 97-110.
- Boersma, P. & Weenink, D. (2005). Praat: doing phonetics by computer (version 4.3.27) [computer programme], Retrieved from <http://www.praat.org/>.
- Booij, G. (1995). *The Phonology of Dutch*. Oxford: Oxford University Press.
- Gordon, M., Barthmaier, P. & Sands, K. (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association*, 32: 141-174.
- Gussenhoven, C. (1999). Illustrations of the IPA: Dutch. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press; 74-77.
- Hamann, S. & Sennema, A. (2005). Voiced labiodental fricatives or glides - all the same to Germans? In: Hazan, V. & Iverson, P. (eds.) *Proceedings of the Conference on Plasticity in Speech Processing*. London: UCL; 164-167.
- Jassem, W. (1979). Classification of fricative spectra using statistical discriminant functions. In: Lindblom, B. & Öhman, S. (eds.) *Frontiers of Speech Communication Research*. New York: Academic Press; 77-91.
- Jessen, M. (1998). *Phonetics and phonology of tense and lax obstruents in German*. Amsterdam: Benjamins.
- Kohler, K. (1995). *Einführung in die Phonetik des Deutschen*. Berlin: Erich Schmidt.
- Kohler, K. (1999). Illustrations of the IPA: German. *Handbook of the International Phonetic Association*. Cambridge: Cambridge University Press; 86-89.

- Scherer, G. & Wollmann, A. (1985). *Englische Phonetik und Phonologie*. Berlin: Erich Schmidt Verlag.
- Wängler, H.-H. (1974). *Grundriss einer Phonetik des Deutschen*. Marburg: N.G. Elwert Verlag.
- Wiese, R. (1996). *Phonology of German*. Oxford: Oxford University Press.

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

**Jana Brunner**

*Humboldt-Universität zu Berlin & Zentrum für Allgemeine Sprachwissenschaft, 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 communicational 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 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 the same degree of coarticulation, which can be seen as a kind of variability, for stops in two Australian 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 / $\epsilon$ / 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/, /I/ and / $\epsilon$ /. 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

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

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

palate, and for the triangle as

$$A_{orig} = \frac{ad}{2}. \quad (2)$$

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

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

for the flat palate and

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

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

$$A_{diff} = A_{orig} - A_{var} \quad (5)$$

$$A_{diff} = ax \quad (6)$$

and for the triangle (the domeshaped palate):

$$A_{diff} = A_{orig} - A_{var} \quad (7)$$

$$A_{diff} = \frac{ax}{2} \quad (8)$$

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 reasonable 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 experimental 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 between 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 recruitment 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.

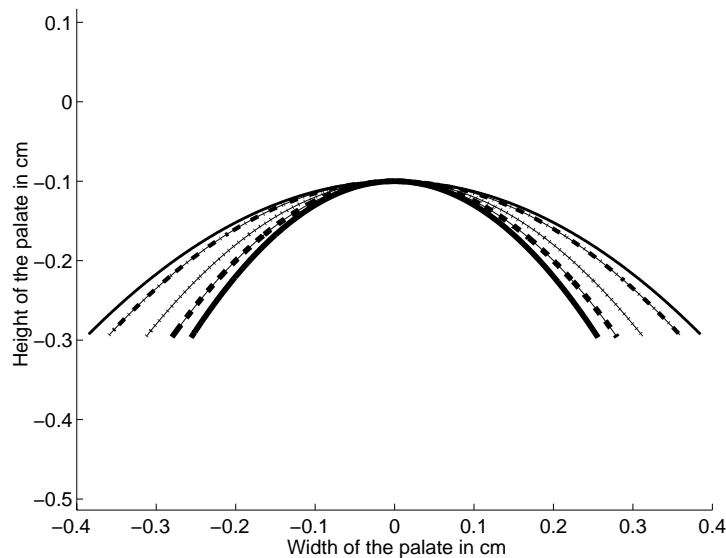
The aim of the simulations was to investigate acoustic variability as a function of the palate shape while articulatory variability was held constant. Basically, three vowels, /a/, /i/ and /u/, were simulated with five palates differing in curvature. The tongue position of the vowels was changed slightly 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  $\alpha$  (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

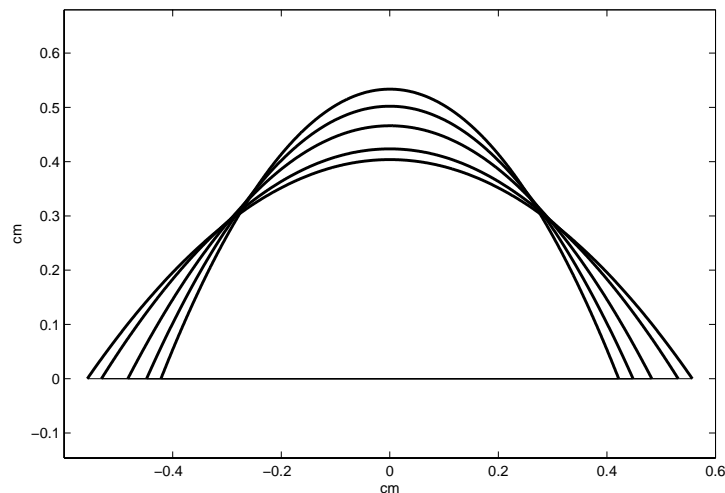


the sagittal distance between the tongue and the palate (for  $\beta$  see Perrier et al. (1992)). As can be estimated from the formula, the lower  $\alpha$  the more curved the palate is: For low  $\alpha$ -values the area becomes smaller, which, given that the sagittal distance stays the same, has to be a result of a palate which is more curved. For the present purpose the following  $\alpha$  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  $\alpha$  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.

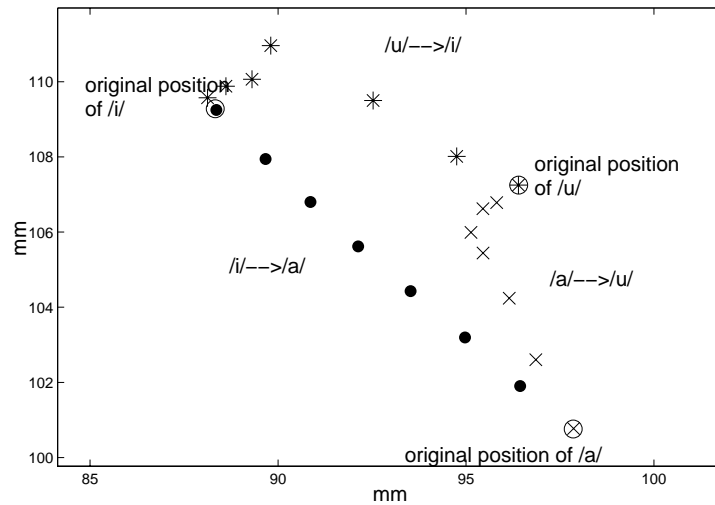
## 2.2 Simulations of "original" and "deviated" vowel positions

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.

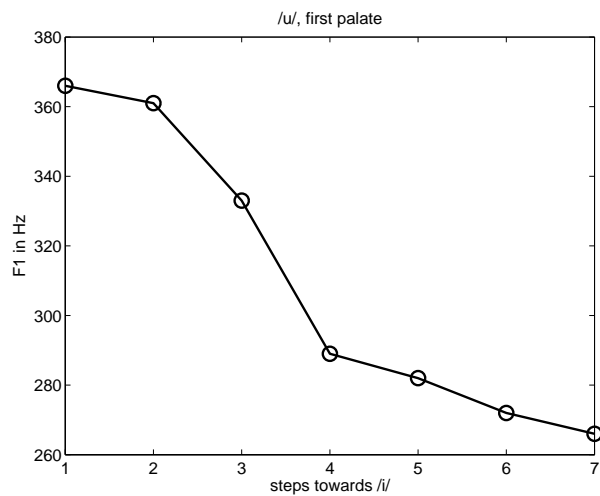
seventh.

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.

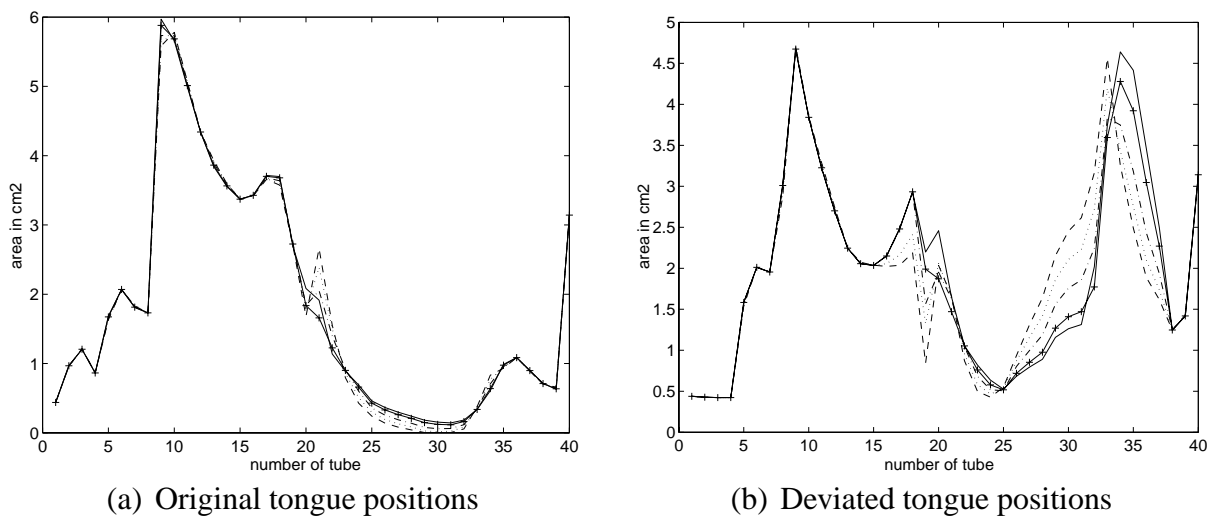
### 3 Results I: Simulations

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 intervals 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.



**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**

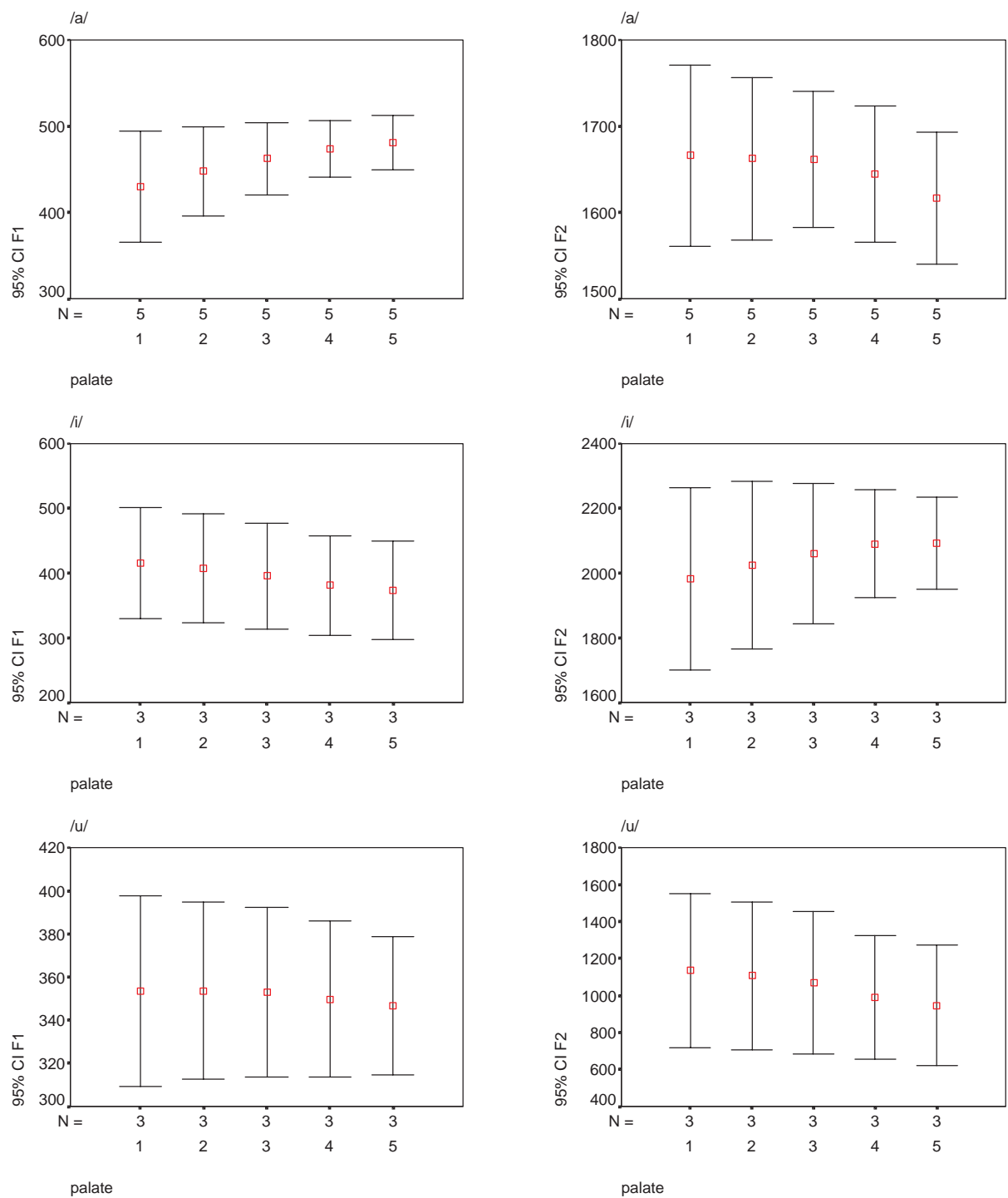
Electropalatography (EPG) allows to investigate linguo-palatal contact. The subject wears an artificial palate made of acryl with 62 electrodes on it (EPG 3.0, Reading system). Each electrode is connected to the system over a small wire. During the recording the subject is holding a further electrode 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 registered 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/, /ç/ and /j/, and the vowels /i/, /e/, and /u/, with their lax counterpart /ɪ/, /ɛ/ and /ʊ/. The sounds were chosen 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/), /'fafa/, /'mici/ (for the Polish speakers /ma'cina/), /'jaja/, /'titi/, /'titi/, /'tu:tu/, /'tutu/, /'te:tə/ and /'tətə/ (for the English speakers /'tətər/ or /'tətə/. The carrier phrases differed from language to language:

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

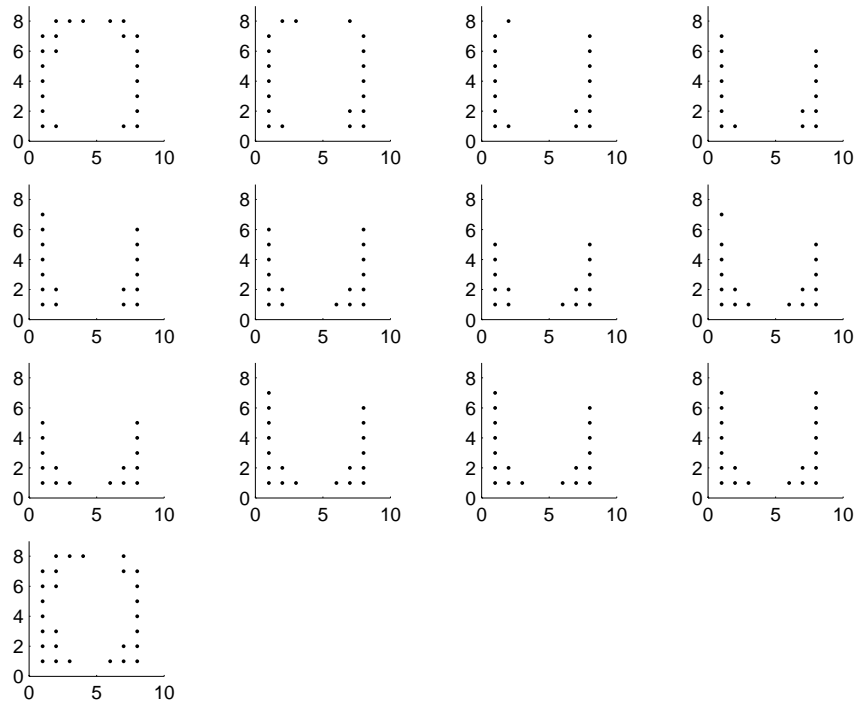
- friction onset and offset for the fricatives /s/, /ʃ/ and /ç/
- sonorant onset and offset for the sonorant /j/ as the points in the middle of the formant transitions between the surrounding vowels and the sonorant.
- onset and offset of the second formant for the vowels.

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

$$poc = \frac{noc * 100}{62} \quad (9)$$

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/.

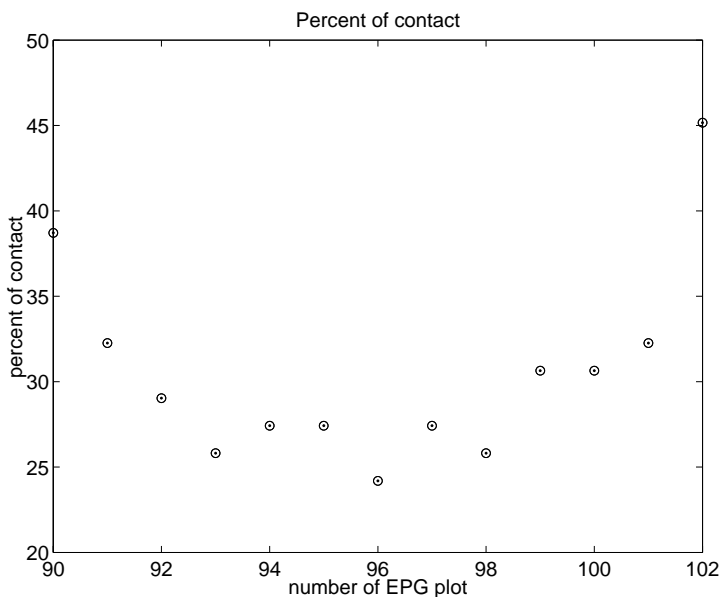
These calculations were carried out for the 30 repetitions of each item. A problem for further calculations was that the length of the segments differed depending on the the velocity of speaking. Therefore, the segments had to be made the same length. This was done by a spline interpolation on 20 points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

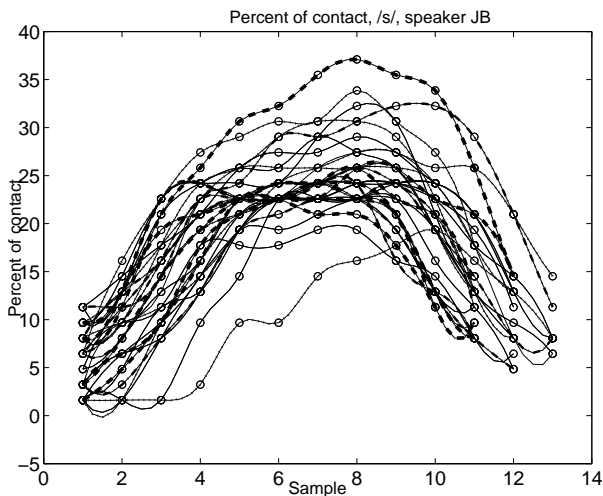
$$v = \frac{s}{\bar{x}} \quad (10)$$

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

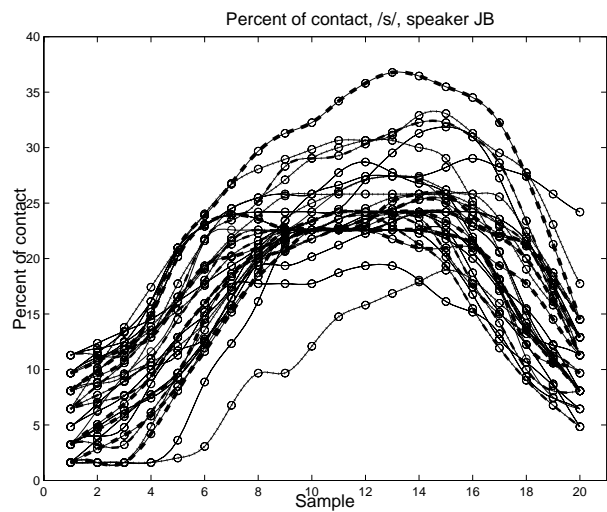




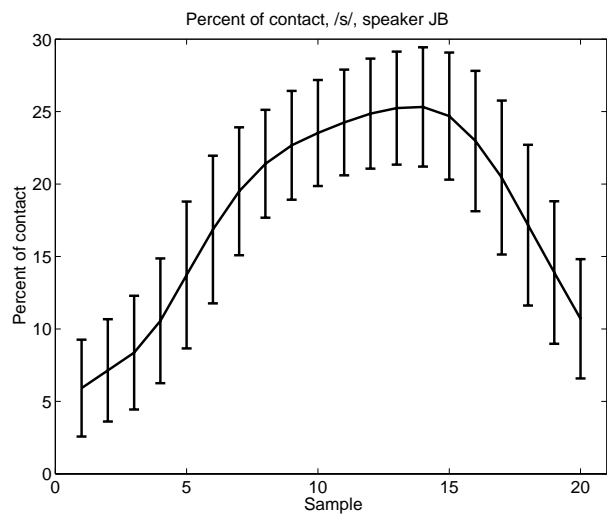
**Figure 8:** Percent of contact for the EPG frames of /u/ shown in figure 7 (with spline interpolation for easier understanding)



**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.



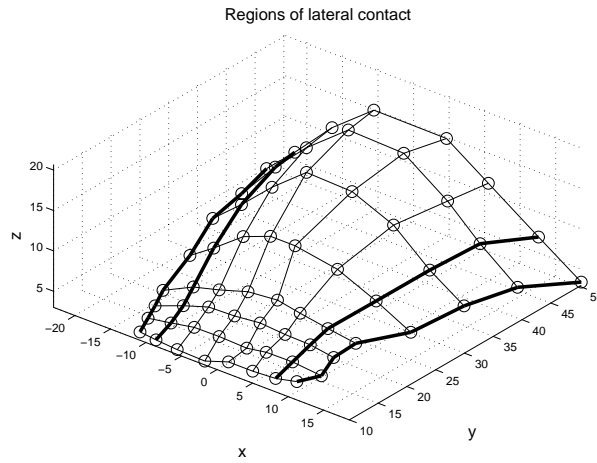
**Figure 11:** Mean values and standard deviation of the percent of contact of 30 repetitions of /s/

### 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

$$li = \frac{noc * 100}{30} \quad (11)$$

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.

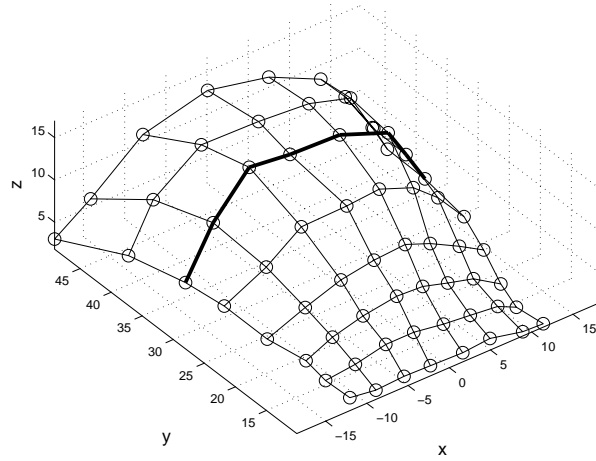
Same as for the percent of contact, an interpolation on 20 points was carried out in order to make all the 30 repetitions the same length. Afterwards the mean value of each of the 20 sample points was calculated. From these 20 numbers a mean was calculated as average lateral contact.

### 4.4 Determination of the palate shape

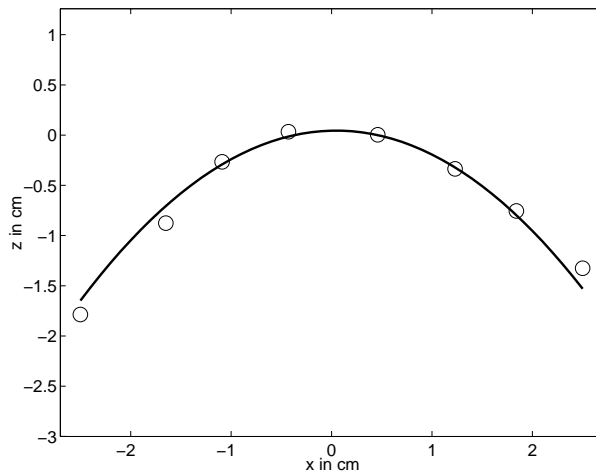
In order to investigate the relationship between palate shape and variability one needs to find a method to determine the curvature of the palate. Because an image of each palate existed in form of the EPG palates these were taken as a basis for this step.

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  $\alpha$  (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

$$y = ax^2 + b \quad (12)$$

$\alpha$  was calculated as (Perrier et al. (1992))

$$\alpha = \frac{\frac{4}{3}}{\sqrt{|a|}} \quad (13)$$

A high  $\alpha$  value corresponds to a flat palate and a low value to a domeshaped palate.

#### 4.5 Calculation of the correlation

The correlations (Pearson) between  $\alpha$  and variability, variability and lateral contact and  $\alpha$  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  $\alpha$  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  $\alpha$  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.

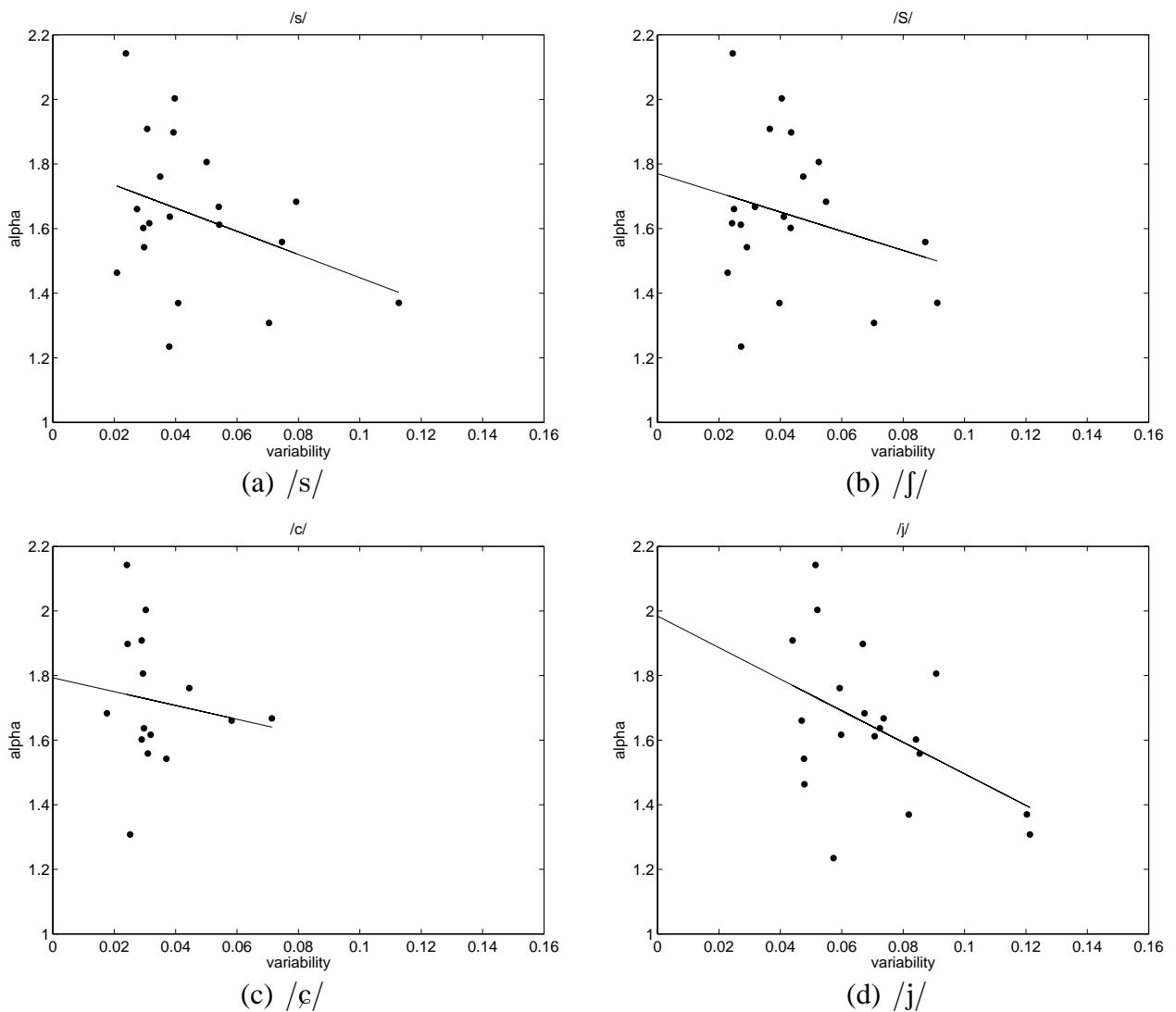
The results differ very much with respect to the item. They are best for palatal vowels and sonorants, not as good for the fricatives and worst for the back vowels.

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

item	variability - $\alpha$	variability - lat. contact	lateral contact - $\alpha$
/ç/	-.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)
/I/	-.559* (.030)	-.342 (.212)	.487 (.066)
/u/	-.001 (.998)	-.679** (.001)	-.110 (.644)
/ū/	-.011 (.969)	-.589* (.021)	-.195 (.486)
/e/	-.382 (.160)	-.705** (.003)	.524* (.045)
/ε/	-.291 (.293)	.080 (.777)	.400 (.139)

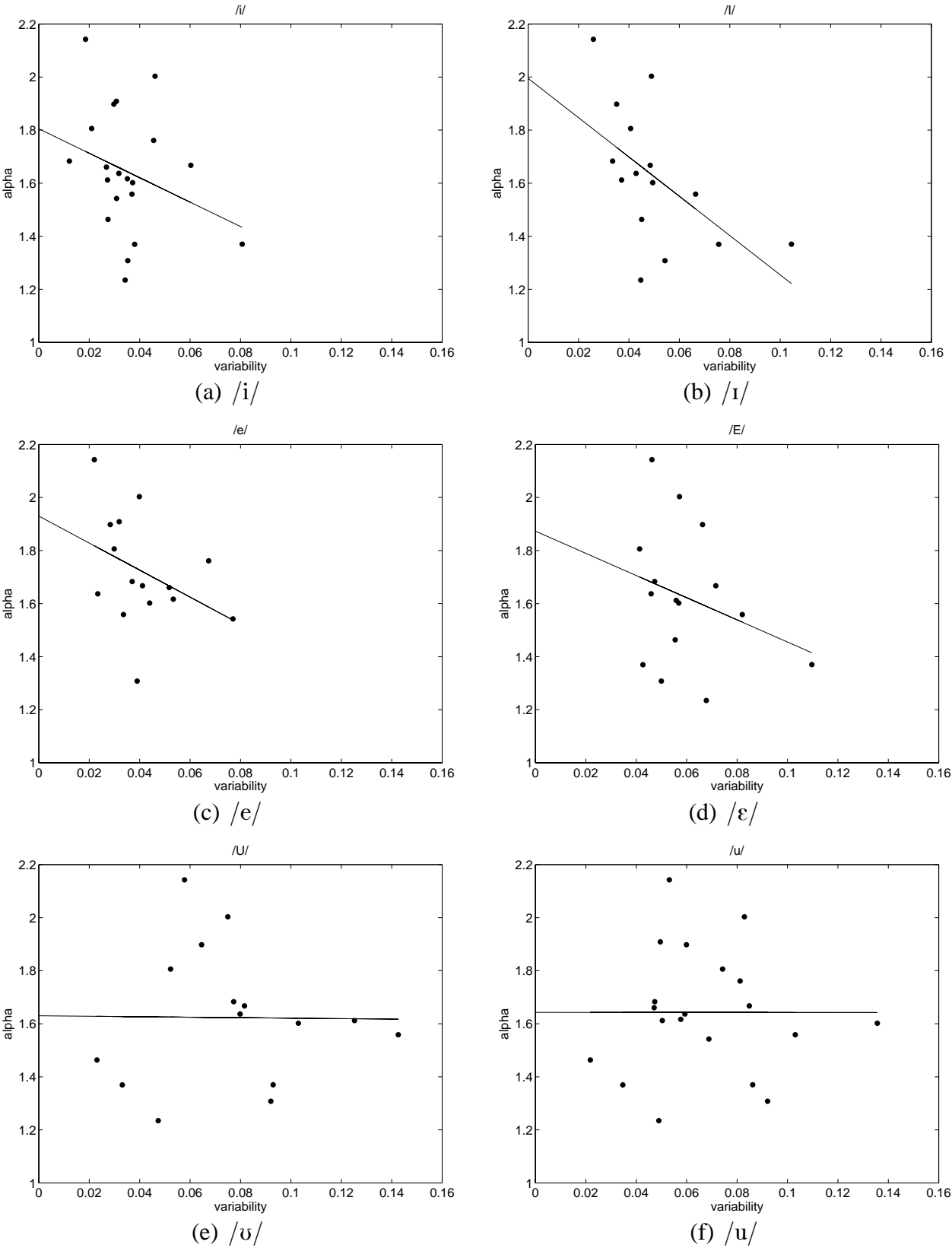
There are two possible reasons for the lack of significance of the correlation between  $\alpha$  and variability: Either the sample is too small and one needs to

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  $\alpha$  is plotted against variability.



**Figure 15:** Correlation between variability (abscissa) and  $\alpha$  (ordinate) for the consonants

What can be seen is that speakers with a flat palate (a high  $\alpha$ ) 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.



**Figure 16:** Correlation between variability (abscissa) and  $\alpha$  (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 explanation. 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.

### **Acknowledgements**

This work is supported by a grant from the German Research Council (Po 334/4-1). We would like to thank Olessia Panzyga and Anke Busler at the Zentrum für Allgemeine Sprachwissenschaft for acoustical segmentation and for carrying out palate measurements. Thanks to Jörg Dreyer for carrying out most of the recordings as well as for setup and maintenance of the EPG system. Furthermore, thanks to the subjects in Berlin and at the QMUC Edinburgh.

### **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](http://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.

# (Non)Retroflexivity of Slavic Affricates and Its Motivation. Evidence from Polish and Czech <č>.

**Marzena Żygis**

*Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany*

---

The goal of this paper is two-fold. First, it revises the common assumption that the affricate <č> denotes /tʃ/ for all Slavic languages. On the basis of experimental results it is shown that Slavic <č> stands for two sounds: /tʃ/ as e.g. in Czech and /tʂ/ as in Polish.

The second goal of the paper is to show that this difference is not accidental but it is motivated by perceptual relations among sibilants. In Polish, /tʃ/ changed to /tʂ/ thus lowering its sibilant tonality and creating a better perceptual distance to /tɕ/, whereas in Czech /tʃ/ did not turn to /tʂ/, as the former displayed sufficient perceptual distance to the only affricate present in the inventory, namely, the alveolar /ts/. Finally, an analysis of Czech and Polish affricate inventories is offered.

---

## 1 Introduction

In the Slavic tradition, the affricate <č> is tacitly or explicitly assumed to be /tʃ/ for all Slavic languages. (see e.g. de Bray 1951, Comrie & Corbett 1993). In this paper I revise the affricate inventories of Polish and Czech showing that the symbol <č> stands for the palatoalveolar /tʃ/ in Czech and the retroflex /tʂ/ in Polish. This conclusion is based on the experimental results presented in the paper.

Second, it will be explained *why* the two languages Polish and Czech, which belong to the same Slavic family, differ in the quality of the affricates. It will be argued that the arrangement of affricates in individual Slavic languages is not accidental but is rather dependent on perceptual relations between the affricates. Slavic inventories clearly show a tendency to optimize perceptual contrast among the sibilants. If the inventory is complex, i.e., consisting of at least one (denti-)alveolar and two postalveolar affricates, then one of the postalveolar affricates is of low sibilant tonality. This is motivated by the principle of contrast optimization: the retroflex affricate displays more

perceptual distance to other affricates than, for example, a palatoalveolar affricate, see also Żygis (2003a).

In simple sibilant systems, i.e., consisting of one (denti-)alveolar and one postalveolar affricate, the latter is almost always a palatoalveolar [tʃ] because the perceptual distance between the two sounds is sufficient and an optimal contrast already exists. In fact, this is the case in Czech.

Furthermore, it will be shown that perception played the underlying role in forming affricate systems. The otherwise unexplainable context-free rules, as e.g. /tʃ/ → [tʃ] are straightforwardly accounted for if perceptual relations among phonemes are taken into consideration. This will be shown by analyzing Polish and Czech inventories from a diachronic point of view.

The study is organised as follows. In section 2 the assumptions made for the purposes of the present study are outlined. In section 3 coronal stop inventories, including affricates of Polish and Czech, are analyzed. Section 4 analyses a diachronic development of these two languages. The experimental results presented in section 5 reveal a clear difference between Czech and Polish postalveolar affricates and provide an explanation for these results. A Dispersion-Theory-account of the experimental findings is proposed in section 6. Finally, in section 7, the main conclusions are summarised.

## 2 Necessary assumptions and comments

For the purposes of the present study the following assumptions have been made.

- 1) First of all, it has been assumed that affricates of a low sibilant quality are denoted as retroflexes [ʈʂ]. This symbol is, however, not entirely adequate as the sound shows great articulatory variability in terms of place of articulation and the shape of the tongue, see Żygis (2005). The only stable characteristic of this sound is the involvement of the tongue tip as the main articulator. This fact, together with the ‘postalveolarity’ of this sound, induced classification as the retroflex from an articulatory point of view; see Keating (1991) and Hamann (2003) for more discussion on this point.
- 2) The crucial point for the present study is the fact that sibilants display a different perceptual quality in terms of their sibilant tonality. It is assumed that specifications of the feature [sibilant tonality], as given below, express the contrast between the sibilants. The specifications are mainly based on perceptual impressions and acoustic results, as presented in the experimental part of the study:

	[t͡s]	[t͡ʃʲ]	[t͡ʃ]	[t͡ɕ]	[t͡s]
sibilant	[low]	[low raising]	[middle]	[middle-high]	[high]
tonality:					

These affricates exhibit different perceptual distances with each other. This point is discussed in section 6.

- 3) There is also a phonological piece of evidence which could potentially help in the identification of retroflexes. It has been argued that retroflexes avoid the following high front vocoides (Bhat 1973), or even that they are not followed by high vocoids due to the incompatibility of two articulatory gestures: the curled-up tongue tip is in conflict with the high and raised tongue tip of the front vowels (Hamann 2003). Consequently, a test demonstrating that the sounds under question are not followed by a high vowel /i/ would provide evidence in favor of their retroflex character. In fact, such a test has been used for various languages (see examples provided in Hall 1997b :48) including Slavic languages, see also Hamann (2004). In the present study, I will not apply the test for the purposes of the retroflex identification. My decision is based on the following arguments:

- (i) As already mentioned, there is a lot of articulatory variation in the production of postalveolar sounds, and the prevailing majority of x-rays of the potential retroflexes does not demonstrate a typical curled-up tongue tip. Instead, the tongue tip is often placed at the alveolar ridge and the tongue blade together with the tongue dorsum is flat. Consequently, the incompatibility of the articulatory gestures does not hold for these group of sounds.
  - (ii) in some Slavic languages the sounds in question are indeed followed by [i] and not by [ɪ], see, for example, the rule called Retraction in Polish (Rubach 1984). However, this rule also affects other coronal sounds including dental stops, fricatives, and the trill [r].
  - (iii) from a diachronic perspective, the [ɪ] vs. [i] distribution goes back to depalatalisation processes which affected all *palatalised* sounds, including labials, and in some Slavic languages velar sounds. Thus, the process cannot be explained by the incompatibility of the articulatory gestures of the curled tongue tip and the raised tongue blade of [i], but deserves a more general explanation.
- 4) Following Rubach (1994), LaCharité (1993), Kim (1997), Clements (1999), and Kehrein (2002), I assume that affricates are phonologically

strident stops. They form a natural class with other coronal stops, which are also included as a subject of the present investigation. It is shown that the coronal stops, despite forming a natural class with affricates, are not directly influential on sibilant systems. This is due to their different acoustic/perceptual properties, which do not directly compete with the properties of the sibilant frication, as discussed in the experimental part of the present study.

- 5) Finally, it should be noted that the present account considerably differs from articulatory-based accounts of sibilant systems, as e.g. Hume (1994) or Hall (1997a). Both approaches are discussed in Żygis (2005) in detail.

### 3 Affricates in Slavic languages

For the purposes of the present study fourteen present-day Slavic languages have been investigated. Besides Czech and Polish, recordings from Belorussian, Bulgarian, Croatian, Kashubian, Russian, Macedonian, Serbian, Slovak, Slovenian, Upper and Lower Sorbian and Ukrainian were taken (between 2 and 5 speakers of each language). The data were investigated acoustically and perceptually. In addition, articulatory descriptions of the affricates, including x-ray tracings available in the literature, were considered.

The investigation showed that affricate systems underlie the perceptually based principle in (1):

(1)

If the inventory is complex, i.e., consisting of at least one (denti-)alveolar and two postalveolar affricates or a strongly palatalized /tʲ/, then one of the postalveolar affricates displays a low sibilant tonality.

Perceptual relations are also responsible for shaping simple sibilant systems. It is argued that:

(2)

In simple sibilant systems, i.e., consisting of one (denti-)alveolar and one postalveolar affricate, the latter is often a palatoalveolar [tʃ].

Note that the if-principle in (1) applies to complex systems only, in which one of the postalveolar affricates must be of low tonality. However, this principle does not exclude the possibility that, in simple systems, the postalveolar sibilant *can* also be of low sibilant tonality. This is because the perceptual distance between the affricates can be extended. In simple systems, more perceptual space is available, see Żygis (2005) for more discussion.

In the following I will focus on Polish and Czech sibilant affricate inventories including coronal stops.

### 3.1 Standard Polish and its dialects

Polish shows a complex contrast in coronal inventories, as depicted in (3).<sup>1</sup>

#### (3) Standard Polish<sup>2</sup>

	denti-alveolar		retroflex	alveolo-palatal
stop	t	d		
affricate	$\widehat{ts}$	$\widehat{dz}$	$\widehat{t_s}$ $\widehat{d_z}$	$\widehat{t_c}$ $\widehat{d_c}$
fricative	s	z	$\text{ʂ}$ $\text{ʐ}$	$\text{ɕ}$ $\text{ʑ}$

The retroflex status of the Polish postalveolar affricates  $\widehat{t_s}$  /  $\widehat{d_z}$ , as is argued in the present study, has not been investigated according to the best of my knowledge. In Slavic tradition these affricates are either transcribed as [č] [ž] (see Benni 1931, Wierchowska 1971, Rubach 1984), [č̣], [dẓ̌] (see Gussmann 1980, Szpyra 1995), or [tʃ̣], [dʒ̣] in IPA terms (see Dukiewicz & Sawicka 1995, Jassem 2003).

By way of contrast, in a non-Slavic tradition researchers have pointed out the retroflex character of the Polish sibilants; but their studies were limited to fricatives; cf. Keating (1993), Ladefoged & Maddieson (1996), Hall (1997a), and Hamann (2003). Only one study by Stevens & Blumstein (1975) considers the Polish affricate  $\widehat{d_z}$  as an example of a retroflex sound, albeit even there its properties are not discussed in detail.

In the following I provide articulatory and perceptual evidence showing that the Polish affricate under consideration is not a palatoalveolar  $\widehat{t_j}$ , but that it exhibits some characteristics of the retroflex  $\widehat{t_s}$ . While the articulatory and perceptual aspects will be discussed in the present section, the acoustic arguments are provided in section 5 in which the experimental results are discussed.

As far as the articulatory aspect of the Polish postalveolar affricate is concerned, its stop and fricative components display retroflex characteristics.

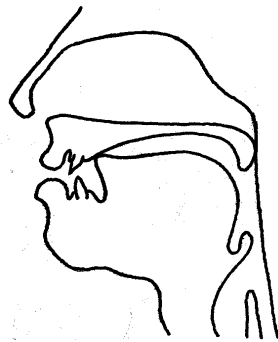
<sup>1</sup> It should be noticed that Polish contrasts retroflex affricates and sequences of stops followed by fricatives, e.g., [tʃi] ‘three’ vs.  $\widehat{[t_s i]}$  ‘whether.’

<sup>2</sup> It should be stressed that some scholars also assume that Polish has palatalized dentals /tʲ/ /dʲ/ in its phonemic inventory, see, e.g., Bethin (1992). Others maintain that palatalized stops occurring on the surface are underlying sequences of stops followed by /j/ e.g. /tj/, /dj/; see Rubach (1984). As it will be shown by the experimental results, this difference bears no effect on the present investigation.



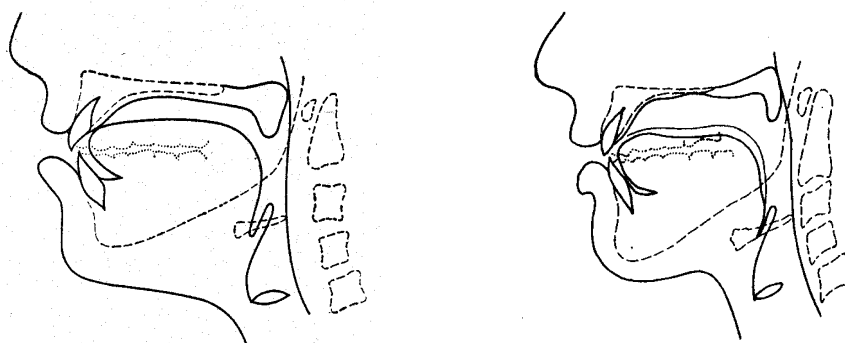
Biedrzycki (1974) provides an x-ray tracing of the Polish stop component of  $[\text{t͡ɕ}]$ , which leaves no doubt that the stop component also shares features characteristic of typical retroflex stops: the tongue tip is extended out from the tongue body and raised. It touches the alveolar ridge or even the area behind it. In addition the tongue body is raised and thus the sound is velarized; see Figure 1. The fricative component is not provided by Biedrzycki (1974).

A very similar x-ray tracing to the one presented in Figure 1 is provided by Ostaszewska & Tambor (2001:40), although it is not said explicitly which affricate component is presented by the frame.



**Figure 1:** Stop component of Polish  $[\text{t͡ɕ}]$  (Biedrzycki 1974: 22).

Wierzchowska (1971:163) provides another x-ray tracing of the stop component of the postalveolar affricate. It is shown in Figure 2a whereas in Figure 2b the fricative component is presented; see Wierzchowska (1980:64).

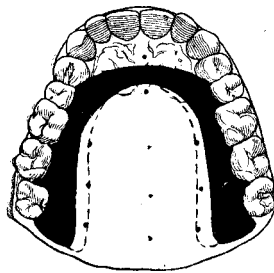


**Figure 2:** a. Stop component of Polish  $[\text{t͡ɕ}]$       b. Fricative component of Polish  $[\text{t͡ɕ}]$

Although the tip is not curled up in Figure 2a Wierzchowska (1971:163) notices

that the difference between Polish coronal stop, [t] and the stop component of [t͡ɕ], is that the tongue tip is positioned higher in the latter than in the former case. As displayed in Figure 2a the tongue tip touches the alveolar ridge, whereas in the case of [t] the tongue tip is positioned behind the teeth.<sup>3</sup> Note, however, that it cannot be maintained, on the basis of the x-ray frames in Figure 2 that the affricate [t͡ɕ] is articulated at a posterior place of articulation (as typically occurs with retroflexes), but rather it is articulated at the alveolar place. It also displays a sublingual cavity which is characteristic of retroflexes. As far as the fricative part is concerned, Wierzchowska provides the same x-ray tracing as for the corresponding fricative, which is described as apical and produced at the (denti-) alveolar place of articulation. According to the definition of retroflexes adopted for the present study the fricative part of the Polish [t͡ɕ] can also be classified as retroflex.

Benni (1931) provides a palatogram of the stop component of Polish [t͡ɕ], showing that the tongue tip is positioned farther back at the rear of the alveoli, see Figure 3.



**Figure 3:** Stop component of Polish [t͡ɕ] (Benni 1931:14)

From an impressionistic perceptual point of view, Polish affricates denoted as [t͡ɕ] in the present study, are considered without exception to be hard, especially when they are compared with affricates of other Slavic languages like e.g. Russian; see for example, de Bray (1951). The hardness of these sounds is acoustically mirrored by prominent lower frequencies which are characteristic for retroflexes. This point will be experimentally analyzed in section 5 in great detail.

<sup>3</sup> Wierzchowska (1971:164) also notices that the stop component as shown in Figure 2a occurs in sequences before fricatives [ɕ] [ʐ] which do not create an affricate, e.g. [t͡ɕ]y ‘three.’

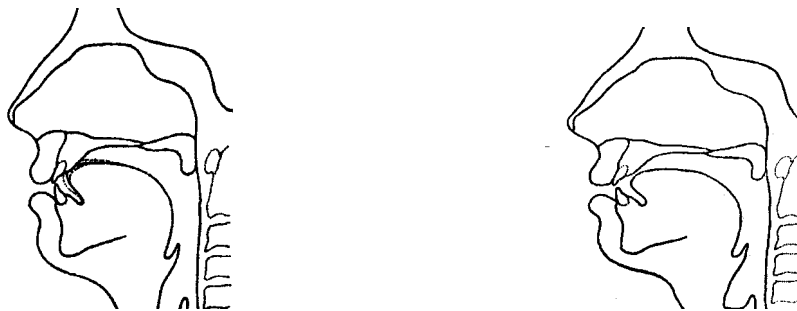
### 3.2 Czech

Czech belongs to languages having a simple affricate contrast:  $\widehat{ts}/$  vs.  $\widehat{tʃ}/$ . The two affricates form a natural class with the alveolar  $/t/$  and the palatal  $/c/$ . The inventory is shown in (4).<sup>4</sup>

#### (4) Czech

	alveolar		palatoalveolar		palatal	
stop	t	d			c	ɟ
affricate	$\widehat{ts}$		$\widehat{tʃ}$	( $\widehat{dʒ}$ )		
fricative	s	z	ʃ	ʒ		

From the articulatory evidence, it is however far from obvious whether the postalveolar affricates are indeed palatoalveolars. In Figure 4a an x-ray tracing of the plosive component of  $\widehat{tʃ}$  and in Figure 4b an x-ray tracing of the fricative  $[ʃ]$  is shown.



**Figure 4:** a. Czech  $\widehat{tʃ}$  (Palková 1994: 235)

b. Czech  $[ʃ]$  (Palková 1994:229)

Palková (1994:235) states that in the closure phase the tongue tip is situated at the rear of the alveolar ridge. This is confirmed both by a palatogram and a linguagram provided in Figure 5.

<sup>4</sup> However, there are differences in describing the particular places of articulation. With respect to  $/t/$  and  $/d/$ , Short (1993:535) assumes the dento-alveolar place of articulation, while de Bray (1951:439) and Stadnik (1998:387) the dental. There is also no consensus with respect to  $/c/$  and  $/ɟ/$ : Whereas Stadnik (1998) and Short (1993) assume that they are palatal sounds, others consider them to be palatalized alveolars  $/t^j/$ ,  $/d^j/$ ; see de Bray (1951), Palková (1994). In the present study, they will be presented as palatals following the experimental study by Machač & Skarnitzl (2004).



**Figure 5:** Palatogram and linguagram of the closure of Czech [tʃ] (Palková 1994:235)

In the release phase of the closure, a constriction similar to that of [ʃ] is created. The lips are protruded. Palková (1994:235) also observes that the affricate is hard from a perceptual point of view. However, the low spectral prominence between 1.4 and 2 kHz which would suggest the low sibilant tonality refer only to the corresponding fricative. This point is not confirmed by the experiment results presented 5, where it is shown that the COGs of the palatoalveolar [tʃ] are higher (above 3 kHz on average).

As far as the perceptual aspect of Czech [tʃ] is concerned, Lehr-Spławiński & Stieber (1957:40) maintain that the pronunciation of the fricatives [š] [ž] can be as hard as the Polish corresponding sounds, but it often happens that the sounds are articulated in a semi-soft way. This is especially noticeable – as Lehr-Spławiński & Stieber (1957) observe – when Czechs speak Polish. With respect to the corresponding affricate [č] the situation is different as far there is a no option: the affricate is always semi-soft and differs from the Polish [tʃ]. The experimental results presented in section 5 indicate that the Czech postalveolar affricate considerably differs from Polish [tʃ] and should be classified as [tʃ̠].

In conclusion, Czech postalveolar affricates correspond to the IPA palatoalveolar [tʃ̠]. It also appears that other stops such as /t/, and especially the palatal /c/, do not have a direct influence on creating affricate inventories, albeit creating a natural class with them. This point is confirmed by experimental results, presented in section 5.

#### **4 Czech and Polish affricates from a diachronic point of view**

This section deals with the emergence of affricates and their development in two selected Slavic languages, Czech and Polish. The choice of these two neighbouring languages is motivated by the fact that they have developed different affricate contrasts: a two-way contrast in Czech and a three-way contrast in Polish. The inventories, together with coronal stops, are repeated in

(5) for convenience. Note that in Polish the retroflex  $\widehat{t\mathfrak{s}}$  is proposed according to the assumptions made in the present study.

(5)	Czech	t	c	$\widehat{ts}$	$\widehat{t\mathfrak{s}}$	
	Polish	t		$\widehat{ts}$	$\widehat{t\mathfrak{s}}$	$\widehat{t\mathfrak{c}}$

In the following it will be shown that the Polish postalveolar affricate <č> was  $\widehat{t\mathfrak{s}}$  and later changed to  $\widehat{t\mathfrak{s}}$  in order to create a more optimized contrast to  $\widehat{t\mathfrak{c}}$ . Furthermore, it will be argued that the emergence of the palatal /c/ in Czech did not have a significant impact on its affricate system:  $\widehat{t\mathfrak{s}}$  did not change to the retroflex  $\widehat{t\mathfrak{s}}$  because the perceptual contrast to the already existing affricate  $\widehat{ts}$ , as well as other stops, was sufficient.

From a diachronic point of view the issues listed in (6) are the main points of interests for the present study. They have been listed chronologically.

(6) Development of affricate system:

- (i) The emergence and development of  $\widehat{t\mathfrak{s}}$  in Czech,
- (ii) The emergence and development of  $\widehat{t\mathfrak{s}}$  in Polish
- (iii) The emergence of /c/ in Czech and  $\widehat{t\mathfrak{c}}$  in Polish

As far as (i) is concerned, the emergence of  $\widehat{t\mathfrak{s}}$  goes back to the Proto-Slavic First Velar Palatalization (1<sup>st</sup>VP) according to which /k/, /g/, and /x/ changed to  $\widehat{t\mathfrak{j}}$ , [ʒ], and [ʃ] before front vowels; see the rule presented in (7). The process would have been accomplished in or about the 6<sup>th</sup> or 7<sup>th</sup> century (Stieber 1969:67)

(7) 1st Velar Palatalisation (Stieber 1969:66)

/k, g, x/ →  $\widehat{t\mathfrak{j}}$ , [ʒ, ʃ]/\_ī, ĭ, ē, ě<sup>5</sup>

Stieber (1957: 93) observes that <č> of present-day Czech is not as soft as the Proto-Slavic  $\widehat{t\mathfrak{j}}$ , but still softer than the corresponding Polish sound. A similar observation is made by Carlton (1991). Therefore, I assume that in terms of IPA the 1st Velar Palatalization of /k/ produced palatalized palatoalveolar  $\widehat{t\mathfrak{j}}$ .

Since the process of the 1<sup>st</sup>VP occurred in Proto-Slavic, palatoalveolar  $\widehat{t\mathfrak{j}}$  was also an ancestor of Polish retroflex affricate  $\widehat{t\mathfrak{s}}$ , see (6iii). However, around the 16<sup>th</sup> century, the palatalized  $\widehat{t\mathfrak{j}}$  originating from 1<sup>st</sup> Velar Palatalization was hardened and converted to the retroflex  $\widehat{t\mathfrak{s}}$ , e.g.  $\widehat{t\mathfrak{j}i}sto$  vs.  $\widehat{t\mathfrak{s}i}sto$  ‘clean’ (Rospond 1971:91).

<sup>5</sup> The symbols /ī ē/ stand for short /i, e/, while /ě, ĭ/ for long /e, i/.

In the light of the facts presented above, a question arises as to how we can explain the difference in the development of Protoslavic  $\widehat{tʃ}$  in Czech and Polish.

The answer, as it is argued in the present study, is provided by a different development of Protoslavic  $/tʃ/$  and its perceptual impact on the already existing affricates. In Polish the Protoslavic  $/tʃ/$  originating from  $/tj/$  was converted to the alveolo-palatal  $[tʃ]$  around the 13<sup>th</sup> century; for instance,  $i[dʲ]e[tʃ]e \rightarrow i[dʒ]e[tʃ]e$  (Stieber 1962:63). Subsequently,  $[tʃ]$  was phonemized, and since then it has formed an integral part of Polish consonantal inventory. Hence, until the 16<sup>th</sup> century  $\widehat{tʃ}$  was found along side  $\widehat{tʃ}$  in a Polish phonemic inventory, and then the latter changed to  $\widehat{tʃ}$ .

In contrast to Polish,  $/tʃ/$  was not affricatized in Czech. Instead, it had gradually changed to the palatal  $[c]$  and around the end of the 14<sup>th</sup> century it entered the phonemic inventory of Czech (Lamprecht, Šlosar & Bauer 1977). Since then it has co-occurred with  $\widehat{tʃ}$ .

The motivation for the differences in the development of  $\widehat{tʃ}$  becomes clearer if we consider the acoustic/perceptual properties of  $[c]$  and  $[tʃ]$ , which will be discussed in section 5 in more detail. It will be shown that in contrast to  $[tʃ]$ , the palatal  $[c]$  does not share the fricative-like properties with affricates. Therefore, the former, and not the latter, was directly involved in the formation of the affricate system. Since  $[tʃ]$  (and not  $[c]$ ) was perceptually close to  $\widehat{tʃ}$ , the latter sound changed to  $\widehat{tʃ}$  in order to create more perceptual distance from  $[tʃ]$ . In the Czech system, this change was not required, because the perceptual distance between the already existing affricates had not been changed by the entrance of the new phoneme  $/c/$ .

In summary, it has been observed that the Czech and Polish affricate  $\widehat{tʃ}$  did not develop in a parallel manner: whereas the palatoalveolar  $\widehat{tʃ}$  changed to the retroflex  $\widehat{tʃ}$  in Polish, it remained palatoalveolar in Czech. This discrepancy can be argued to be attributed to an asymmetrical development of the Protoslavic  $/tʃ/$ : while in Czech  $/tʃ/$  had developed into palatal stop  $/c/$ , in Polish it converted to an alveolo-palatal affricate  $\widehat{tʃ}$ . This difference played a significant role in the development of sibilant affricates in these languages.

## **5 Phonetic investigations: Experimental results**

In this section, phonetic evidence underpinning the assumptions made in previous sections will be empirically demonstrated. The aim of this section is two-fold. Firstly, it will be experimentally shown that the Slavic affricates are indeed palatoalveolars as commonly assumed. Secondly, it will be shown what influence other phonemes of the same coronal natural class (the stops  $/t/$  and  $/c/$ ,

as well as the affricates  $\widehat{ts/}$  and  $\widehat{t\phi/}$  have on the postalveolar affricate  $\widehat{tʃ/}$ , and thus on the shape of the sibilant inventory.

The study is limited to the two Slavic languages, Czech and Polish, whose relevant (voiceless) stop contrasts are repeated in (8) for convenience. Note that the Polish retroflex  $\widehat{tʂ/}$  has already been assumed in this study. This assumption requires, however, further acoustic underpinnings.

(8)	Czech	alveolar t $\widehat{ts}$	palatoalveolar $\widehat{tʃ}$	palatal c
	Polish	dento-alveolar t $\widehat{ts}$	retroflex $\widehat{tʂ}$	alveolo-palatal $\widehat{t\phi}$

The languages in (8) have been chosen for the following reasons: The place of articulation of Czech  $\widehat{tʃ/}$ , denoted mostly as <č>, is by no means clear from the descriptions available in the literature; see 3.2. In the same vein, the corresponding Polish postalveolar affricate is repeatedly reported as the palatoalveolar  $\widehat{tʃ/}$ , contrary to what is argued in the present study, see 3.1. Furthermore, the presence of the palatal stop /c/ in Czech on the one hand, and the alveolo-palatal affricate  $\widehat{t\phi/}$  in Polish on the other is important because it gives a possibility for proving to what extent these sounds might influence the postalveolar affricates, in the sense that the latter convert to retroflexes.

Three predictions are made for the purposes of the present study. They are listed in (9).

(9) Predictions:

- (i) The Czech postalveolar affricate is  $\widehat{tʃ/}$ , while the Polish corresponding affricate is the retroflex  $\widehat{tʂ/}$ .
- (ii) The Czech palatal stop /c/ does not have any significant impact on the shape of the affricate sibilant inventory.
- (iii) The Polish alveolopalatal affricate  $\widehat{t\phi/}$  plays an essential role in creating the sibilant system.

In order to test the predictions in (9) the recordings of four native speakers of Czech (two females, MM, BM and two males, RS and MK) and four native speakers of Polish (two females, MR, MZ and two males, SL, CZ) were made. The speakers were asked to read the items listed in (10) five times embedded in the following carrier sentences: ‘*Powiedziala X do ciebie*’, ‘I said... to you’ in Polish and ‘*Predal jsem X. Petrovi*’ ‘I passed X onto Peter.’ in Czech. Note that the capital letter in (10) denotes a stressed syllable.

(10) Experimental items

Czech	Ata	$\widehat{Atsa}$	$\widehat{At\mathfrak{s}a}$	Aca	
Polish	Ata	$\widehat{Atsa}$	$\widehat{At\mathfrak{s}a}$	$\widehat{At\mathfrak{c}a}$	$\widehat{At\mathfrak{s}^ja}$

It has to be noted that the Polish item *Atʃʲa* in which an allophone [tʃʲ] occurs has been considered for reasons of comparison to the Czech *Atʃ*.

The recordings were made at a sample rate of 22.05 kHz. The items were further analysed with PRAAT (version 4.2.21). For statistical calculations SPSS (version 11.0.) was used.

In order to test the predictions in (9) five acoustic parameters were investigated, listed in (11). Most parameters in (11) refer to a ‘frication phase’. In the case of an affricate the frication phase comprises the whole fricative component. In items such as *Ata* and *Aca* the frication phase refers to a brief period starting after the burst and ending at the starting point of fundamental frequency.

(11) Parameters:

- (i) The duration of the closure and of the frication phase
- (ii) The amplitude of the frication phase
- (iii) The transition of the vowel formants F2 and F3 preceding and following the consonant
- (iv) The centre of gravity values of the frication phase
- (v) The correspondence of the frequency of the highest-amplitude spectral cue at the release of the burst, and at the steady-state part of the fricative to the formant frequencies of the following vowel

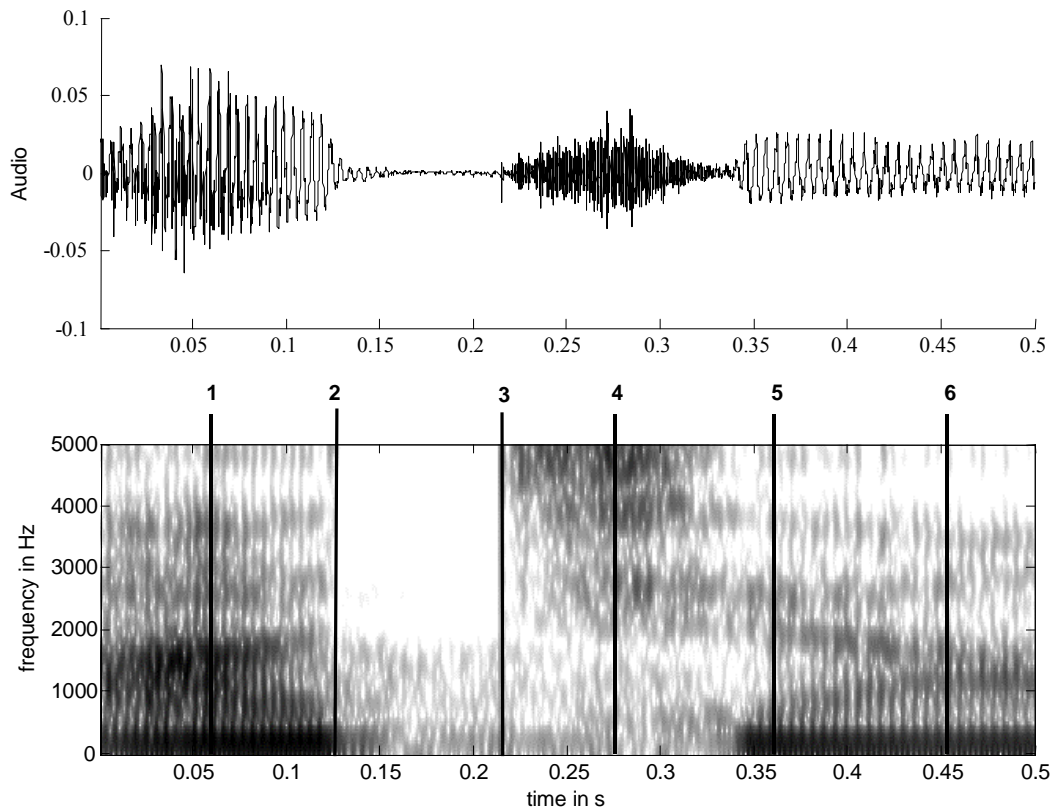
For the calculations of the parameters in (11) the following six temporal landmarks were extracted:

(12) Points of investigation:

- (1) The steady state of the vowel preceding the consonant,
- (2) The end of the formants of the vowel preceding the consonant,
- (3) The burst,
- (4) The steady state of the frication,
- (5) The beginning of the formants of the following vowel,
- (6) The steady state of the following vowel.



All six places in (12) are exemplified on the spectrogram of Polish [at̪ɕa] in Figure 6.



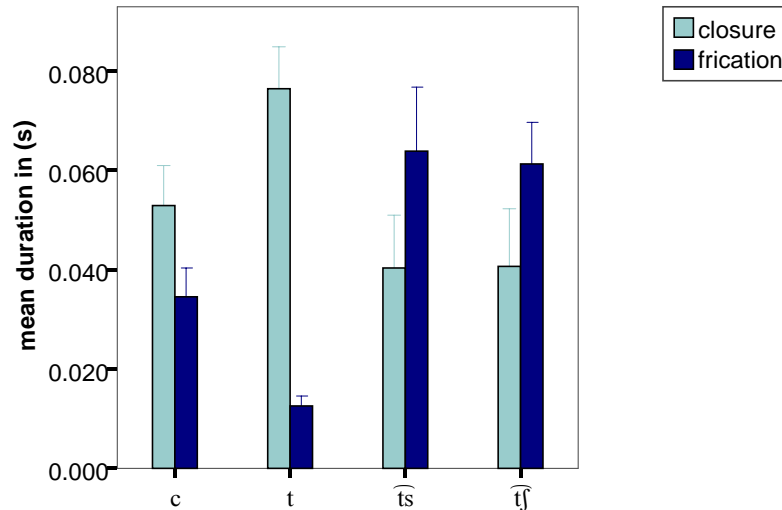
**Figure 6:** Oscillogram and spectrogram of [at̪ɕa] as pronounced by a Polish native speaker.

In the following the experimental results will be presented. The order of the presentation is in accordance with the parameters listed in (11).

**Parameter (i):** The duration of the closure and of the frication phase

In the investigation of the parameter (i), the duration of the closure (from 2 to 3 in Figure 6) and the duration of frication (from 3 to 5 in Figure 6) were measured.

Figure 7 shows mean duration values of the closure and frication phase as obtained from four Czech speakers. The differently coloured error bars (with  $\pm 1.0$  standard deviation) stand for mean duration of the closure and frication as indicated by the legend on the right. They are assigned to the appropriate consonants as indicated on the horizontal axis.

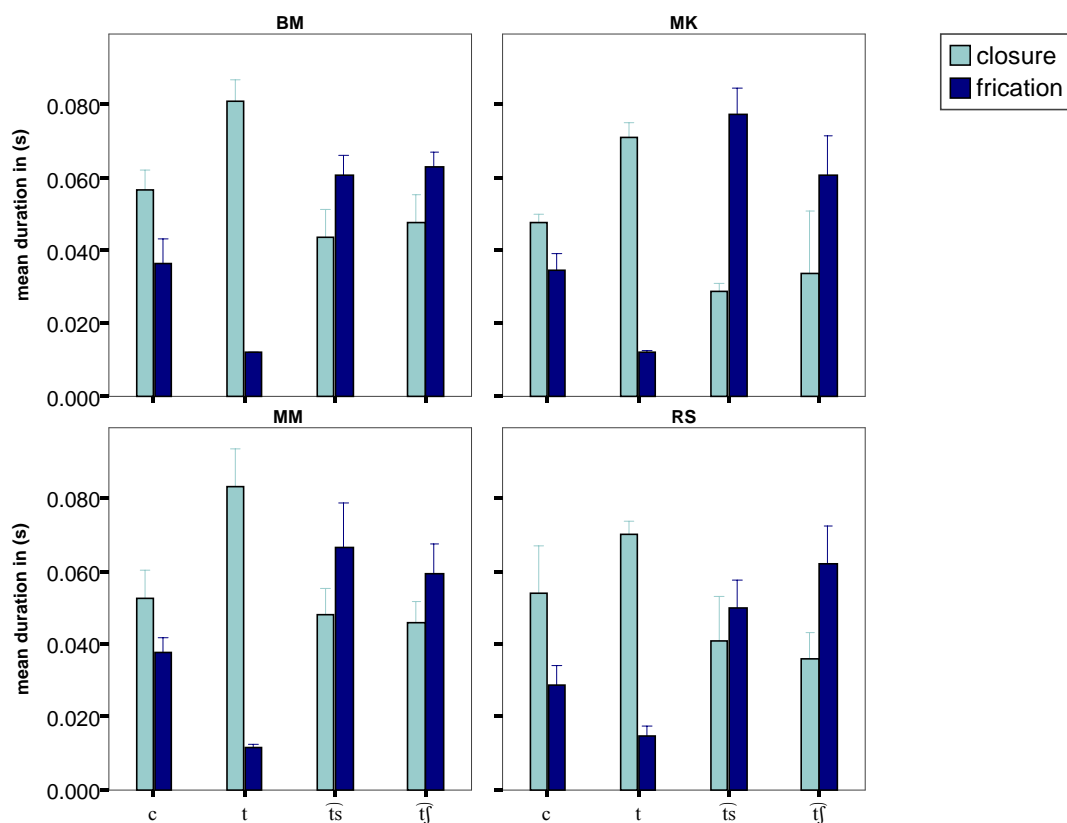


**Figure 7:** The average duration of closure and frication in Czech consonants.

The results displayed in Figure 7 show that the frication phase of the affricates  $\widehat{tj}$  and  $\widehat{ts}$  is longer than the closure phase. As far as  $\widehat{tj}$  is concerned, its mean closure duration amounts to ca. 40.8 ms followed by a 61.3 ms frication. Very similar results are obtained for the affricate  $\widehat{ts}$ : 40.3 ms vs. 63.7 ms. Conversely, the closure phase of the stops are longer than their releases.

A one-factorial ANOVA calculated for every consonant separately with *duration* as dependent variable and *closure&frication*, i.e. the two affricate components, as an independent variable shows a significant effect for all consonants in Figure 7 with respect to the difference between the closure and frication durations:  $/t/$   $F(1,39) = 1098.947$   $p < .001$ ,  $/c/$   $F(1,39) = 66.072$   $p < .001$ ,  $\widehat{ts}/$   $F(1,39) = 39.465$   $p < .001$ ,  $\widehat{tj}/$   $F(1,39) = 43.023$   $p < .001$ . In addition, the differences in closure duration of the two affricates, as well as the differences in their frication duration, are not significant.

Figure 8 presents the results split by speaker.



**Figure 8:** The average duration of closure and frication split according to Czech speakers.

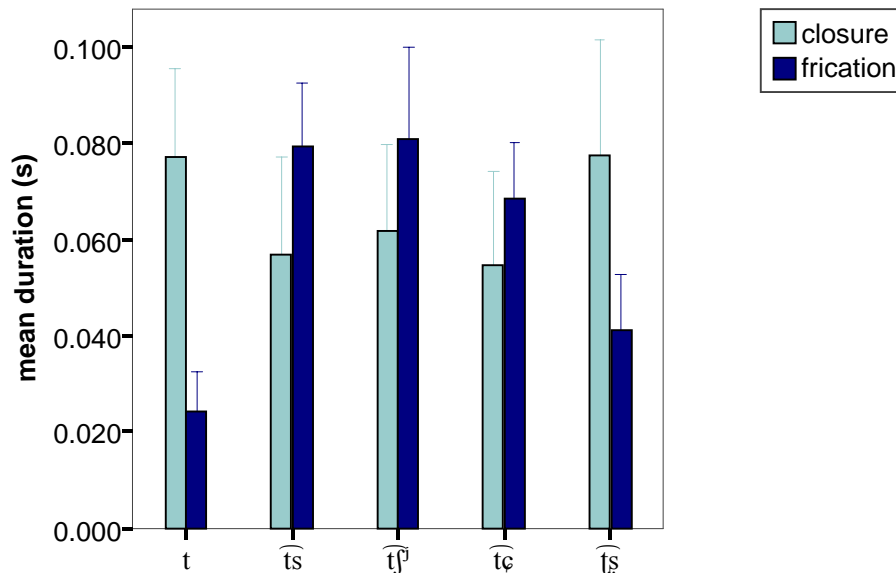
All Czech speakers show an asymmetry: closure is shorter than frication in affricates  $/ts̃/$  and  $/tʃ̃/$ , whereas a reverse pattern is found in the stops  $/t/$  and  $/c/$ . From a statistical point of view, the relation between the closure and frication duration is significant for almost all items. The only exception is the item  $/ts̃/$  as produced by speaker RS where the closure is shorter than the frication but the difference is not statistically significant. The detailed statistic calculations are given in Table 1.

**Table 1:** Statistical analysis of the relation between closure and frication duration in Czech consonants split by speakers

	/t/	/c/	/t͡s/	/t͡ʃ/
speaker BM	F(1,9)=767.410 p<.001	F(1,9)=25.157 p<.01	F(1,9)=16.725 p<.01	F(1,9)=16.288 p<.01
speaker MM	F(1,9)=236.247 p<.001	F(1,9)=15.756 p<.01	F(1,9)=8.344 p<.05	F(1,9)=9.111 p<.05
speaker MK	F(1,9)=916.053 p<.001	F(1,9)=33.665 p<.001	F(1,9)=203.195 p<.001	F(1,9)=8.899 p<.05
speaker RS	F(1,9)=971.253 p<.001	F(1,9)=16.971 p<.01	F(1,9)=1.971 n.s.	F(1,9)=21.973 p<.01

As far as the palatoalveolar /t͡ʃ/ is concerned, its closure duration does not significantly differ from the frication of /t͡s/ for each speaker. A similar conclusion can be drawn with respect to frication duration of /t͡ʃ/ and /t͡s/ with the only exception noted in the results of speaker MK where the frication in /t͡ʃ/ is significantly shorter than in /t͡s/, p<.05.

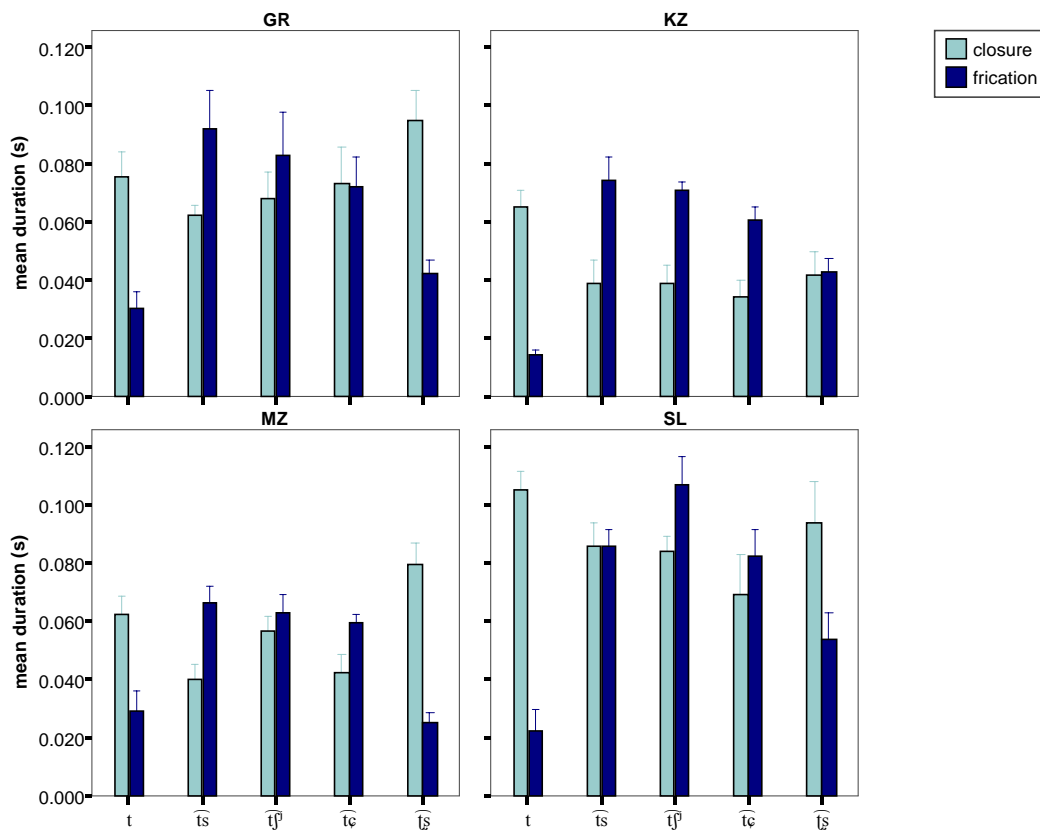
Let us compare these results to those obtained from Polish. Figure 9 presents the averages of closure and frication duration of Polish consonants for all four speakers together.



**Figure 9:** The average duration of closure and frication in Polish consonants.

The duration differences between closure and frication are significant for all items presented in Figure 9: /t/  $F(1,39) = 136.774$   $p < .001$ , /t͡ɕ/  $F(1,39) = 7.222$   $p < .05$ , /t͡s/  $F(1,39) = 17.072$   $p < .001$ , /t͡ʂ/  $F(1,39) = 37.195$   $p < .001$ , /t͡ɕʲ/  $F(1,39) = 10.374$   $p = .003$ .

The most interesting result is probably the relation between the closure and frication in the postalveolar affricate, which I denoted /t͡ʂ/. This affricate, in contrast to others presented in Figure 9, displays a longer closure duration than frication duration. Its mean closure duration amounts to 77.5 ms, while its mean frication duration is 41.2 ms. Other affricates show a reverse pattern: /t͡ɕ/ 54.9ms vs. 68.6ms; /t͡s/ 56.8 ms vs. 79.5 ms, /t͡ɕʲ/ 61.9 vs. 80.8 ms. Only in the case of /t/ is the closure longer than the frication: 77ms vs. 24.3ms. From a statistical point of view the difference in closure duration is significant between /t͡ʂ/ and /t͡ɕ/ ( $F(4,99)=5.914$   $p < .05$ ) as well as between /t͡ʂ/ and /t͡s/ ( $F(4,99)=5.914$   $p < .05$ ). The differences between /t͡ʂ/ and /t͡ɕʲ/ as well as /t͡ʂ/ and /t/ with respect to the closure duration are not significant. As far as frication duration is concerned the only non-significant difference is the one between /t͡ʂ/ and /t/. Other affricates show a longer duration than /t͡ʂ/ does, which is highly significant ( $F(4,99)=71.205$   $p < .001$ ).



**Figure 10:** The average duration of closure and frication split according to Polish speakers.

Similar results are obtained in the pronunciation of the individual speakers, as shown in Figure 10. In the pronunciation of three speakers (GR, SL and MZ), the closure in [t͡ɕ] lasts longer than the fricative part of the affricate. These differences are significant; see Table 2. The only speaker who does not show this difference is speaker KZ in whose pronunciation of [t͡ɕ] the closure and frication are of almost the same duration and do not show any significant statistical effect. Still, the frication is short which is important for drawing conclusions with respect to articulatory characteristic of this sound, see below.

In the case of other affricates the frication is always longer than the closure phase although this effect is not always significant. Table 2 shows statistical calculation results as achieved for individual speakers.

**Table 2:** Relation between closure and frication duration in Polish consonants split by speaker from a statistical point of view.

	/t/	/c/	/t͡ɕ/	/t͡ɕ͡/	ʈ͡ʂ
speaker GR	F(1,9)=101.657 p<.001	F(1,9)=.026 n.s.	F(1,9)=22.927 p<.01	F(1,9)=119.445 p<.001	F(1,9)=3.457 n.s.
speaker SL	F(1,9)=359.850 p<.001	F(1,9)=3.187 n.s.	F(1,9)=.000 n.s.	F(1,9)=27.645 p<.01	F(1,9)=22.900 p<.01
speaker KZ	F(1,9)=371.067 p<.001	F(1,9)=62.861 p<.001	F(1,9)=51.570 p<.001	F(1,9)=.190 n.s.	F(1,9)=108.776 p<.001
speaker MZ	F(1,9)=359.850 p<.001	F(1,9)=3.187 n.s.	F(1,9)=.000 n.s.	F(1,9)=27.645 p<.01	F(1,9)=22.900 p<.01

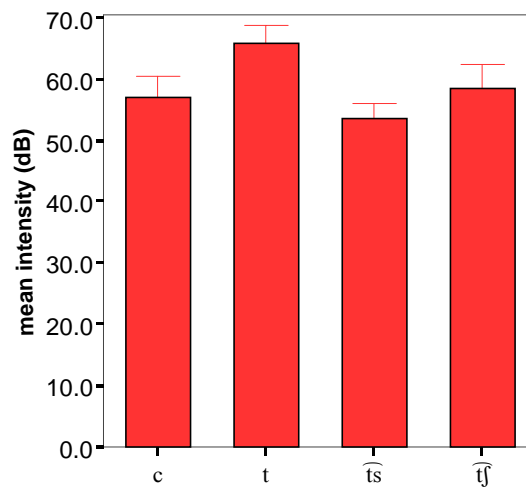
In summary, the investigation of the Polish and Czech postalveolar affricate, commonly denoted as <č> or /ʈ͡ʂ/ in IPA terms, shows that the two sounds are essentially different with respect to the closure and frication duration: whereas the closure phase in Czech /ʈ͡ʂ/ is significantly shorter than the frication phase, a reverse pattern is observable in the corresponding Polish sound - its closure lasts significantly longer than its frication component. This property also distinguishes the sound from other Polish affricates.

The results point to an important articulatory difference between the postalveolar affricates. The Czech affricate /ʈ͡ʂ/ is articulated with the tongue blade whereas the corresponding Polish sound is articulated with the tongue tip (also by speaker KZ). This is essential for classifying the latter sound as retroflex.

Finally, the results confirm that there is a difference between the Czech [tʃ] and Polish [tʃ]. Both the closure and frication duration are longer in the latter case which is attributed to the secondary palatalisation of the Polish sound. (closure 40.8 ms vs. 61.9 ms; frication 61.3 ms vs. 80.8 ms)

**Parameter (ii):** The amplitude of the frication phase.

Parameter (ii) includes the average of frication amplitude calculated from the end of the burst until the starting point of fundamental frequency. Figure 11 presents the results calculated for all four Czech speakers. For reasons of transparency, the results will be interpreted with focus on postalveolar affricates.

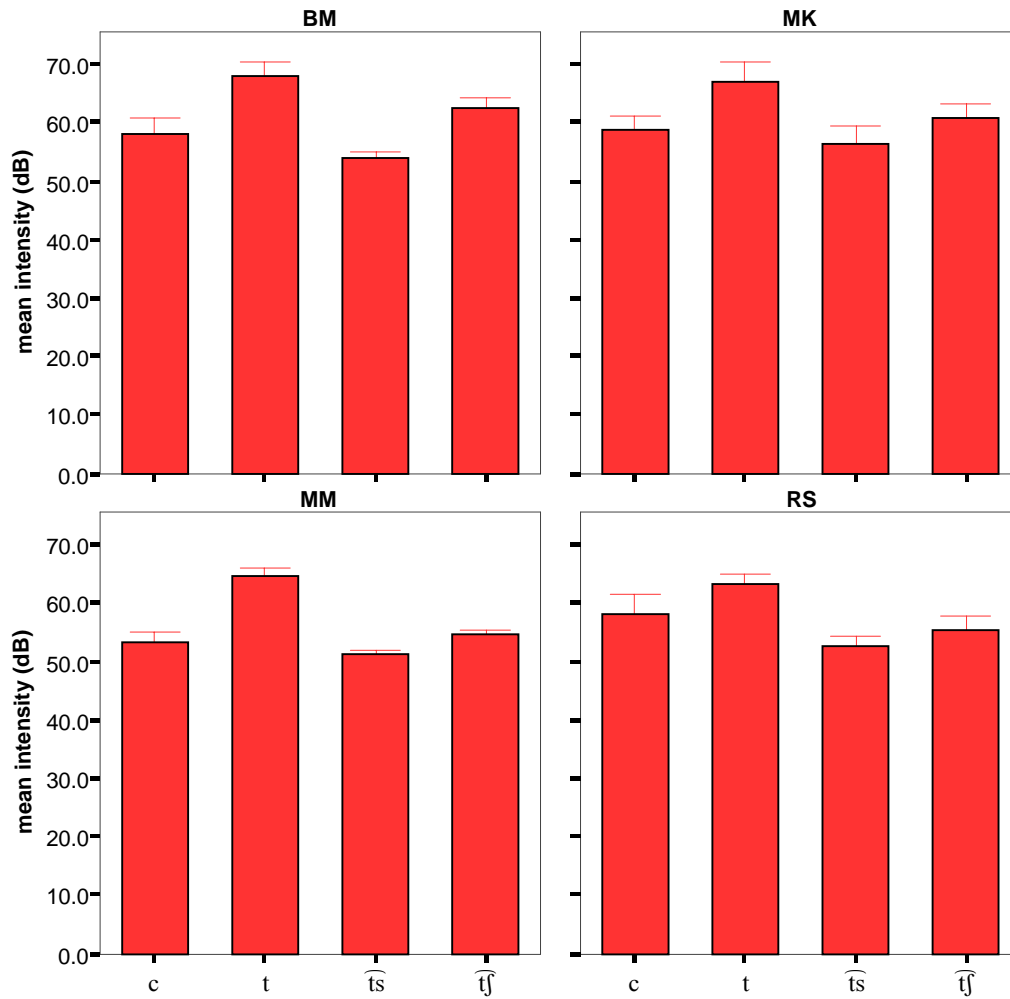


**Figure 11:** The amplitude average of frication in Czech

All consonants investigated were followed by the vowel [a] facilitating the comparison of the amplitude. The calculations of the amplitude of [a] following /t/, /c/, /ts/ and /tʃ/ show that independently of the item, it amounts to ca. 70 dB for every speaker. The only significant amplitude difference has been found between the frication in /ts/ and /t/ for speaker BM (69.84 dB vs. 74.45 p<.05; F(3,19) = 4.392).

The results presented in Figure 11 indicate that the average amplitude of /tʃ/ is significantly higher than the average amplitude of /ts/ (F(3,79) = 50.255 p<.001) and the average amplitude of /t/ (p<.001). There is no significant difference between the frication amplitudes of /tʃ/ and /c/.

Figure 12 shows average amplitudes of the frication of the consonants split according to Czech speakers.



**Figure 12:** The amplitude average of the frication phase split by speaker.

The results in Figure 12 show that the amplitude of  $/tʃ/$  is always higher than that of  $/ts/$  but the difference is statistically significant for two speakers only (BM and MM, see Table 3). With respect to  $/t/$  the difference is significant for every speaker and with respect to  $/c/$  almost always not significant; see again Table 3.

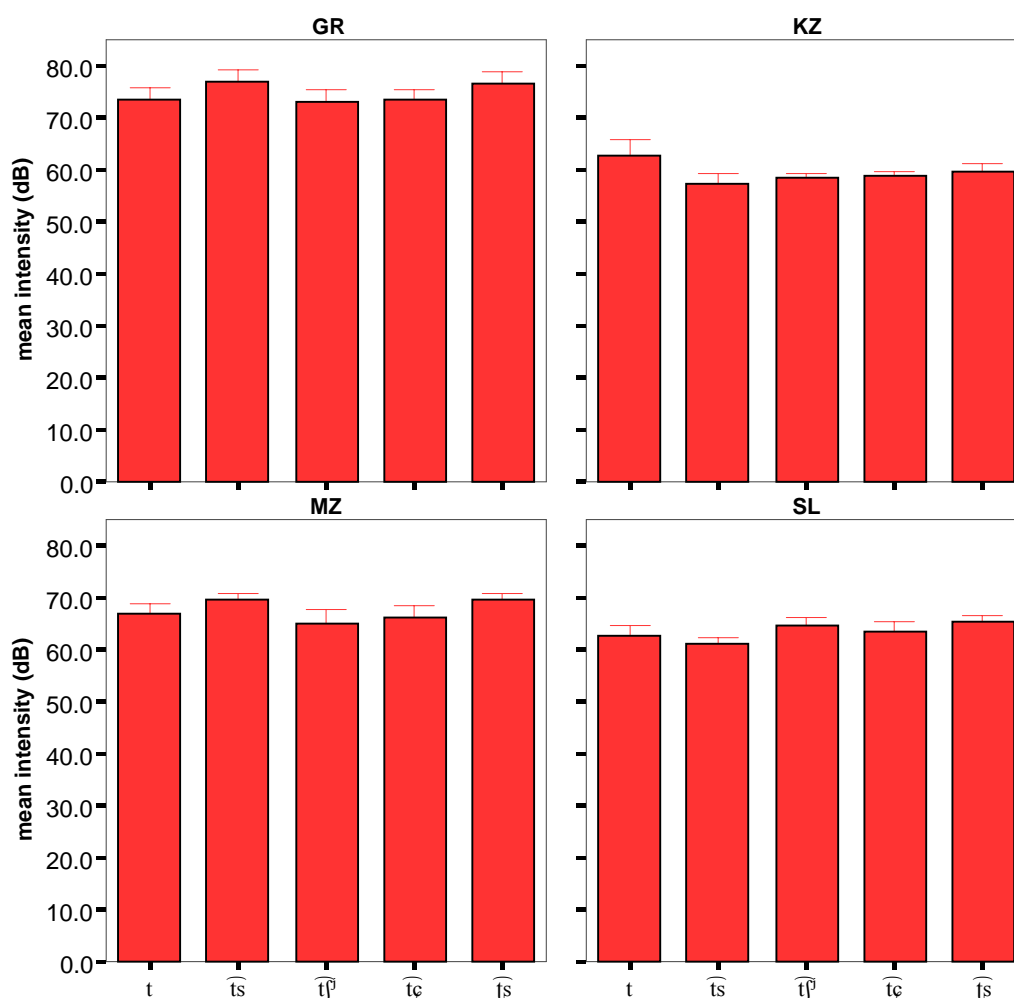


**Table 3:** Mean amplitude of the consonantal frication phase split by Czech native speakers from a statistical point of view.

		ʧ vs. t	ʧ vs. c	ʧ vs. ʦ
speaker BM	F(3,19)=41.175	p<.01	p<.05	p<.001
speaker MK	F(3,19)=13.264	p<.05	n. s.	n.s.
speaker MM	F(3,19)=115.230	p<.001	n. s.	p<.01
speaker RS	F(3,19)=16.656	p<.01	n.s.	n.s.

As far as Polish is concerned, the amplitude of the Polish retroflex /ʧ/ does not differ from the amplitude of other consonants. The results are presented in Figure 13. It should be noted that the amplitude of the following [a] was not dependent on the consonant under investigation. No significant effect has been found between the items as produced by individual speakers. However, the mean average amplitude of [a] for all items of every speaker shows some significant effects. The mean [a] amplitude calculated for speaker GR (mean 85.75 dB) was significantly higher than the mean [a] amplitude calculated for speaker KZ (p<.001, mean 73.47 dB), speaker MZ (p<.001, mean 77.91 dB) and speaker SL (p<.001, mean 74.65 dB). Significant differences have also been found between speakers MZ and KZ (p<.001) as well as MZ and SL (p<.001, F(3,99)=197,721). The difference in amplitude between speakers KZ and SL is not significant.

Due to the significance of effects found in [a] amplitude I dispensed with presenting the amplitude averages attained for all speakers together. Figure 13 presents the results for each speaker separately.



**Figure 13:** The average amplitude of the frication split by the speakers.

For speakers GR and KZ the amplitude of the consonant  $\widehat{t\text{ʂ}}$  does not show significant differences with respect to other consonants presented in Figure 13. For speaker MZ the only significant difference in amplitude is that between  $\widehat{t\text{ʂ}}$  and  $\widehat{t\text{ʃ}}$  ( $F(4,25)= 6.091$   $p<.05$ ) and for speaker SL it is between  $\widehat{t\text{ʂ}}$  and  $\widehat{t\text{ɕ}}$  ( $F(4,24)=5.964$   $p<.01$ ).

In summary, the investigation of frication amplitude does not show significant effects in Polish consonants. Hence, this parameter does not appear to be helpful in stating the differences among the affricates and between the Polish  $\widehat{t\text{ʂ}}$  and the corresponding Czech affricate.

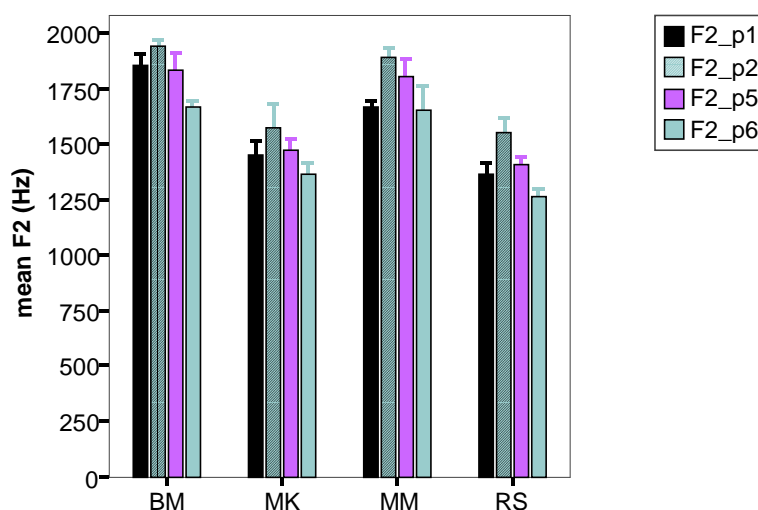
**Parameter (iii):** The shape of the vowel formants F2 and F3 preceding and following the consonant

As far as the formant transitions is concerned, the main point of interest is the second formant (F2) and the third formant (F3). While F2 characterizes the horizontal shape of the tongue, F3 is especially important for proving the possible retroflexivity the Polish palatoalveolar sound in question. If the F2 of the following vowel were falling, then it would indicate the transition from the palatal position. characteristic for palatalized segments or palatals which are produced with the tongue blade or tongue dorsum; see for example Ladefoged & Maddieson (1996:364).

In addition, the retroflex character of the sounds can also be postulated by looking at F3 transitions. There is a general consensus in the literature concerning a common acoustic cue of retroflexes which is a falling F3, due to the further rearwards (but still coronal) place of articulation; see for example Stevens & Blumstein (1975), Narayanan & Kaun (1999).

In order to analyze the spectral shape of the vowels preceding and following the consonant under consideration, formants of the vowel segments were measured semi-automatically by means of Linear Predictive Coding (LPC). For the formant analysis the software PRAAT (version 4.3) was used. Prior to formant analysis the sounds were downsampled to 11 KHz for female, and 10 KHz for male speakers to maintain the spectral structure of the first five formants only. The LPC was then calculated by using the following parameters: pre-emphasis frequency 50 Hz, analysis window duration 0.0256s, time step 0.001s and a prediction order of 13 for female, and 12 for male speakers. LPC spectra were calculated at four time instants (1, 2, 5, 6 in Figure 1) that were manually derived prior to calculation of the spectra. Maximally five peaks from a LPC spectrum derived by peak picking were temporarily considered as formants. As in some cases one formant value could not be detected by the peak-picking algorithm, the five temporary formant values were checked for every spectrum and manually corrected if necessary in order to determine the final formant frequencies.

Figure 14 shows the average values of the second formant in items including the palatoalveolar affricate  $/tʃ/$  as calculated for four Czech speakers. The bars represent mean F2 frequency at four different points, as described by the legend (e.g. f2\_p1 stands for the mean value of the second formant at point (1)). On the horizontal axis the initials of individual speakers are shown.

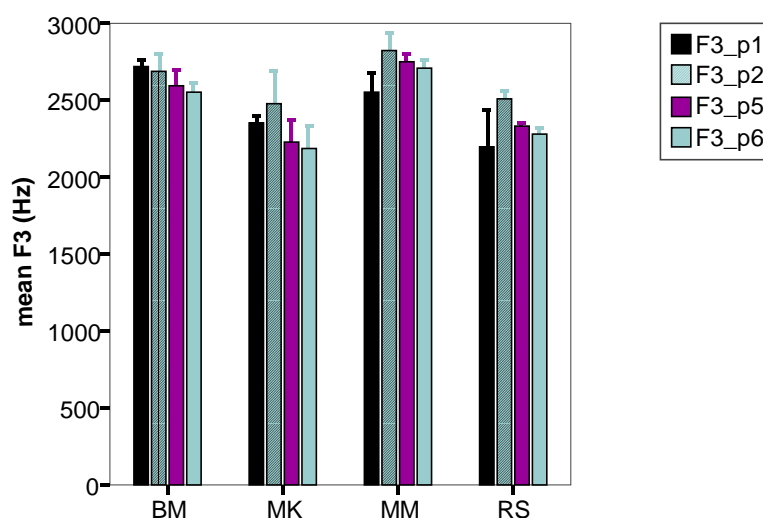


**Figure 14:** The average values of F2 as obtained for [tʃ] by Czech speakers.

The results shown in Figure 14 show regularities in F2 shape in the pronunciation of all four speakers. The second formant of the vowel preceding the consonant is rising and falling when it occurs after the consonant. The rising F2 is statistically significant for two speakers MM ( $F(3,19) = 13.416$   $p < .01$ ) and RS ( $F(3,19) = 29.477$   $p < .001$ ). The falling F2 is significant for three speakers: MM  $p < .05$ , RS  $p < .01$ , BM  $p < .01$  ( $F(3,19) = 21.744$ ).

In addition, a clear difference between F2 of male and female pronunciation is visible in the sense that the former is considerably lower than the latter. Hence, the differences between speaker BM and MM with respect to the F2 at all measurement points (f2\_p1, f2\_p2, f2\_p5, f2\_p6) are not significant. Similarly, the differences in F2 for speakers RS and MK are not significant, with the only exception concerning f2\_p1, where the difference is highly significant  $p < .001$ ,  $F(3,19) = 95.326$ . All other differences are highly significant.

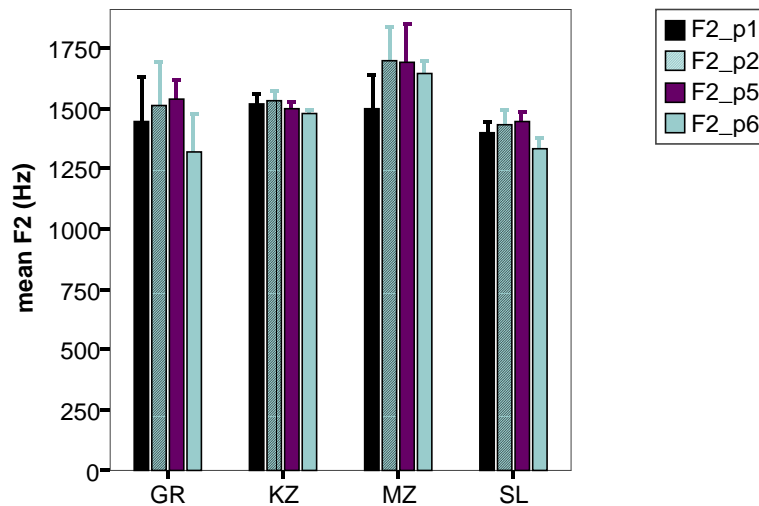
Figure 15 shows the mean values of F3 at the same four points as in the case of F2 calculated for each Czech speaker separately.



**Figure 15:** The average F3 values as obtained for [tʃ] by Czech speakers.

As far as the third formant is concerned, no regularities in its shape can be stated. The only significant difference has been found in the pronunciation of speaker MM: F3 of the preceding vowel is rising ( $F(3,19) = 6.955$   $p < .01$ ).

Figure 16 shows the shape of the second formant calculated for each Polish speaker separately.

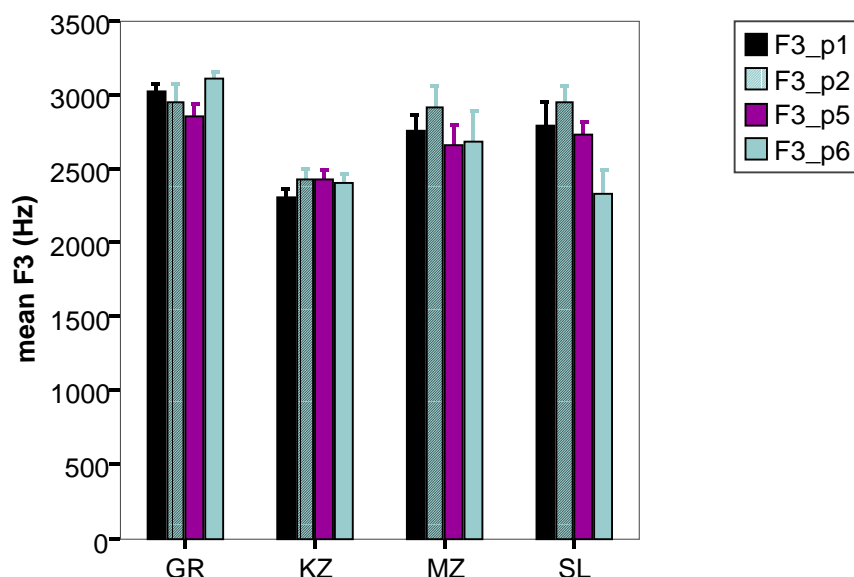


**Figure 16:** The average of F2 as obtained for [tʃ] by Polish speakers.

The results presented in Figure 16 show only two significant effects on F2 shape: for speaker MZ for the preceding vowel ( $F(3,19) = 5.936$   $p < .05$ ) and for speaker SL for the following vowel ( $F(3,19) = 5.345$   $p < .05$ ).

With these results obtained, a conclusion may be drawn that in the pronunciation of the Polish postalveolar affricate [tʂ] the formants of the preceding and following vowel remain pretty stable.

Figure 17 shows the shape of the second formant calculated for each Polish speaker separately.

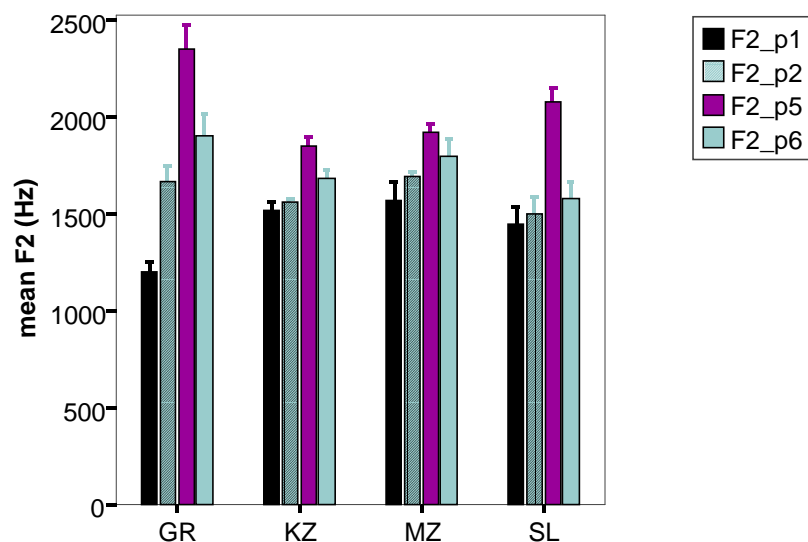


**Figure 17:** The average of F3 as obtained for [tʂ] by Polish speakers.

The results presented in Figure 17 are significant in only one case. Speaker SL shows a significant difference between the beginning and steady state of the following vowel: F3 is falling  $F(3,19) = 16.104$   $p < .005$ .

In summary, the investigation of vowel formants reveals that in the case of the Czech [tʃ] F2 of the preceding vowel is raising whereas F2 of following vowel is falling, which is a typical pattern for sounds produced with the raised and fronted tongue blade. The corresponding Polish affricate shows pretty stable formants of the flanking vowels. Hence, the two sounds are different with respect to F2. The shape of F3 in both languages is stable, in the sense that it does not show rising or falling effects.

If we compare the Polish [tʂ] and the Czech [tʃ] to the Polish palatalized palatoalveolar [tʃʲ], it turns out that the Czech [tʃ] and the Polish [tʃʲ] share some properties as far as the formants of the surrounding vowels are concerned. The results of F2 are shown in Figure 18.



**Figure 18:** The average values of F2 as obtained for [tʃ] by Polish speakers.

In contrast to F2 of Polish [tʃ], the investigation of the F2 of [tʃʲ] reveals its falling shape in the vowel following the consonant. This effect is statistically highly significant in the pronunciation of three speakers (GR  $F(3,19) = 121.628$   $p < .001$ , KZ  $F(3,19) = 70.877$   $p < .001$ , SL  $F(3,19) = 55.329$   $p < .001$ ). As far as the shape of F2 of the vowel preceding [tʃʲ] is concerned, its rising shape is significant only in the pronunciation of the speaker GR  $p < .001$ . The rising F2 makes the Polish [tʃʲ] more similar to the Czech [tʃ] which independently confirms the raised tongue blade in the production of the two sounds and the difference between the Czech [tʃ] and the Polish [tʃ].

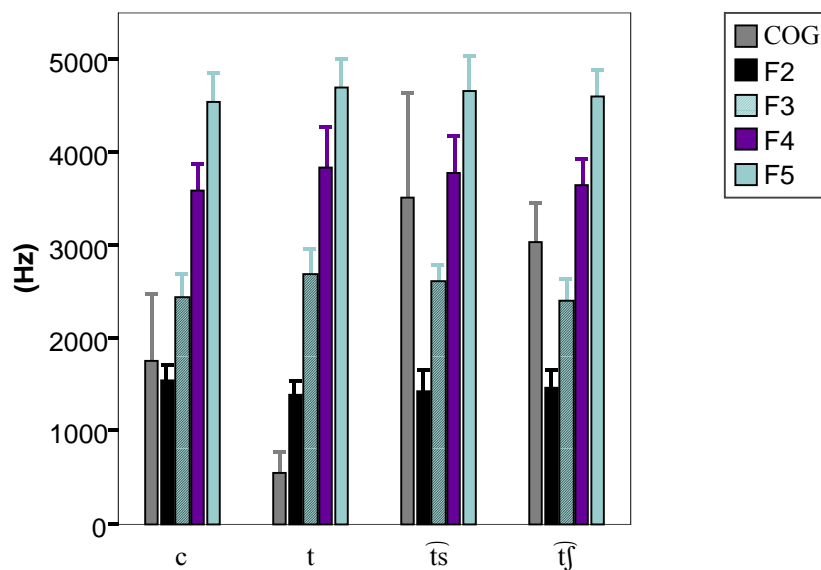
Finally, the investigation of F3 does not show significant effects apart from with speaker SL, whose F2 rises into the consonant and also rises from the consonant into the following vowel. Both effects are slightly significant  $F(3,19) = 7.035$ ,  $p < .05$ .

**Parameter (iv):** The center of gravity values of the frication phase

The (non)retroflexivity of the fricative part of an affricate (as well as fricatives) can be also inferred from measurements of the spectral mean, i.e., the center of gravity values (COG); see Jassem (1979), Nittrouer, Studdert-Kennedy & McGowan (1989), Gordon, Barthmaier & Sands (2002). With regard to articulation, the COG correlates to the size of the front cavity: The smaller the cavity, the higher the COG values. Consequently, if the supralaryngeal constriction is located at more posterior places, the front cavity is larger and the spectral mean is therefore lower. Lower COG values are expected for those retroflexes which display a relatively large front cavity.

The center of gravity values (COG) were calculated for the fricative portion of affricates and in the case of stop and vowel sequences [ta] or [ca] for the frication phase between the burst and the beginning of the following vowel. The fricative portion or frication phase respectively was extracted by a 25.6 ms long Hanning window centered on a time instant (point 4 in Figure 6) manually derived prior to the cog analysis. At first the spectrum was calculated by means of an overall spectral analysis (Fourier transform) over the frication portion. Then the center of gravity of the spectrum was calculated with the "power" setting 'p=2'.

Figure 19 presents mean COG values of frication in relation to the mean formant values of the following vowel (F2, F3, F4, F5).

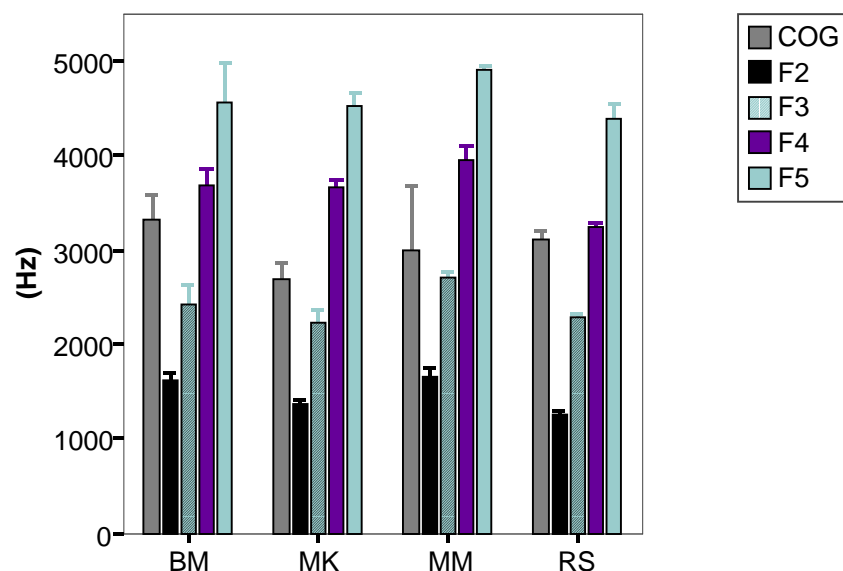


**Figure 19:** The average of COG values of frication in relation to the formant values of the following vowel as obtained for Czech.

Results presented in Figure 19 show different COG values with respect to the formant values of the following vowel. The lowest COG value is obtained for [t], followed by [c]. Much higher COGs are displayed by the fricative component of the affricates [ts] and [tj], whereby the highest COGs are shown by [ts]. The COGs of [tj] are situated between the third and the fourth formant. The differences here are highly significant: COG vs. F3  $p < .001$ , COG vs. F4  $p < .001$   $F(4,99) = 326.140$ . This is in contrast to [ts] COGs which are as high as the fourth formant (the difference is not significant).

In Figure 20 the results are split according to speakers. Note that all results refer to [tj].





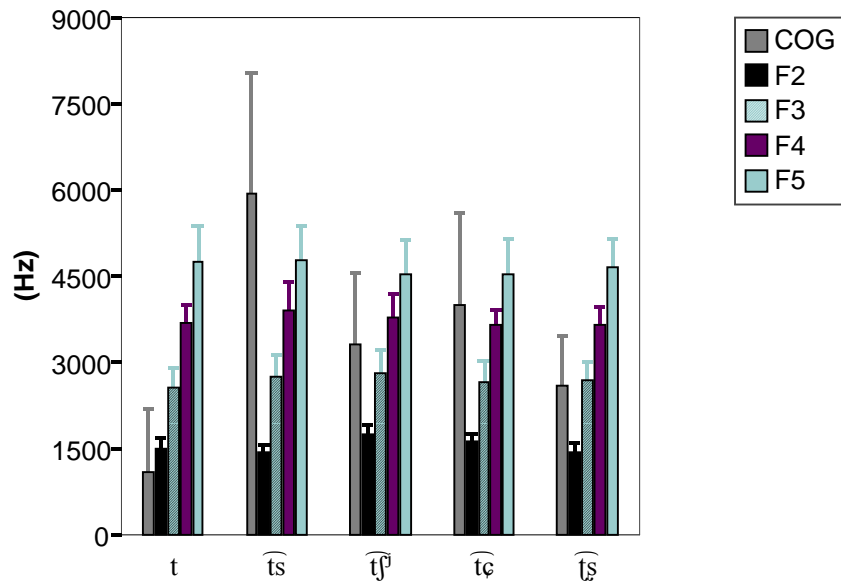
**Figure 20:** The average of COG values and formant values as obtained for [tʃ] by Czech speakers.

Speaker BM shows a significant effect of the COG value in relation to F3  $p < .001$  but not in relation to F4 ( $F(4,24) = 101.521$ ). Similar results are obtained for speaker RS: COG vs. F3  $p < .001$ , COG vs. F4 not significant ( $F(4,24) = 847.890$ ). In the case of speaker MK the differences are highly significant: the COG value is higher than F3 ( $p < .001$ ) but lower than F4 ( $p < .001$ ,  $F(4,24) = 512.544$ ). In the pronunciation of speaker RS the COG is not significant with respect to F3 and significantly lower with respect to F4  $p < .01$  ( $F(4,24) = 73.932$ ).

Figure 21 presents mean COG values of Polish consonants in relation to the formants of the following vowel at its steady state.<sup>6</sup>

The results presented in Figure 21 indicate that the COG values of [tʃ] are the lowest among Polish affricates. The COGs of [tʃ] are not higher than the third formant and the relation between them is not significant ( $F(4,99) = 118,168$ ). The relation to other formants is highly significant  $p < .001$ .

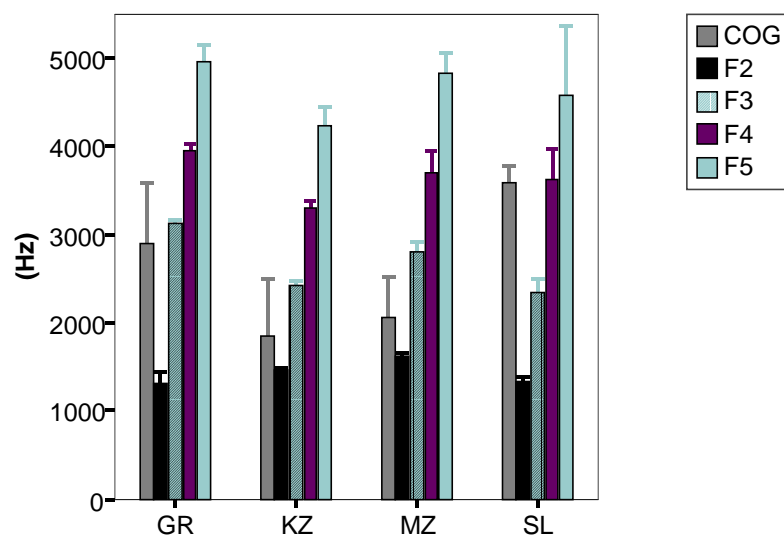
<sup>6</sup> It should be noted that COG values of the frication of [tʃ] show a very high standard deviation for both Polish and Czech affricates. In fact, this result mirrors great differences found among native speakers. It is also partly ascribed to the difficulty in extracting frication from the burst because the two components could not often be differentiated.



**Figure 21:** The average of COG values of frication in relation to the formant values of the following vowel as obtained for Polish.

The COG of [t] is as high as for F2 (the difference is not significant here), whereas the COGs of [tʂ] relate to F3 (no significant difference). Furthermore, the COGs of [tʃ] are higher than F3 but lower than F4. The difference is neither significant with respect to F3 nor to F4. Finally, the highest COGs are achieved by [ts] and are higher than F5. The difference is slightly significant ( $F(4,104) = 62.705$   $p < .05$ ).

Figure 22 presents the COGs of the fricative part of [tʂ] split by speakers.



**Figure 22:** The average COG and formant values as obtained for [tʂ] by Polish speakers.

Splitting the results by speaker reveals that the COG values of [tʂ] are lower in comparison to Czech [tʃ] in two cases. In the pronunciation of speakers KZ and MZ the COGs are not higher than the second formant of the following vowel from a statistical point of view (the relation between the COG and F2 is not significant). The COG of [tʂ] in the pronunciation of speaker GR relates to the third formant (no significant effect has been found in this relation). Finally, the pronunciation of [tʂ] by speaker SL shows rather high COG values - as high as the fourth formant.

In summary, the investigation of center of gravity values shows that Czech postalveolar affricates display higher COGs than the corresponding Polish sounds. The results indicate that during the articulation of the Polish sound, the front cavity is larger than in Czech. This is, however, attested for two Polish speakers. The two other speakers show higher COGs which suggests the variability in the size of the front cavity. Czech speakers show less variability and the COG values are higher which is in agreement with the expectations.

#### **Parameter (v): spectral peaks of the burst and frication**

The final parameter investigated in the present study included the correspondence of the frequency of the highest-amplitude spectral peak at the burst and at the steady-state part of the fricative to the formant frequencies of the following vowel. Implementing such a strategy makes possible a cross-speaker comparison; cf. Stevens (1989), Ohde & Stevens (1983), Hedric & Ohde (1993), Kim (2001).

Stevens (1989) states that in the case of [ʃ], its highest-amplitude spectral peak occurs at about the same frequency as the third formant of the following vowel [a]; see also Hedric & Ohde (1993). The alveolar [s] displays its highest spectral peak at about frequency of the fifth or higher formant of the following vowel [a].

According to Stevens (1989:26) the highest amplitude peak in its relation to the formant of the following vowel reflects the size of the front cavity. In the case of the longer front cavity the highest spectral peak is lower in relation to the following vowel formants. Since the retroflex is expected to have the largest front cavity due to its place of articulation and possible rounding, its highest spectral peak should be the lowest in comparison to the highest spectral peak of [s], [ʃ] or [ɕ].

The same strategy can be applied to affricates, as Kim (2001) suggests. Since an affricate consists of an oral closure and fricative release, and both, as claimed by Stevens (1993), can be manipulated independently, the highest spectral peak can be stated in its relation to the following vowel independently

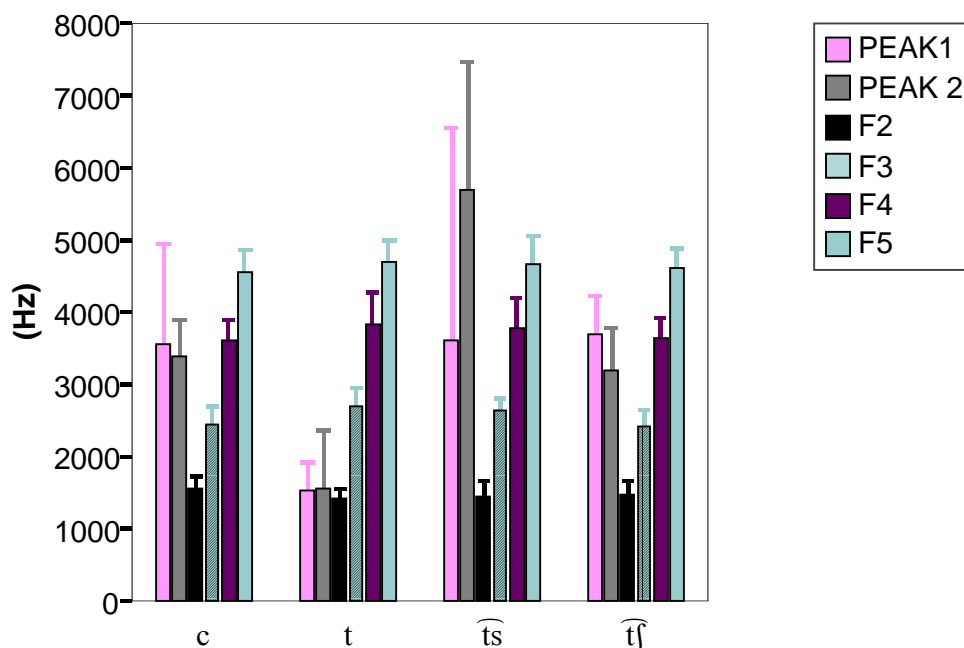
for both parts of the affricate. In experimental practice, this means that the highest spectral peak can be compared to the formants of the following vowel at (i) the release burst of the stop and (ii) the steady-state portion of the fricative. The steady-state portion of a fricative starts at least 20-30 ms after the release of the stop. This has been postulated by Stevens (1993) and adopted by Kim (2001) for the investigation of the Korean affricate [t͡s].

In the following a similar procedure will be applied for the investigation of Polish and Czech stops and affricates. In contrast to parameter (iii), the peaks will be determined for both the burst and frication.

For measurement purposes, the cursor was placed at three different points of the spectrogram of the item investigated: at the burst, i.e. point (3) in Figure 6, at the steady state portion of the frication, i.e. point (4) in Figure 6, and at the steady state portion of the following vowel, i.e. point (6) in Figure 6.

The formant frequencies of the following vowel were obtained in exactly the same way as presented for parameter (iii) above. The peak-picking algorithm objectively identified the frequency peaks of the burst and the frication. Only the frequency of the highest peak was saved.

Figure 23 present the results obtained for Czech. The bars illustrate the averages of the highest burst peaks (PEAK 1), and the highest frication peaks (PEAK 2), as well as the average formant values of the following vowel [a] in its steady state (F2 = the second formant, F3 = the third formant, F4 = the fourth formant, F5 = the fifth formant). The results show the averages for all four Czech speakers.



**Figure 23:** The correspondence of the highest peaks of the burst and frication in relation to the formants of the

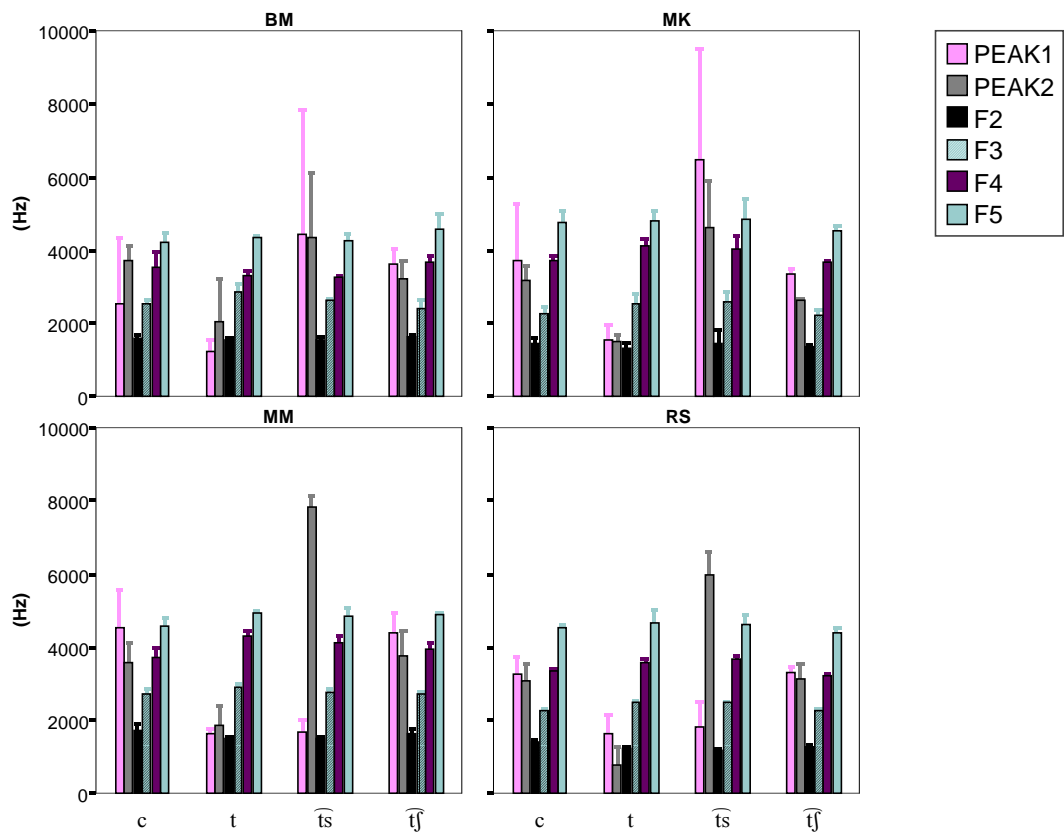
following consonant for all Czech speakers.

The following discussion will be limited to the postalveolar affricate [tʃ], the main point of interest for the present study. A Scheffé test comparing both peak values (PEAK1 and PEAK2) to four formants of the following vowel reveals that:

(i) PEAK 1 does not significantly differ from F4, whereas it is significantly higher than F2, and F3 and lower than F5 ( $p < .001$ ),  $F(5,119) = 157.665$ ,

(ii) PEAK 2 is placed between the third and the fourth formant. It is significantly higher than F2, F3 ( $p < .001$ ) and lower than F4 ( $p < .05$ ) and F5 ( $p < .001$ ),  $F(5,119) = 157.665$ .

Figure 24 presents the same parameters as obtained by individual Czech speakers.



**Figure 24:** The correspondence of the highest peaks of the burst and frication in relation to the formants of the following consonant for all Polish speakers.

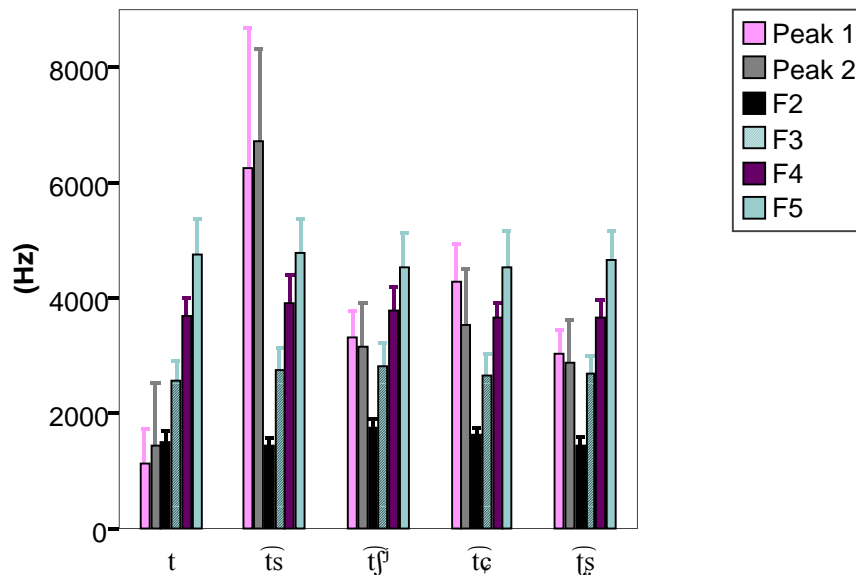
As far as the spectral peaks of [tʃ] are concerned, in the pronunciation of three speakers PEAK 1 and PEAK 2 reach almost the same frequency as the fourth

formant. In the pronunciation of speaker MM, PEAK 1 does not even significantly differ from F5. Lower peaks are observed in one case only: this is speaker MK whose PEAK 1 and PEAK 2 are higher than F3 but lower than F4. Table 4 shows the statistical details about the relation of the two spectral peaks to the formants of the following vowel.

**Table 4:** Statistical calculations obtained for [tʃ] by Czech speakers.

		F2	F3	F4	F5	PEAK1
speaker MK	PEAK1	p<.001	p<.001	p<.05	p<.001	
F(5,29)= 540.759	PEAK2	p<.001	p<.001	p<.001	p<.001	n.s.
speaker BM	PEAK1	p<.001	p<.001	n.s.	p<.01	
F(5,29)=49.342	PEAK2	p<.001	p<.05	n.s.	p<.001	p<.001
speaker MM	PEAK1	p<.001	p<.001	n.s.	n.s.	
F(5,29)= 40.072	PEAK2	p<.001	p<.05	n.s.	p<.01	n.s.
speaker RS	PEAK1	p<.001	p<.001	n.s.	p<.001	
F(5,34)=154.131	PEAK2	p<.001	p<.001	n.s.	p<.001	n.s.

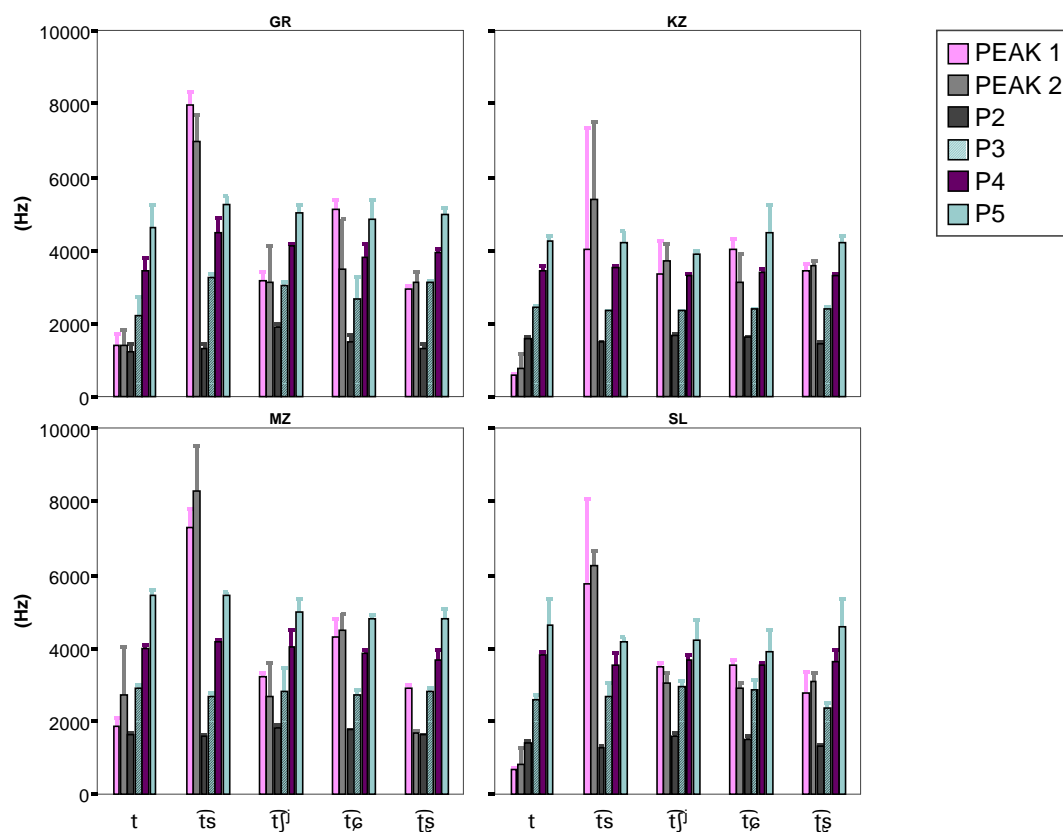
The results obtained for Polish are shown in Figure 25. Again, the following discussion will be limited to [tʃ].



**Figure 25:** The correspondence of the highest peaks of the burst and frication in relation to the formants of the following consonant for all Polish speakers

A post-hoc Scheffé test reveals that the highest spectral peak of the burst (PEAK 1) and the fricative part of the affricate [tʂ] (PEAK 2) are both as high as the third formant, since the differences between PEAK 1 vs. F3 and PEAK 2 vs. F3 are not significant. The two peaks are higher than F2 ( $p < .001$  for both PEAKS), lower than F4 (PEAK 1  $p < .01$ , PEAK 2  $p < .001$ ) and lower than F5 ( $p < .001$  for both PEAKS,  $F(5,119) = 116.149$ ).

If we split the results by speaker we obtain the relations as presented in Figure 26.



**Figure 26:** The correspondence of the highest peaks of the burst and the frication in relation to the formants of the following consonant for Polish speakers

Splitting the results according to speakers does not lead to similar effects as shown by Figure 25. Although in the pronunciation of three speakers (GR, MZ, SL) PEAK 1 is as high as F3 from a statistical point of view, other effects are attested as well. For example, for speaker SL PEAK1 is not significantly lower than F4. Speaker KZ does not show any significant effects in the relation between PEAK 1 and F4, but in his pronunciation PEAK 2 is even significantly higher than F4. For speaker MZ, on the other hand, PEAK 2 is as low as F2. Table 5 shows a more detailed picture on statistical calculations.

**Table 5:** Statistical calculations obtained for [tʂ] by Polish speakers

		F2	F3	F4	F5	PEAK1
speaker GR F(5,29)=278.248	PEAK1	p<.001	n.s.	p<.001	p<.001	
	PEAK2	p<.001	n.s.	p<.001	p<.001	n.s.
speaker MZ F(5,29)=308.511	PEAK1	p<.001	n.s.	p<.001	p<.001	
	PEAK2	n.s.	p<.001	p<.001	p<.001	p<.001
speaker KZ F(5,29)=334.872	PEAK1	p<.001	p<.001	n.s.	p<.001	
	PEAK2	p<.001	p<.001	p<.05.	p<.001	n.s.
speaker SL F(6,34)=32.806	PEAK1	p<.01	n.s.	n.s.	p<.001	
	PEAK2	p<.001	n.s.	n.s.	p<.01	n.s.

In summary, the investigation of the spectral peaks of the burst and frication phase does not show significant differences between Czech and Polish postalveolar affricates. This suggests a high variability of the front cavity size, an observation partly confirmed by COG measurements.

Finally, the investigations of different acoustic parameters have revealed significant differences between the Czech [tʃ] and the corresponding Polish sound. It has been shown that there is an essential difference between the closure duration and the frication duration. While in the Czech [tʃ], the frication is significantly longer than the closure, the Polish postalveolar affricate [tʂ] shows a reverse pattern: a long closure followed by a short frication. This indicates that the affricate is an apical sound because its release, i.e. the fricative part, lasts for a short time only. In the case of the Czech [tʃ], the fricative part is of considerably longer duration because the tongue blade takes longer to separate from the prepalate. Another parameter, which also shows consistent differences between the affricates under consideration, is that of the F2 of the following vowel, which has a rising shape in the Czech [tʃ] and shows stability in Polish [tʂ].

Another parameter, i.e. the amplitude of the frication phase, appears not to be helpful in determining the places of articulation of sibilants. All Polish consonants show nearly the same amplitude (without significant effects).

An average calculation of center of gravity shows a clear difference between Polish and Czech postalveolar affricates, albeit not confirmed for each speaker individually.

Finally, the correspondence of the highest spectral peaks to the formants of the following vowel show rather a large variability, and only partly confirm the differences between the two affricates. This result does not only indicate a



variability of the front cavity size for Czech and Polish but it also independently confirms that the articulatory gestures are not necessarily stable. This point is discussed in Żygis (in progress).

## 6 A DT-Analysis

In the following, the development of the Czech  $\widehat{t\mathfrak{f}}$  and the Polish  $\widehat{t\mathfrak{s}}$  will be analyzed in terms of Dispersion Theory. Two types of constraints are involved in the present analysis: markedness constraints, and faithfulness constraints. The markedness constraints are grounded in the articulatory and perceptual properties of the sounds under consideration. The faithfulness constraints regulate the relation between the underlying and phonetic representation of the items investigated.

The faithfulness constraints insure a faithful parsing of features of underlying representation to the phonetic surface. For the present analysis, it is assumed that the faithfulness constraints evaluate the post-lexical mapping; see Kiparsky (1988), Padgett & Żygis (2003).

A constraint which is involved in the present analysis is IDENT<sub>sibilant</sub> presented in (13).

(13) IDENT<sub>sibilant</sub> : Sibilant features agree on the lexical and post-lexical level.

The articulatory markedness constraints follow from the scale presented in (14). This scale shows that the secondarily palatalized  $\widehat{t\mathfrak{f}}]$ , the retroflex  $\widehat{t\mathfrak{s}}$  and the alveolopalatal  $\widehat{t\mathfrak{c}}$  are articulatorily more complex than the palatoalveolar  $\widehat{t\mathfrak{f}}$ ; see also Padgett & Żygis (2003) for a fricative scale. I do not attempt to rank  $\widehat{t\mathfrak{f}}]$ ,  $\widehat{t\mathfrak{s}}$ , and  $\widehat{t\mathfrak{c}}$  with respect to each other because in my view there is not enough detailed information available about the differences in the articulatory complexity of these sounds.

(14) Articulatory complexity scale:  $\widehat{t\mathfrak{f}}]$ ,  $\widehat{t\mathfrak{s}}$ ,  $\widehat{t\mathfrak{c}}$  >  $\widehat{t\mathfrak{f}}$

According to the scale in (14), there is a markedness ranking implying that  $\widehat{t\mathfrak{f}}]$ ,  $\widehat{t\mathfrak{s}}$ ,  $\widehat{t\mathfrak{c}}$  are more marked than  $\widehat{t\mathfrak{f}}$ . The ranking of the markedness constraints is presented in (15).

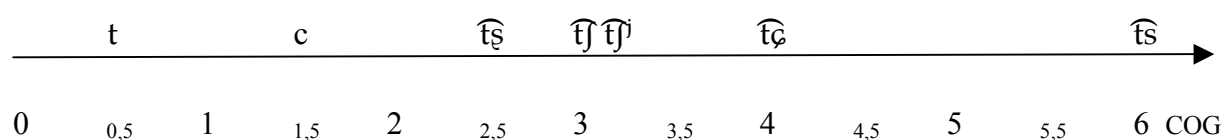
(15) \*ART complexity  $\widehat{t\mathfrak{f}}]$ ,  $\widehat{t\mathfrak{s}}$ ,  $\widehat{t\mathfrak{c}}$  < \*ART complexity  $\widehat{t\mathfrak{f}}$

Besides the articulatorily based constraints, there are perceptually grounded constraints which play an important role in the analysis of sibilant systems. These constraints are based on different acoustic parameters. The present

analysis will be limited to one acoustic parameter, namely the center of gravity (COG). I assume that in a complete perceptual analysis which expresses ‘hardness’ (low sibilant tonality) or ‘softness’ (high sibilant tonality) of the sounds, other parameters also have to be considered. Since the aim of the present analysis is to show a basic mechanism of phoneme change due to perceptual distinctiveness, I will focus on COG, see Zygis (2003b, 2005) for a more detailed discussion on this issue. In Padgett & Zygis (2005), the results of a perceptual experiment concerning the sibilant fricatives are presented and discussed.

The COG constraints are based on the COG scale as displayed in (16). The values assigned to each phonetic symbol approximately correspond to the averaged COGs obtained experimentally.

(16) COG scale



The constraints which regulate the distance between the sibilants are called Minimal Distance constraints, as introduced by Flemming (1995). Their aim is to maximize the auditory distinctiveness the sounds. The constraints are displayed in the format ‘Dimension:distance’ which indicates the distance between the segments along a given dimension. For example, if we took into consideration the scale in (16), then ‘MINDIST=COG:1’ would require a distance of 1 between the given stops on the COG dimension. This constraint is satisfied by, e.g., [tʃ] vs. [tɕ], [tʂ] vs. [tɕ] and others. At the same time, it is violated by, e.g., [tʃʲ] vs. [tʂ] or [tʃ] vs. [tʃʲ]. Other Minimal Distance constraints, e.g. ‘MINDIST=COG:2’ or ‘MINDIST=COG:3’ require a distance of 2 or 3, respectively.

In the following, it will be shown how the interaction of ‘markedness’ and ‘faithfulness’ constraints leads to the selection of the optimal candidates. The present analysis is limited to Polish and Czech.

The tableau in (17) presents the Proto-Slavic sibilant affricate inventory after the 1<sup>st</sup> Velar Palatalisation. It should be noted that the presentation order of the sibilants is crucial. IDENT<sub>sibilant</sub> evaluates the relation between input segments and the corresponding output segments displayed under the input (in the same column).

(17)

$\widehat{ts}$ $\widehat{t}^j$	IDENT <sub>sibilant</sub>	MINDIST=COG:1
$\widehat{ts}$ $\widehat{t}^{\text{c}}$	*!	
$\widehat{ts}$ $\widehat{t}^j$	*!	
$\widehat{ts}$ $\widehat{t}^{\text{s}}$	*!	
$\widehat{t}^{\text{c}}$ $\widehat{ts}$ $\widehat{t}^j$		

Although none of the candidates listed in (17) violates MINDIST=COG:1, the pair  $[\widehat{ts}]$  vs.  $[\widehat{t}^j]$  is selected as optimal, as it satisfies the high-ranking IDENT<sub>sibilant</sub>.

The optimal inventory  $/\widehat{ts} \ \widehat{t}^j/$  existed in Polish until approximately the 13<sup>th</sup> century when the alveolopalatal  $[\widehat{t}^{\text{c}}]$  emerging from the palatalised stop  $[t^j]$  entered the sibilant inventory (see 4. for details). The situation is illustrated by the tableau in (18).

(18)

$\widehat{ts}$ $\widehat{t}^{\text{c}}$ $\widehat{t}^j$	IDENT <sub>sibilant</sub>	MINDIST=COG:1
$\widehat{ts}$ $\widehat{t}^{\text{c}}$ $\widehat{t}^j$	*!	*
$\widehat{t}^{\text{c}}$ $\widehat{ts}$ $\widehat{t}^j$		*
$\widehat{ts}$ $\widehat{t}^{\text{c}}$ $\widehat{t}^{\text{s}}$	*!	
$\widehat{ts}$ $\widehat{t}^{\text{c}}$ $\widehat{t}^{\text{c}}$	*!	*
$\widehat{ts}$ $\widehat{t}^j$ $\widehat{t}^j$	*!*	*
$\widehat{ts}$ $\widehat{t}^j$ $\widehat{t}^{\text{s}}$	*!	*
$\widehat{ts}$ $\widehat{t}^j$ $\widehat{t}^{\text{s}}$	*!*	*

The optimal candidate  $/\widehat{ts} \ \widehat{t}^{\text{c}} \ \widehat{t}^j/$  is the only one which does not violate the high-ranking IDENT<sub>sibilant</sub>. It does violate MINDIST=COG:1 due to the perceptual distance between  $/\widehat{t}^{\text{c}}/$  vs.  $/\widehat{t}^j/$  which amounts to 0.5 on the COG scale. This inventory existed in Polish from the 13<sup>th</sup> until the 16<sup>th</sup> century. In the 16<sup>th</sup> century,  $/\widehat{t}^j/$  changed to  $[\widehat{t}^{\text{s}}]$ . I conclude that this change was motivated perceptually since the perceptual distance between  $/\widehat{t}^j/$  and  $/\widehat{t}^{\text{c}}/$  was not optimal.

In terms of a constraint ranking, the highest position of IDENT<sub>sibilant</sub> was taken by MINDIST=COG:1,5. This situation is illustrated by the tableau in (19) where all possible candidates are listed. The only sibilant which does not change in (19) is  $[\widehat{ts}]$ . I assume that  $[\widehat{ts}]$  had to be stable for at least two reasons. Firstly, its COG is the highest from all sibilants, and the only possibility of changing  $[\widehat{ts}]$  is having lower COGs and thus being closer to other sibilants. Secondly, the properties of the  $[\widehat{ts}]$  frication are perceptually prominent and  $[\widehat{ts}]$  creates

optimal distance from other sibilants. For these reasons I assume that the stability of [ʈs] is assured by the high-ranking IDENT [ʈs] which I will not list in the following tables for reasons of simplification.

(19)


ʈs    ʈʂ    ʈʃ	MINDIST = COG:1	IDENT <sub>sibilant</sub>
ʈs    ʈʂ    ʈʃ	*!	*
ʈs    ʈʂ    ʈʃ	*!	
☞ ʈs    ʈʂ    ʈʂ		*
ʈs    ʈʂ    ʈʂ	*!	*
ʈs    ʈʃ    ʈʃ	*!	**
ʈs    ʈʃ    ʈʃ	*!	*
ʈs    ʈʃ    ʈʂ	*!	**
ʈs    ʈʃ    ʈʂ		**
ʈs    ʈʂ    ʈʃ	*!	**
ʈs    ʈʂ    ʈʃ	*!	*
ʈs    ʈʂ    ʈʂ	*!	**
ʈs    ʈʂ    ʈʂ		**!
ʈs    ʈʃ    ʈʃ	*!	**
ʈs    ʈʃ    ʈʃ	*!	*
ʈs    ʈʃ    ʈʂ	*!	**
ʈs    ʈʃ    ʈʂ		**!

In the tableau in (19) only two candidates do not violate the high-ranking MINDIST= COG:1,5, namely, /ʈs ʈʂ ʈʂ/ and /ʈs ʈʂ ʈʂ/, which are actually the same. However, the latter inventory violates IDENT<sub>sibilant</sub> twice, whereas the former violates it ones, thereby being selected as the optimal candidate. This is because in the first inventory, one change took place from /ʈʃ/ to [ʈʂ]. It seems that the selected inventory is stable as it still persists in present-day Polish.

As far as Czech is concerned, the Proto-Slavic ancestor of the present Czech sibilant systems was the pair /ʈs, ʈʃ/, see tableau (17). Such a situation

lasted till the 14<sup>th</sup> century when the palatalised stop /tʲ/ changed to the palatal [c] and the latter was finally phonemised. This is displayed by the tableau (20).

(20)

$\widehat{ts}$ $\widehat{tʲ}$ c	IDENT <sub>sibilant</sub>	MINDIST=COG:1
$\widehat{ts}$ $\widehat{tʲ}$ $\widehat{tʲ}$	*!	*
 $\widehat{ts}$ $\widehat{tʲ}$ c		
$\widehat{ts}$ $\widehat{tʲ}$ $\widehat{tʃ}$	*!	*
$\widehat{ts}$ $\widehat{tʲ}$ $\widehat{tʃ}$	*!	*
$\widehat{ts}$ $\widehat{tʲ}$ $\widehat{tʃ}$	*!*	*
$\widehat{ts}$ $\widehat{tʲ}$ $\widehat{tʃ}$	*!*	*
$\widehat{ts}$ $\widehat{tʲ}$ $\widehat{tʃ}$	*!*	*
$\widehat{ts}$ $\widehat{tʲ}$ c	*!	

The optimal candidate  $\widehat{ts}$   $\widehat{tʲ}$  c/ violates neither IDENT<sub>sibilant</sub> nor MINDIST=COG:1. Note that the distance between  $\widehat{ts}$ / and  $\widehat{tʲ}$ / amounts to 3 on the COG scale and the distance between  $\widehat{tʲ}$ / and /c/ is 1.5. Hence, the systems seem to be relatively stable as far as the perceptual relations are concerned. Indeed, the only difference which took place was the depalatalisation of  $\widehat{tʲ}$ / to  $\widehat{tʃ}$ /. I assume that this change was primarily motivated by the articulatorily complexity of the secondarily palatalised  $\widehat{tʲ}$ /, banned by the constraint in (15). This is illustrated by the tableau in (21).

(21)

$\widehat{ts}$ $\widehat{tj}$ c	*ART [ $\widehat{tj}$ , $\widehat{ts}$ , $\widehat{t\check{c}}$ ]	MINDIST=COG1	IDENT
$\widehat{ts}$ $\widehat{tj}$ $\widehat{tj}$	*!	*	*!
$\widehat{ts}$ $\widehat{tj}$ c	*!		
$\widehat{ts}$ $\widehat{tj}$ $\widehat{ts}$	*!*	*	*!
$\widehat{ts}$ $\widehat{tj}$ $\widehat{t\check{c}}$	*!*	*	*!
$\widehat{ts}$ $\widehat{tj}$ $\widehat{tj}$		*!	*!*
$\widehat{ts}$ $\widehat{tj}$ $\widehat{tj}$	*!	*	*!*
$\widehat{ts}$ $\widehat{tj}$ $\widehat{ts}$	*!	*	*!*
$\widehat{ts}$ $\widehat{tj}$ c			*!

Two candidates  $\widehat{ts}$   $\widehat{tj}$   $\widehat{tj}$ / and  $\widehat{ts}$   $\widehat{tj}$  c/ do not violate the high-ranking \*ART [ $\widehat{tj}$ ,  $\widehat{ts}$ ,  $\widehat{t\check{c}}$ ] but only the latter is selected as optimal. This is due to the next constraint MINDIST=COG1,5 which is violated by two identical segments  $\widehat{tj}$   $\widehat{tj}$ / in the former inventory. These segments show, in fact, a merge of  $\widehat{tj}$ / and  $\widehat{c}$ / into [ $\widehat{tj}$ ] which would be an unexpected change.

In summary, the analysis proposed above shows that the changes in Polish and Czech affricate systems are not accidental. They can be seen to be clearly motivated by perceptual relations among the affricates. In addition, the analysis shows that articulatory complexity also plays a role in creating sibilant systems.

## 7 Conclusions

Slavic sibilant inventories underlie the principle in (22).

(22) Slavic sibilant systems:

In complex sibilant systems which include more than one postalveolar affricate or a strongly affricated  $\widehat{t^j}$ /, one of the affricate has a low sibilant tonality.

Special attention was paid to two selected Slavic languages, Polish and Czech, which display considerable differences in their coronal inventories. A diachronic study of the two inventories has contributed to the understanding of the role of perceptual relations for shaping the affricate systems in Czech and Polish. The

ProtoSlavic secondarily palatalized  $\widehat{t\mathfrak{s}}$  converted to  $\widehat{t\mathfrak{f}}$  in Czech, and to the retroflex  $\widehat{t\mathfrak{s}}$  in Polish. This discrepancy has been argued to have had a fundamental effect on the asymmetrical development of the ProtoSlavic  $/tj/$ ; in Czech  $/tj/$  had developed into the palatal stop  $/c/$ , and in Polish  $/tj/$  had converted to an alveolo-palatal affricate  $\widehat{t\mathfrak{c}}$  as  $\widehat{t\mathfrak{f}}$  converted to  $\widehat{t\mathfrak{s}}$ .

In the experimental part of the study it has been shown that the perceptual relations expressed in terms of acoustic parameters were of particular importance for the development of sibilant affricates in these two languages. The results have revealed a clear difference between the Czech and Polish affricate which is often assumed to be the same palatoalveolar affricate  $\widehat{t\mathfrak{f}}$  in both languages. Whereas the Czech affricate is indeed a palatoalveolar  $\widehat{t\mathfrak{f}}$ , the Polish postalveolar affricate can be classified as a retroflex  $\widehat{t\mathfrak{s}}$ . It has been also shown that stops such as  $/tj/$  and  $/c/$  do not have a direct perceptual impact on the affricate inventories, despite forming a natural class with them.

Finally, an analysis of Polish and Czech sibilant systems has been offered in the framework of Dispersion Theory.

### Acknowledgments

This paper has benefited from the advice and comments of Bernd Pompino-Marschall and Jaye Padgett. I am also grateful to Susanne Fuchs, Christine Mooshammer, and Ralf Winkler for their helpful comments. Ralf was also very helpful in issues concerning acoustic measurements.

### References

- Bethin, Ch. Y. (1992). *Polish Syllables. The Role of Prosody in Phonology and Morphology*. Columbus, Ohio: Slavica Publishers.
- Bhat, D.N.S. (1973). Retroflexion: an areal feature. In *Working Papers on Language Universals*. 27-67.
- Biedrzycki (1974). *Abriß der polnischen Phonetik*. Warszawa: Wiedza Powszechna.
- Carlton, T.R. (1991). *Introduction to the Phonological History of the Slavic Languages*. Slavica Publishers: Columbus, Ohio.
- Clements, G. N. (1999). Affricates as noncontoured stops. In: Fujimura, O., Joseph, B.D. & B. Palek (eds.). *Item Order in Language and Speech. Proceedings of LP'98*. Prague: The Karolinum Press. 271-299.
- Comrie, B. & Corbett, G.G. (eds.) (1993). *The Slavonic Languages*. London: Routledge.
- Flemming, E.S. (1995/2002). *Auditory Representations in Phonology*. London: Routledge.

- Gordon, M., Barthmaier, P. & K. Sands (2002). A cross-linguistic acoustic study of voiceless fricatives. *Journal of the International Phonetic Association* 32: 141-171.
- Gussmann, E. (1980). *Studies in Abstract Phonology*. Cambridge, Mass: MIT Press.
- Hall, T. A. (1997a). The Historical Development of Retroflex Consonants in Indo-Aryan. *Lingua* 101: 203-221.
- Hall, T. A. (1997b). *The Phonology of Coronals*. Amsterdam: Benjamins.
- Hamann, S. (2003). *The Phonetics and Phonology of Retroflexes*. Utrecht: LOT.
- Hamann, S. (2004). Retroflex fricatives in Slavic languages. *JIPA* 34, 53- 67.
- Hedrick, Mark S. & Ralph N. Ohde (1993) Effect of relative amplitude on perception of place of articulation. *JASA* 94: 2005-2026.
- Hume, E. (1994). *Front Vowels, Coronal Consonants and their Interaction in Nonlinear Phonology*. London: Garland.
- Jassem, W. (1979). Classification of fricative spectra using statistical discriminant functions. In: Lindblom B. & S. Öhman (eds.). *Frontiers of Speech Communication Research*. New York: Academic Press. 77-91.
- Jassem, W. (2003). Illustrations of the IPA. *Polish. Journal of the International Phonetic Association* 33: 103-107.
- Jongman, A., Wayland, R. & S. Wong (2000). Acoustic characteristics of English fricatives. *JASA* 108: 1252-1263.
- Keating, P. (1991). Coronal Places of Articulation. In C. Paradis & J.-F. Prunet (eds.) *Phonetics and Phonology. The Special Status of Coronals. Internal and External Evidence*, 29 – 48. New York: Academic Press.
- Keating, P. (1993). Phonetic Representation of Palatalization versus fronting. *UCLA Working Papers in Phonetics* 85:6-21.
- Kehrein, Wolfgang (2002). *Phonological Representation and Phonetic Parsing. Affricates and Laryngeals*. Tübingen: Max Niemeyer Verlag.
- Kim, H. (1997). The phonological representation of affricates: evidence from Korean and other languages. PhD Dissertation, Cornell University, Ithaca New York.
- Kim, H. (2001). The place of articulation of the Korean plain affricates in intervocalic position: an articulatory and acoustic study. *Journal of the International Phonetic Association* 31: 227-257.
- Kiparski, P. (1988). *Paradigm effects and opacity*. Ms. Stanford University.
- LaCharité, D. (1993). *The Internal Structure of Affricates*. Unpublished PhD dissertation, University of Ottawa.
- Ladefoged, P. & I. Maddieson (1996) *The Sounds of the World's Languages*. Oxford: Blackwell.
- Lamprecht, A., Šlosar, D. & J. Bauer (1977). *Historický vývoj češtiny*. Praha: SPN.



- Lehr-Spławiński, T. & Z. Stieber (eds.) (1957). *Gramatyka historyczna języka czeskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Machač, P. & R. Skarnitzl (2004). Selected acoustic properties of the Czech palatal plosives. In: Vích, R. (ed.) *13th Czech-German Workshop on Speech Processing*. Prague. 29-35.
- Narayanan, S & A. Kaun (1999). Acoustic modeling of Tamil retroflex liquids. *Proceedings of the International Congress of Phonetic Sciences 1999*, San Francisco, 2097-2110.
- Nitttrouer, S., Studdert-Kennedy, M. & R.S. McGowan (1989). The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. *Journal of Speech and Hearing Research* 32: 120-132.
- Ohde, Ralph N. & Kenneth Stevens (1983). Effect on burst amplitude on the perception of stop consonant place of articulation. *JASA* 74: 706-714.
- Ostaszewska, D. & J. Tambor (2001). *Fonetyka i fonologia współczesnego języka polskiego*. Warszawa: Wydawnictwo Naukowe PWN.
- Padgett, J. & M. Żygis (2003). The Evolution of sibilants in Polish and Russian. *ZAS Papers in Linguistics* 32: 155-174.
- Padgett, J. & M. Żygis (2005). *A Perceptual Study of Polish Fricatives, and its Relation to Historical Sound Change*. Ms.
- Palková, Z. (1994). *Fonetika a fonologie češtiny*. Praha: Univerzita Karlova.
- Rospond, S. (1971). *Gramatyka historyczna języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Rubach, J. (1984). *Cyclic and Lexical Phonology. The Structure of Polish*. Dordrecht: Foris.
- Rubach, J. (1994). Affricates as Strident Stops in Polish. *Linguistic Inquiry* 25: 119-144.
- Short, D. (1993). Czech. In: B. Combrie & G.G. Corbett (eds.): *The Slavonic Languages*. London: Routledge. 455-532.
- Stadnik, E. (1998). Phonemtypologie der slawischen Sprachen und ihre Bedeutung für die Erforschung der diachronen Phonologie. *Zeitschrift für Slavistik* 43(4): 377-400.
- Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics* 17: 3-45.
- Stevens, K. N. & Blumstein, S. E. (1975). Quantal aspects of consonant production and perception: a study of retroflex stop consonants. *Journal of Phonetics* 3: 215-233.
- Stieber, Z. (1957). *Gramatyka historyczna języka czeskiego*. Warszawa: PWN.
- Stieber, Z. (1962). *Rozwój fonologiczny języka polskiego*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Stieber, Z. (1969). *Zarys gramatyki porównawczej języków słowiańskich. Fonologia*. Warszawa: Państwowe Wydawnictwo Naukowe.
- Szpyra, J. (1995). *Three Tiers in Polish and English Phonology*. Lublin: Wydawnictwo Uniwersytetu Marii Curie-Skłodowskiej.
- Wierzchowska, B. (1971). *Wymowa polska*. Warszawa: Państwowe Zakłady Wydawnictw Szkolnych.

- Wierzchowska, B. (1980). *Fonetyka i fonologia języka polskiego*. Wrocław, Warszawa: Zakład Narodowy im. Ossolińskich: Wydawnictwo Polskiej Akademii Nauk.
- Zygis, M. (2003a). The Role of Perception in Slavic Sibilant Systems. In: Kosta, P., Blaszcak, J., Frasek, J., Geist, L. & M. Zygis (eds.). *Investigations into Formal Slavic Linguistics*. Berlin: Peter Lang Verlag. 137-154.
- Zygis, M. (2003b). Phonetic and phonological aspects of Slavic sibilant fricatives. *ZAS Papers in Linguistics* 32: 175-212.
- Zygis, M. (2005). Representation of Slavic sibilants in terms of distinctive features. Chapter 5. *Habilitationsschrift*. Ms.
- Zygis, M. (in progress). Contrast Optimization in Slavic Sibilant Systems. Ms.

# Phonetics or Phonology: Asymmetries in Loanword Adaptations - French and German Mid Front Rounded Vowels in Japanese

**Katrin Dohlus**

*Kobe University, Japan / Humboldt-Universität zu Berlin, Germany*

---

It is one of the most highly debated issues in loanword phonology whether loanword adaptations are phonologically or phonetically driven. This paper addresses this issue and aims at demonstrating that only the acceptance of both a phonological as well as a phonetic approximation stance can adequately account for the data found in Japanese. This point will be exemplified with the adaptation of German and French mid front rounded vowels in Japanese. It will be argued that the adaptation of German /œ/ and /ø/ as Japanese /e/ is phonologically grounded, whereas the adaptation of French /œ/ and /ø/ as Japanese /u/ is phonetically grounded. This asymmetry in the adaptation process of German and French mid front rounded vowels and further examples of loans in Japanese lead to the only conclusion that both strategies of loanword adaptation occur in languages. It will be shown that not only perception, but also the influence of orthography, of conventions and the knowledge of the source language play a role in the adaptation process.

---

## 1 Introduction

Japanese with its five-vowel system of short and long /i, u, e, o, a/ does not have a contrast between front unrounded and front rounded vowels – it only allows the unmarked front unrounded vowels. For that reason I investigated how Japanese adapts front rounded vowels in loanwords from French and German.

Whereas the *high* front rounded vowels are adapted as /ju/ in loanwords from German as well as from French (Dohlus 2004, 2005), the *mid* front rounded vowels reveal an interesting asymmetry in Japanese: German /œ/ and /ø/ are adapted as Japanese /e/, but French /œ/ and /ø/ are adapted as Japanese /u/.

The first thing that comes to mind is that the different adaptation forms in Japanese are caused by differences between the German and French mid front rounded vowels. However, a comparison between both source vowels refutes this assumption.

Phonologically, German and French /œ/ and /ø/ are identical. In both languages these sounds carry the combination of the phonological features [-high], [coronal], [labial], and [lax] or [tense], respectively<sup>1</sup>. A comparison of acoustic features also shows a high similarity between German and French /œ/ as well as /ø/. Delattre (1965) for instance gives identical F1 and F2 values for the tense vowels (F1: 375 Hz, F2: 1600 Hz) and similar values for the lax vowels (German [œ] F1: 500 Hz, F2: 1550 Hz, French [œ] F1: 550 Hz, F2: 1400 Hz). There are of course differences between German and French in a broader context, for instance in the vowel inventory and in terms of stress or rhythm. However, we will see in the paper that German and French mid front rounded vowels are similarly perceived as /u/ by speakers of Japanese. Instead of being caused by differences between the German and French source vowels, I will argue that the asymmetry in the adaptation forms is grounded on the application of different adaptation strategies: The adaptation of German mid front rounded vowels as /e/ in Japanese is an example of phonological approximation, whereas the adaptation of French mid front rounded vowels as /u/ is an example of phonetic approximation.

## 2 Approaches to Loanword Phonology

Before discussing the problem of the asymmetry in the adaptation patterns of German and French mid front rounded vowels in Japanese, I want to give a short overview of different approaches to loanword adaptation. The current literature distinguishes two main positions – the phonological and the phonetic approximation stance.

### 2.1 Phonological Approximation Stance

LaCharité & Paradis (2005: 223) argue “that loanword adaptation is overwhelmingly phonological” (see also Paradis & LaCharité 1997, Danesi 1985, Lovins 1975). Their major claim is that loanword adaptation is based on the identification of *phoneme categories* of the *source* language and that phonetic approximation plays only a minor role. This presupposes that

---

<sup>1</sup> For a detailed definition of the features [coronal] (“involving a constriction formed by the front of the tongue”) and [labial] (“involving a constriction formed by the lower lip”) see Clements and Hume (1995).

borrowers are bilingual and have an extended knowledge of the source language (LaCharité & Paradis 2005). Borrowers correctly perceive the phonological categories of the source language, they “accurately identify L2 [source language] sound categories; that is, they operate on the mental representation of an L2 sound, not directly on its surface phonetic form” (LaCharité & Paradis 2005: 223, see also Jacobs & Gussenhoven 2000). Hence, according to the phonological approximation stance, perception of foreign sounds is faithful and sounds are only altered in production, where the phonological contrasts of the source language are preserved to the greatest extent possible.

The following examples, taken from LaCharité & Paradis (2005) support the phonological approximation stance.

(1) Examples of phonological approximation<sup>2</sup>

(1a) English voiced stops in Spanish:

A comparison of VOT indicates that Spanish voiceless sounds overlap with English voiced sounds (both have a VOT of 0-30 msecs). This is underlined by the misperception of English voiced sounds as voiceless by Spanish learners of English. However, English loanwords with voiced stops are not adapted as voiceless in Spanish, but as the phonologically identical category [voiced] (LaCharité & Paradis 2005). Similarly, English /b/ is adapted as /b/ in French despite being acoustically closer to French /p/ (LaCharité & Paradis 2005).

(1b) English high lax vowels in Spanish:

English [ɪ] and [ʊ] are phonetically closest to the Spanish phonemes /e/ and /o/. Despite this phonetic closeness, English [ɪ] and [ʊ] are adapted as /i/ and /u/ in Spanish, because they are phonologically identical to the phoneme category of the English source vowels (LaCharité & Paradis 2005).

(1c) English [θ] in Italian Calabrese:

English [θ] is perceptually closest to Calabrese Italian /f/, but it is adapted as /f/ in only a minority of adaptations (2/64 words). In the majority of cases, /t/ is chosen for the representation of English [θ] in Calabrese Italian (62/64 words) (LaCharité & Paradis 2005).

---

<sup>2</sup> The adaptation forms in the following examples appear to be caused by the influence of orthography. However, LaCharité & Paradis (2005: 237) state that orthography plays only a limited role in their database: “Despite what is often believed, the clear influence of orthography is generally weak”. I will discuss the influence of orthography in section 4.4.

These examples show that despite the existence of a phonetically identical or closer sound, the phonologically identical sound of the borrowing language is chosen. They thus indicate that the adaptations are phonologically driven and that “phonetic approximation cannot be held responsible for the adaptation” (LaCharité & Paradis 2005: 235). Foreign sounds are adapted as native sounds that preserve the phonological contrasts of the source language to the greatest extent possible.

## 2.2 *Phonetic Approximation Stance*

Peperkamp & Dupoux (2003), Vendelin & Peperkamp (2004), Kenstowicz (2005) and others hold the opposite standpoint, namely that adaptation is solely determined by acoustic and perceptual factors. Peperkamp & Dupoux (2003: 368) propose “that indeed all adaptations apply in perception and that they are always phonetic in nature”. Adaptation is driven purely by auditory perception, and “a given input sound will be mapped onto the closest available phonetic category” (Peperkamp & Dupoux 2003: 368). ‘Available phonetic category’ hereby indicates that the perception and categorisation of foreign sounds is language-specific: “With respect to nonnative sounds, this mapping is of course massively unfaithful, since the phonetic categories to which these sounds are mapped in the foreign language can simply be absent from the native one” (Peperkamp & Dupoux 2003: 368). Phonological features of a sound in the source language do not play any role and may not even be known to the borrower.

That adaptation is driven phonetically is shown by examples in which 1) a foreign sound is adapted as the phonetically closest native sound irrelevant of the existence of a sound that is identical or closer to the source phoneme in phonological terms, and 2) phonologically identical sounds are adapted differently in a given language because of minimal phonetic differences. The following examples are taken from Vendelin & Peperkamp (2004).

### (2) Examples of phonetic approximation

#### (2a) English /v/ in Cantonese:

Cantonese does not have the voiced fricative /v/, only its voiceless counterpart /f/. However, English /v/ is not adapted as the phonologically closest phoneme /f/, but as the acoustically most similar Cantonese /w/.

#### (2b) Adaptation of English and French /n/ into Japanese

English and French word-final /n/ are adapted differently in Japanese, English /n/ as the Japanese moraic nasal /n/, French /n/ as a nasal geminate

followed by an epenthetic vowel, *-nnu*. This, Vendelin & Peperkamp (2004) argue, is due to the phonetic differences between English /n/ (no release) and French /n/ (release and longer duration) which are perceived by Japanese listeners.

These examples demonstrate “that loanword adaptations are not due to the phonological grammar, but rather to perceptual processes involved in the decoding of nonnative sounds” (Peperkamp & Dupoux 2003: 367).

This section showed that there are two approaches to loanword adaptation. In the following section I will analyse the data from Japanese in phonological as well as phonetic terms and argue that German /œ/ and /ø/ are adapted on phonological grounds, but French /œ/ and /ø/ on phonetic grounds.

### 3 The Asymmetry in the Adaptation of German and French /œ/ and /ø/ in Japanese

#### 3.1 German /œ/ and /ø/ in Japanese

As the examples in (3) illustrate, German /œ/ and /ø/ are adapted as /e/ in Japanese (for sources see Appendix *Sources of loanword data*).

(3) Adaptation of German /œ/ and /ø/ → Japanese /e/ (41/41 words) <sup>3</sup>				
<u>Ö</u> kumene	[œku'me:nə]	→	<u>e</u> kumêne	‘area of settlement’
R <u>ö</u> ntgen	[ʰrœntgən]	→	r <u>e</u> ntogen	‘X-ray’
G <u>ö</u> ethe	[ʰgø:tə]	→	g <u>e</u> te	Goethe (personal name)
Schr <u>ö</u> der	[ʰʃrø:də]	→	shur <u>e</u> dâ	Schröder (personal name)

<sup>3</sup> Transcription of the Japanese data follows the Hepburn system. Vowels are pronounced as in Italian or German. Consonants are pronounced as in English (<g> is always pronounced as [g]). Please note particularly the following conventions:

- macrons mark long vowels
- <y> is pronounced as the front glide [j]
- double consonants are geminates, e.g. <kk> is pronounced as [kː].

As the examples in (3) show, vowel epenthesis is very common in loanwords in Japanese. With the exception of the moraic nasal and geminated consonants, consonant clusters and final consonants are disallowed in Japanese. In order to avoid consonant clusters and final consonants a vowel is inserted in Japanese. This epenthetic vowel is usually the default-vowel /u/ (e.g. Schröder → shurêdâ). If /t/ or /d/ precedes /u/, the quality of the consonant is altered (/t/ + /u/ → [tsu], /d/ + /u/ → [(d)zu]), therefore, /o/ is inserted after /t/ and /d/, as in the example Röntgen → rentogen (for more details on vowel epenthesis in loanwords in Japanese see Lovins 1973).

### 3.1.1 Phonological Analysis of German /œ/ and /ø/ → Japanese /e/

As can be seen in (3), German /œ/ and /ø/ are delabialised and adapted as the front *unrounded* vowel /e/ in Japanese. The adapted vowel maintains the features [-high] and [coronal], but loses the feature [labial] of the source vowel. The loss of [labial] is less crucial, because lip rounding and labiality play only a minor role in Japanese (Dohlus 2004) and are redundant in the description of the Japanese 5-vowel system. A comparison of the phonological features of input and output is presented in (4).

#### (4) Comparison of input and output features

Input German /œ/ and /ø/ →		Output Japanese /e/
[-high]	✓	[-high]
[coronal]	✓	[coronal]
[labial]	☹	

Here we see clearly that the distinctive features for vowel height and frontness are preserved, thus the phonological features of the source language are preserved to the greatest extent possible. The adaptation of German /œ/ and /ø/ as /e/ in Japanese is therefore a phonologically grounded adaptation.<sup>4</sup>

### 3.1.2 Phonetic Analysis of German /œ/ and /ø/ → Japanese /e/

In order to see whether the adaptation of German /œ/ and /ø/ as /u/ in Japanese is based on perception, an experiment was performed with the aim to find out as which Japanese vowel Japanese listeners perceive German mid front rounded vowels. /CVn/ syllables with varying onsets were used as stimuli, and the perception of lax [œ] as well as tense [ø:] were tested in two conditions, in citation form and in sentence condition. I asked my subjects (24 students from Kansai area) to write down in the Japanese syllabary what they heard (for details on the experiment see Dohlus 2005). The results, summarized in the following table, show that German /œ/ and /ø/ are overwhelmingly perceived as /u/ by Japanese speakers.

---

<sup>4</sup> LaCharité & Paradis (2005: 226) measure phonological closeness by the number of steps which have to be taken in order to make the sound permissible in the borrowing language. In the framework of Optimality Theory, phonological closeness is determined by the ranking of faithfulness constraints (see Dohlus 2004 for a phonological analysis of the adaptation of front rounded vowels into Japanese).



**Table 1:** Perception of German [œ] and [ø:] by Japanese speakers

German [œ] and [ø:]		/e/	/u/	/a/	Others	Total
Citation Form Condition	[œ]	2.0%	<b>74.8%</b>	20.2%	3.0%	100%
	[ø:]	-	<b>98.6%</b>	1.0%	0.4%	100%
Sentence Condition	[œ]	2.0%	<b>74.8%</b>	20.2%	3.0%	100%
	[ø:]	-	<b>93.4%</b>	-	6.6%	100%

These results are not consistent with the adaptation pattern of German mid front rounded vowels as Japanese /e/, as we find it in established loans. Thus, perception cannot account for the adaptation form of German /œ/ and /ø/ in Japanese. In order to make the results easier to understand I am going to describe the characteristics of Japanese /u/ briefly.

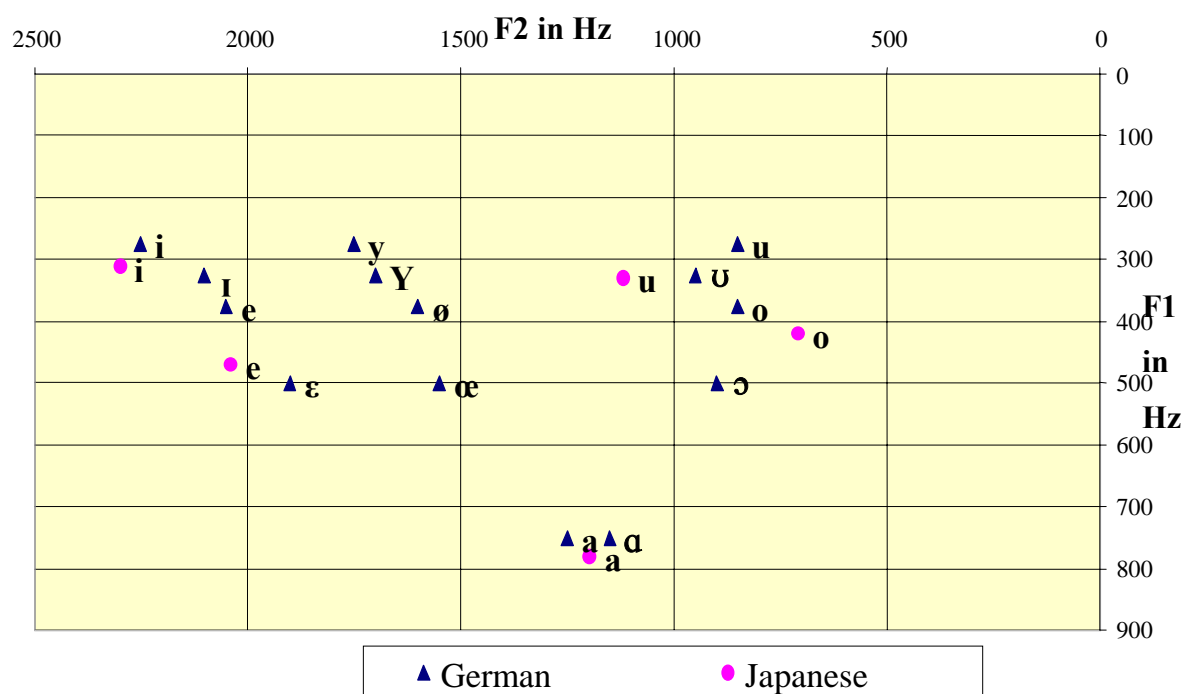
### 3.2. *Characteristics of Japanese /u/ and the Issue of Perceptual Similarity*

Phonologically, Japanese /u/ is a back rounded vowel. Several phenomena in Japanese show that Japanese /u/ behaves as a back vowel: First, /u/ can follow the palatal glide, which only precedes back vowels (Kubozono 2002: 81), second, /u/ patterns as a back vowel in vowel coalescence (Kubozono 1999: 102), and third, /u/ takes the velar glide as the homorganic glide to break hiatus (Kubozono 2002: 84). As Japanese does not contrast back rounded and unrounded vowels, we can assume that Japanese /u/ is the universally unmarked *rounded* vowel (Calabrese 1995: 383, Kubozono 1999: 21ff.).

However, phonetically Japanese /u/ is fairly fronted and therefore a rather centralised vowel (Honma 1985: 103, Kubozono 1999: 36f.). This fronting of Japanese /u/ further results in a weakening of its lip rounding (Kubozono 1999: 37). The following chart, comparing German and Japanese vowels in terms of their first and second formants, illustrates the fronting of Japanese /u/.

Figure 1 shows that the phonetically centralised Japanese /u/ is fairly close to German front rounded vowels in terms of F2. However, perceptual similarity cannot be measured reliably by acoustic features alone. This can be seen in Figure 1 above, where acoustic similarity to German mid front rounded vowels can be stated for Japanese /u/ as well as for Japanese /e/. It cannot explain why Japanese mainly perceive /u/, but hardly ever /e/ in the case of German mid front rounded vowels. This shows that it is not acoustic features alone, but the weighting of several acoustic cues that determines perception. Studies have shown (e.g. Rochet 1995, Escudero & Boersma 2004) that speakers of different languages or dialects identify vowels with identical formant frequency values differently. This demonstrates that perception is highly language-dependent:

“Crosslinguistically, the attention paid to the cues that signal a contrast varies between adult speakers of different languages” (Escudero & Boersma 2004: 552).



**Figure 1:** Formant Frequencies of German and Japanese vowels (utterances of male speakers), from: Delattre 1965 (German) and Imaishi 1997 (Japanese)

The chart above only reflects acoustic features, but not their cue-weighting in Japanese. Thus, it is a task for further research to explore experimentally which cues Japanese speakers use to which extent, and how, based on this cue-weighting, they divide their perceptual vowel space. It can be expected that Japanese /u/ perceptually overlaps with the mid (and high) front rounded vowels. A good example of such a study is Rochet (1995), who investigated the asymmetry in the perception and adaptation of French /y/ in Portuguese and American English. Portuguese speakers replace /y/ by /i/, whereas American English speakers replace it by /u/. Rochet’s experiment shows that the difference in the perception of French /y/ by speakers of Portuguese and of American English is based on “how these subjects *perceive* and categorize the high vowel continuum in their respective languages” (Rochet 1995: 385). In the case of Portuguese, French /y/ falls into the perceptual space of the /i/ category, whereas it falls into the /u/ category for American English speakers.

To conclude, the results of the perceptual experiment have shown that the adaptation of German /œ/ and /ø/ as /e/ in Japanese is not based on perception and thus not phonetically, but phonologically grounded.

### 3.3 French /œ/ and /ø/ in Japanese

As the examples in (5) demonstrate, French /œ/ and /ø/ are adapted as /u/ in Japanese.

- (5) French /œ/ and /ø/ → /u/ (13/17 words)<sup>5</sup>
- |                       |                |   |                         |                     |
|-----------------------|----------------|---|-------------------------|---------------------|
| fle <u>u</u> ret      | [flœ:'rɛ]      | → | fur <u>u</u> re         | 'foil (fencing)'    |
| entreprene <u>u</u> r | [ãtrəprə'nœ:r] | → | antorupurun <u>u</u> ru | 'enterpriser'       |
| pot-au-f <u>eu</u>    | [pɔto'fø]      | → | po to f <u>u</u>        | 'Pot-au-feu (dish)' |
| charme <u>u</u> se    | [ʃar'mø:z]     | → | sharum <u>u</u> zu      | 'fashionable cloth' |

#### 3.3.1 Phonological Analysis of French /œ/ and /ø/ → Japanese /e/

The French mid front rounded vowels are assimilated as high back vowels in Japanese. In (6) we see that only the feature [labial] – redundant in Japanese – is preserved, but the features [-high] and [coronal] of the French input are lost. Therefore, the adaptation of French /œ/ and /ø/ as /u/ in Japanese is not a phonologically grounded adaptation.

#### (6) Comparison of input and output features

Input French /œ/ and /ø/	→	Output Japanese /u/
[-high]	☹	[+high]
[coronal]	☹	[dorsal]
[labial]	✓	[labial]

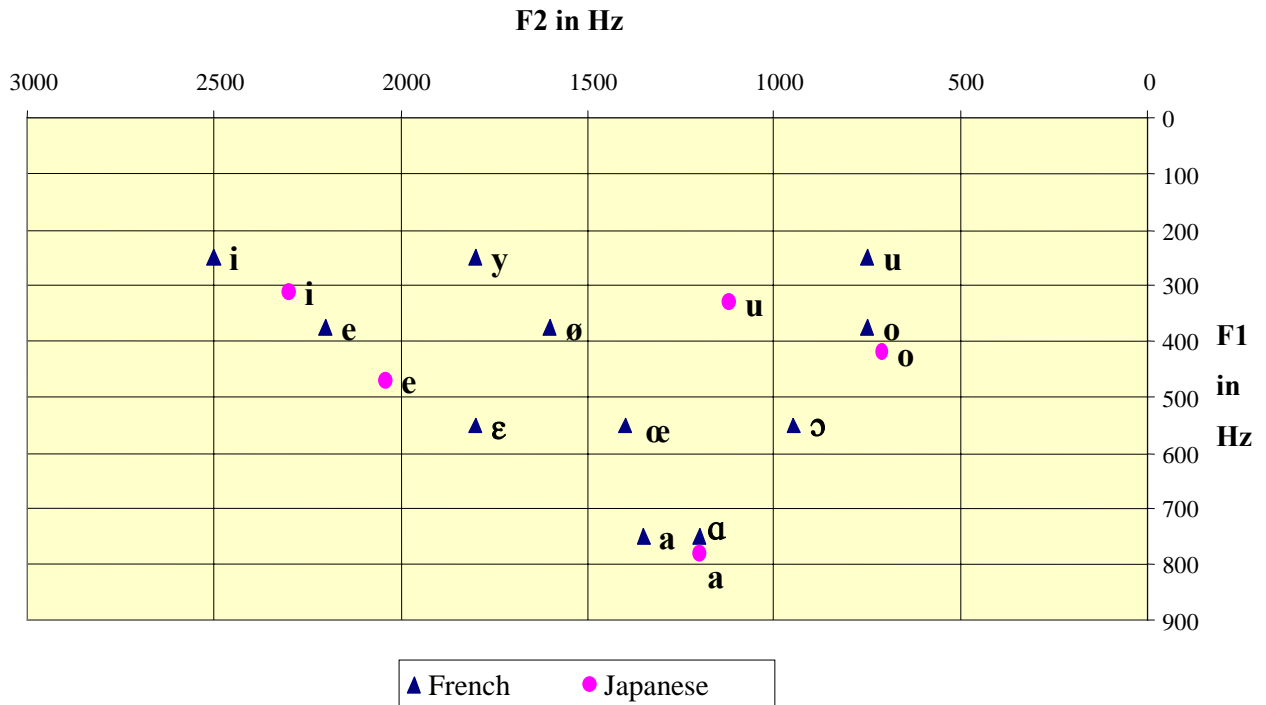
#### 3.3.2 Phonetic Analysis of French /œ/ and /ø/ → Japanese /e/

As the perceptual experiment described above has shown, Japanese /u/ appears to be the sound that is perceptually closest to German /œ/ and /ø/. The same results could be expected for French /œ/ and /ø/. Indeed, Shinohara (1997), who asked her Japanese subjects to convert French words into Japanese, showed that French /œ/ and /ø/ are perceptually closest to Japanese /u/. All of her three

<sup>5</sup> The other four words containing /œ/ or /ø/ in French are adapted as follows: 1) /œ/ → /o/ (*hors-d'œuvre* [ɔr'dœ:vʁə] → *ôdoburu* 'side-dish', 2) /œj/ → /iju/ *cerfeuil* [sɛr'fœj] → *serufiyyu* 'chervil', *millefeuille* [mil'fœj] → *mirufiyyu* 'pie-like cake', and 3) /jø/ → /iju/ *faux camaïeu* [fokama'jø] → *fô kamaiyu* 'colourless'.

Note that nasalised vowels are adapted as a sequence of oral vowel and nasal consonant (/Vn/). On epenthetic vowels see footnote 3 in section 3.1.

subjects represented the French vowel as /u/ (16/16 words, in a few cases /o/ or /ju/ were given as further responses), e.g. *seul* → *suru*, *neutre* → *nu(u)toru*.



**Figure 2:** Formant Frequencies of French and Japanese vowels (utterances of male speakers), from: Delattre 1965 (French) and Imaishi 1997 (Japanese)

In figure 2, which compares French and Japanese vowels in terms of F1 and F2, we find a plausible explanation for the perception of French mid front rounded vowels as /u/ in Japanese, namely the high F2 values of Japanese /u/. However, as said before (see section 3.2), formant frequency values alone cannot account for perceptual similarity, because language-specific weighting of several acoustic cues determines perception.

#### 4 Phonetically or Phonologically Grounded Adaptation?

The asymmetry in the adaptation patterns of German and French mid front rounded vowels in Japanese shows that we find different adaptation strategies in Japanese.

German /œ/ and /ø/ → Japanese /e/ is a phonologically driven adaptation. The phoneme categories of the sound in the source language are - irrelevant of phonetic characteristics - maintained to the greatest extent possible in Japanese.

In contrast, the pattern of French /œ/ and /ø/ → Japanese /u/ constitutes an example of phonetic approximation. The phonological features that the sound carries in the source language are irrelevant, the features [-high] and [coronal] of

the input vowels are lost. Rather, acoustic features and perceptual similarity determine the output form. As seen above, Japanese /u/ appears to be the sound that Japanese listeners perceive when hearing mid front rounded vowels.

In order to understand the asymmetry better, I investigated further examples of phonetically or phonologically driven adaptations. As this section will show, data from Japanese offer further examples for both adaptation strategies.

#### **4.1 Further Examples of Phonetic Approximation**

In Japanese, we find a number of further examples of phonetic approximation, see for instance (7) and (8).

##### **(7) Word-final /n/ in Japanese**

###### **(7a) English /n/ (no release) → Japanese /n/**

cotton	[ <sup>l</sup> kɒtn]	→	<i>kotton</i> <u>n</u>	‘cotton’
line	[lain]	→	<i>rain</i> <u>n</u>	‘line’

###### **(7b) French /n/ (release and longer duration) → Japanese *nnu* (geminated nasal plus epenthetic vowel)**

parisienne	[pari <sup>l</sup> zjɛn]	→	<i>parijennu</i>	‘woman from Paris’
Cannes	[kan]	→	<i>kannu</i>	Cannes (place name)

The minimal differences in the source languages, namely that French word-final /n/ is, in contrast to English word-final /n/, characterised by a release and longer duration, are perceived by speakers of Japanese and reflected in the adaptation forms. These phonetic details are perceived, because Japanese differentiates the single nasal /n/ and the geminated nasal followed by an epenthetic vowel in its vocabulary.

A second example is the adaptation of the English low front vowel /æ/. Whereas English /æ/ is usually adapted as /a/ in Japanese (see (8a)), sequences of a velar consonant and /æ/ are adapted as a sequence of velar consonant, front glide and /a/, namely *kya* ([kja]) and *gya* [gja] in Japanese (see (8b)). Whether the English vowel /æ/ is adapted as a single vowel or as a sequence of front glide and vowel in Japanese appears to depend on the absence or presence of a preceding velar consonant.

(8) Adaptation of English /æ/ in Japanese

(8a) English /æ/ → Japanese /a/

<u>p</u> an	[pæn]	→	<u>p</u> an	‘pan’
<u>b</u> ag	[bæg]	→	<u>b</u> aggu	‘bag’
<u>t</u> actics	[ˈtæktɪks]	→	<u>t</u> akutikkusu	‘tactics’
<u>n</u> apkin	[ˈnæpkin]	→	<u>n</u> apukin	‘napkin’
<u>h</u> at	[hæt]	→	<u>h</u> atto	‘hat’

(8b) Velars preceding /æ/: English /kæ/ and /gæ/ → Japanese /kja/ and /gja/

<u>c</u> at	[kæt]	→	<u>k</u> yatto	‘cat’
<u>c</u> amp	[kæmp]	→	<u>k</u> yam <u>p</u> u	‘camp, camping’
<u>g</u> ang	[gæŋ]	→	<u>g</u> yan <u>g</u> u	‘gang’
<u>g</u> allery	[ˈgæləri]	→	<u>g</u> yararî	‘gallery’

The asymmetry in the adaptation forms of (8a) and (8b) is probably caused by the phonetic differences in the pronunciation of particularly American English, where the “velar stop contact is particularly sensitive to the nature of an adjacent vowel (especially a following vowel). Thus, when a front vowel follows, e.g. /i:/ in *key*, *geese*, the contact will be made on the most forward part of the soft palate and may even overlap onto the hard palate” (Gimson & Cruttenden 1994: 153). I assume that the fronting of the velars is perceived by speakers of Japanese and reflected in the adapted forms by the insertion of a front glide. This again is plausible because Japanese phonemically differs between syllables like /ka/ or /ga/ on the one hand and /kja/ or /gja/ on the other hand.

## 4.2 Adaptation is Phonetically Driven

The examples above support the assumption that loanword adaptation is phonetically driven and based on perception. Indeed, this is most reasonable to me. First, the opposite standpoint, namely the phonological approximation stance, assumes that the phoneme categories of the source language are correctly identified. However, a large number of studies have shown that perception of foreign sounds is not faithful, but heavily influenced by one’s native language (e.g. Best & Strange 1992, Dupoux *et al.* 1999, Rochet 1995). Borrowers are confronted with sounds of a foreign language, and thus their perception, I

assume, is not faithful<sup>6</sup>. Rather, they map these sounds onto their native phonetic categories – as argued by the phonetic approximation stance.

To conclude, I argue that loanword adaptation is phonetic in nature and based on the (often unfaithful) perception of foreign sounds. However, as the adaptation of German /œ/ and /ø/ as Japanese /e/ has shown, there are examples of phonological approximation. A few more examples are presented in the next subsection.

### **4.3 Further Examples of Phonological Approximation**

Further examples of phonological approximation are for instance the adaptation of syllabic /r/ (see (9)) or schwa (see (10)) in Japanese. In the case of schwa for instance, Japanese /u/ is the acoustically and perceptually closest sound, but German schwa is adapted as the phonologically identical /e/ in the majority of cases.

(9) German Syllabic –(e)r [ɐ] → Japanese /eru/:

Kaiser	[ <sup>l</sup> kaizɐ]	→	<i>kaizeru</i>	‘title of German emperor’
Kocher	[ <sup>l</sup> kɔxɐ]	→	<i>kohheru</i>	‘portable cooker’

(10) German schwa [ə] → German /e/ (not /u/):

Abend	[ <sup>l</sup> a:bənt]	→	<i>âbento</i>	‘evening concert/movie’
Eishaken	[ <sup>l</sup> aɪsha:kən]	→	<i>aisuhaken</i>	‘piton’

Examples presented in (9) and (10) immediately raise the question of whether these adaptation forms are not simply orthographically-based adaptations.

I do not think that they constitute purely orthographically based adaptations, first, for the reason that the German or French writing system differs from the Japanese one, and second, for the reason that phonological similarity between input and output can be observed. However, the great influence that orthography has on adaptation cannot be denied.

---

<sup>6</sup> Even if one assumes the ideal case that the borrower has close-to-native-proficiency in the source language, it does not change the situation significantly, because the borrowed form will sooner or later hit a bilingual with less proficiency or a monolingual, as also Peperkamp & Dupoux (2003: 369, footnote 2) point out: “it might very well be the case that the bilinguals who introduce these loanwords pronounce them as in the source language and that the adaptations are subsequently done by the monolingual population”.

#### 4.4 *The Role of Orthography*

Vendelin & Peperkamp (2005) demonstrate how orthography influences adaptation. In an experiment they tested the adaptation of English words by French speakers in an oral-only and an oral-written condition. The results show that subjects relied purely on perception in the oral-only condition. In the oral-written condition though, subjects applied a grapheme-to-phoneme correspondence that they had acquired in foreign-language classes (Vendelin & Peperkamp 2005).

This experiment points out the problem: written forms give information or hints on the phoneme in the source language. Irrespective of unfaithful perception, written forms offer the possibility to correctly identify the phoneme categories of the source language. To give a simple example: a Japanese speaker who cannot perceive the difference between English /r/ and /l/ knows with some minimal knowledge of English spelling very well whether he is confronted with /r/ or /l/.

Thus, orthography enables faithful perception due to hinting to the source phoneme and as a consequence triggers phonological approximation. Perception becomes secondary if one can reliably identify the source sounds due to the presence (or knowledge) of the source's written form.

As written forms always played a major role in Japanese – the focus of foreign-language learning has been on translations, there is little contact with native speakers in Japan, and foreign-language classes are still mainly grammar/translation oriented – it is not surprising to find a large number of examples of phonological approximation in Japanese.

#### 4.5 *Problem of Conventions*

A second major issue in dealing with loanwords are conventions. In the case of Japanese, we still find a variety of adaptation forms for German mid front rounded vowels in the 19<sup>th</sup> century. Yazaki (1964: 170) for instance lists 29 different adaptation forms for the name of the German author Goethe ([<sup>l</sup>gø:te]). However, with the 20<sup>th</sup> century, adaptation forms became standardised by the publishing of loanword dictionaries, by conventions of the *Kokugo Shingikai* (National Language Inquiry Commission) and also by foreign-language classes.

Conventions are grapheme-to-phoneme-correspondences based on written forms and thus trigger phonological approximation.



#### **4.6 Knowledge of the Source Language**

Knowledge of the source language<sup>7</sup> provides information on the phonological contrasts of the source language. It may trigger the application of an already established adaptation pattern, as acquired in foreign-language classes. The claim that knowledge of the source language influences perception, and thus also loanword adaptation, is supported by a number of studies (e.g. LaCharité & Paradis 2005, Silverman 1992). For instance, the perception of English rhotics by Japanese listeners differs according to the English proficiency of the Japanese: “those Japanese speakers with little or no exposure to spoken English classified the English onset rhotic on phonetic grounds, while those with more experience classified it on phonological grounds” (LaCharité & Paradis 2005: 245). Similarly, the perception of English voiced stops in Spanish varies in dependence on their knowledge of English: “monolingual Spanish speakers classify English stops on phonetic grounds, leading to their (mis)identification as voiceless, in accordance with the VOT values of Spanish. However, as English proficiency improves, the VOT value boundary approaches that of English monolinguals, with the classification performance of Spanish-English bilinguals being comparable to that of English monolinguals” (LaCharité & Paradis 2005: 247). Studies like these constitute clear evidence that perception and adaptation are strongly influenced by a listener’s knowledge of the source language.

#### **4.7 Conclusion: Phonetics or Phonology**

As argued before, I assume that loanword adaptation is basically phonetically grounded: foreign sounds are mapped onto the closest phonetic categories of the borrowing language. However, for a phonetically grounded adaptation, we need sufficient oral input. If there is a lack of oral input or the possibility of ‘faithful perception’ due to the presence of written forms, then this triggers phonological approximation. Thus, we do find phonological approximation if there is a lack of direct contact with native speakers, a major influence of written media and conventions, and knowledge of the foreign language that is based on grammar/translation-oriented foreign-language classes. These points quite well reflect the situation we find in Japan.

---

<sup>7</sup> By knowledge of the source language, I mainly mean abstract ‘classroom’ knowledge on the source language’s phonology that is supported by knowledge on the written forms, and not proficiency to the extent that the speaker has acquired the phonetic categories of the source language.

## **5 Why Different Adaptation Strategies for German and French Mid Front Rounded Vowels?**

The section above explained why both phonetic as well as phonological approximation may occur in the borrowing process. However, it still leaves the question of why the adaptation of German and French mid front rounded vowels into Japanese differs. There are two possible explanation for this asymmetry.

### ***5.1 Roles of German and French Loans in Japanese***

One possible explanation for this divergence in the adaptation process is grounded on different roles that German and French loans played in Japanese.

German loans, which entered Japanese mainly from the end of the 19<sup>th</sup> century on, are words from the fields of medicine, philosophy, chemistry, and outdoor and ski sports. Most of the German loans in Japanese are academic terms used in sciences only. This indicates that German words almost certainly reached Japan in the context of the studies of sciences, and thus mainly via written media. German was also extensively used in higher education (Loveday 1996, Kuze 1976), which was based on written materials and might have fastened the process of standardisation of the adapted forms.

In contrast, French loans, entering Japanese mainly from mid 19<sup>th</sup> century, are words from the fields of fashion, French cuisine, arts, dancing and military (Yazaki 1964). French loans appeared to have played a great role in everyday-communication (Steinberg 1996) and thus French most likely provided more oral input than German.

### ***5.2 Differences in the Spelling of German and French***

A second explanation for the asymmetry in the adaptation of German and French mid front rounded vowels in Japanese is related to the spelling systems of both source languages. German spelling quite faithfully reflects the pronunciation, from a written form the pronunciation of a word (and with it the phoneme categories of a sound) are easy to recognize. For instance, seeing the grapheme <ö> in a written form tells one even with low knowledge of German that this is the mid front rounded vowel. This might explain the high number of phonologically driven adaptation forms in the case of German.

In contrast, French spelling is rather difficult, not faithfully reflecting the pronunciation. For instance, the phonemes /œ/ and /ø/ are transcribed by the diagraphs <eu> or <œu>. One needs a higher knowledge of French than is needed in the case of German in order to identify phonemes correctly from written forms in French.

## 6 Conclusion

This paper investigated the asymmetry in the adaptation patterns of German and French mid front rounded vowels into Japanese. It was shown that the adaptation of German /œ/ and /ø/ as Japanese /e/ is a phonological approximation, but the adaptation of French /œ/ and /ø/ as Japanese /u/ a phonetic approximation. In this context I argued that loanword adaptations are basically phonetically grounded, but that a lack of oral input and a large influence of written media trigger phonological approximation.

To conclude, I think that the process of borrowing is far too complex to be accounted for by phonetic approximation only. Although I assume adaptations to be fundamentally based on language-specific perception and thus to be phonetic in nature, secondary factors such as a borrower's knowledge of the source language, orthography and conventions trigger phonological approximation.

## References

- Best, C.T. & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics* 20: 305-330.
- Calabrese, A. (1995). A constraint-based theory of phonological markedness and simplification procedures. *Linguistic Inquiry* 26-3: 373-463.
- Clements, G.N. & Hume, E.V. (1995). The internal organization of speech sounds. In: Goldsmith, J.A. (ed.) *The Handbook of Phonological Theory*. Cambridge, Mass. and Oxford, UK: Blackwell, 245-306.
- Danesi, M. (1985). *Loanwords and phonological methodology*. Toronto: Didier.
- Delattre, P. (1965). *Comparing the Phonetic Features of English, French, German and Spanish: An Interim Report*. Heidelberg: Julius Groos.
- Dohlus, K. (2004). The adaptation of German front rounded vowels into Japanese. *Kobe Papers in Linguistics* 2004-4, 1-19.
- Dohlus, K. (2005). On the asymmetry in the adaptation pattern of German umlaut in Japanese. *Journal of the Phonetic Society of Japan* 2005-1: 39-49.
- Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C. & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 25: 1568-1578.
- Escudero, P. & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Education* 26: 551-585.

- Gimson, A.C. & Cruttenden, A. (1994). *Gimson's Pronunciation of English*. London et al.: Edward Arnold.
- Honma, Y. (1985). *Acoustic Phonetics in English and Japanese*. Tokyo: Yamaguchi.
- Imaishi, M. (1997). *Nihongo onsei no jikkenteki kenkyû* [Experimental Research on Japanese Phonetics]. Osaka: Izumi.
- Jacobs, H. & Gussenhoven, C. (2000). Loan phonology: perception, salience, the lexicon and optimality theory. In: Dekkers, J., van der Leeuw, F. & van de Weijer, J.: *Optimality Theory: Phonology, Syntax, and Acquisition*, Oxford: Oxford University Press, 193-210.
- Kenstowicz, M. (2005). The phonetics and phonology of Korean loanword adaptation. To appear in: Rhee, S-J. (ed.) *Proceedings of First European Conference on Korean Linguistics*.
- Kubozono, H. (1999). *Nihongo no Onsei* [The Sound System of Japanese]. Tokyo: Iwanami.
- Kubozono, H. (2002). Prosodic structure of loanwords in Japanese: Syllable structure, accent and morphology. *Journal of the Phonetic Society of Japan* 6: 79- 97.
- Kuze, Y. (1976). *Gairaigo Zatsugaku Hyakka* [Miscellaneous Studies on Loanwords in Japanese]. Tokyo: Shin-Jinbutsuoraisha.
- LaCharité, D. & Paradis, C. (2005). Category preservation and proximity versus phonetic approximation in loanword adaptation. *Linguistic Inquiry* 36-2: 223-258.
- Loveday, L. J. (1996). *Language Contact in Japan: A Socio-linguistic History*. Oxford: Claredon Press.
- Lovins, J.B. (1973). Loanwords and the phonological structure of Japanese. Unpublished PhD dissertation. Chicago Illinois: University of Chicago.
- Paradis, C. & LaCharité, D. (1997). Preservation and minimality in loanword adaptation. *Journal of Linguistics* 33, 379-430.
- Peperkamp, Sh. & Dupoux, E. (2003). Reinterpreting loanword adaptations : the role of perception. *Proceedings of the 15<sup>th</sup> International Congress of Phonetic Sciences*, 367-370.
- Rochet, B.L. (1995). Perception and production of second-language speech sounds by adults. In: Strange, W. (ed.) *Speech Perception and Linguistic Experience*. Timonium, MD: York Press, 379-410.
- Shinohara, Sh. (1997). Analyse phonologique de l'adaption japonaise des mots étranger. PhD dissertation. Paris 3.
- Silverman, D. (1992). Multiple scansions in loanword phonology: evidence from Cantonese. *Phonology* 9: 289-328.
- Steinberg, J. D. (1996). Lexical Borrowing and Modernization in China and Japan. PhD dissertation. University of California, Los Angeles.
- Vendelin, I. & Peperkamp, Sh. (2004). Evidence for phonetic adaptation of loanwords: an experimental study. *Actes des Journées d'Etudes Linguistique*, 129-131.

- Vendelin, I. & Peperkamp, Sh. (2005). The influence of orthography on loanword adaptations. *Lingua* (in press)
- Yazaki, G. (1964). *Nihon no Gairaigo* [Japanese loanwords]. Tokyo: Iwanami.

## **Appendix**

### **Sources of loanword data**

The following dictionaries, travel guides and place-name dictionary are my main source of German and French loanwords in Japanese. As I investigated established loans in Japanese, I mainly relied on loanword dictionaries. However, data of loanwords that contain mid front rounded vowels in the source language are rare. Therefore, I also included data from travel guides, a place-name dictionary and a few proper names as they appeared in the news (internet).

- Arakawa, S. (1941). *Gairaigojiten* [Loanword dictionary]. Tokyo: Kadokawa.
- Arakawa, S. (1967). *Gairaigojiten* [Loanword dictionary]. Tokyo: Kadokawa.
- Chikyu no arukikata (1995/96). *Doitsu* [Germany]. Tokyo: Daiyamondo.
- Chikyu no arukikata (1997/98). *Furansu* [France]. Tokyo: Daiyamondo.
- Eriagaido (1997). *Yôroppa* [Europe]. Tokyo and Osaka: Shobunsha.
- Sanseido (1923). *Nihon Gairaigojiten* [Japanese Loanword dictionary]. Toyko: Sanseido.
- Sanseido (1977). *Konsaisu Chimeijiten* [Concise Dictionary of Place Names]. Tokyo: Sanseido.
- Shogakukan (1998). *Reibun de yomu katakanago no jiten: A Dictionary of loanwords*. Tokyo: Shogakukan.
- Umegaki, M. (1978). *Gairaigojiten* [Loanword dictionary]. Tokyo: Tôkyôdô.

# On the complex nature of speech kinematics

**Susanne Fuchs**

*Zentrum für Allgemeine Sprachwissenschaft, Berlin, Germany*

*Institut de la Communication Parlée, UMR CNRS 5009, Institut National Polytechnique de Grenoble & Université Stendhal, Grenoble, France*

**Pascal Perrier**

*Institut de la Communication Parlée, UMR CNRS 5009, Institut National Polytechnique de Grenoble & Université Stendhal, Grenoble, France*

---

Studying kinematic behavior in speech production is an indispensable and fruitful methodology in order to describe for instance phonemic contrasts, allophonic variations, prosodic effects in articulatory movements. More intriguingly, it is also interpreted with respect to its underlying control mechanisms. Several interpretations have been borrowed from motor control studies of arm, eye, and limb movements. They do either explain kinematics with respect to a fine tuned control by the Central Nervous System (CNS) or they take into account a combination of influences arising from motor control strategies at the CNS level and from the complex physical properties of the peripheral speech apparatus. We assume that the latter is more realistic and ecological. The aims of this article are: first, to show, via a literature review related to the so called '1/3 power law' in human arm motor control, that this debate is of first importance in human motor control research in general. Second, to study a number of speech specific examples offering a fruitful framework to address this issue. However, it is also suggested that speech motor control differs from general motor control principles in the sense that it uses specific physical properties such as vocal tract limitations, aerodynamics and biomechanics in order to produce the relevant sounds. Third, experimental and modelling results are described supporting the idea that the three properties are crucial in shaping speech kinematics for selected speech phenomena. Hence, caution should be taken when interpreting kinematic results based on experimental data alone.

---

## **1 Studying kinematic behaviour: Evidence from experiments and models**

Based on a number of previous articles in the motor control and speech production literature, this article intends to show the complex nature of speech

kinematics and the difficulty of its interpretation in terms of speech motor control. Interpretations discussed here either explain kinematics with respect to fine tuned control by the Central Nervous System (CNS) or they take into account a combination of influences arising from motor control strategies at the CNS level and from the complex physical properties of the peripheral motor system. We hypothesize that the latter is more realistic and ecological. We also suggest that speech motor control makes use of specific physical properties of the speech production apparatus to achieve its goals. These physical properties are vocal tract limitations, aerodynamics, and biomechanics. By means of experimental and modelling results evidence will be provided that aerodynamics, vocal tract limitations, and biomechanics<sup>1</sup> play a crucial role and shape the kinematics of speech.

### ***1.1 Controversies in motor control: Examples investigating the 1/3 power law***

In this section we will mainly focus on the ‘1/3 power law’ in order to show by means of a well-known characteristic how kinematic behaviour has been interpreted in very different ways with respect to its underlying motor control mechanism.

The 1/3 power law has been extensively described by Viviani and colleagues (e.g. Viviani & Terzuolo 1982, Viviani & Schneider 1991, Viviani & Flash 1995) based on experimental data of subjects tracing or perceiving planar movements. Evidence has been given that there is a relationship between the degree of curvature and the speed of movement, with slower movements in the more curved parts and faster movements in the less curved parts of the trajectory. The power law can be described by the formula:

$$V(t) = k * R(t)^{\beta}$$

where  $V$  is the tangential velocity,  $R$  is the radius of the curvature,  $k$  is a velocity gain factor, and  $\beta$  has been estimated from experimental data as being close to 1/3. Therefore, the rule was called the 1/3 power law (1/3 corresponds to the movement in a Cartesian coordinate system – if movements are described in an angular space than it becomes the 2/3 power law).

Viviani and colleagues have not only shown that this law systematically applies for planar human arm movements, they also demonstrated that it deeply influences the perception of these movements. Indeed, if artificial movements do

---

<sup>1</sup> Biomechanics as such are not a peculiarity of speech motor control, they are also found in limb systems. However, biomechanics are included here since the tongue is a muscular hydrostat and has very specific biomechanical properties.

not respect this law, human subjects will not perceive them correctly. Since the power law can be discussed with respect to the action-perception interaction in speech (although it has not been intensively studied for speech), it is worth asking whether it is inherent to the human motor system or whether it is specifically controlled by the Central Nervous System (CNS).

Viviani and Flash (1995) defended the hypothesis that this rule is used by the CNS in order to optimise gestures (jerk minimization, see below), a strategy to limit the excess of degrees of freedom. In addition, these authors rejected the hypothesis that the  $1/3$  power law could be explained by muscle properties and/or muscle dynamics since it also holds under isometric conditions (Massey et al. 1992), where the subjects were asked to draw planar patterns by grasping a 3D handle and pushing on it without moving their arms.

However, Gribble and Ostry (1996) noted that even under isometric conditions, muscle properties can affect force development and, additionally, that under these conditions it is difficult to separate between centrally planned strategies and effects due to the periphery. They suggested that muscle properties can account for the  $1/3$  power law. Thus, using a biomechanical planar arm model they simulated elliptical tracings of arm movement. The model included the shoulder and the elbow joints and is generally controlled by means of shifting equilibrium points (Feldman 1986, Feldman et al. 1990). Simulations were run with control signals specifying an equilibrium shift at a constant tangential velocity (i.e. no relation between curvature and speed existed in the modelled control signals). They also simulated arm trajectories under an isometric condition and circular movements with control signals of constant speed and radius of curvature. In all the different conditions Gribble and Ostry found the simulated movements respected the  $1/3$  power law relationship. The authors propose that *“muscle properties and dynamics can play a significant role in the emergence of this relationship”* (p. 2859).

Lebedev, Tsui and Gelder (2001) tried to explain the  $1/3$  power law by means of the principle of least action, a principle known in theoretical physics. They report: *“From the principle of least action it follows that the CNS does not impose the power law directly, but follows the strategy of accomplishing the desired goal in a preset time with the minimum mechanical work required”* (p. 50). From their point of view the CNS follows a strategy minimizing the amount of mechanical work. A similar principle - although not related to the  $1/3$  power law - was introduced to speech by Nelson (1983). He pointed out that the kinematic movement of speech articulators would be the result of a centrally controlled optimisation process, aiming at minimising the jerk, the third derivative of displacement over time. However, he also noted that the resulting velocity profiles are similarly bell-shaped as in a simple undamped linear mass-spring system with constant stiffness.



Another explanation for the  $1/3$  power law has been given by Harris and Wolpert (1998) who suggested that it would be the result of a strategy minimizing positional variability due to signal dependent neuronal noise which is neurobiologically more plausible than a centrally planned strategy minimizing the jerk.

The power law has been found in movement production and to be a determinant in perception (for speech production first results are reported by Tasko and Westbury 2004).

To summarise, the  $1/3$  power law describes the kinematic relation between speed of movement and degree of curvature for different human motor systems and it is integrated in human perception processes. Interpretations of the law are manifold. On the one hand researchers tried to show that this law can be explained by a centrally planned mechanism, e.g. minimizing the jerk, which would involve a complex internal representation of the motor system in the CNS. On the other hand, it has been shown that the  $1/3$  power law may purely be the consequence of muscle properties and dynamics and there is no need for a complex control of the phenomenon. Similar controversies have also been observed with respect to articulatory movements in speech production, which have been described by means of central control strategies or by means of specific characteristics of the speech production apparatus.

### ***1.2 Controversies in speech motor control: Kinematic variations due to speech rate and loudness differences, and their underlying control***

One of the most cited references investigating speech rate effects on kinematics is Adams, Weismer and Kent (1993). The authors recorded 5 speakers by means of the x-ray microbeam system producing a single sentence several times at 5 different speech rates. Changes in speech rate had a different impact on movement duration for opening and closing gestures, and for lower lip and tongue movements. Additionally, the number of velocity peaks as well as parameter  $c$ , an index of the velocity profile's geometry ( $c=V_{max}/V_{mean}$ ), increased with decreasing speech rate and measures of the symmetry of the velocity profile changed across speaking rate. Adams, Weismer and Kent discussed these kinematic results with respect to the following motor control principles suggested in the literature:

- (1) Opening and closing gestures are differently controlled and reflect different muscle synergies (based on Gracco's kinematic results and discussion in 1988).
- (2) Asymmetries of the velocity profile due to changes in speech rate are a consequence of feedback mechanisms providing spatial information about the articulator (based on Bullock and Grossberg's VITE model 1988).

With decreasing speech rate the potential role of this feedback mechanism increases and would result in a more asymmetrical velocity profile. Changes in symmetry may also reflect a shift of the motor control system from an open-looped control (=without using feedback information) to a closed-loop control (=using feedback information) with slower movements.

(3) In agreement with Wiencke, Janssen and Belderboss (1987) the greater number of velocity peaks at slower rates may be a universal mechanism of speech motor control. Adams, Weismer and Kent also suggested in accordance with Milner and Ijaz' (1989) findings for hand movements that multiple peaks could originate from overlapping submovements, since it may be difficult to generate longer movements with one motor command only.

Thus, these authors clearly interpreted kinematic changes due to speech rate as a consequence of active control mechanisms from the CNS.

Although some of the hypotheses mentioned here may be true, we would like to note the following two points: First, most people who compared opening and closing gestures in repetitive CV-syllables (including Gracco 1988) used oral stops in their speech material. The main goal in oral stop production is to produce an airtight seal for the oral closure and a following perceptually salient burst. It has been hypothesized that the articulatory movement is planned towards a target located above the actual vocal tract limit (e.g. for bilabials see Löfqvist and Gracco 1997, for alveolars see Fuchs et al. in press). In terms of stability and simplicity, such a control strategy seems to be extremely efficient in comparison with the control of a fine positioning. The impact of the lower lip on the upper for bilabials or of the tongue at the palate for alveolars is likely to influence the velocity profile of the closing gesture, but has no or less influence on the opening gestures.

Second, double velocity peaks can also occur for instance in a single /y/-/u/ movement without any underlying submovements. Payan and Perrier (1997) found that the origin of such a double peaked pattern is due to muscle anatomy inducing a certain time sequencing of the activation/deactivation of the Styloglossus and the Genioglossus posterior muscles. Therefore, any inference from kinematics alone about the underlying motor control mechanisms should be considered with caution.

McClean and Clay (1995) studied the relation between lower lip kinematics and their underlying single motor unit activity by means of EMG. They proposed that at least three different mechanisms may contribute to an increase in movement velocity: changes of the rate of firing motor units, changes in motor unit recruitment, and changes of the stiffness of the relevant articulator. Their aim was to observe the first two - firing rate and recruitment patterns of

single lower lip motor units simultaneously with the corresponding lip kinematics under varying speech rate conditions and phonetic structure. Three subjects were recorded by means of EMG of the following muscles: orbicularis oris inferior (OOI, active during lip closing), depressor labii inferior (DLI, active during lip opening) and mentalis (MENT, active during lip closing). The speech material consisted of repetitive CV syllables with C being /p/, /v/ or /f/ and V being /æ/ or /ʌ/. The firing rate was defined as the number of spikes per second, and spikes have been determined operationally by means of a threshold criterion. Kinematic results for lower lip movements at a higher speech rate exhibit differences with respect to opening and closing gestures. In general, closing gestures showed a significant increase in velocity whereas opening gestures did not. However, in the opening gestures differences in the average number of spikes per syllable were observed for the DLI in dependence of the vowel context. The average number of spikes were positively correlated with the amplitude of the velocity peaks in the kinematic signal. A similar correlation could not be found for the closing gestures. The authors suggest that an increase in speech rate from very slow to fast is associated with an increase of the firing rate of single motor units. According to these authors and contrary to Adams et al.'s (1993) suggestion, increase in speech rate is produced without changing the control strategy from multiple submovements at a slow rate to a unique movement at a fast rate, at least for opening gestures. According to McClean (personal communication), this finding sheds also new light on the interpretation of the variation of the number of velocity peaks with speech rate as observed by Adams et al. (1993): it is related to the motor units firing rate that determines the overlap between the successive parts of the movement associated with each motor unit activation.

In a follow-up study McClean and Tasko (2003) proposed again: *“Although the relationship between neural input to motoneurons and kinematics is extremely complex, kinematic analysis can provide a partial window to the neural processes underlying speech production”* (p. 1388). They investigated average lower lip and jaw muscle activities (mentalis MENT, depressor labii inferior DLI, masseter MAS (jaw opener)), anterior belly of digastric ABD (jaw opener) by means of broad-field EMG recordings simultaneously with kinematic data by means of EMA. Speech rate varied in 5 conditions and intensity in 2 different levels. Results for variations of loudness showed a strong positive correlation between muscle activation level with mean movement speed and movement distance. Concerning variations in speech rate similar results could not be found. Most consistently across speakers a general negative correlation of muscle activation levels with movement duration has been detected.

In summary, observations of kinematics with varying speech rate are manifold. Authors often discuss kinematic results with respect to concepts from motor

control in general. One of the most cited references has been mentioned, discussing lip and tongue kinematics with respect to different underlying control strategies counting for variations in speech rate (Adams et al. 1993).

There is little experimental evidence in the literature directly linking kinematics of speech movements and their underlying motor unit/s activity. McClean and Clay (1995) and McClean and Tasko (2003) investigated variations of speech rate and loudness by simultaneously recording lip and jaw movements and lip and jaw muscle activity. Parts of their findings provide evidence on the link between kinematics and muscle activity and demonstrate that no particular control strategy is necessary when switching from slow to fast (for opening gestures in single motor units McClean & Clay 1995) or from normal to loud speech (in multiple motor units McClean and Tasko 2003). In order to further understand speech motor control mechanisms it seems therefore indispensable either to directly investigate kinematics and muscle activity, or/and to compare experimental data with simulations using different motor control models generating muscle activation. Other appropriate methodologies would be to perturb the motor system and compare compensatory movements with the unperturbed condition or to compare normal and pathological speech. The two latter will not be taken into account here.

### ***1.3 The implementation of motor control models in speech: Some notes, our assumptions and methodology***

Speech production involves the precise control of fast articulatory actions in a task specific manner and is therefore characterised by neural and muscular activities. However, there are at least two particular properties of the speech production mechanism which seem to be speech specific in comparison to other human motor systems (e.g. arm or eye movements):

- (1) the most flexible articulator, the tongue, moves in a narrow space delimited by the palate, pharyngeal walls, teeth, cheeks and lips, and
- (2) expiratory air coming from the lungs, passing the glottis propagates through the vocal tract with certain characteristics depending on the changing vocal tract configurations and the corresponding changes in the perturbation of air.

The consequences of these two properties are challenging since the speech motor control system may integrate vocal tract limitations or certain aerodynamic information in the planning of sounds. Hoole et al. (1998) for instance, investigated the potential role of aerodynamics onto kinematics in order to explain the forward movements during oral closure in velar stop production (for further discussion on this topic see 2.3.2.). By means of EMA and intraoral pressure measurements they carried out an experiment where three speakers pronounced

the target words in an egressive and in an ingressive condition. Although forward movement for velars considerably decreased in the ingressive condition, it was not eliminated. The authors suggest “*that the motor planning system may be anticipating the aerodynamic forces and planning movement trajectories to take advantage of the direction and magnitude of the force vector*” (p.136).

We suppose that during speech acquisition the infant develops simple internal models (for a review on the simplicity versus complexity of internal representations, also called ‘internal models’ see Perrier in press). It establishes relations between (1) motor commands and perceptual outputs, (2) motor commands and proprioception (including the limits of tongue movement due to vocal tract boundaries), (3) motor commands and aerodynamics (in particular the magnitudes of subglottal pressure), and (4) biomechanical properties of the articulators such as mass, inertia, muscle force directions etc. These relations are mapped with respect to their auditory consequences and within the learning process integrated in planning sounds. Given the fast nature of some articulatory movements, which are sometimes below 50 ms, such internal models are required since this duration is below the minimum delay necessary for the cortex to monitor the ongoing speech act. We also assume that many kinematic patterns are consequences of the physical properties of the complex speech apparatus tuned by the CNS. They are not due to complex internal models that would precisely determine articulatory movements at each point of the trajectory. In this framework we want to test to what extent speech characteristics can be explained by simple control strategies and simple internal models. Our methodology is therefore to compare experimental results with results from simulations by means of a complex biomechanical model, and by controlling targets, not trajectories. Of particular interest for the current work is the hypothesis that vocal tract limitations, aerodynamics and biomechanics can affect the kinematics of articulatory movements. Their potential role will be discussed below.

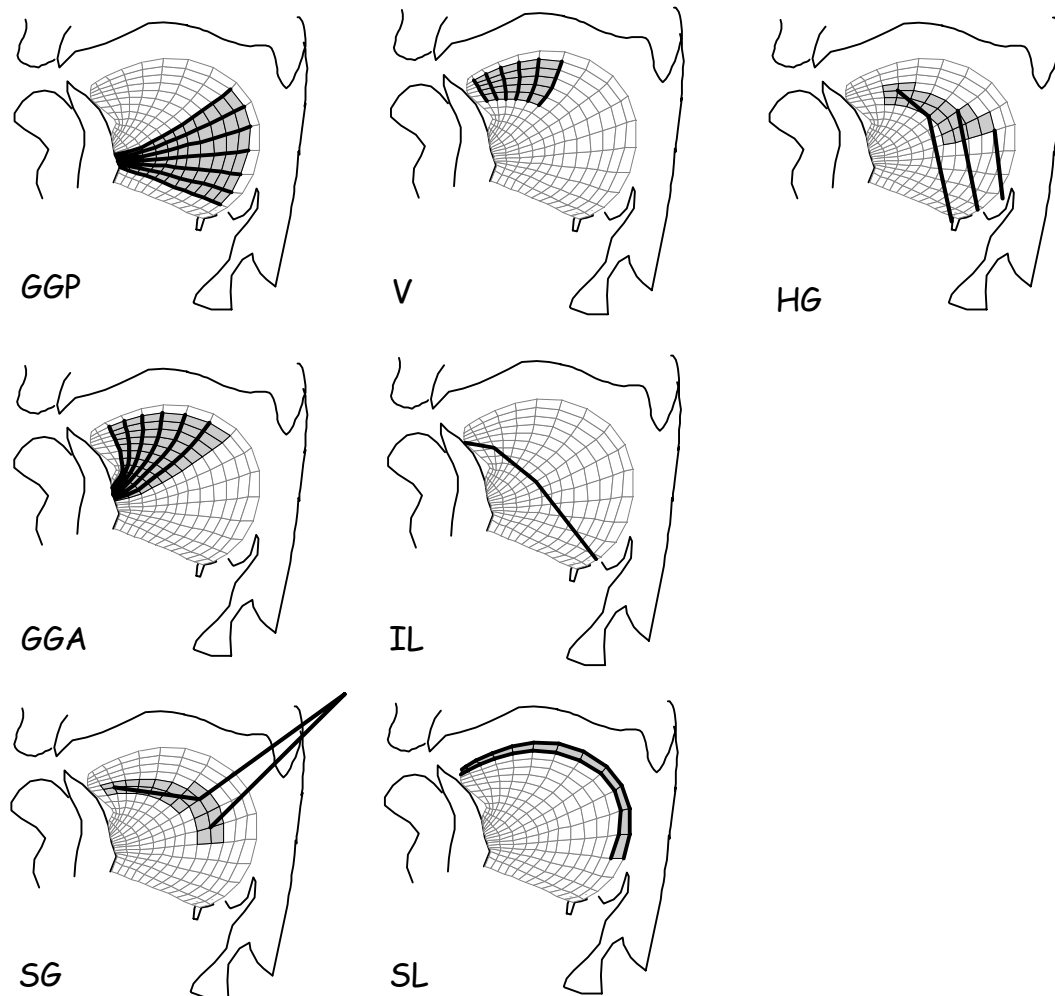
## **2 Potential underlying factors of speech kinematics**

Before providing examples on how biomechanics, aerodynamics and vocal tract limits can shape speech movements, we like to introduce briefly the model we used.

### **2.1 Introduction to the complex peripheral model**

Since we assume a complex peripheral speech apparatus plus a simple control due to the CNS and our methodology is to compare experimental data with data from simulations of a model, a complex biomechanical tongue model has been built previously.

Elastic properties of tongue tissues are accounted for by finite element (FE) modelling. Muscles are modelled 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. Curves (see Figure 1) representing the contours of the lips, palate and pharynx in the midsagittal plane are added to specify the limits of the vocal tract.



**Figure 1:** Defined tongue muscles of the FE structure, black lines correspond to macrofibres: GGP (genioglossus posterior), GGA (genioglossus anterior), SG (styloglossus), V (verticalis), IL (inferior longitudinalis), SL (superior longitudinalis), HG (hyoglossus)

The jaw and the hyoid bone are represented in this plane by static rigid structures to which the tongue is attached. Changes in jaw height can be simulated through a single parameter that modifies the vertical position of the whole FE structure. The model is controlled according to the  $\lambda$  model (Feldman 1986) that specifies for each muscle a threshold length,  $\lambda$ , where active force

starts. If the muscle length is larger than  $\lambda$ , 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  $\lambda$  space (for more details see Payan & Perrier 1997, Perrier et al. 2003). Muscle forces are applied to the FE structures via macrofibers that specify the insertions and the main force directions of each muscle (see black lines in Figure 1).

The modelling of the fluid-wall interaction and the action of the tongue at the palate are integrated in the complex biomechanical model. The first implies at each time step the specification of: (1) the area function from the sagittal distances generated by the 2D tongue model and their coronal correspondence (see Perrier et al. 1992), and (2) the volume velocity of the airflow through the vocal tract. These steps are followed by the computation of: (1) the distribution of pressure within the vocal tract, and (2) the pressure forces at each node of the tongue model, which are then added to the muscle forces to calculate the global forces shaping the tongue.

The area function is computed using an adapted version of the original  $\alpha\beta$ -model (Heinz and Stevens 1965), where  $\beta=1.5$ . In order to provide realistic vocal tract cross-sectional areas,  $\alpha$  varies from the glottis to the lips according to a division of the vocal tract into 7 sections and to the value of the sagittal distance (Perrier et al. 1992).

Flow velocity and pressure distribution can be calculated with a flow model (Pelorson et al. 1995) based on a simple 2D potential flow theory, accounting for viscous losses as a perturbation of the inviscid solution. In addition, flow separation effects within a constriction are taken into account. For the sake of simplicity, the flow separation position is estimated as the point downstream of the constriction where the cross-sectional area is 20% larger than the minimum area of the constriction (Pelorson et al. 1995). The flow model is driven by a single parameter: the pressure difference  $\Delta P = P_0 - P_{out}$ , where  $P_0$  and  $P_{out}$  are respectively the pressure past the glottis and at the lips.

The action of the tongue at the palate is modelled in two steps: (1) detection of contact of the tongue at the palate, and (2) generation of the resulting contact forces by means of the penalty method (Marhefka & Orin 1996). More specifically, if a node of the FE structure moves beyond the limit (palate) a repulsion force is generated to move this node back as a non-linear function of the penetration distance (for further details on the method, see Perrier et al. 2003).

The model is currently under development in order to improve it from two dimensions to three dimensions (Gerard et al. in press).

## **2.2 Aerodynamics**

Precise articulatory movements on their own do NOT produce speech sounds, the propagation of air with a certain density and speed of particles through the vocal tract is required. The source (e.g. vocal fold vibrations) and the vocal tract interact with each other during the production of different sounds.

Experimental evidence from simultaneous aerodynamic and articulatory movement measures are rather rare. However, kinematic results have often been interpreted with respect to aerodynamics, especially in explaining phenomena like devoicing in oral stops or voicing/voicelessness in consonant clusters. The same holds true for aerodynamic results, where underlying articulatory movements have been inferred without any experimental evidence. One exception is for instance the study of Svirsky et al. (1997) who observed tongue displacement in relation to intraoral pressure changes. Both measurements were used to assess the validity of a tongue compliance model. Results for tongue displacement for the consonants in /aba/, /apa/, and /ama/, and differences in intraoral pressure were investigated to shed some light on the question whether cavity enlargement is an active or passive mechanism. Cavity enlargement has been discussed as one mechanism to sustain the transglottal pressure difference, necessary for the production of vocal fold vibrations during oral closure. It turned out that the magnitudes of peak tongue dorsum displacement recorded by means of EMA were significantly larger during the production of voiced bilabials compared to smaller magnitudes in voiceless bilabials, even though the intraoral pressure was higher for the voiceless. It seemed surprising that such tongue dorsum differences occurred during a bilabial when surrounded by the same unrounded vowel context. The displacement was close to zero in the sequence involving the nasal. Svirsky et al. reported: *“It is interesting to observe that the relatively sharp, fast downward tongue dorsum displacements during /apa/ or /aba/ were generally close to the rise in intraoral pressure”*(p.565). Using a lumped parameter circuit model Svirsky et al. estimated tongue compliance and found much higher values for the voiced stops than for the voiceless. However, relaxation of the tongue for the voiced stop did not explain all the results. Hence, the authors proposed a combination of intentional relaxation of tongue muscles with an active displacement of the tongue. For the voiceless they suggested an active stiffening process of the tongue. In the context we discuss here their findings also provide evidence that intraoral pressure, i.e. aerodynamics, can affect tongue displacement since differences do not occur in the nasal (Nasals obviously do not involve a high intraoral pressure, since the air can escape via the nasal cavity). However, a separation between changing tongue positioning due to pressure and/or due to an active mechanism



cannot be based on their results and the amount of change the pressure causes is rather speculative.

Another example for the potential influence of aerodynamics on tongue kinematics are looping patterns, which will be discussed in more details in section 2.3.2. (see the contribution of Hoole et al. 1998, Perrier et al. 2000a).

## **2.3 Biomechanics**

Two different examples will be given here in order to point out the potential role of biomechanics for the speech motor control process and its implications for kinematics. The first one is related to the explanation of the limited degrees of freedom of tongue movements in vowel production<sup>2</sup>, and the second to the explanation of the so called ‘looping patterns’ in velar stop production.

### *2.3.1 Degrees of freedom for tongue movements during speech production*

The production of speech requires the simultaneous control of at least thirty different muscles. However, at the same time classical articulatory descriptions of vowel production are limited to a small number of parameters such as high versus low, front versus back for the tongue, and rounded versus spread for the lips. Hence, the understanding of speech motor control requires a reduction of the dimensionality from the muscle control space to a more functional, speech-related control space. The functional, speech-related control space will hereafter be called the degrees of freedom of the vocal tract. The reduction in dimensionality is a desired aim since it allows generalisations independent of speakers’ specific mechanisms. Even more broadly, it might show to what extent specific muscles are coordinated to produce meaningful sounds in the different languages. Previous work on the degrees of freedom were mainly based on statistic analyses of kinematic data. Harshman et al. (1977), Jackson (1988), Nix et al. (1996) and Hoole (1999) applied a PARAFAC analysis to x-ray or EMA data for English, Icelandic, and German. Maeda (1990) ran a guided principal component analysis (PCA) on x-ray data of French. Sanguineti and colleagues (Sanguineti et al. 1997, 1998) used the same corpus as Maeda, but additionally they provided a projection of the data set in a modelled muscle space by means of a biomechanical model of the tongue, jaw and hyoid bone. These authors were not only able to present a reduction in dimensionality, but

---

<sup>2</sup> It should be noted that parts of this work (but with different simulations) have been presented at ICSLP Beijing, see Perrier et al. (2000b).

also a description of the muscular correlates of the degrees of freedom during vowel production.

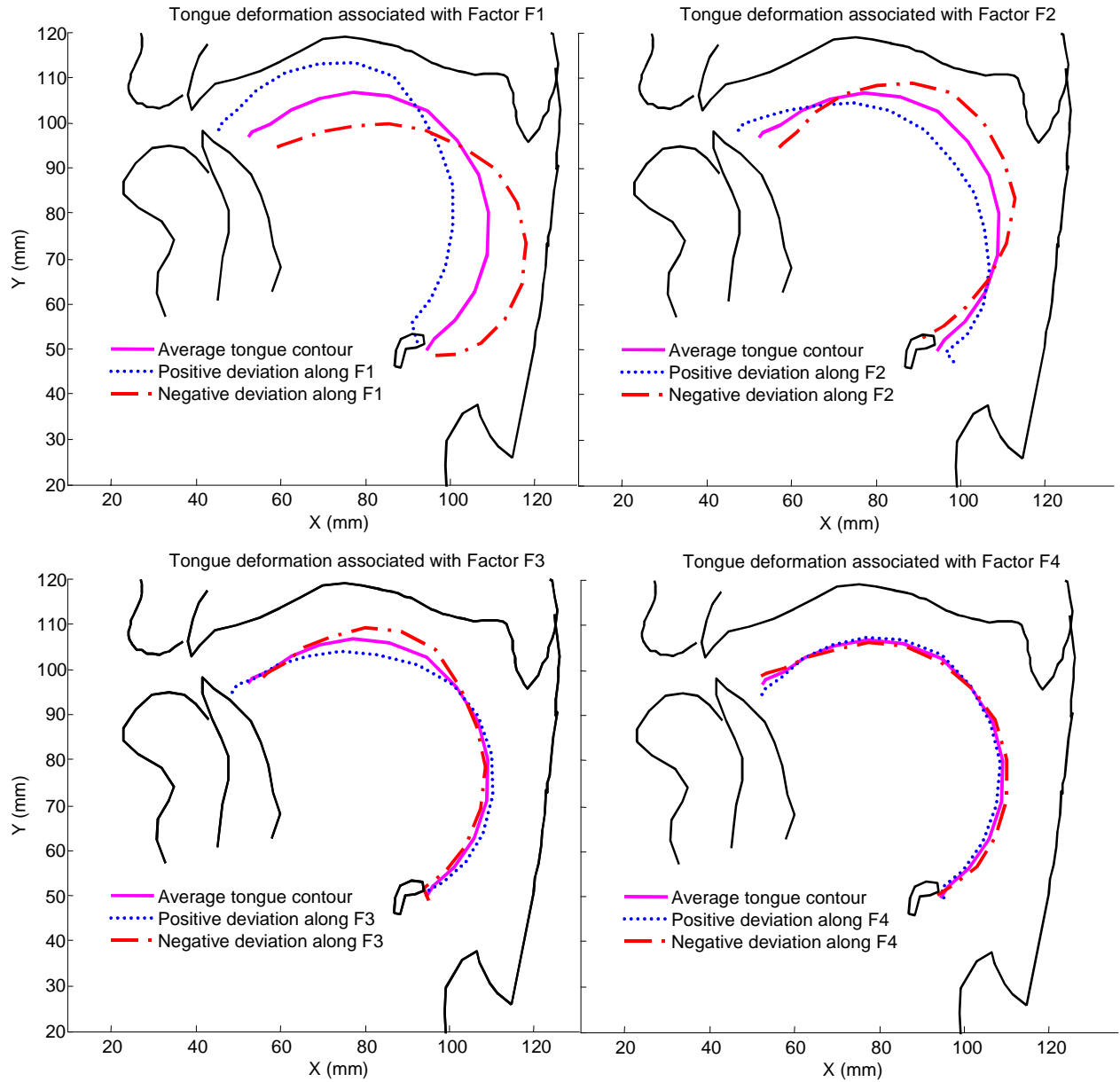
Although four different languages were analysed in these studies, most of the results presented show that more than 90% of the variance observed in the tongue shapes can be ascribed along two main degrees of freedom: (1) a movement of the tongue body along a high-front to low-back axis (called ‘front raising’ in Harshman et al. 1977) and (2) a bunching of the tongue along a high-back to low-front axis (called ‘back raising’ in Harshman et al., 1977). Jackson (1988) found that the number of degrees of freedom were language specific, i.e. different for English and Icelandic. However, his PARAFAC analysis was then proved to be degenerate by Nix et al. (1996), who reanalysed the same data set.

The results of these studies lead to questions about the origin of the two main degrees of freedom: are they learned, speech-specific actions, or are they due to basic properties of the speech production mechanism? In the following we will explore the hypothesis that the two main degrees of freedom have their origin in the anatomical and biomechanical properties of the speech production apparatus. Toward this aim, the bio-mechanical model of the tongue was used to generate a large set of tongue configurations, on which a PCA was ran in order to extract the main axes of deformation.

First results were presented in Perrier et al. (2000b). They were based on a gaussian sampling of the motor control space with the commands around the rest position as an average vector. These simulations were limited to the analysis of tongue configurations during vowel production, excluding those which were too close to the palate. In this paper we propose an extension of the previous work, covering a very broad range of tongue shapes. We adopted a uniform sampling method and included tongue configurations in slight contacts with the palate. In doing so our simulations cover the whole range of tongue shapes that can be generated by the model. Thus, 9000 tongue configurations were simulated and analysed with the classical PCA procedure (see Perrier et al. 2000b for details). The results of the PCA are depicted in figure 2 for a variation of  $\pm 1$  standard deviation around the mean value along each of the principal axes. The first and second factors clearly correspond to the typical front and back raising patterns. The third factor can be associated with a vertical downward movement of the tongue body and results for the fourth factor are rather marginal.

In the majority of the studies based on statistical analyses of articulatory data, more than 90% of the variance observed for a subject were described by the first two factors, while in our study 3 factors are necessary to reach approximately the same level of description. Results are as follows: the first factor explains 69 % of the variance, the first two factors 88 %, the first three factors 96 % and the first four factors 99 %. The slightly greater number of factors is in agreement with Nix et al.’s (1996) findings, which showed that

when the tongue shapes of 6 speakers were analysed together, 4 factors were necessary to reach the same level of description in comparison to the 2 factors extracted from the data of a single subject. Since our data were generated from a variety of random muscle commands relevant for vowel production, they may be more general, analogous to the combined data from 6 speakers.



**Figure 2:** Tongue deformations based on a PCA for the first four factors (from upper left to lower right), solid line: average contours, dotted lines: positive deviations from the average, dashed-dotted lines: negative deviations from the average; for further details see text

We conclude that the degrees of freedom in vowel production extracted from our simulations for French, and found in several studies for German, Icelandic, and English are due to the anatomical and biomechanical properties of the tongue and therefore not language-specific. Speech motor control uses these degrees of freedom to determine and differentiate speech articulations with respect to the various sounds of a language.

### *2.3.2 On looping patterns*

In a series of papers (e.g. Houde 1968, Ohala 1983, Mooshammer et al. 1995, Hoole et al. 1998, Löfqvist and Gracco 2002, Geng et al. 2003, Perrier et al. 2000a, Perrier et al. 2003, Brunner et al. 2004, Brunner 2005) researchers were interested in explaining the striking movement trajectories occurring during velar stop production. The trajectories have been called ‘looping patterns’ since they are reminiscent of ellipses. Loops can be found during V1CV2-sequences with C being a velar stop. Depending on the surrounding vowel context with V1 being a back vowel and V2 a front vowel one would expect a forward movement during the oral closure, simply as a consequence of coarticulation. Such forward sliding movements are however also found for V1=V2 as for instance in /aka/ where one could assume comparable movements towards oral closure and back.

The explanations for the phenomenon are manifold: due to aerodynamics, biomechanics<sup>3</sup> or cost minimization. Aerodynamics is in most cases mentioned, but for different reasons: Houde (1968) assumed that the forward movement of the tongue along the palate in a voiced velar stop would be due to the increased intraoral air pressure. Ohala (1983) attributed looping patterns to a strategy enlarging the oral cavity in order to maintain voicing for the voiced velar stop. Mooshammer et al. (1995) rejected this hypothesis since they found larger forward movements for the voiceless in comparison to the voiced stops. In order to test the impact of intraoral pressure onto tongue kinematics quantitatively, Hoole et al. (1998) observed looping patterns in normal and ingressive speech. Although they found smaller loops in ingressive speech, they were also directed forwards so that an increased intraoral pressure can not capture the whole phenomenon alone. Modeling work has been carried out by Perrier et al. (2000a) using a combination of a biomechanical model and an airflow model. They investigated looping patterns in low back and high front vowel contexts and found that biomechanics have a major impact on the kinematic patterns while aerodynamics play a negligible role when the velar stop is produced during low

---

<sup>3</sup> Since our previous work on loops was mainly related to biomechanics (Perrier et al. 2003), we have included the example at this point. However, it could also be included at the section on aerodynamics or vocal tract limits.

back vowel context. For /aki/ and /iki/ sequences the authors mentioned comparable influences of biomechanics and aerodynamics on the loops. These patterns were sensitive to the onset of pressure rise in the closing gesture and to the amount of pressure. With an earlier onset of the pressure rise and with a higher pressure, larger movement amplitudes were simulated.

A totally different perspective explaining looping patterns has been given by Löfqvist and Gracco (2002). They state that neither aerodynamics nor biomechanics alone would account for the observed patterns. Hence they suggest that loops are a result of a general motor control principle - the cost minimization process. This principle is associated with holding the third derivative of the movement, the jerk, as small as possible; this corresponds to a general smoothing strategy (Hogan 1990). According to the cost minimization principle, the whole trajectory of the loops would be controlled by the CNS.

In contradiction to this idea, Perrier et al. (2003) simulated looping patterns by means of a biomechanical model without any cost minimization strategy: Consonants and vowels in VCV-sequences have been specified in terms of targets. The consonant was always /k/ and for the two surrounding vowels /a, i, u/ have been used. Based on the findings of their simulations they suggested that biomechanical properties of the tongue explain looping patterns for all sequences where the first vowel was /a/ or /u/, independent of the second vowel. When the first vowel was /i/ they found variable forward or backward loop patterns, depending on the position of the target specified for /i/. This finding was consistent with the variability of experimental data reported in the literature. Consequently, no central processes seem to be necessary to control the whole trajectory of these sequences.

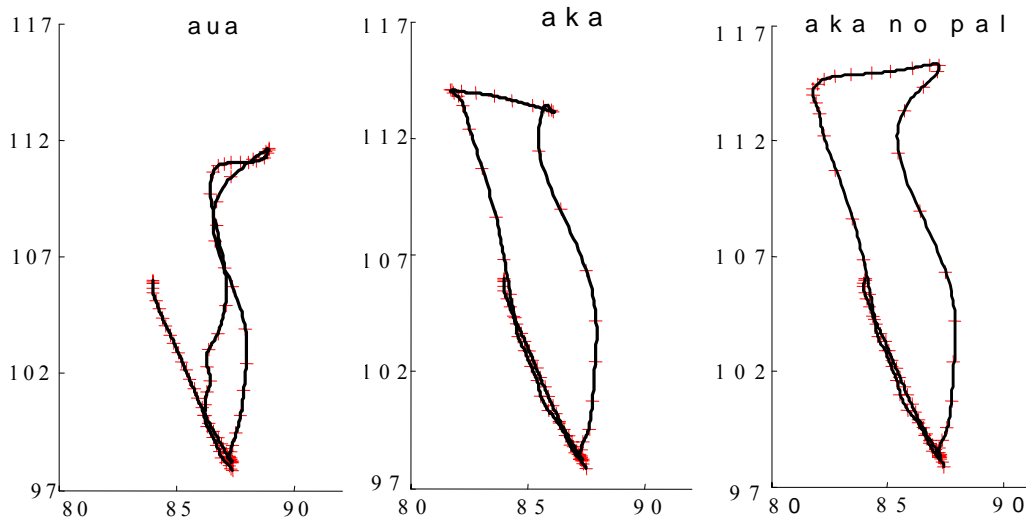
A generally accepted explanation for the combination of underlying factors and their strength contributing to these kinematics is still missing. In order to further support the biomechanical explanation of looping patterns we assume that loops may not only be found in the production of velar stops in e.g. /a/-context, but also in any other movement directed to the velar region. Thus, a sequence such as /aua/ should also show looping patterns to a certain extent. We therefore simulated 3 different sequences: /aua/, /aka/ with the impact of the tongue at the palate included in the model, and /aka/ with no palate in place, i.e. no impact of the tongue at the palate. The muscle activation patterns are given in table 1.

In all cases (see figure 3) it can be observed that in the upper part of the trajectory slight forward movements occur. The size of the loop is clearly larger for the /aka/-sequence than for /aua/. This is consistent with Perrier et al.'s (2003) findings that the relative position of the consonant and first vowel target has an incidence on the size of the loop. Finally, the trajectory goes further back in the absence of the palate (compare /aua/ and /aka/ without the palate). This

movement is due to the major influence of the Styloglossus that pulls the tongue back high in the velar region.

**Table 1:** Muscle activation patterns for the three simulations: - no activation, + slight activation, ++ clear activation)

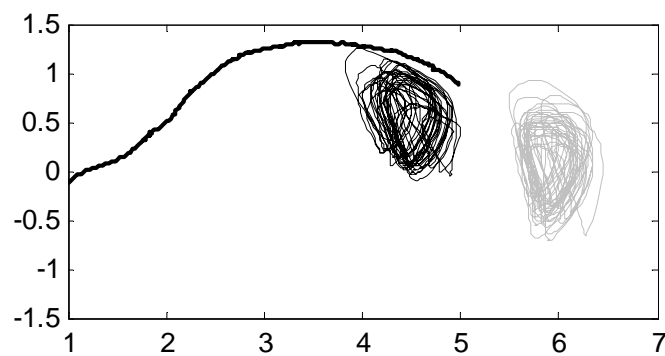
	<b>GGP</b>	<b>GGA</b>	<b>HYO</b>	<b>SG</b>	<b>VER</b>	<b>SL</b>	<b>IL</b>
<b>/a/</b>	-	+	+	-	-	-	+
<b>/k/</b>	++	-	-	++	-	-	-
<b>/u/</b>	+	-	-	++	-	-	-



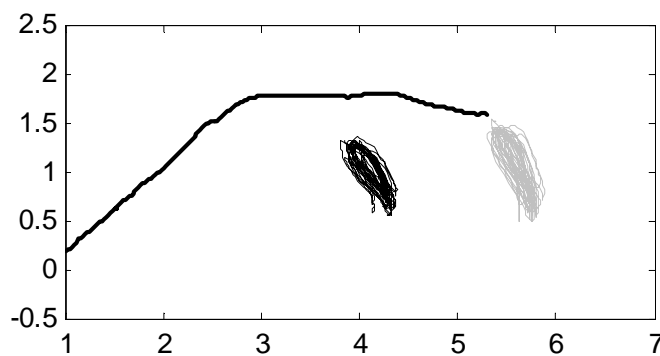
**Figure 3:** Trajectories of the simulated sequences in mm: /a u a/, /a k a/ (with palate) and /a k a/ (without palate) – from left to right; all simulations start at a rest position (at 83.5mm, 106mm) move downwards to the first /a/ than upwards for the /k/ or /u/, forwards and downwards again for the second /a/

We would like to note briefly another factor which might influence the shape of the looping patterns and maybe due to the anatomical properties of the relevant speaker. Figure 4 plots EMA trajectories for two speakers repetitively producing /ka/-sequences for a period of 10 seconds. Their task was to realise the syllables as quickly and as intelligible as possible (for methodology see Hartinger 2005). The x-y coordinates of the tongue back coil (in grey) and the tongue dorsum coil (in black) are displayed. The first and the last tokens are discarded for visualisation purpose. The bold black line on top corresponds to the palate trace. Speaker 1 clearly exhibits larger looping patterns for n=39

repetitions, especially due to the larger forward movement of the tongue in comparison to speaker 2 (n=35 repetitions). It seems implausible to explain the speaker dependent differences due to differences in speech rate since these are minor. Additional to possible biomechanical and aerodynamic factors which may contribute to the different looping patterns, one can also notice differences in the palate shape for the two, with speaker 1 exhibiting a dome shaped palate (see figure 4) and speaker 2 a flat palate shape from a sagittal perspective (see figure 5). The palate shapes for the two speakers are not only known due to the palate trace of the EMA recording, but they have also been analysed on the basis of their EPG palates. It is hypothesised that the variations in palate shape, the planned consonant target (for speaker 1 it is further backward), and the angle of incidence between tongue trajectory and palate contour are partly responsible for the different trajectories (for the general idea of the latter see Brunner et al. 2005). However, this hypothesis needs further verification by implementing different palate shapes in the biomechanical model.



**Figure 4:** Articulatory trajectories during repetitive /ka/- productions for speaker 1. black: tongue dorsum coil, grey: tongue back coil



**Figure 5:** Articulatory trajectories during repetitive /ka/- productions for speaker 2, black: tongue dorsum coil, grey: tongue back coil

## **2.4 Limits due to vocal tract borders**

It is one of the peculiarities of speech that the most flexible articulator the tongue, moves in a narrow space delimited by soft tissues (lips, cheeks, soft palate, pharyngeal walls) and hard tissues (the hard palate and the teeth)<sup>4</sup>. We mainly focus on the upper limit for the tongue's action, the palate and assume that:

(1) these vocal tract borders influence kinematic patterns and their token variability especially for those sounds which are realised very close to the vocal tract borders such as high front vowels (see Mooshammer et al. 2004, Brunner et al. this volume).

(2) the tongue's action at the palate is taken into account in the speech motor control process in terms of limiting the degrees of freedom for tongue movement and supporting the tongue's shaping. As far as consonant production is concerned, Stone (1991) for instance suggested that some tongue shapes, particularly those in the production of alveolar fricatives, could not be produced by a free-standing tongue position, i.e. without the palate as a reference.

In previous studies (Fuchs et al. 2001, Fuchs et al. in press) we investigated the production strategies of alveolar stops and fricatives by means of simultaneous EMA and EPG recordings. For alveolar stops versus fricatives, two different control strategies were hypothesized: a target above the contact location for alveolar stops resulting in a collision of the tongue tip at the palate as opposed to a precise positioning of the tongue at the lateral margins of the palate for alveolar fricatives. Results for both strategies were evident in tongue tip kinematics and tongue palate contact patterns. The large deceleration peak in /t/ during the closing gesture in comparison to a smaller peak in the preceding opening gesture supports the hypothesis for a collision of the tongue tip at the palate (in agreement with Hoole 1996, Fuchs et al. 2001). Additionally, the movement amplitude and the velocity for the closing gesture in /t/ were larger as opposed to the alveolar fricative, although the closing gesture duration was significantly shorter (/a/-context). The stop also showed more anterior palatal contact patterns than the fricative which may be interpreted as a consequence of the collision of the tip against the palate in comparison to a precise positioning. Further evidence for this hypothesis was provided by measuring the amplitude of movement during the acoustically defined closure or constriction. The tip sensor moved to a greater extent for the stop than for the fricative.

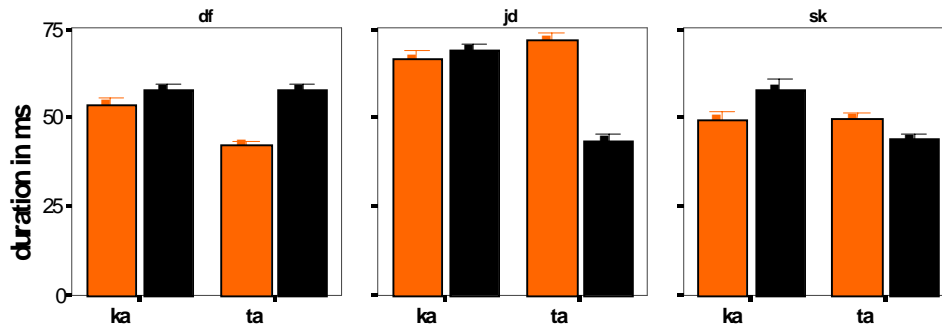
Based on our previous results we suppose that the collision of the tongue at the palate has an impact on the duration of the deceleration phase of the closing gesture, i.e. the stronger the impact, the shorter the deceleration phase,

---

<sup>4</sup> The effects of soft tissues on tongue kinematics may be different in comparison to effects due to an action of the tongue at hard tissues.



resulting in an asymmetrical velocity profile. Since the acceleration phase should not be affected and the deceleration duration shortens, the profile should become skewed to the right. In order to test this hypothesis we carried out the following experiment<sup>5</sup>. Three speakers were recorded by means of EMA producing repetitive CV-syllables (/ta/ and /ka/) as fast and as intelligible as possible within a 10s time interval. On average between 35 and 40 syllables were produced. So far only closing gestures have been taken into account. The acceleration and the deceleration duration for the tongue back sensor in /ka/ and for the tongue tip sensor in /ta/ were measured in the tangential velocity signal. The acceleration phase was defined as the duration between closing gesture onset (velocity minimum) and the velocity peak and the deceleration duration as the time interval between peak velocity and closing gesture offset (following velocity minimum). Figure 6 shows the results of this measurement.

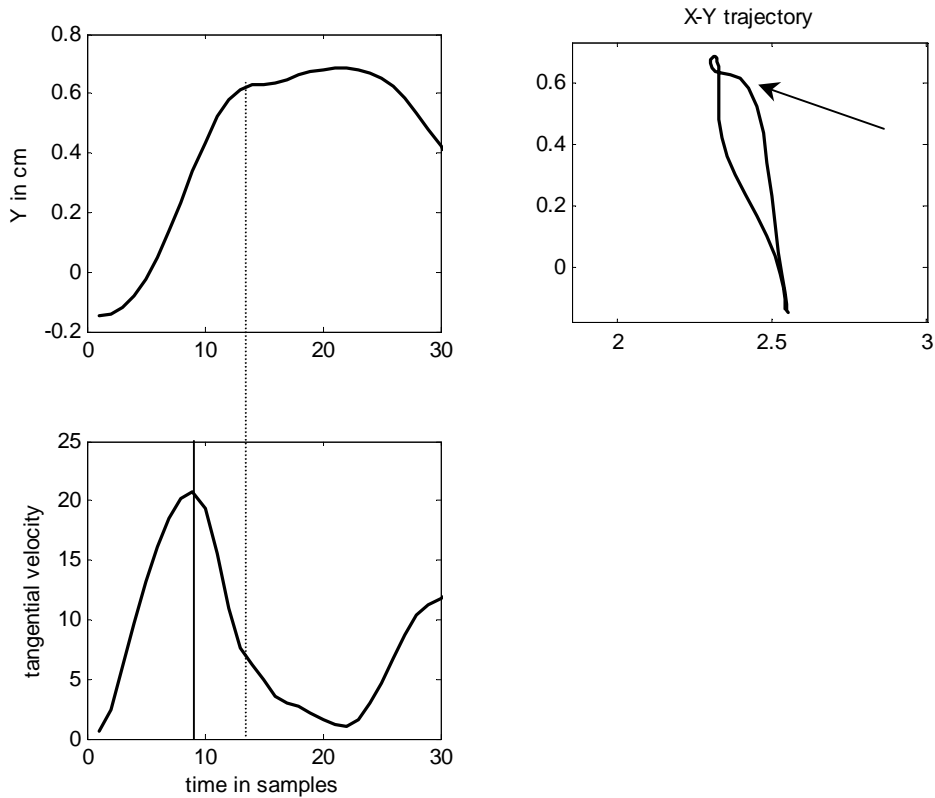


**Figure 6:** Means of acceleration and deceleration duration in ms with  $\pm 1$  standard error for the three subjects (df, jd and sk from left to right), grey bars correspond to the acceleration phase and black bars to the deceleration phase; left two bars /ka/, right two bars /ta/

At first glance, the results in figure 6 do not support the predicted patterns. For /ka/ none of the subjects shows the differences we supposed, since the deceleration duration is longer for sk, and rather similar to the acceleration duration for jd and for df. For /ta/ the results for two subjects (jd and sk) are in agreement with our assumptions, but df shows the reverse.

When looking into the details it becomes apparent that speaker df produces a small loop in the alveolar stop (see figure 7, right upper graph), i.e. the longer deceleration phase is due to a small forward sliding of the tip at the alveolars, starting at the marked dotted line.

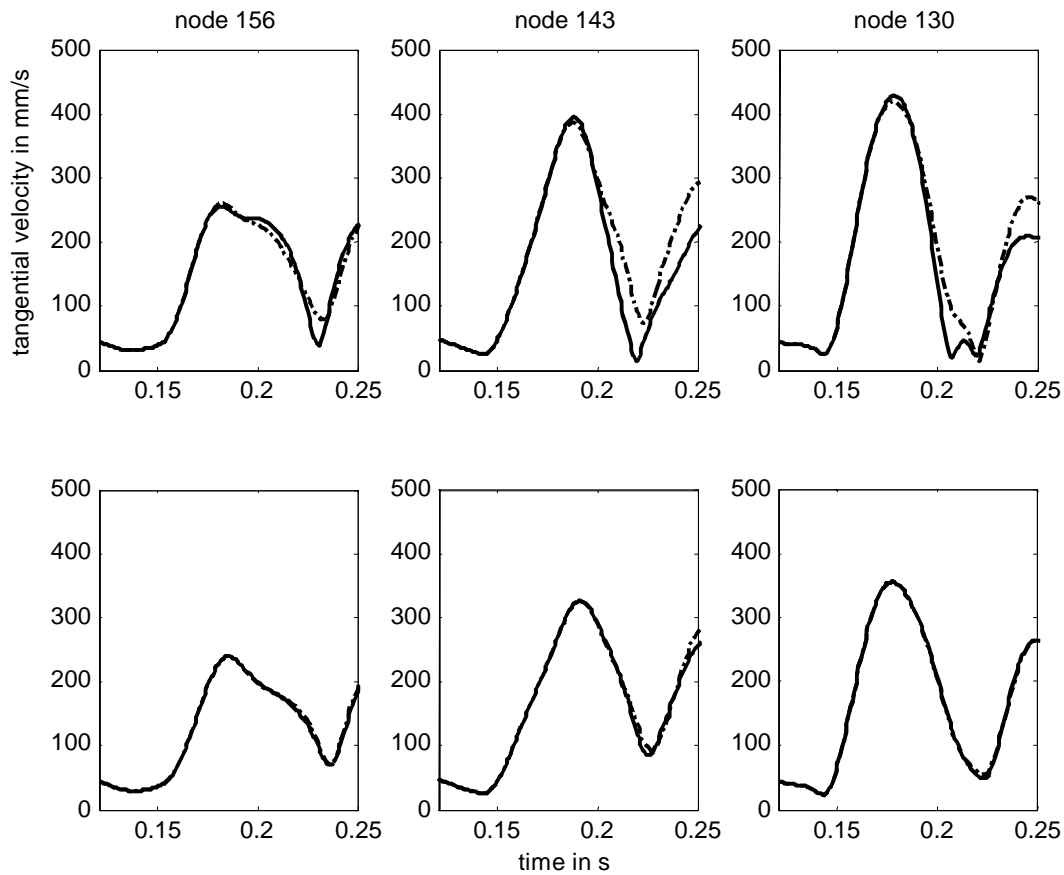
<sup>5</sup> Thanks to Mariam Hartinger and Jörg Dreyer for the recordings.



**Figure 7:** Example for DF's /ta/ production, left upper graph: upward vertical tongue tip movement, right upper graph: XY trajectory, left lower graph: tangential velocity profile, bold line: velocity peak, dotted line: beginning of forward movement corresponds to the array in the right upper graph

The corresponding tangential velocity profile (left lower graph) decelerates more slowly after tongue-palatal contact was made. The deceleration phase can be divided into two different parts, one where the tongue makes first contact with the palate and the second, where it continues to move along the palate in forward direction. Here the deceleration phase becomes longer than the acceleration. It is a typical pattern in the results for the velar stop and to some extent in the production for the alveolar stop for DF.

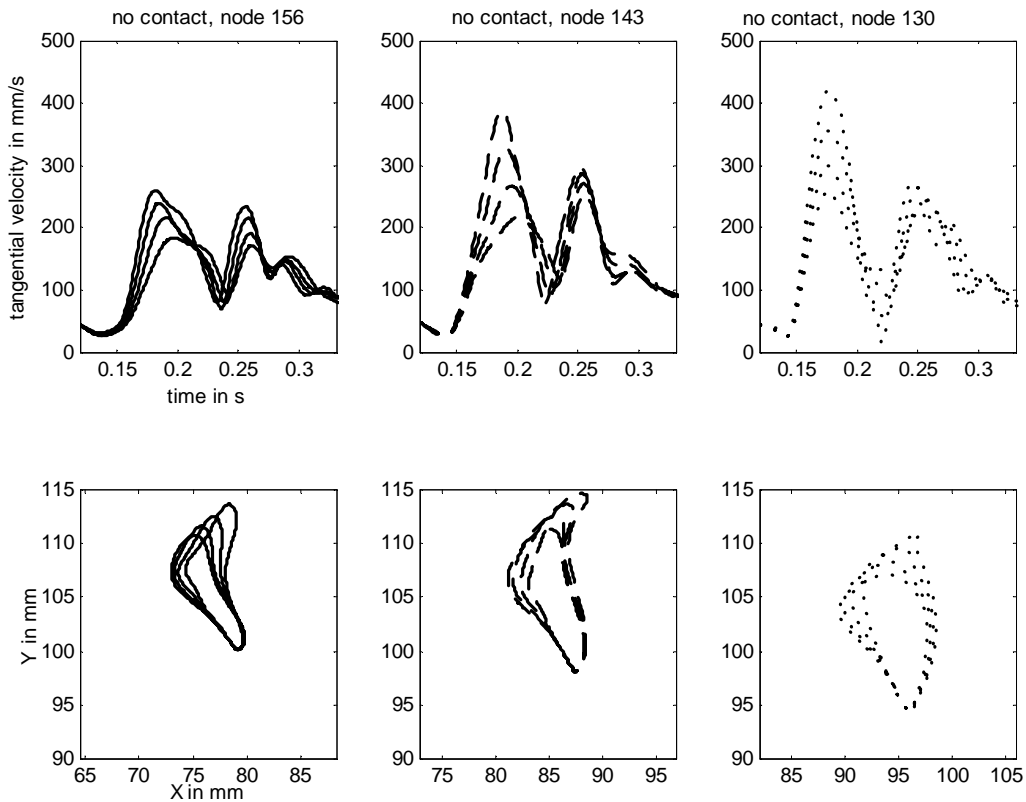
We used the tongue model in order to study the impact of the tongue's movement at the palate. The model has the advantage that simulations can be run with and without the palate in place and both conditions can be compared. Since the model is most realistic for /ka/ we chose the velar sequence and simulated 100 sequences with varying Styloglossus activity (lambda values between 61 and 91 in 0.3 steps; the lower the value, the greater the activity and the likelihood for an impact). We used two conditions, one with palate in place and one without the palate.



**Figure 8:** Simulated data showing the tangential velocity of the closing gesture for three different nodes (left: 156, middle: 143, right: 130 corresponding to: further front (156) to a further back (130) articulation in the vocal tract; bold lines: simulations with the palate in place, dashed-dot lines: simulations with no palate; upper track: high SG activity (73), lower track: lower SG activity (79)

Figure 8 shows a comparison of tangential velocities between simulations with and without the palate in place. In the upper track a higher Styloglossus activity has been chosen than in the lower track. It corresponds to a target which

is higher above the palate. These simulations clearly exhibit an impact of vocal tract limits on tongue kinematics when comparing simulations with and without palate. Differences are less strong for the more anterior articulation (node 156) in comparison to the more posterior articulation (node 130). The deceleration phase is noticeably shorter due to the impact. In the simulations with lower Styloglossus activity differences are rather marginal and won't be taken into account.



**Figure 9:** Tangential velocities in mm/s (upper track) and XY-trajectories in mm (lower track) for simulations with increasing Styloglossus activity (91, 85, 79, 73); no palate in place, 3 columns correspond to the 3 nodes (156, 143, 130) from left to right, see text for further details

Figure 9 exhibits some further interesting results for the simulations without the palate (similar effects exist also in the simulations with the palate and are therefore not included here): (1) higher SG activity coincides with higher peak velocity, with an overall shorter duration and further backward movement during the closing gesture. (2) Asymmetries in the velocity profile of the closing gestures (first velocity profile in the upper tracks) vary with respect to the different nodes. Node 156 shows a left shaped pattern with a longer deceleration phase than the acceleration phase. Nodes 143 and 130 which are further

backward in their placement show rather symmetrical velocity profiles with some variations depending on the SG activity.

This means that the speaker dependent differences we found in the experimental data may be a consequence of biomechanics (muscle cocontraction) and differences in the tongue back sensor placement, with sk having the tongue back sensor located more anteriorly than speakers jd and df. Additionally, they do not contradict our hypothesis that the impact of the tongue at the palate reduces the deceleration phase since all the simulations with a non-negligible impact (see figure 8) exhibit this pattern. However, the velocity profile serving as a reference may be asymmetrical due to biomechanical reasons and not bell-shaped as supposed. Therefore, the deceleration duration is not generally shorter than the acceleration duration.

### **3 Conclusions**

In this paper, we have presented a number of examples showing that the interpretation of speech kinematics in terms of underlying speech motor control is not straightforward. We have demonstrated that the parameters such as peak velocity, asymmetry of the velocity profile, and trajectory curvature, which have frequently been used in the literature to infer hypotheses about the underlying speech motor control strategies, are in fact the result of complex and non linear combinations of different factors. These factors are obviously linked with high level motor control strategies including optimal planning and listener oriented control, but they are also linked with physical phenomena such as speech articulators' muscle anatomy, biomechanics and dynamics, mechanical interactions between articulators (tongue-palate or tongue-teeth contacts), and interactions between airflow and soft tissues. As exemplified by the controversial debates about the origin of the so-called 1/3 power law corresponding to a non linear relation between tangential velocity and trajectory curvature of human arm movements, this statement holds not only true for speech, but also for every kind of skilled human movement. However, it is particularly acute for speech, since speech production necessitates the control of a complex motor system, coupling hard bodies, soft tissues and aerodynamics under time-varying mechanical constraints via the coordination of more than 30 potentially independent muscles. It involves the control of movements realized in sometimes very short durations, discarding any potential on-line feedback mechanism going up to the cortex.

And last but not least, speech production is a peculiar motor activity which essence is semiotic. Hence, hypotheses about speech motor control must

not only give an account of the observed kinematics, but also of the link between the observed kinematics and the underlying semiotic code.

Studies of speech kinematics have been the main basis of speech production research for many years. They have been extremely fruitful and have allowed to develop major hypotheses about speech motor control and its interaction with speech perception in the semiotic framework of speech communication.

To continue this kind of investigation is a necessity and justifies the remarkable effort that many of our colleagues have put in the development and enhancement of new data acquisition techniques using the most recent developments in physical measurement technologies. In parallel, investigations in the broad domain of human motor control have made noticeable progress, in particular in the domain of learning, of internal representations, and in the way to integrate low-level short delay feedback loops. Interpretations of speech kinematics in terms of motor control have been inspired from these findings. However, the main trend in speech production studies has been to relate observations of speech kinematics directly to high level motor control strategies involving complex internal models. They have often overseen the important role of physics in shaping the patterns of speech kinematics.

With this paper we propose that the physical properties of the peripheral speech production apparatus should be put into the center of our investigations in order to account for the complex nature of speech kinematics. We suggest that the complex nature of speech kinematics is for a large part due to the complex peripheral speech apparatus and that it may not systematically be found in higher level motor control strategies.

## **Acknowledgements**

This work has been supported by a grant from the German Research Council (DFG) GWZ 4/8-1, P.1. Thanks to Joe Perkell for the collaboration on a previous version of the degrees of freedom in vowel production, to Mariam Hartinger for recording the EMA data on repetitive syllables, for Phil Hoole, Christian Abry, and the editors of this volume to provide useful comments on an earlier version of this paper.

## **References**

Adams, S.G., Weismer, G. & Kent, R.D. (1993): Speaking rate and speech movement velocity profiles. *Journal of Speech and Hearing Research*, 36: 41-54.

- Bullock, D. & Grossberg, S. (1988): Neural dynamics of planned arm movements: Emergent invariants and speed-accuracy properties during trajectory formation. *Psychological Review*, 95: 49-90.
- Brunner, J., Fuchs, S., Perrier, P. & Kim, H.-Z. (2004): Correlation between angle of incidence and sliding patterns of the tongue along the palate in Korean velar stops. *147<sup>th</sup> Meeting of the Acoustical Society of America, New York* [abstract & poster].
- Brunner, J. (2005): Supralaryngeal mechanisms of the voicing contrast in velars. *ZAS Papers in Linguistics* 39.
- Brunner, J., Fuchs, S. & Perrier, P. (2005): The influence of the palate shape on articulatory token-to-token variability. *ZAS Papers in Linguistics*, 42: 43-66
- Feldman, A.G. (1986): Once more on the Equilibrium-Point Hypothesis ( $\lambda$  Model) for motor control. *Journal of Motor Behavior*, 18(1): 17-54.
- Feldman, A.G., Adamovich, S.V., Ostry, D.J. & Flanagan, J.R. (1990): The origin of electromyograms – explanations based on the equilibrium point hypothesis. In Winters, J. & Woo, S.: *Multiple muscle systems: Biomechanics and movement organization*. Springer Verlag: New York: 195-213.
- Fuchs, S., Perrier, P. & Mooshammer, C. (2001): The role of the palate in tongue kinematics: an experimental assessment in VC sequences from EPG and EMMA data. *Proceedings of Eurospeech Aalborg*, 3: 1487-1490.
- Fuchs, S., Perrier, P., Geng, C. & Mooshammer, C. (in press): What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruents. Harrington, J. and Tabain, M. (Eds.): *Towards a better understanding of speech production processes*. Psychology Press: New York.
- Geng, C., Fuchs, S., Mooshammer, C., Pompino-Marschall, B. (2003): How does vowel context influence loops? *Proceedings of the 6<sup>th</sup> Speech Production Seminar Sydney*: 1-6.
- Gerard, J., Perrier, P. & Payan, Y. (in press): 3D biomechanical tongue modelling to study speech production. In Harrington, J. and Tabain, M. (eds.): *Towards a better understanding of speech production processes*. Psychology Press: New York.
- Gracco, V.L. (1988): Timing factors in the coordination of speech movements. *The Journal of Neuroscience*, 8: 4628-4639.
- Gribble, P.L. & Ostry, D.J. (1996): Origins of the power law relation between movement velocity and curvature: Modeling the effects of muscle mechanics and limb dynamics. *Journal of Neurophysiology*, 76(5): 2853-2860.
- Harris, C. M., & Wolpert, D. M. (1998): Signal-dependent noise determines motor planning. *Nature*, 394: 780–784.
- Harshman, R. A., Ladefoged, P. N., & Goldstein, L. (1977): Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, 62: 693–707.
- Hartinger, M. (2005): *Untersuchungen der Sprechmotorik von Poltereren mit Hilfe der elektromagnetischen mediosagittalen Artikulographie (EMMA)*. Unpublished PhD thesis at Martin Luther University Halle (Saale).

- Heinz, J.M. & Stevens, K.N. (1965): On the relations between lateral cineradiographs, area functions and acoustic spectra of the speech. *Proceedings of the 5<sup>th</sup> International Congress of Acoustic*, A44.
- Hogan, N. (1990): Mechanical impedance of single- and multi-articulator systems. In Winters, J.M. & Woo, S.L.-Y.(eds.): *Multiple muscle systems. Biomechanics and movement organization*. Springer: Berlin, New York: 149-164.
- Hoole, P. (1996). Theoretische und methodische Grundlagen der Artikulationsanalyse in der experimentellen Phonetik. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)*, 34: 3-173.
- Hoole, P. (1998): Modelling tongue configuration in German vowel production. *Proceedings of the 5<sup>th</sup> ICSLP Sydney*, 5: 1867-1870.
- Hoole, P. (1999): On the lingual organization of the German vowel system. *Journal of the Acoustical Society of America*, 106(2): 1020-1032.
- Hoole, P., Munhall, K. & Mooshammer, C. (1998): Do air-stream mechanisms influence tongue movement paths? *Phonetica*, 55(3): 131-146.
- Houde, R. (1968): A study of tongue body motion during selected consonant sounds. *Speech Communications Research Laboratory, Santa Barbara, SCRL Monograph 2*.
- Jackson, M.T.T. (1988): Analysis of tongue positions: Language-specific and cross-linguistic models. *Journal of the Acoustical Society of America*, 84(1): 124-143.
- Lebedev, S., Tsui, W.H. & Van Gelder, P. (2001): Drawing movements as an outcome of the principle of least action. *Journal of Mathematical Psychology*, 45: 43-52.
- Löfqvist, A. & Gracco, V.L. (1997): Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech, Language, and Hearing Research*, 40(4): 877-893.
- Löfqvist, A. & Gracco, V. L. (2002): Control of oral closure in lingual stop consonant production. *Journal of the Acoustical Society of America*, 111(6): 2811–2827.
- Maeda, S. (1990): Compensatory articulation during speech: evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In Hardcastle, W.J. & Marchal, A. (eds.): *Speech production and speech modelling*. Dordrecht: Kluwer Academic Publishers: 131-150.
- Marhefka, D. W. & Orin, D. E. (1996): Simulations of contact using a non-linear damping model. *Proceedings of IEEE International Conference on Robotics and Automation, Minneapolis, MN*, 2: 1662–1668.
- Massey, J., Lurito, J., Pellizzer, G. & Georgopoulos, A. (1992): Three-dimensional drawings in isometric conditions: relations between geometry and kinematics. *Experimental Brain Research*, 88: 685-690.
- McClean, M.D., Clay, J.L. (1995): Activation of lip motor units with variations in speech rate and phonetic structure. *Journal of Speech, Language and Hearing Research*, 38: 772-782.



- McClean, M.D. & Tasko, S.M. (2003): Association of orofacial muscle activity and movement during changes in speech rate and intensity. *Journal of Speech, Language and Hearing Research*, 46: 1387-1400.
- Milner, T.E. & Ijaz, M. (1990): The effect of accuracy constraints on three-dimensional movement kinematics. *Neuroscience*, 35: 365-374.
- Mooshammer, C., Hoole, P. & Kühnert, B. (1995): On loops. *Journal of Phonetics*, 23: 3–21.
- Mooshammer, C., Perrier, P., Fuchs, S., Geng, C. and Pape, D. (2004). An EMMA and EPG study on token-to-token variability. *Arbeitsberichte Institut für Phonetik und digitale Sprachverarbeitung Universität Kiel (AIPUK)*, 36: 46-63.
- Nelson, W.L. (1983): Physical principles for economies of skilled movements. *Biological Cybernetics* 46: 135-147.
- Nix, D. A., Papcun, G., Hogden J., & Zlokarnik, I. (1996): Two cross-linguistic factors underlying tongue shapes for vowels. *Journal of the Acoustical Society of America*, 99: 3707-3717.
- Ohala, J.J. (1983): The origin of sound patterns in vocal tract constraints. In MacNeilage, P.F. (ed.): *The Production of Speech*. Springer Verlag: New York, Heidelberg, Berlin: 189-216.
- Ostry, D.J. & Feldman, A.G. (2003): A critical evaluation of the force control hypothesis in motor control. *Experimental Brain Research*, 153: 275-288.
- 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.
- Pelorson, X., Hirshberg, A., Wijnands, A.P.J. & Bailliet H.M.A. (1995): Description of the flow through in-vitro models of the glottis during phonation. *Acta Acustica*, 3: 191-202.
- Pelorson X., Liljencrants J., Kröger B. (1995): On the aeroacoustics of voiced sound production. *Proceedings of the 15<sup>th</sup> International Congress on Acoustics, Trondheim, Norway*, 4: 501-504.
- Perrier P., Boe 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., Perkell, J., Zandipour, M., Pelorson, X., Coisy, V. & Matthies, M. (2000a): An attempt to simulate fluid-walls interactions during velar stops. In *Proceedings of the 5<sup>th</sup> Seminar on Speech Production: Models and Data, Kloster Seeon*: 149-152.
- Perrier, P., Perkell, J., Payan, Y., Zandipour, M., Guenther, F. & Khalighi, A. (2000b): Degrees of freedom of tongue movements in speech may be constrained by biomechanics. *Proceedings of the ISCLP Beijing*, 2: 162-165.
- Perrier, P., Payan, Y., Zandipour, M. and Perkell, J. (2003): Influences of tongue biomechanics on speech movements during the production of velar stop consonants: A modeling study. *Journal of the Acoustical Society of America*, 114(3): 1582-1599.

- Perrier P. (in press) About speech motor control complexity. In Harrington, J. and Tabain, M. (eds.): *Towards a better understanding of speech production processes*. Psychology Press: New York.
- Sanguineti, V., Laboissière, R., & Payan, Y. (1997): A control model of human tongue movements in speech. *Biological Cybernetics*, 77(1): 11–22.
- Sanguineti, V., Laboissière, R., & Ostry, D.J. (1998): A dynamic biomechanical model for neural control of speech production. *Journal of the Acoustical Society of America*, 103(3): 1615-1627.
- Stone, M. (1991): Toward a model of three-dimensional tongue movements. *Journal of Phonetics*, 19: 309-320.
- Svirsky, M., Stevens, K., Matthies, M., Manzella, J., Perkell, J. and Wilhelms-Tricarico, R. (1997) Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America*, 102: 562-571.
- Tasko, S.T. & Westbury, J.R. (2004): Speed–curvature relations for speech-related articulatory movement. *Journal of Phonetics*, 32: 65-80.
- Viviani, P., & Terzuolo, C. (1982) Trajectory determines movement dynamics. *Neuroscience*, 7, 431-437
- Viviani, P. & Schneider, R. (1991): A developmental study of the relationship between geometry and kinematics in drawing movements. *Journal of Experimental Psychology. Human Perception and Performance*, 17: 198-218.
- Viviani, P. & Flash, T. (1995): Minimum-jerk, two-thirds power law, and isochrony: converging approaches to movement planning. *Journal of Experimental Psychology. Human Perception and Performance*, 21: 32-53.
- Wiencke, G., Janssen, P. & Belderbos, H.(1987): The influence of speaking rate on the duration of jaw movement. *Journal of Phonetics*, 15:111-126.

# Articulatory variability of clutterers

**Mariam Hartinger**

*Zentrum für Allgemeine Sprachwissenschaft (ZAS), Berlin, Germany*

**Christine Mooshammer**

*Christian-Albrechts-Universität Kiel, Germany*

---

In order to investigate the articulatory processes of the hasty and mumbled speech of clutterers, the kinematic variability was analysed by means of electromagnetic midsagittal articulography (EMMA). In contrast to stutterers, clutterers improve their intelligibility by concentrating on their speech task. Variability is an important criterion in comparable studies of stuttering and is discussed in terms of the stability of the speech motor system. The aim of the current study was to analyse the spatial and temporal variability in the speech of three clutterers and three control speakers. All speakers were native speakers of German. The speech material consisted of repetitive CV-syllables and foreign words, because clutterers have the most severe problems with long words which have a complex syllable structure. The results showed a higher quotient of variation for clutterers in the foreign word production. For the syllable repetition task, no significant differences between clutterers and controls were found. The extremely large and variable displacements were interpreted as a strategy that helps clutterers to improve the intelligibility of their speech.

---

## 1 Introduction

Speech production underlies a certain amount of natural variability, a property that it shares with all other human motor behaviour. Hence, a specific motor task cannot be replicated in an absolutely identical way. This kind of token-to-token variability has been attributed to neural noise corrupting the control signals (Perkell & Nelson 1985). Knowledge about normal variability of speech motor abilities is limited and the cut-points between normal and so-called pathological motor behaviour is arbitrary (van Lieshout et al. 2004). In speech therapy it is common to classify the behaviour of speakers as pathological if it is different. This is the reason why the variability in the speech production of stutterers "... has always been a hot topic in research ..." (van Lieshout et al. 2004, 329).

In this current study, our aim is to investigate the spatial and temporal kinematic variability of the fluency disorder cluttering. Up to now no unique definition exists for this fluency disorder and the transition between “normal” and “pathological” is not fixed. For example, the speech rate of clutterers is described as extraordinarily fast and hasty. Normal speakers, however, are also capable of speaking very quickly but their speech is usually still intelligible and not classified as pathological. Fast speakers do not exhibit as many speech errors as clutterers (Sick 2000). In the literature, there are hardly indications of how many speech errors like elisions, substitutions etc. are “normal” and how many are typical for cluttering. For normal speakers Levelt (1991) describes 1 error per 1000 words.

The symptoms of cluttering also seem to be rather divers. Scherer (2003) denotes it individual and situational variability. In particular, in emotional situations the symptoms appear very clearly. However, the more clutterers concentrate on their own speech, the more intelligible their utterances become. The improvement of fluency by concentration is an important criterion to differentiate between clutterers and stutterers. Up to now there is no unique definition of cluttering.

As was pointed out by Caruso et al. (1988) a higher variability in stuttering reflects an inherently unstable situation, that is based on underlying neuromotor control problems. Up to now the etiology of cluttering is non-specific. Impairments of the central nervous system (Wirth 1994) were as well discussed as the disability of formulating speech (Braun 1999) or as the auditory processing deficit (Molt 1996). Following the conclusion of van Riper (1990) it is possible that not only stuttering but also cluttering is caused by tiny lags and disruptions in the timing of the speech production process. The reasons for cluttering, as well as the intra-individually inconstant pathology and experimental studies on stuttering, discussed above lead to the expectation of a higher articulatory variability for clutterers as compared to controls. Therefore, our hypothesis is that clutterers exhibit higher temporal and spatial token-to-token variability of articulatory gestures as compared to control speakers.

## **2 Methods**

Electromagnetic midsagittal articulography (EMMA) was used for recording the articulatory abilities of three clutterers and three controls. In comparison with the X-ray technology EMMA is a medically harmless experimental method which allows the measurement of two-dimensional articulatory movements simultaneously with a high spatial and temporal resolution.

In the field of speech pathology, EMMA has been used to investigate the fluency disorder stuttering (e.g. van Lieshout et al. 1993, Ward 1997, McClean

& Runyan 2000, McClean et al. 2004), as well as analyses of swallowing disorders (Kretschmer 1996) and motor disorders like dysarthria (Jaeger et al. 2000) and aphasia (Katz et al. 1990).

## **2.1 Subjects**

Six German native speakers took part in this experiment. There were two male and one female clutterers and two male and one female normal speakers. The clutterers were diagnosed during speech therapy and had no other speech disorders. The subjects were between 21 and 36 years old. In the following the subjects will be identified with initials that refer to the group (P = clutterer<sup>1</sup>, N = Normal speaker) and the gender (M = male, W = female) they belong to.

## **2.2 Material**

The subjects were instructed to produce two different tasks: repetitive CV-sequences and test words with carrier phrases. The first task was to repeat simple CV-sequences as fast and intelligible as possible within a 10 sec interval. The syllables consisted of /pa/, /ta/ and /ka/; and each syllable train was repeated 2 times.

For the second task foreign words were embedded in the carrier sentence „Sage ... bitte“ (say ... please) in order to elicit more natural utterances. The words consisted of a minimum of 5 and a maximum of 8 syllables. Foreign words were chosen because clutterers have special problems with longer words with complex syllable structures. All words contained the sequence /nali/ in which the final vowel /ɪ/ or /i/ was either stressed and lax as in /dimenziona<sup>1</sup>lɪstɪʃ/ or unstressed and tense as in /dimenzionali<sup>1</sup>zi:rən/ (bold letters indicate lexically stressed vowels). The vowel /a/ was realised as pre-stressed 1 (p1) or pre-stressed 2 (p2). Each of the 5 word pairs appeared in the p1- and the p2-condition.

Each word was repeated 10 times in randomised order. Overall each speaker produced 100 sentences (50 x [na<sup>1</sup>lɪ], p1, 50 x [nali], p2). Because during the recording session of speaker NM1 one sensor came off, only 35 repetitions of each word were used for further analysis.

The subjects were asked to speak as fast and at the same time as intelligible as possible. The aim of this instruction was to check how sensitive clutterers are to the perception of their own intelligibility.

---

<sup>1</sup> in German = “Polterer”

### 2.3 EMMA recording

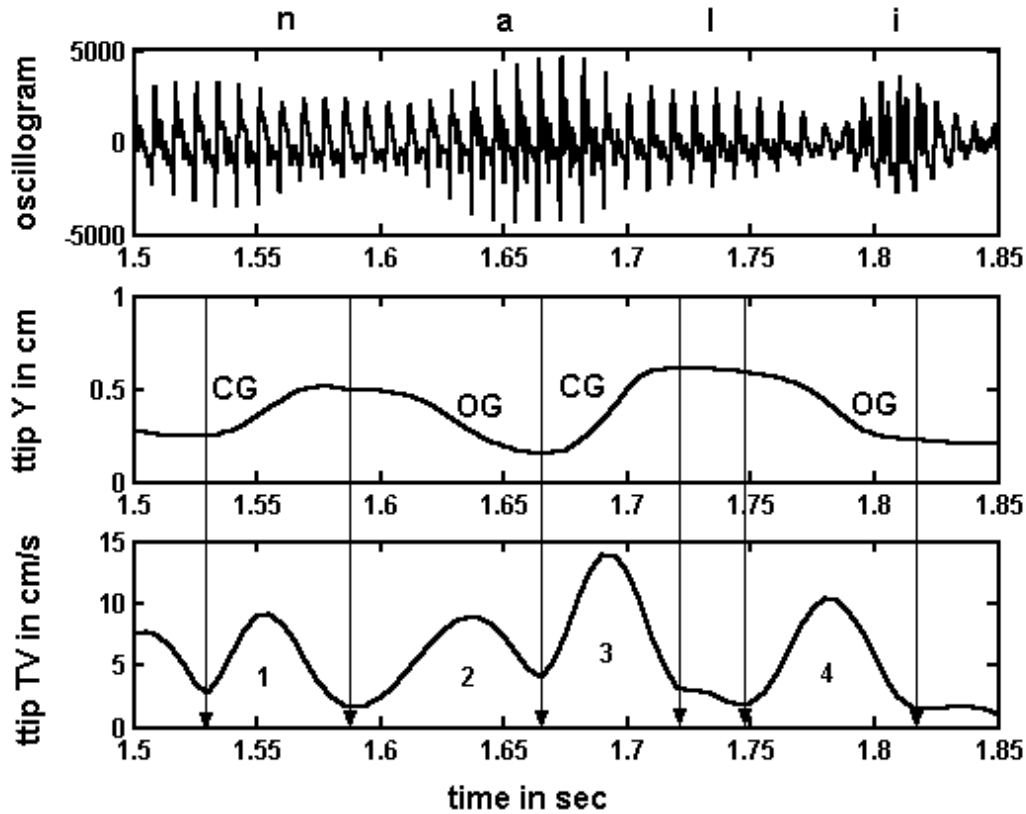
For the present study kinematic data were recorded by means of the Articulograph AG 100 (Carstens Medizinelektronik Göttingen 1992). During the recording the speaker wears a plexiglass helmet to which three transmitter coils are attached in an exactly defined distance to each other. They generate electromagnetic fields with three different frequencies. A sensor coil moving in this electromagnetic field induces a current, the strength of which is approximately inversely proportional to the distance between the transmitter and the receiver coils, with an accuracy of about a couple of tenth millimeters.

Sensors were attached midsagittally to the jaw, the lower lip and four tongue positions with a distance of approximately 1cm to each other. The tongue tip sensor was placed about 1cm behind the tongue tip. Reference sensors were fixed to the gums of the upper incisors and to the nasion. Two more coils were used to measure the occlusion plane. After recording, the resulting movement signals were rotated to the occlusion plane and the origin of the new coordinate system was located at the lower edge of the upper incisors. For data analysis, the movement signals were differentiated and all movement, velocity and acceleration signals were smoothed with a cutoff frequency of 20Hz. The tangential velocity signal was calculated from the x- and y-movement of the sensors.

### 2.4 Data segmentation

Figure 1 represents an EMMA display with tongue tip movements during the production of /nali/. The acoustic signal is shown in the upper panel, in the second panel the vertical movement of the tongue tip is represented, the third panel shows the tangential velocity signal. During the articulation of the sequence /nali/ 4 gestures were produced. The vertical lines in the figure indicate the first closing gesture (CG) towards /n/. The following gestures are the opening gesture (OG) towards the vowel /a/, the second closing gesture towards /l/ and the second opening gesture towards /i/.

By means of the tangential velocity signal (TV in cm/sec), shown in the third panel in figure 1, the start and end points of each gesture were defined by the left and right minimum, surrounding the velocity peak.



**Figure 1:** EMMA-display with the articulatory labelling criteria for the foreign words; panel 1: oscillogram, panel 2: vertical movement of the tongue tip sensor and panel 3: tangential velocity, vertical lines mark the 4 gestures

For the syllable repetition task one opening gesture towards /a/ and one closing gesture towards the stops /p/, /t/ or /k/ was measured for each syllable.

### 3 Results

For the analysis we focussed on the opening gestures from the nasal to the vowel /a/. All statistical calculations are based on the 10 repetitions of each foreign word.

#### 3.1 Qualitative analysis of movement patterns

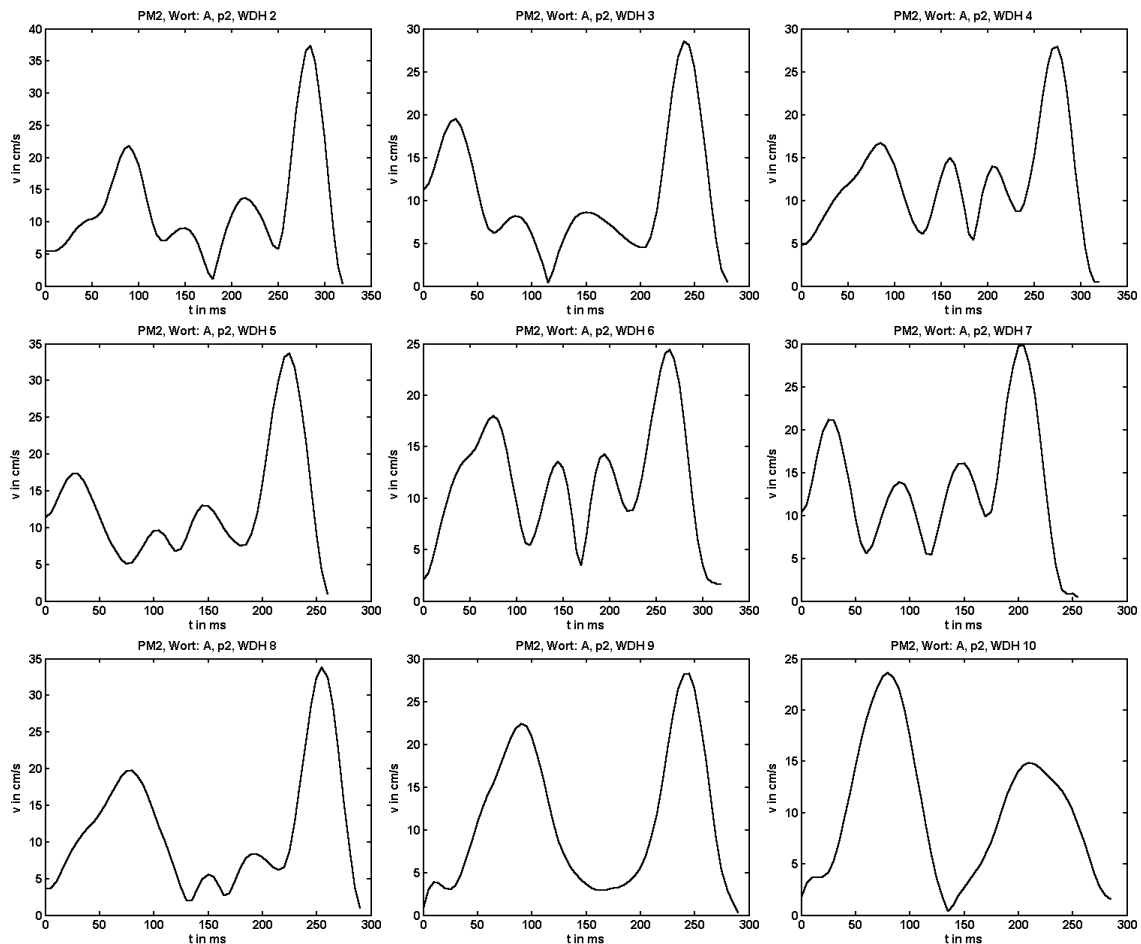
As can be seen from table 1 a great amount of the data of the clutterers could not be analysed due to reduced gestures and target undershoot. An extreme amount of reductions- presumably because of high speech rate and mumbled speech- was found for speaker PM2, for whom it was impossible to analyse nearly half of his data in the p1-condition. This reduction phenomenon is presented in figure 2.

In line with the question of the perception of their own intelligibility one clutterer (PM2) sometimes noticed his fast speech rate. Speaker PW3 hardly noticed elisions e.g. realising [kɒnstitsjona'listɪʃ] instead of [kɒnstɪtutsjona'listɪʃ].

**Table 1:** Frequency of data that could not be analysed of /nali/ in the p1- and p2-condition

condition	Controls			Clutterers		
	NM1	NM2	NW3	PM1	PM2	PW3
p1		1 (2%)	1 (2%)	7 (14%)	21 (42%)	
p2	2 (5.7%)	13 (26%)	1 (2%)	16 (32%)	11 (22%)	6 (12%)

The following figures demonstrate the reduction phenomenon, exemplified for the tongue tip tangential velocity signal during /nali/ of speaker PM2 for nine of the ten repetitions of the word “emotionalisieren” in the p2-condition.



**Figure 2:** Tangential velocities of the tongue tip movement of PM2 during the nine repetitions of /nali/ in „emotionalisieren“ in the p2-condition



In comparison to the segmented /nali/ shown in figure 1 it is noticeable that the second and third gesture always presents target undershoot, i.e. the two medial velocity peaks are much smaller than the first and the fourth, whereas in figure 1 the opposite is the case. In the last two displays these gestures were completely reduced.

### 3.2 Quantitative assessment of articulatory variability

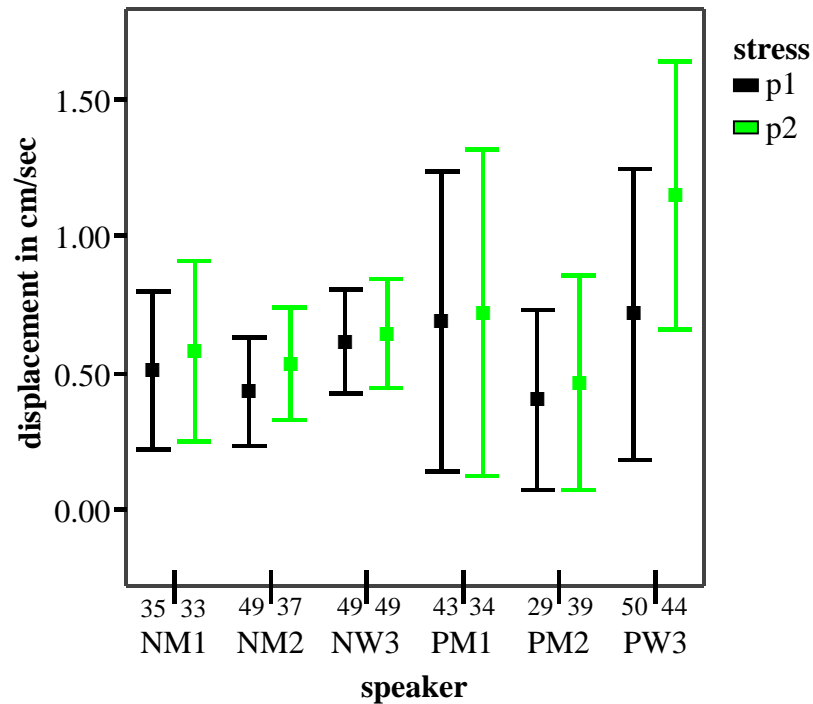
Based on means and standard deviations the coefficient of variation (CV) was calculated for the durations and amplitudes of the opening gesture (OG) from the nasal towards the vowel /a/ in /nali/ in the p1- and p2-condition [ $CV = (\text{standard deviation} * 100) / \text{mean}$ ]. Table 2 shows a higher coefficient of variation for the clutterers compared to the control group, e.g. the values of PM2 scatter with 19.7% and 22.6% over a much wider range from the mean than normal speakers with a maximum of 12.3%.

**Table 2:** Coefficient of variation (CV) in percent for the duration of the OG of /nali/ in the p1-/p2-condition

speaker	Controls			Clutterers		
	NM1	NM2	NW3	PM1	PM2	PW3
CV in %						
p1	9.5	12.3	9.9	19.0	19.7	13.9
p2	10.6	8.9	8.9	21.8	22.6	11.0

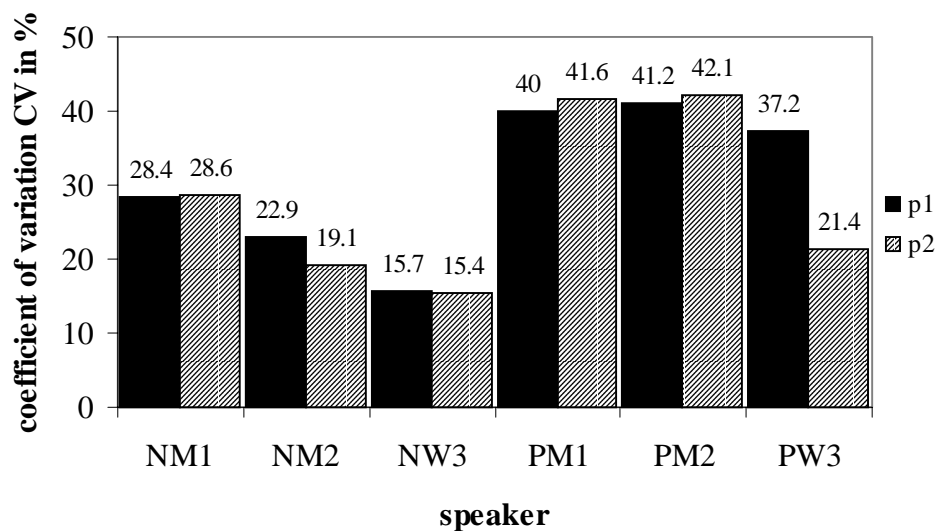
### 3.3 Displacements

The following figure also shows the frequency of the analysed foreign words. Speaker PM2 who produced only 29 instead of 50 analysable repetitions shows a smaller standard deviation than the other clutterers. However, the coefficient of variation (figure 4) indicates that his amplitudes vary to a similar amount as speaker PM1. Comparing the standard deviations of clutterers and controls in figure 3, it becomes clear that PM1 and PW3 produced the opening gesture with larger amplitudes which were also much more variable than the amplitudes of controls.



**Figure 3:** Standard deviation of the displacements of the OG of /nali/ in the p1/p2-condition

Figure 4 shows the coefficient of variation for the amplitudes of the OG of /nali/ in the p1-/p2-condition. As can be seen here, especially the results of the male clutterers are much more variable than speakers of the control group.



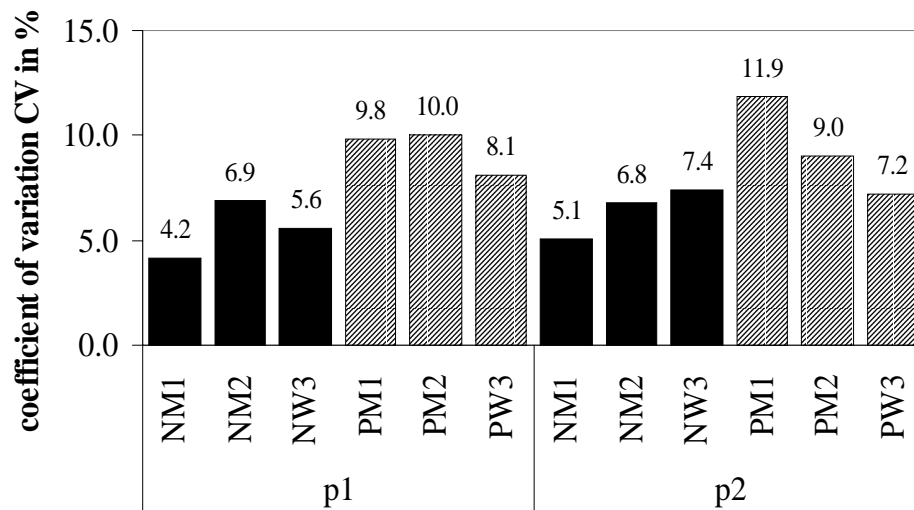
**Figure 4:** Coefficient of variation (CV) in percent for the displacements of the OG of /nali/ in the p1-/p2-condition

### 3.4 Syllable sequences

For the syllable repetition task, clutterers and controls did not differ significantly in the variability of the spatial and temporal domain.

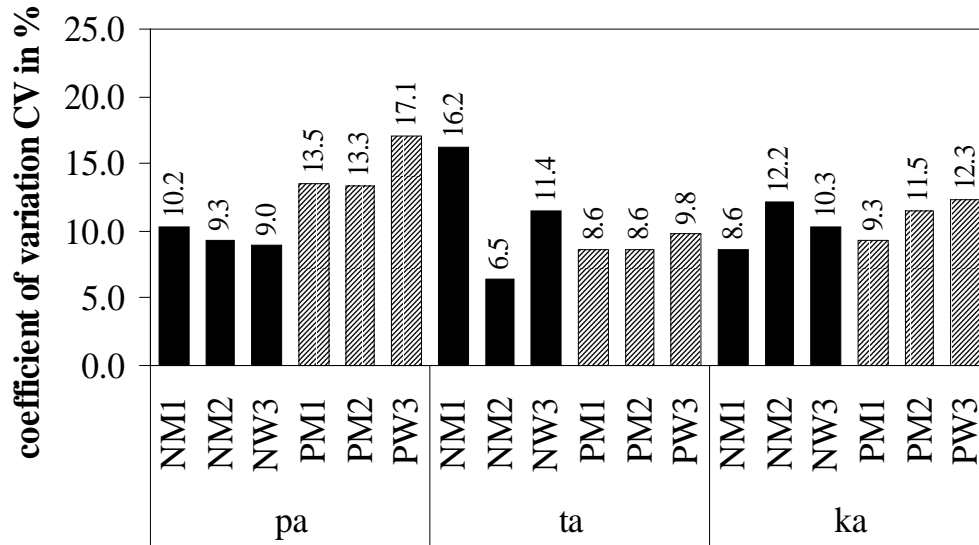
### 3.5 Quantitative assessment of acoustic variability

Based on acoustic labelling the durations of the whole /nali/ sequences were calculated. As can be seen in figure 5 the coefficient of variation is higher for the group of clutterers than for normal speakers. Only PW3 in the p2-condition shows a similar value as the female control speaker.



**Figure 5:** Coefficient of variation (CV) in percentage for the duration of the sequence /nali/

In agreement with the kinematic results, the syllable repetitions showed no prominent differences in the acoustic temporal variability between the two groups. Figures 6 illustrates the coefficient of variation, calculated for the syllable duration of /pa/, /ta/ and /ka/. Only in the production of /pa/ the durations are more variable.



**Figure 6:** Coefficient of variation (CV) in percentage for the duration of the syllable repetitions /pa/, /ta/ and /ka/

#### 4 Discussion

Remarkably high coefficients of variation were found for the spatial and temporal data of clutterers. Against expectations clutterers showed not only variable displacements but also very large articulatory movements. On the one hand the displacements were partly larger than in normal speakers and were interpreted as strategy for improvement of the intelligibility. However, on the other hand more instances of movement data had to be excluded for the clutterers because of gestural reduction and extreme target undershoot.

Concerning the temporal results, again a higher variability was found for the clutterers. However, in comparison to the displacements the durations varied to a smaller degree. This means, that clutterers exhibit more variation in the spatial than in the temporal domain.

More variable results were found only for the fluent utterances of the clutterers. Their speech production altogether showed obviously a higher variability due to speech errors, break-ups within words and incomplete articulation.

The results of the present study are consistent with the findings on stuttering (Zimmermann 1980, Caruso et al. 1988, van Lieshout et al. 1993, Gracco 1994, Jäncke 1994, Ward 1997). The increased variability in stutterers was interpreted as a consequence of the usage of different articulatory strategies in order to avoid disfluencies (Ward 1997). For achieving the instructed intelligibility in the present study clutterers also used different strategies. They varied the speech rate and the displacements in order to speak clearly. In

comparison to normal speakers they might be more uncertain about their own speech production because they are aware of their problems and often do not know how to improve their speech quality. They control and vary their speech again and again. For one person (PW3) a high number of break-ups within words was observed. This might indicate that she always expects speech errors and therefore she stops speaking although no errors occurred.

Zimmermann (1980) explains the instability with an imbalance in the afferent-efferent nerve impulses and the critical spatial-temporal relationship in the command execution. This is one reason why research and treatment of cluttering should focus more strongly on the processes in the central nervous system. Not only for producing intelligible speech but also for the execution of other gross and fine movements is it necessary that the motor system with the motor cortex, basal ganglia, cerebellum works without any disruptions. Before reaching the muscles the commands run through the cerebellum, where the movements are modulated. The basal ganglia determine parameters like displacement, direction, velocity and strength of the movements.

In order to understand the processes in the brain it would be a worthwhile investigation to use functional magnetic resonance imaging (fMRI) for clutterers. A very recent study with right-handed stutterers has shown a right-hemispheric hyperactivity (Neumann et al. 2005). After speech therapy they observed a more widespread activation in the speech and language areas on the left hemisphere.

Both, in the articulatory and acoustic analysis of the present study, a higher variability for clutterers could not be observed for the syllable repetitions. The speech production system seems to work more stable during the syllable repetition task than during the articulation of foreign words with complex consonant clusters. Stop-vowel sequences are the most frequent syllable structures in general and are also the first linguistic units during the speech development of human beings (for a recent overview see Ackermann et al. 2005). This could explain why clutterers produced the syllable repetition task with ease.

An additional explanation of the observed variability patterns is the effect of speech rate. According to Fitts' (1954) law an increase in speech rate causes an increase in movement variability. Ward (1997) who analysed utterances in a normal, fast and slow speech rate confirmed this relationship. In the present study a higher speech rate for clutterers was hypothesized, but this did not turn out to be true in all cases. Furthermore clutterers showed a high temporal variability which makes it difficult to discern the effect of speech rate and intrinsic variability. There must be other explanations for the higher coefficients of variation.

Another point of view concerns the linguistic complexity and is well documented in van Lieshout et al. (2004). Findings from Smith & Kleinow (2000) showed a greater movement variability for all speakers, not only for stutterers in syntactically more complex utterances. In terms of the present results it might be possible that the production of the foreign words were more difficult for the clutterers than for normal speakers. In a previous study van Lieshout (1995) explained the complexity of long words from a speech production view. According to him the brain has to prepare more production units for long words than for short words. While formulating, storing or executing these commands there are a lot of sources of errors. It should also be considered that complex words exhibit less common articulatory and prosodic patterns (van Lieshout 1995). This seems to be the case for the described test corpus. Because clutterers have problems in general with long and complex words further investigations should follow this question, for instance by means of fMRI.

For speech therapy is it advisable to practise a very clear speech with large articulatory movements. This strategy seems easier to realise for clutterers than the task to reduce speech rate.

## References

- Ackermann, H., Hertrich, I., Mathiak, K. (2005) Neurobiologische Grundlagen der Sprachlautwahrnehmung: Klinische und funktionell-bildgebende Befunde. *Sprache, Stimme, Gehör*, 29, (3): 112-120.
- Braun, O. (1999) *Sprachstörungen bei Kindern und Jugendlichen: Diagnostik, Therapie, Förderung*. Stuttgart: Kohlhammer.
- Carstens (1992) Articulograph AG100. *Elektromagnetisches Artikulations-Meßsystem*. Benutzerhandbuch. Carstens Medizinelektronik GmbH.
- Caruso, A. J., Abbs, J. H., Gracco, V. L. (1988) Kinematic analysis of multiple movement coordination during speech in stutterers. *Brain*, 111: 439-455.
- Fitts, P. M. (1954) The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, 47: 381-391.
- Gracco, V. L. (1994) Some organizational characteristics of speech movement control. *Journal of Speech and Hearing Research*, 37: 4-27.
- Jaeger, M., Hertrich, I., Stattrop, U., Schönle, P.-W., Ackermann, H. (2000) Speech disorders following severe traumatic brain injury: Kinematic analysis of syllable repetitions using electromagnetic articulography. *Folia Phoniatrica et Logopaedica*, 52: 187-196.
- Jäncke, L. (1994) Variability and duration of voice onset time and phonation in stuttering and nonstuttering adults. *Journal of Fluency Disorders*, 19: 21-37.

- Katz, W., Machetanz, J., Orth, U., Schönle, P.-W. (1990) A kinematic analysis of anticipatory coarticulation in the speech of anterior aphasia subjects using electromagnetic articulography. *Brain and Language*, 38, (4): 555-575.
- Kretschmer, I. M. (1996) *Untersuchungen zur Analyse von Sprech- und Schluckbewegungen mit Hilfe der elektromagnetischen Artikulographie*. dissertation, Universität Tübingen.
- Levelt, W. J. M. (1991) *Speaking*. 2. edition, Cambridge, Massachusetts: MIT Press.
- McClean, M. D., Runyan, C. M. (2000) Variations in the relative speeds of orofacial structures with stuttering severity. *Journal of Speech, Language and Hearing Research*, 43, (6): 1524-1531.
- McClean, M. D., Tasko, S. M., Runyan, C. M. (2004) Orofacial movements associated with fluent speech in persons who stutter. *Journal of Speech, Language, and Hearing Research*, 47, (2): 294-303.
- Molt, L. F. (1996) An examination of various aspects of auditory processing in clutterers. *Journal of Fluency Disorders*, 21: 215-225.
- Neumann, K., Preibisch, C., Euler, H. A., von Gudenberg, A. W., Lanfermann, H., Gall, V., Giraud, A.-L. (2005) Cortical plasticity associated with stuttering therapy. *Journal of Fluency Disorders*, 30: 23-39.
- Perkell, J.S., Nelson, W.L. (1985) Variability in production of the vowels /i/ and /a/. *Journal of the Acoustical Society of America*, 77: 1889-1895.
- Scherer, A. (2003) Poltern und Stottern als Ausdruck der emotionalen Befindlichkeit: ein Erfahrungsbericht aus sprachtherapeutischer Sicht. *Sprache, Stimme, Gehör*, 27: 88-91.
- Sick, U. (2000) Spontansprache bei Poltern. *Forum Logopädie*, 4, (14): 7-16.
- Smith, A., Kleinow, J. (2000) Kinematic correlates of speaking rate changes in stuttering and normally fluent adults. *Journal of Speech, Language and Hearing Research*, 43, (2): 521-36.
- van Lieshout, P. H. H. M. (1995) *Motor planning and articulation in fluent speech of stutterers and nonstutterers*. dissertation, Nijmegen Institute for Cognition and Information.
- van Lieshout, P. H. H. M., Alfonso, P. J., Hulstijn, W., Peters, H. F. M. (1993) Electromagnetic articulography (EMA) in stuttering research. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 31: 215-224.
- van Lieshout, P. H. H. M., Hulstijn, W., Peters, H. F. M. (2004) Searching for the weak link in the speech production chain of people who stutter: A motor skill approach. In: B. Maassen, R. Kent, H. Peters, P. H. H. M. van Lieshout & W. Hulstijn (eds.) *Speech Motor Control in Normal and Disordered Speech*. Oxford: Oxford University Press: 313-355.
- van Riper, C. (1990) Final thoughts about stuttering. *Journal of Fluency Disorders*, 15: 317-318.

- Ward, D. (1997) Intrinsic and extrinsic timing in stutterers' speech: Data and implications. *Language and Speech*, 40, (3): 289-310.
- Wirth, G. (1994): *Sprachstörungen, Sprechstörungen, Kindliche Hörstörungen*. 4. edition, Köln: Deutscher Ärzte-Verlag.
- Zimmermann, G. (1980) Stuttering: A disorder of movement. *Journal of Speech and Hearing Research*, 23: 122-136.



- Ward, D. (1997) Intrinsic and extrinsic timing in stutterers' speech: Data and implications. *Language and Speech*, 40, (3): 289-310.
- Wirth, G. (1994): *Sprachstörungen, Sprechstörungen, Kindliche Hörstörungen*. 4. edition, Köln: Deutscher Ärzte-Verlag.
- Zimmermann, G. (1980) Stuttering: A disorder of movement. *Journal of Speech and Hearing Research*, 23: 122-136.

## Die motorische Funktionsprüfung bei oralen Tumoren

(Testing of motor functions in cases of intra-oral cancer)

Sabine Koppetsch

Zentrum für Allgemeine Sprachwissenschaft (ZAS), Humboldt Universität zu Berlin (HU)

Cancer in the oral cavity is one of the most common carcinomas worldwide. Besides various therapeutic treatments, surgical resection of the tumor plays the most important role. The principles and the extend of surgical treatment depend on state and histologic type of the malignoma, localization of the carcinoma and infiltration of adjacent structures. The resulting loss of anatomic structures such as parts of the jaw, the tongue and the floor of the mouth leads to various forms of functional oral disorders. Mastication, swallowing, mandibular motor functions, speech production and sense of taste are affected as well as the aesthetic appearance. The therapy of intraoral malignoma leads to a deterioration of postoperative quality of life. Besides efforts to ensure the survival of the patient, therapy is focusing more and more on the endeavour to simultaneously maintain the quality of life of tumor patients. That requires us to plan medical care with the aim to achieve a maximal retention of function. On the other hand, specialised speech therapy procedures are important post-operatively in order to train functional and articulatory skills (Städtler, 1989). This is, however, only possible if the extent of post-operative functional changes is known. In order to investigate the patient's oral skills, a motoric questionnaire, which makes such a specific, systematic test possible, has been developed at Zentrum für Allgemeine Sprachwissenschaft.

Maligne Tumore der Mundhöhle und der Zunge stehen weltweit an sechster Stelle aller Krebserkrankungen (Becker, 1997; Werner, 2000). Neben einer Reihe therapeutischer Behandlungsmöglichkeiten nimmt die chirurgische Resektion der Tumore eine wichtige Stellung ein. Auf Grund der häufig sehr ausgedehnten Befunde führt der resektionsbedingte Verlust anatomischer Strukturen im Bereich des Kiefers, des Mundbodens oder der Zunge oft zu Störungen aller oraler Funktionen und Funktionsabläufe. Bei vielen Patienten sind das Kauvermögen, das Schlucken, das Sprechen, die Sensibilität, die Geschmacksempfindung, aber auch die Ästhetik im Kopf- und Halsbereich betroffen (Schröder, 1985; Grimm, 1990; Panje & Morris, 1995; Reuther & Bill, 1998; Lenarz & Lesinski-Schiedat, 2001). Orale Tumore haben daher einen massiven Einfluss auf die postoperative Lebensqualität der betroffenen Patienten. Neben dem Bemühen das Überleben der Patienten zu sichern, nimmt daher das Bestreben die Lebenssituation der Patienten zu verbessern einen zunehmend wichtigeren Platz ein. Hierzu gehört zum einen, das medizinische Vorgehen so zu planen, dass ein maximaler Funktionserhalt angestrebt wird. Zum anderen ist postoperativ das gezielte sprachtherapeutische

Vorgehen wichtig um funktionelle und artikulatorische Fähigkeiten gezielt schulen zu können (Städtler, 1989). Dies ist jedoch nur möglich, wenn die postoperativen funktionellen Veränderungen bekannt sind. Um eine Prüfung der oralen Fähigkeiten zu ermöglichen, wurde am Zentrum für Allgemeine Sprachwissenschaft ein Motorischer Bogen entwickelt, der eine gezielte und systematische Überprüfung ermöglicht.

## 1 Einleitung

Um die sich verändernden oralen Fähigkeiten von Tumorpatienten prä- und postoperativ erfassen zu können, wurde am Zentrum für Allgemeine Sprachwissenschaft (ZAS) ein Motorischer Bogen entwickelt, der es ermöglicht, die für das Kauen, Schlucken und Sprechen relevanten oralen Bewegungen zu erfassen (Koppetsch 2004). Sein langfristiger Einsatz ermöglichte es, einen Überblick über die prä- und postoperativen motorischen oralen Fähigkeiten der Patienten zu erhalten und Veränderungen erfassen zu können. Der Bogen wurde in der Mund-Kiefer-Gesichtschirurgie des Universitätsklinikums Rudolph Virchow, in der Mund-Kiefer-Gesichtschirurgie des Universitätsklinikums Benjamin Franklin und an der Klinik und Poliklinik für Mund-, Kiefer- und Gesichtschirurgie, Klinikum rechts der Isar der TU München im Rahmen der Patientenversorgung getestet.

## 2 Methodik

### 2.1 Patienten

Die motorische Überprüfung erfolgte bei Patienten, bei denen ein Tumor (Plattenepithelkarzinom - PECA, Ameloblastom) im Bereich des Mundbodens (mb) oder der Zunge (z) diagnostiziert wurde (Tabelle 1).

Die medizinische Versorgung der Patienten war sehr unterschiedlich und abhängig von Tumorgröße und Infiltration in das umliegende Gewebe. Es erfolgte entweder eine Tumoresektion (res), eine Deckung des entstandenen Gewebedefektes (rek) und/oder eine Radio- (ra) oder Chemotherapie (ch).

Von anfänglich 25 Patienten konnten 19 Patienten durchgängig an der Überprüfung teilnehmen. Sechs Patienten schieden während der laufenden Studie aus gesundheitlichen Gründen aus.

Tabelle 1: Patientenübersicht

	Patient	Alter	Diagnose	Lokalisation	ch	ra	res	rek
1.	mb1m	45	PECA	mb	-	-	x	x
2.	mb2m	48	PECA	mb	x	x	x	-
3.	mb3m	41	PECA	mb	-	-	x	x
4.	mb4m	59	PECA	mb	x	x	x	x
5.	mb5f	73	Ameloblastom	mb	-	-	x	x
6.	mb6m	33	PECA	mb	x	x	-	-
7.	mb7f	44	PECA	mb	x	x	x	x
8.	mb8m	60	PECA	mb	x	x	-	-
9.	mb9f	56	PECA	mb	x	x	x	x
10.	mb10f	69	PECA	mb	x	x	x	-
11.	z1f	37	PECA	z	x	x	x	x
12.	z2m	60	PECA	z	-	-	x	x
13.	z3m	49	PECA	z	-	-	x	x
14.	z4m	34	PECA	z	-	x	x	x
15.	z5m	54	PECA	z	x	x	-	-
16.	z6m	56	PECA	z	x	x	-	-
17.	z7m	30	PECA	z	x	x	x	x
18.	z8m	64	PECA	z	-	-	x	x
19.	z9m	63	PECA	z	x	x	-	-

### 2.2 Korpus







Der motorische Test beschränkte sich zunächst lediglich auf die Überprüfung oraler Abläufe, so dass nur die Bewegungen der Zunge, der Lippen und des Unterkiefers überprüft wurden. Mit Fortschreiten der Studie wurde jedoch deutlich, dass die motorischen Beeinträchtigungen der Patienten sich auf Grund von ausgedehnten Tumorbefunden nicht nur auf den oralen Bereich begrenzten, sondern wesentlich umfassender waren. Da festgestellt werden konnte, dass eingeschränkte Kopf- und Halsbewegungen auch Einfluss auf die Bewegungen des Unterkiefers haben und somit die Lautrealisation beeinflussen, wurde der Motorische Bogen während der laufenden Studie kontinuierlich weiter entwickelt, so dass auch Bewegungen im Kopf-, Schulter- und Halsbereich erfasst wurden (Abb. 1). Um ein systematisches Vorgehen zu ermöglichen, wurde der motorische Untersuchungsbogen in verschiedene Abschnitte unterteilt:

- Beurteilung der Artikulationsorgane,
- Beurteilung der Kiefer-, Lippen- und Zungenbewegungen,
- Beurteilung der Kopfbewegungen,
- Beurteilung der Schulterbewegungen.

Weitere Beobachtungen oder Besonderheiten können im Bogen ergänzend festgehalten werden.

# Untersuchungsbogen - Motorik

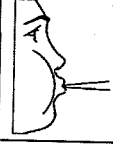






Befunderhebung bei Glossektomie - Patienten - Blatt 1

Name				präoperativ	o	Datum
				postoperativ	o	
Anomalien der Sprechorgane	o ja	o nein	sonstiges:			
Zahnstellung / Biss						
Zunge						
Lippen						
Nase						
Kiefer						
<b>KIEFER</b>						
Ruhestellung						
Seitwärtsbewegung li. / re.						
Kreisen des Unterkiefers						
Stellung beim Schlucken						
sonstige Anmerkungen:						
<b>LIPPEN</b>						
Aussehen						
Ruhestellung						
	Lippen weit öffnen		Lippen leicht öffnen		Lippen pressen	
normal		1		1		1
beeinträchtigt		2		2		2
nicht möglich		3		3		3
asymmetrisch		4		4		4
verlangsamt		5		5		5
schmerzhaft		6		6		6
verkrampft		7		7		7
sonst. Anmerkungen:						
	Mundwinkel hoch ziehen		Mundwinkel runter ziehen		Lippen spitzen	
normal		1		1		1
beeinträchtigt		2		2		2
nicht möglich		3		3		3
asymmetrisch		4		4		4
verlangsamt		5		5		5
schmerzhaft		6		6		6
verkrampft		7		7		7
sonst. Anmerkungen:						


<sup>1</sup> alle Abbildungen nach Müller-Ellermann (o.J.)


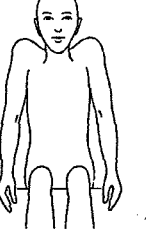
# Untersuchungsbogen - Motorik

Befunderhebung bei Glossektomie - Patienten - Blatt 2

<b>LIPPEN</b>				
	pusten			
	saugen			
	schmatzen			
	schnalzen			
Mundwinkel nach links ziehen		Mundwinkel nach rechts ziehen		
	1	normal	1	
	2	beeinträchtigt	2	
	3	nicht möglich	3	
	4	asymmetrisch	4	
	5	verlangsamt	5	
	6	schmerzhaft	6	
	7	verkrampft	7	
sonst. Anmerkungen:				
Unterlippe über die Oberlippe		Oberlippe über die Unterlippe		
	1	normal	1	
	2	beeinträchtigt	2	
	3	nicht möglich	3	
	4	asymmetrisch	4	
	5	verlangsamt	5	
	6	schmerzhaft	6	
	7	verkrampft	7	
sonst. Anmerkungen:				
<b>ZUNGE</b>				
Aussehen				
Form				
Ruhestellung (bei Mundöffnung)				
Zungenspitze zum Mundwinkel - nach links		Zungenspitze zum Mundwinkel - nach rechts		
	1	normal	1	
	2	beeinträchtigt	2	
	3	nicht möglich	3	
	4	asymmetrisch	4	
	5	verlangsamt	5	
	6	schmerzhaft	6	
	7	verkrampft	7	
Zungenspitze zur Nase/zum Kinn				
Zunge weit rausstrecken				
Lippen lecken oben/unten				
Kreisen im Mund (Zahndamm)				
Kreisen auf den Lippen				
sonst. Anmerkungen:				

# Untersuchungsbogen - Motorik Befunderhebung bei Glossektomie - Patienten - Blatt 3

ZUNGE			
Zunge gegen die Wange drücken - links		Zunge gegen die Wange drücken - rechts	
	1	normal	1
	2	beeinträchtigt	2
	3	nicht möglich	3
	4	asymmetrisch	4
	5	verlangsamt	5
	6	schmerzhaft	6
	7	verkrampft	7
Heben der Zunge gegen einen Widerstand			
sonst. Anmerkungen:			

KOPF / SCHULTERN			
Kopf drehen - nach rechts		Kopf drehen - nach links	
	1	normal	1
	2	beeinträchtigt	2
	3	nicht möglich	3
	4	asymmetrisch	4
	5	verlangsamt	5
	6	schmerzhaft	6
	7	verkrampft	7
sonst. Anmerkungen:			
Kopf pendeln			
beide Schultern heben		beide Schultern senken	
	1	normal	1
	2	beeinträchtigt	2
	3	nicht möglich	3
	4	asymmetrisch	4
	5	verlangsamt	5
	6	schmerzhaft	6
	7	verkrampft	7
Schultern vor			
Schultern zurück			
Schultern kreisen			
sonst. Anmerkungen:			
Allgemeine Bemerkungen:			

## 2.3 Datenerhebung

Die motorischen Überprüfungen wurden möglichst einheitlich durchgeführt und die Untersuchungsergebnisse im Bogen skizziert:

- Überprüfung 1 – präoperativ / zwei Wochen, spätestens ein Tag vor der OP
- Überprüfung 2 – postoperativ / eine Woche bis einen Monat nach der OP
- Überprüfung 3 – postoperativ / drei bis fünf Monate nach der OP
- Überprüfung 4 – postoperativ / sechs bis acht Monate nach der OP.

Den Patienten standen bei Bedarf für jede Aufgabe mehrere Versuche zur Verfügung, da so eine genauere Beurteilung erfolgen konnte. Auf eine zeitliche Vorgabe wurde verzichtet, da sie für die Beurteilung nicht relevant erschien.

## 3 Resultat

Auf Grund der systematischen Datenerhebungen konnten bei den Patienten die funktionellen oralen Fähigkeiten gezielt erfasst und die prä- und postoperativen bzw. prä- und posttherapeutischen Ergebnisse verglichen werden. Es war somit möglich, artikulatorische Veränderungen besser verstehen und das weitere sprachtherapeutische Vorgehen gezielt planen zu können.

### 3.1 Patienten mit einem Tumor im Bereich des Mundbodens

Bei allen Patienten die an einem Tumor im Bereich des Mundbodens erkrankt waren und operativ behandelt wurden, veränderte sich postoperativ die Zungen-, Lippen- und Kieferbeweglichkeit, und somit auch das Kauen, das Schlucken und das Sprechen. Der Schwerpunkt lag dabei im Zeitraum der zweiten und dritten Aufnahme. Dieser Zeitpunkt deckte sich häufig mit dem Beginn der Radiochemotherapie, und führte meist zu einer massiven Verschlechterung aller oraler Abläufe, da Entzündungen der Mundschleimhaut (Mucositis) und eine Verhärtung der Zungenmuskulatur (bedingt durch den gestörten Lymphabfluss), jede orale Bewegung erschwerte.

Auffallend war z.B., dass die Zunge häufig nicht mehr symmetrisch in der Ruhelage gehalten werden konnte (Abb. 1a). Auch das Anheben der Zungenspitze (Abb. 1b) und das Kreisen mit der Zunge im Mundinnenraum gelang oft nur mit Mühe. Das sich die motorischen Einschränkungen auch artikulatorisch äußerten, soll an Hand des Beispiels von Patient mb3m verdeutlicht werden, bei dem nach der Tumoresektion im Bereich des Mundbodens eine Defektdeckung mittels Zungenlappen erfolgte.

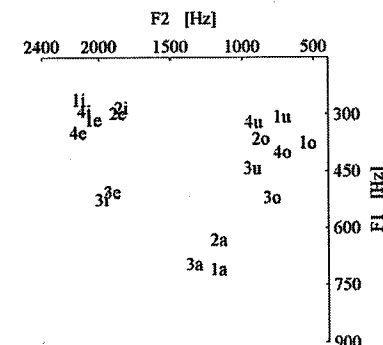


**Abb. 1:** PECA im Bereich des Mundbodens (postoperativ)  
– a. Zungenkörper in Ruhelage; b. maximales Anheben der Zunge (sichtbar der Zungenlappen)

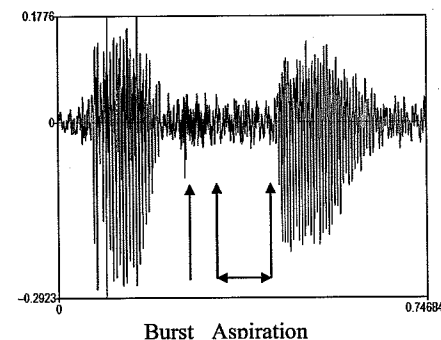
Mittels motorischer Überprüfung wurde deutlich, dass es schwierig war, die Zunge symmetrisch zu heben und vorwärts zu bewegen. Auf Grund dieser Beeinträchtigung konnte während der Lautrealisation z.B. nur mühsam eine Enge (Konstriktion) zwischen Zunge und Gaumen gebildet werden. Diese ist jedoch für die Vokalrealisation der hohen Vokale und auch für die Bildung alveolarer Frikative wichtig. Die postoperativ zentralisiertere Realisation der hohen Vokale /i/ und /u/ (Abb. 2), der Plosive /d, t/ (Abb. 3) und der Frikative /s/ und /ʃ/ ist daher leicht nachzuvollziehen.

Zusätzlich wurden veränderte Kieferbewegungen festgestellt. Besonders die Öffnung des Mundes gelang nur verlangsamt und mitunter leicht asymmetrisch. Allein diese Feststellung ließ die Vermutung zu, dass auch die Realisation des Vokals /a/ verändert sein würde. Sowohl mittels akustischer Analyse (Formantwerte F1 und F2) als auch ohrenphonetisch konnte eine veränderte Realisation festgestellt werden.

Auf Grund der motorischen Überprüfung konnten die Ursachen für die veränderte Lautrealisation klar festgestellt werden. Es war dem Sprachtherapeuten daher möglich, notwendige Übungen für eine weiterführende Funktions- und Artikulationsschulung in einer gezielten Therapieplanung einzubauen.



**Abb. 2:** Formantkarte (F1, F2) der isoliert gebildeten Vokale und ihre Veränderung abhängig vom Zeitpunkt der Aufnahme (1=Aufnahme 1, 2=Aufnahme 2, 3=Aufnahme 3), 4=Aufnahme 4



**Abb. 3:** /iti/ zum Zeitpunkt der dritten Aufnahme – Verschlusslösung mit geringer Intensität und Aspiration nach der Verschlusslösung

### 3.2 Patienten mit einem Tumor im Bereich der Zunge

Bei Patienten, die an einem PECA im Bereich der Zunge (Zungenrand, Zungengrund) erkrankt waren und operativ behandelt wurden, konnten besonders veränderte Zungenbewegungen festgestellt werden. Die Kieferbewegungen waren kaum beeinträchtigt. Die Bewertung des Sprechens, Kauens und Schluckens erfolgte postoperativ wesentlich schlechter als bei den Patienten, die an einem Tumor im Bereich des Mundbodens erkrankt waren.

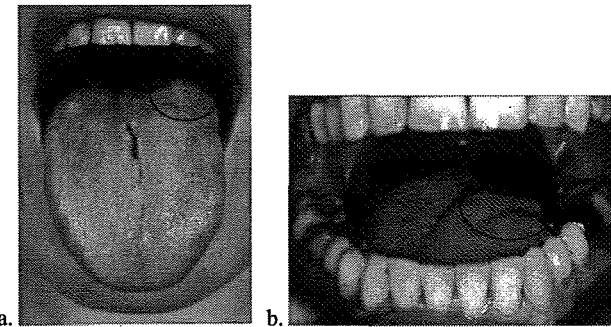


Besonders auffallend war, dass auf Grund der Tumorlokalisation und dem damit verbundenen Resektionsgebiet postoperativ häufig eine veränderte Zungenform festgestellt werden konnte, die Ursache für die veränderten Funktionsabläufe zu sein schien (Abb. 4b, 5a, 6a).

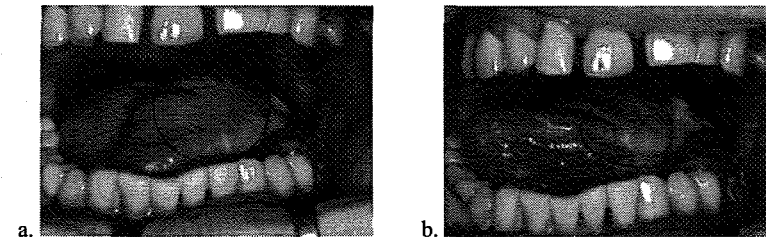
Bei allen postoperativen motorischen Überprüfungen wurde deutlich, dass besonders die Vorwärtsbewegungen und Rückverlagerungen des Zungenkörpers erschwert waren. Bei der Vorwärtsbewegung konnten mitunter z.B. die Alveolen (Zahndamm) oder die Lippen nur erschwert mit der Zungenspitze erreicht werden (Abb. 5b, 6b). Besonders für die Realisation der alveolaren Laute /d, t, s, z, l/ ist jedoch eine Vorwärtsbewegung bei gleichzeitigem Anheben der Zungenspitze notwendig. Die Ursache für die beeinträchtigte Lautrealisation der velaren Laute /x, g, k/ schien die, mittels Motorischem Bogen festgestellte, eingeschränkte Rückverlagerung des Zungenkörpers zu sein.

Die größten Veränderungen sowohl motorisch als auch im Rahmen der Lautrealisation konnten bei Patient z4m (Abb. 6) festgestellt werden, bei dem im Rahmen einer zweiten Operation eine Nachresektion erfolgte und die Zungenspitze entfernt werden musste. Die motorische Überprüfung zeigte, dass der Patient mit der Zunge weder die Zähne noch die Alveolen erreichen konnte, die Lippen konnten nicht abgeleckt werden, die Wangen sowohl links als auch rechts konnten nicht berührt werden (in der Mundhöhle), der gesamte Zungenkörper konnte nur noch gering und asymmetrisch angehoben werden und auch die Mundöffnung war erschwert und schmerzhaft. Diese funktionellen Einschränkungen ließen die Vermutung zu, dass die Lautrealisation gravierend verändert sein würde, und nunmehr nicht nur die alveolaren und velaren Laute betroffen sein würden, sondern die gesamte Lautrealisation. Sowohl ohrenphonetisch als auch mittels akustischer Auswertung bestätigte sich diese Vermutung.

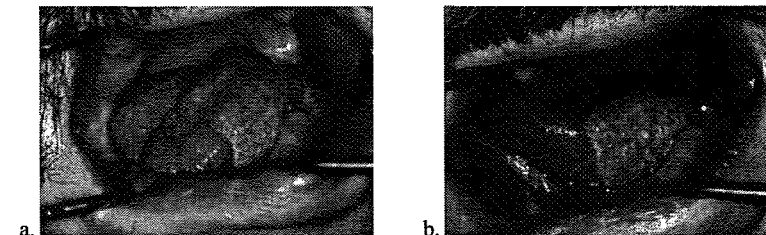
Die unterschiedlichen funktionellen und artikulatorischen Leistungen der Patienten zeigen, dass postoperativ kein allgemeingültiges therapeutisches Vorgehen zum Einsatz kommen kann. Nur eine gezielte Prüfung der funktionellen Fähigkeiten ermöglicht eine individuelle, optimale und gezielte postoperative Förderung des Patienten durch den Sprachtherapeuten.



**Abb. 4:** PECA im Bereich des Zungenrandes (Defektdeckung durch ein Lappentransplantat); a: Zunge präoperativ mit Tumormarkierung; b: Zunge in Ruhelage postoperativ mit Markierung des Resektionsgebietes



**Abb. 5:** PECA im Bereich des Zungenrandes (Defektdeckung durch ein Lappentransplantat); a: Zunge postoperativ in Ruhelage; b: Zunge postoperativ bei dem Versuch den Zungenkörper anzuheben



**Abb. 6:** PECA im Bereich des Zungenrandes (Defektdeckung durch einen Platysmalappen); a: Zunge postoperativ in Ruhelage; b: Zunge postoperativ bei dem Versuch den Zungenkörper anzuheben

### 3.3 Artikulatorische und motorische Veränderungen nach Abschluss der kombinierten Radiochemotherapie

Alle Patienten, die an einem Tumor des Zungenrandes erkrankt waren und mittels Radiochemotherapie behandelt wurden beschrieben ihre Sprechfähigkeit vor Beginn der Radiochemotherapie als sehr gut, unmittelbar nach dem Abschluss jedoch als unbefriedigend. Auch mittels motorischer Überprüfung konnte eine massive Verschlechterung aller oralen Abläufe festgestellt werden. Die Ursache liegt in der häufig vorkommenden bestrahlungsbedingten Mucositis und krankhaftem vermehrtem zähflüssigem Speichelfluss, die jede orale Bewegung und somit auch die Lautrealisation erschwerte. Folgende Einschränkung konnten besonders häufig festgestellt werden:

- der Mund konnte nicht mehr oder nur noch mühsam geöffnet werden,
- jede Lageveränderung der Zunge wurde als schmerzhaft beschrieben und möglichst vermieden,
- es war den Patienten nicht mehr möglich, die Zähne oder Alveolen mit der Zungenspitze zu erreichen,
- das Anheben des Zungenrückens gelang nur erschwert,
- die Rückverlagerung des Zungenkörpers gelang nur sehr langsam.

Sowohl mittels akustischer Auswertungen als auch ohrenphonetisch wurde postoperativ eine massiv veränderte Lautrealisation festgestellt. So wurde z.B.:

- eine Zentralisation der Formantwerte bei der Vokalrealisation (Abb. 7),
- starke Aspirationen bei der Realisation der velaren Plosive,
- fehlende Verschlusslösungen bei der Realisation der Plosive (Abb. 8) festgestellt.

Ohrenphonetisch konnte die Lautrealisation als verwaschen, undeutlich und schwer verständlich beschrieben werden.

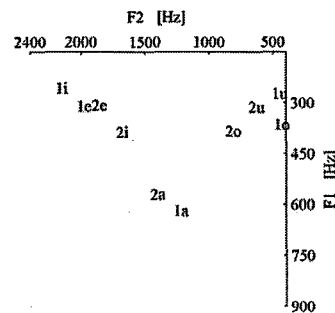


Abb. 7: Formantkarte (F1, F2) der isoliert gebildeten Vokale und ihre Veränderung abhängig vom Zeitpunkt der Aufnahme (1 = Aufnahme 1, 2 = Aufnahme 2)

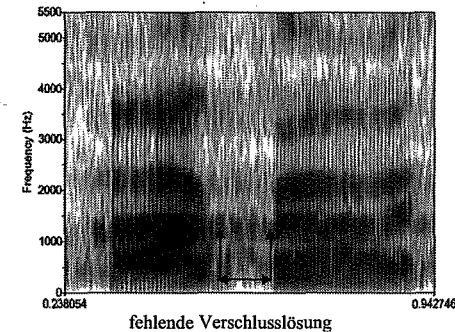


Abb. 8: /ata/ zum Zeitpunkt der zweiten Aufnahme - /t/ gesprochen als stimmhafter Plosiv /d/, kein Burst im gekennzeichneten Bereich sichtbar und keine Verschlusslösung hörbar

## 4 Zusammenfassung / Diskussion

Obwohl die Auswirkungen oraler Tumore von verschiedenen Fachdisziplinen untersucht werden (Forschungen in den Bereichen HNO, MKG, Sprachtherapie etc.) wird den objektivierenden Untersuchungen hinsichtlich der oralen Motorik und der Sprechweise nach operativen Eingriffen häufig noch immer zu wenig Beachtung geschenkt. Doch gerade das Wissen über postoperativ veränderte Funktionsabläufe ermöglicht zum einen eine gezielte Planung des medizinischen Vorgehens um neben dem Überleben auch den maximalen Erhalt der oralen Funktionalität zu berücksichtigen. Zum anderen ist eine gezielte postoperative sprachtherapeutische Betreuung nur möglich, wenn die Ursachen der artikulatorischen Beeinträchtigungen bekannt sind.

Mit Hilfe der vorliegenden Untersuchungen konnte festgestellt werden, dass die oralen funktionellen und artikulatorischen Veränderungen der Patienten postoperativ sehr vielfältig und unterschiedlich ausgeprägt waren. Abhängig von Lokalisation und Größe des Tumors, von chirurgischer und therapeutischer Behandlung variierten die postoperativen funktionellen und artikulatorischen Folgen. Diese spiegelten sich sowohl in einer veränderten Lautrealisation wider, als auch in veränderten Fähigkeiten beim Kauen und Schlucken. Interessant war nunmehr ein Vergleich der Ergebnisse der motorischen Überprüfung mit den fehlgebildeten Lauten:

- War z.B. die Vor- und Rückverlagerung des Zungenkörpers erschwert, konnten besonders Probleme bei der Realisierung von /l, s, z, d, t/ und /g, k, x/ festgestellt werden.
- Bei einer Resektion der Zungenspitze war besonders die Realisation der alveolar gebildeten Laute /l, s, z/ beeinträchtigt bzw. nicht möglich, da keine Enge bzw. kein Verschluss gebildet werden konnte.
- Bei einer Resektion im Bereich des Mundbodens war sowohl die Öffnung des Mundes und damit die Realisation des Vokals /a/ erschwert als auch das Anheben der Zunge beeinträchtigt, das besonders eine veränderte Realisation der Vokale /i/ und /e/ zur Folge hatte.
- Mit Beginn der Radiochemotherapie kam es bei allen Patienten zu einer weiteren Verschlechterung der oralen Motorik (funktionellen Defiziten) und somit auch zu einer weiteren Verschlechterung der Sprechfähigkeit.

Diese klaren Resultate konnten systematisch und sehr detailliert mittels Motorischem Untersuchungsbogen erfasst werden. Obwohl der Einsatz des Untersuchungsbogens präoperativ keine Prognosen über den Umfang und die Schwere der zu erwartenden artikulatorischen Beeinträchtigung ermöglichte, konnten artikulatorisch veränderte Resultate postoperativ besser erklärt werden, und unterstützen ein gezieltes therapeutisches Vorgehen. Sowohl sprachtherapeutische Übungen als auch ein orales Funktionstraining konnten wesentlich gezielter erstellt werden, so dass empfohlen werden kann, vor jeder sprachtherapeutischen Übungsbehandlung zunächst eine Überprüfung der oralen Motorik durchzuführen.

### Danksagung

Die vorliegende Studie konnte auf Grund der Zusammenarbeit der MKG des Universitätsklinikums Benjamin Franklin - Berlin, des Universitätsklinikums Campus Virchow - Berlin, der MKG des Klinikums rechts der Isar der TU München und des Zentrums für Allgemeine Sprachwissenschaft - Berlin erarbeitet werden. Die Studie wurde durch ein Stipendium der Nafög und durch die DFG finanziert. Ein besonderer Dank gilt den Patienten, die diese Studie erst ermöglichen.

### Literatur

- Becker, N. & Wahrendorf, J. (1997) *Krebsatlas der Bundesrepublik Deutschland 1981-1990*. Berlin, Heidelberg: Springer-Verlag.

- Grimm, G. (1990) Geschwülste im Mund- und Kieferbereich. In: N. Schwenzer & G. Grimm eds. *Zahn-Mund-Kiefer-Heilkunde (Lehrbuch zur Aus- und Fortbildung in 5 Bänden)*: 253-356. Stuttgart: Thieme Verlag.
- Koppetsch, S. (2004) *Orofaziale Rekonstruktionen nach Mundboden- und Zungenteilresektion*, Berlin: Wissenschaftsverlag Berlin.
- Lenarz, T. & Lesinski-Schiedat, A. (2001) Ethische Probleme bei der Therapie von Kopf-Hals-Tumoren. *Medizinische Hochschule Hannover, Klinik und Poliklinik für Hals-, Nasen-, Ohrenheilkunde*: online. [http://www.medicin-ethik.ch/publik/ethische\\_probleme.htm](http://www.medicin-ethik.ch/publik/ethische_probleme.htm) (Version: 2003).
- Müller-Ellermann, I. (o.J.) Parkinson-Syndrom Ratgeber für den Patienten, Anleitung für Sprachübungen, Nordmark Arzneimittel GmbH 2082 Uetresen - Info-Service Parkinson
- Panje, W. R. & Morris, M. R. (1995) Chirurgie von Mundhöhle, Zunge und Oropharynx. In: H. H. Naumann ed. *Kopf- und Hals-Chirurgie / Band 1: Gesicht, Nase und Gesichtsschädel, Teil II*: 711-738. Stuttgart, New York: Thieme Verlag.
- Reuther, J. & Bill, J. S. (1998) Plastische und wiederherstellende Mund-Kiefer-Gesichtschirurgie. *Praxis der Zahnheilkunde Band 10/II*. München: Urban & Schwarzenberg.
- Schröder, A. (1985) *Zur Problematik und chirurgischen Technik der Zungen- und Mundbodenrekonstruktion*. Diss.; Humboldt-Universität zu Berlin.
- Städtler, A. (1989) *Zur Sprachbehandlung von Tumorpatienten nach Zungen- und Mundbodenresektion mit plastischer Deckung bzw. Rekonstruktion (Entwicklung einer komplexen Behandlung auf der Grundlage einer Kompensationstherapie)*. Diss.; Humboldt-Universität zu Berlin.
- Werner, J. A. (2000) Krebserkrankungen im HNO-Bereich. *Klinik für Hals-, Nasen- und Ohrenheilkunde der Philipps-Universität Marburg*: online. <http://www.hno-marburg.de/indexger.htm> (Version: 2002).



# Syllable cut and energy contour in vowels: a comparative study on German and Hungarian

**Katalin Mády**

*Pázmány Péter Katolikus Egyetem, Germanisztikai Intézet, Piliscsaba, Hungary*

**Krisztián Z. Tronka**

*Pázmány Péter Katolikus Egyetem, Germanisztikai Intézet, Piliscsaba, Hungary*

**Uwe D. Reichel**

*Institut für Phonetik und Sprachliche Kommunikation, Ludwig-Maximilians-Universität München, Germany*

---

Syllable cut is said to be a phonologically distinctive feature in some languages where the difference in vowel quantity is accompanied by a difference in vowel quality like in German. There have been several attempts to find the corresponding phonetic correlates for syllable cut, from which the energy measurements of vowels by Spiekermann (2000) proved appropriate for explaining the difference between long, i.e. smoothly, and short, i.e. abruptly cut, vowels: in smoothly cut vowels, a larger number of peaks was counted in the energy contour which were located further back than in abruptly cut segments, and the overall energy was more constant throughout the entire nucleus. On this basis, we intended to compare German as a syllable cut language and Hungarian where the feature was not expected to be relevant. However, the phonetic correlates of syllable cut found in this study do not entirely confirm Spiekermann's results. It seems that the energy features of vowels are more strongly connected to their duration than to their quality.

---

## 1 Introduction

The German vowel system is characterised by a correlation of vowel quantity and vowel quality: long vowels are normally tense, while short vowels are lax, cf. [i:] – [ɪ]: *Miete* 'rent' – *Mitte* 'centre', [e:] – [ɛ]: *Weg* 'way' – *weg* 'away' etc.

It has been an object to discussion for decades whether one of both features is predictable from the other and can therefore be regarded as redundant.

One group of phonologists treats the quantity as the primary phonological (or even the only phonologically relevant) feature in this opposition. However, quantity is an accent-phenomenon in German, i.e. long vowels occur mainly in stressed position. An appropriate description must thus assume a set of rules shortening an underlying long vowel in an unstressed syllable in order to provide the correct surface forms, cf. *Musik* [ʊ]<sup>1</sup> [i:] ‘music’ – *Musiker* [u:] [ɪ] ‘musician’ – *musikalisch* [ʊ] [ɪ] [a:] ‘musical’ – *Musikalität* [ʊ] [ɪ] [a] [ɪ] [e:]<sup>2</sup> ‘musicality’. Other phonologists propose that the distinctive feature is rather tenseness. Since this feature remains intact in the alternation above, such an analysis can describe it in a more plausible way without assuming rules changing an underlying feature in the surface representation. However, the assumption of distinctive tenseness is in one respect unsatisfactory: there are several connections between the vowel opposition and prosodic phenomena (quantity, stress, phonotactic equivalence between long vowels, diphthongs and short vowel + consonant combinations etc.) – indicating that this opposition is probably not a segmental one.

Another solution of the problem is based on the assumption of a syllable cut opposition in Standard German. The basic idea of this concept is that stressed short lax vowels are somehow “not perfect” in the sense that they require a postvocalic segment in the same syllable, while short (if unstressed) or long (if stressed) tense vowels do not. The described problems of the other two concurring theories are avoided in this concept since (1) the opposition of abrupt cut (*scharfer Schnitt*) with a lax vowel and smooth cut (*sanfter Schnitt*) with a tense vowel is clearly a prosodic one and (2) temporal differences between the two vowel classes are just concomitant phonetic phenomena (or even side effects) of this higher suprasegmental contrast. Despite of its phonological plausibility, this concept was often rejected in the second half of the 20<sup>th</sup> century – because of the lacking phonetic correlate of the syllable cut in Contemporary German.

In his study, Spiekermann (2000) discussed and investigated all phonetic correlates for vowel segments that had been assumed so far by phonologists from Sievers through Trubetzkoy up to Vennemann and Maas & Tophinke (for references, see Spiekermann, 2000). Spiekermann found that the parameters used to describe energy contours were highly relevant for the contrast abrupt vs.

<sup>1</sup> While prescriptive transcriptions suggest that a short tense [u] is pronounced here, most natives would prefer [ʊ].

<sup>2</sup> According to Northern Standard German and everyday speech. In elaborated speech, the last vowel is pronounced as [ɛ:] instead of [e:].

smooth cut: 'E-Max' (Germ *E-Zahl*, number of energy peaks), 'E-Pos' (position of the energy maximum) and 'E-Hold' (Germ *E-Halt*, difference between energy minimum and maximum divided by the maximum). According to Spiekermann's results, smoothly cut (i.e. tense and long) vowels had more energy peaks that were located further back in the segment, and smoothly cut vowels had a higher intensity level throughout the entire segment than abruptly cut vowels. The tendency for the energy maximum to be located further back in smoothly cut and earlier in abruptly cut vowels lead Spiekermann to the assumption that the main characteristics of the syllable are not to be found in the nucleus-coda transition as proposed by Sievers, but in the onset-nucleus transition.

Spiekermann also tested vowel oppositions in Finnish and Czech that primarily make use of a quantitative opposition and thus are not regarded as syllable cut languages. He found that in all languages, longer durations are associated with a higher E-Max, while E-Pos and E-Hold were more or less indifferent for duration. These values were either located between smooth and abrupt cut in German or were closer to the measures for abrupt cut.

While Spiekermann's results are impressive, there are two main shortcomings in the experimental setup. Firstly, he relied on a relatively small corpus ( $n = 225$ ) that involved all VC combinations of German uttered only once, thus, no statistic analysis could be undertaken. Secondly, his analysis was carried out manually, and the parameters were expressed in three categories instead of metric (i.e. percent) values.

There are strong phonological arguments for the assumption that syllable cut is not crucial for the Hungarian vowel system (Tronka, 2005). First, while a German syllable including a short vowel is only well formed if the vowel is followed by a consonant, short vowels can be syllable final (i.e. they do not require a coda) in Hungarian (eg. *fal* /a/, /u/, 'village'). Second, the relevance of syllable cut was primarily restricted to accented syllables, as it is the only position where vowel quantity is distinctive in German (and most Germanic languages). In Hungarian, however, vowel quantity is independent of word stress (which is always on the first syllable in the word), c.f. *falat* /'falat/ 'mouthful' – *falát* /'fala:t/ 'his/her wall'.

Like German, Hungarian involves seven vowel classes, (/i, y, u, e, ø, o, a/), of which all can be realised long or short (Mády, 2001). The main vowel opposition in Hungarian is durational, while long and short /e/ and /a/ also differ in quality. There is a smaller quality opposition in /o/ and /ø/, where the laxness of the short vowel is mostly explained by dynamic effects, and which most speakers of Hungarian are not aware of (Siptár & Törkenczy, 2000).

Based on the assumption that syllable cut plays a central role for German vowels but it is not relevant for Hungarian, it was hypothesised that the features

E-Max, E-Norm, E-Pos, and E-Hold were relevant for the distinction between long (smoothly cut) and short (abruptly cut) vowels in German. At the same time, long and short vowels were not expected to differ significantly for Hungarian along their energy patterns, but to behave similar to Finnish and Czech.

In order to test Spiekermann's finding for Hungarian, we constructed a pilot study (Tronka, Mády & Reichel, 2006) with slightly modified parameters based on metric instead of ordinal scales (see 2.2). Our results for German were not completely comparable with those in Spiekermann (2000): while smoothly cut vowels included more energy maxima which were located further back in the segment, their overall energy showed a greater minimum-maximum difference than that of abruptly cut vowels. Moreover, exactly the same tendencies were found for Hungarian where syllable cut was not supposed to apply.

However, the German and Hungarian corpora available at that time were not entirely comparable (little overlap of consonant environment), thus we felt it necessary to perform the analysis on a more appropriate speech material. The measures will first be tested for German and Hungarian separately. On the basis of these findings, the results from the two languages will be compared and discussed.

## **2 Material and methods**

### **2.1 Material**

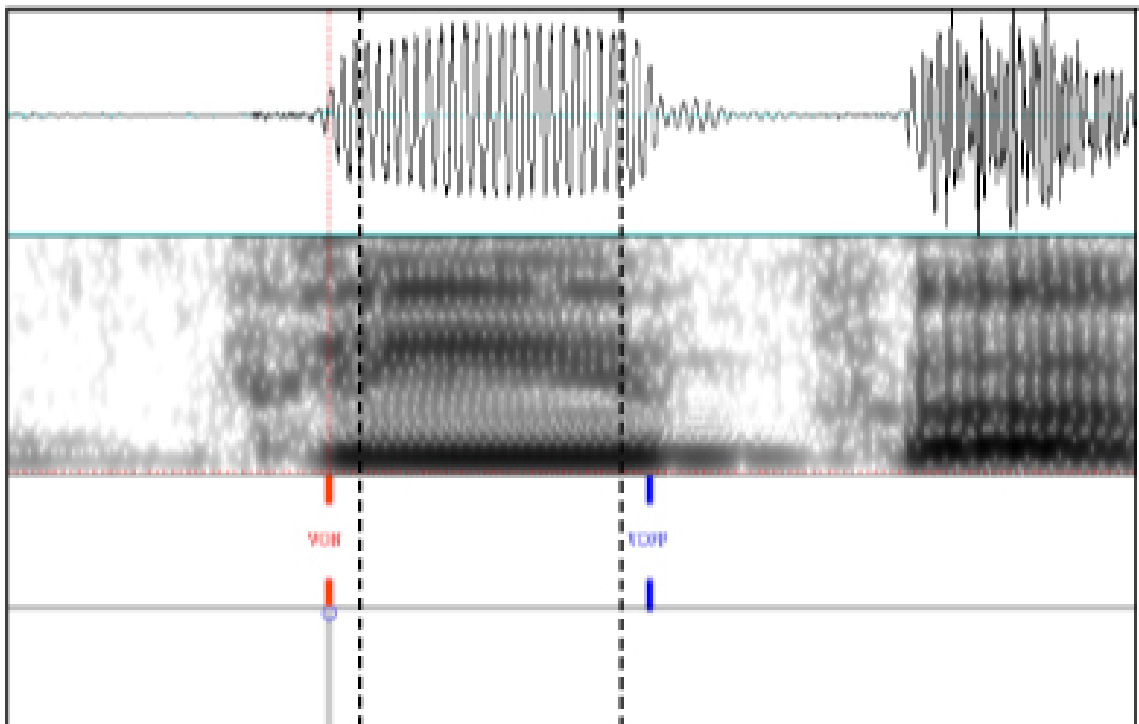
Both the German (4 speakers) and the Hungarian (3 speakers) corpora<sup>3</sup> included /i, y, u, e, ø, o, a/ as short and long vowels in nonsense words embedded in a carrier sentence (including 6 syllables in the German corpus and 9 in the Hungarian one). German words had the structure /C<sub>1</sub>VC<sub>2</sub>ə/ where C<sub>1</sub> and C<sub>2</sub> were stops and had the same place of articulation (PoA) (labial or velar), while C<sub>1</sub> was voiced and C<sub>2</sub> unvoiced. The structure of the Hungarian stimuli was slightly different: the last vowel was /a/ plus C<sub>1</sub> and C<sub>2</sub> were identical. Consonants were varied for PoA (labial, alveolar and velar). Both corpora were balanced for vowel duration and quality and consonant PoA [n(germ) = 1076, n(hung) = 1006].

The speech material was segmented automatically (by the software MAUS from the Department of Phonetics and Speech Communication in Munich) and corrected manually (in Praat 4.3). F2 onset and offset were applied as a boundary marker. As shown in Figure 1, some vowels offer two

---

<sup>3</sup> Both corpora have been recorded at ZAS, Berlin, for articulatory investigations by Electromagnetic Midsagittal Articulography (EMMA).

interpretations of the segment boundary: (1) the first and last appearance of F2 in the segment (including transitions), (2) the first and last appearance of the entire formant structure including F3 and F4 and thus including only the central, relatively steady phase of the vowel. Our preference of the first alternative relies on the assumption, that syllable cut is primarily based on certain requirements of syllable structure and not just on the vowel, thus, juncture (before and after the nucleus) will probably play a central role in its physical manifestation. Thus, it seems convincing to concentrate on the entire vowel duration.



**Figure 1:** Segmentation technique shown by the example of the Hungarian item /pipɒ/. (1) Left and right segment boundary: onset and offset of F2, (2) dotted lines between them: onset and offset of entire formant structure (including F3 and F4).

## 2.2 *Methods*

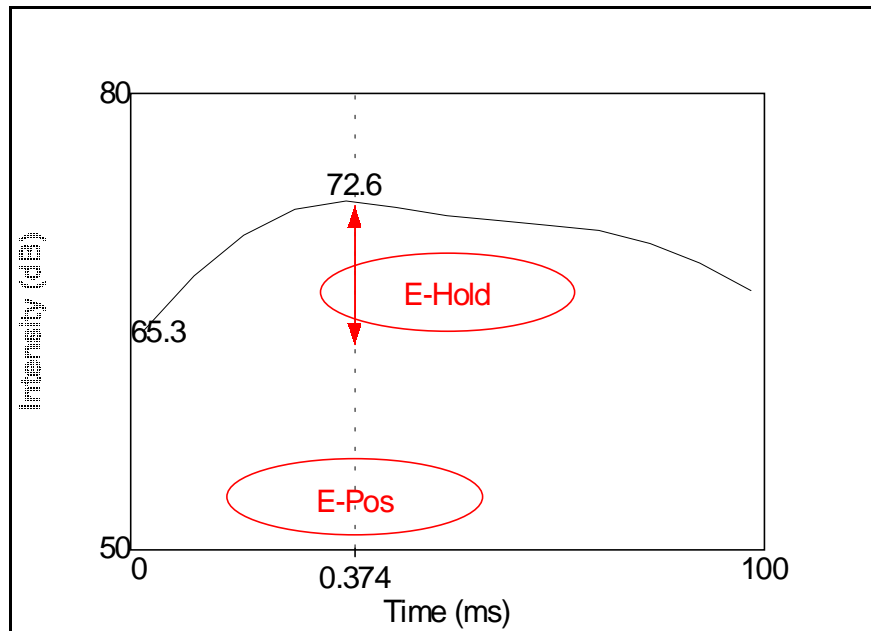
Smoothed energy contours were calculated by applying overlapping Hanning windows of 20 ms on the rectified oscillogram in order to remove glottal closure peaks.

Following measures, based on Spiekermann (2000) but not in full accordance with this study, were then derived from these contours (see also Figure 2):

- **E-Max:** the absolute number of maxima,
- **E-Norm:** E-Max normalised to the length of the contour (number of peaks divided by number of samples),
- **E-Pos:** the relative position of the last maximum within the contour,
- **E-Hold:** the ratio of the difference between the absolute maximum and minimum with respect to the maximum.

In contrast to Spiekermann we did not only calculate E-Max in absolute terms, because a positive relation between contour length and the number of maxima within this contour is somewhat self-explaining, and as vowels in smoothly cut syllables tend to be longer than in abruptly cut ones, the former will trivially show more energy peaks than the latter.

In order to cancel out this durational effect, we divided the energy peak number by the length of the energy contour. Furthermore, we avoided the loss of information due to data quantification Spiekermann carried out for E-Pos, for which he divided the vowel into 9 segments, and for the quotient E-Hold which had been categorised in 3 different classes. Instead of categorising E-Pos, we directly calculated the relative position of the last maximum with respect to vowel length, and also for E-Hold no classification was done. Therefore in our study the features E-Pos and E-Hold are not ordinally but metrically scaled.



**Figure 2:** Energy contour of a vowel segment with the parameters E-Max = 1, E-Norm = 0.0007, E-Pos = 0.374, and E-Hold = 0.095, duration = 100 ms.

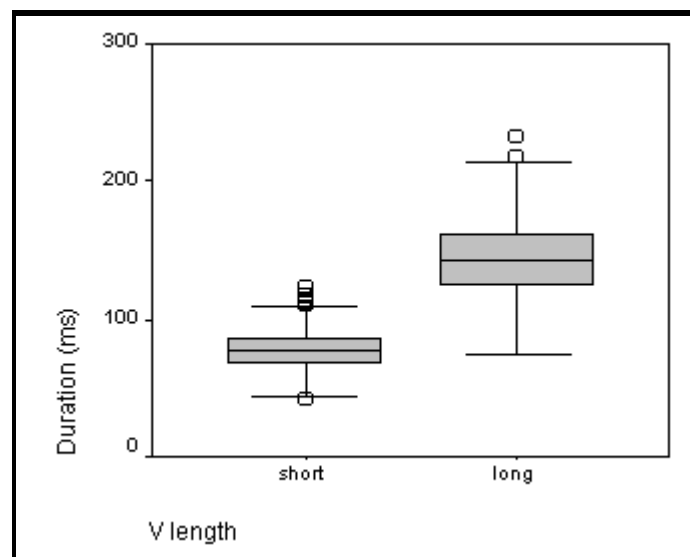
The measure E-Pos was also modified: in Spiekermann's analysis, only vowels with exactly one energy maximum were included in the analysis. However, if the distinctive character of syllable cut is based on the state of the energy level in the moment when the vowel is cut by the following consonant, then the position of the last energy maximum is relevant, no matter how many peaks were counted before. Thus, we calculated E-Pos as the position of the last maximum, but for reasons of compatibility, E-Pos was also calculated for vowels with one maximum.

### 3 Results

#### 3.1 German vowels

##### 3.1.1 Vowel length

German long ('l', smoothly cut) and short ('s', abruptly cut) vowels differed significantly for duration with slight overlap of the peripheral values. Short vowels tend to have more outliers (more than 1.5 interquartile distances but less than 3) towards long vowels than the other way around. In other words, while there is a relative contrast between long and short vowels, a given duration value cannot be directly associated with smooth or abrupt cut, and abruptly cut vowels seem to be less attached to a short duration than smoothly cut vowels to a longer one.

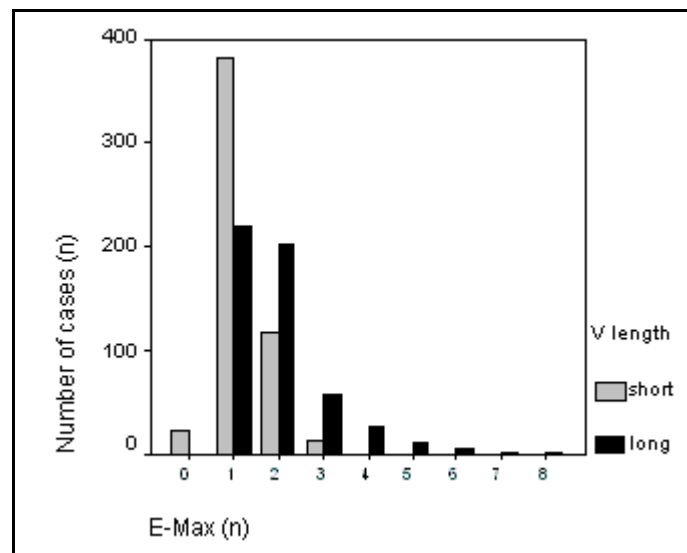


**Figure 3:** Duration for short (s) and long (l) German vowels. Mean(s) = 78 ms, SD(s) = 13 ms, mean(l) = 144 ms, SD(l) = 27 ms.

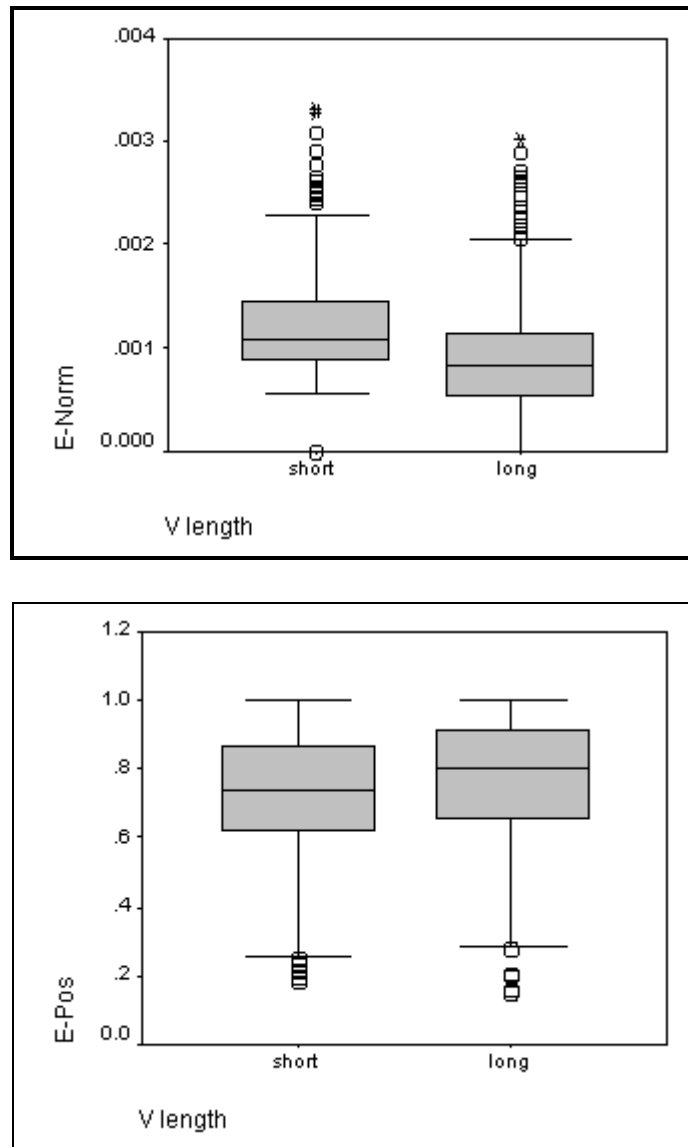
The parameters duration, E-Max, E-Norm, E-Pos and E-Hold were tested for correlation. None of the energy measures showed a linear correlation. Therefore, Spearman's Rho was calculated over all parameters. No strong correlation (higher than  $\rho = 0.6$ ) was found between any of the parameters. Duration was correlated positively with E-Max, E-Pos and E-Hold, but negatively with E-Norm, i.e. longer vowels had relatively fewer energy peaks than short ones.

The significance of duration and the energy measures was tested by a t-test for two independent samples ( $\alpha \leq 0.05$ , two-tailed). Most data units did not meet the condition of a normal distribution for an ANOVA, but they were large enough to perform a Welch test ( $n > 50$ ) that does not require normally distributed and homogenous samples.

The difference for all tested variables between smoothly and abruptly cut vowels was highly significant. Long vowels had more energy peaks (E-Max). However, the relative number of energy maxima (E-Norm) was smaller for longer vowels, i.e. they were less dense in long vowels than in short ones (Figure 4a,b). The last energy maximum (E-Pos) was located further back in the vowel segment, as was also found in Spiekermann (2000).



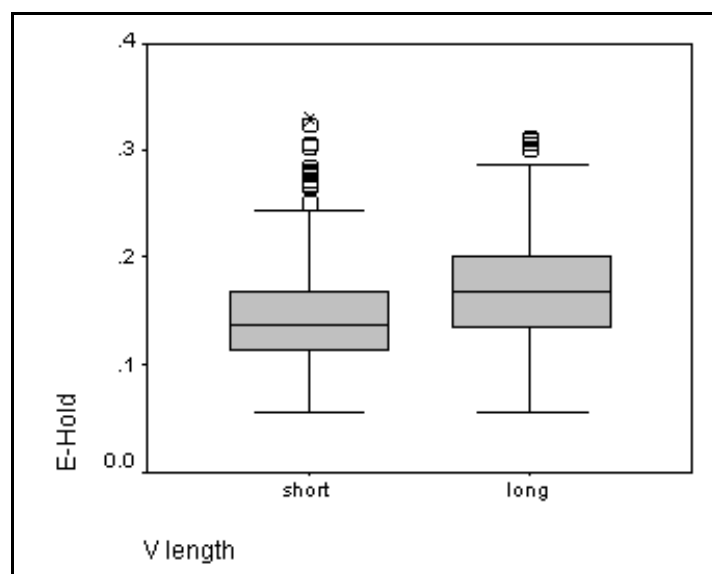




**Figure 4a–c:** (a) E-Max, (b) E-Norm, (c) E-Pos of short and long German vowels. Box plots: black line: median, upper and lower box, upper and lower whisker: 25% of cases (interquartiles), respectively.

However, we obtained results different from those of Spiekermann (2000) regarding the overall energy level in the segment. While he found a high intensity level in long vowels throughout the entire vowel segment (E-Hold), our results show exactly the opposite pattern: according to the t-test, the difference between intensity maximum and minimum in long vowels is significantly larger than in short ones. While in Spiekermann's study, long vowels often had an E-Hold of less than 5% (0.05 in the present scaling). As shown in Figure 5, such a

small difference was not found in any segment in our data, the smallest E-Hold being 0.055.<sup>4</sup>

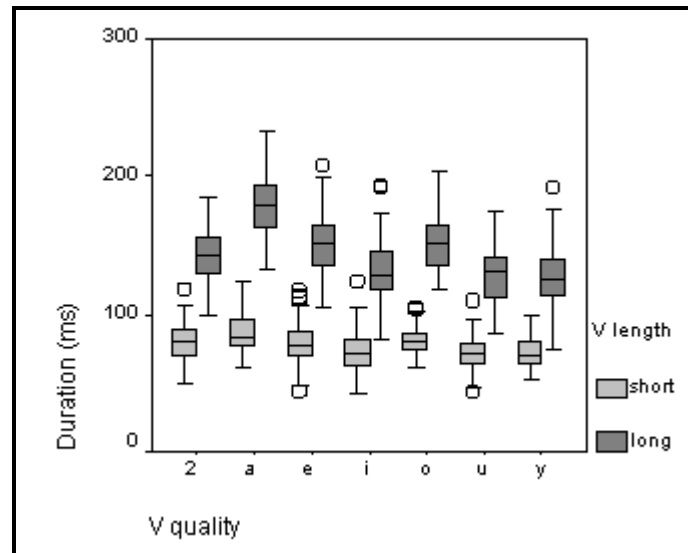


**Figure 5:** E-Hold of short and long vowels in German.

### 3.1.2 Vowel classes

Duration for each vowel quality (long vs. short) was significantly different (Figure 6). The fact that high vowels are the shortest and low vowels the longest segments corresponds to general tendencies regarding intrinsic duration: high vowels tend to be shorter than low ones in most languages (Kassai, 1998).

<sup>4</sup> This might be a consequence of different segmentation guidelines from those used in our corpora. If not the onset and offset of F2, but the entire visible formant structure was used as boundary markers of the vowel, the segment duration is probably shorter and thus, differences within this domain are smaller (see 2.1 and Figure 1).

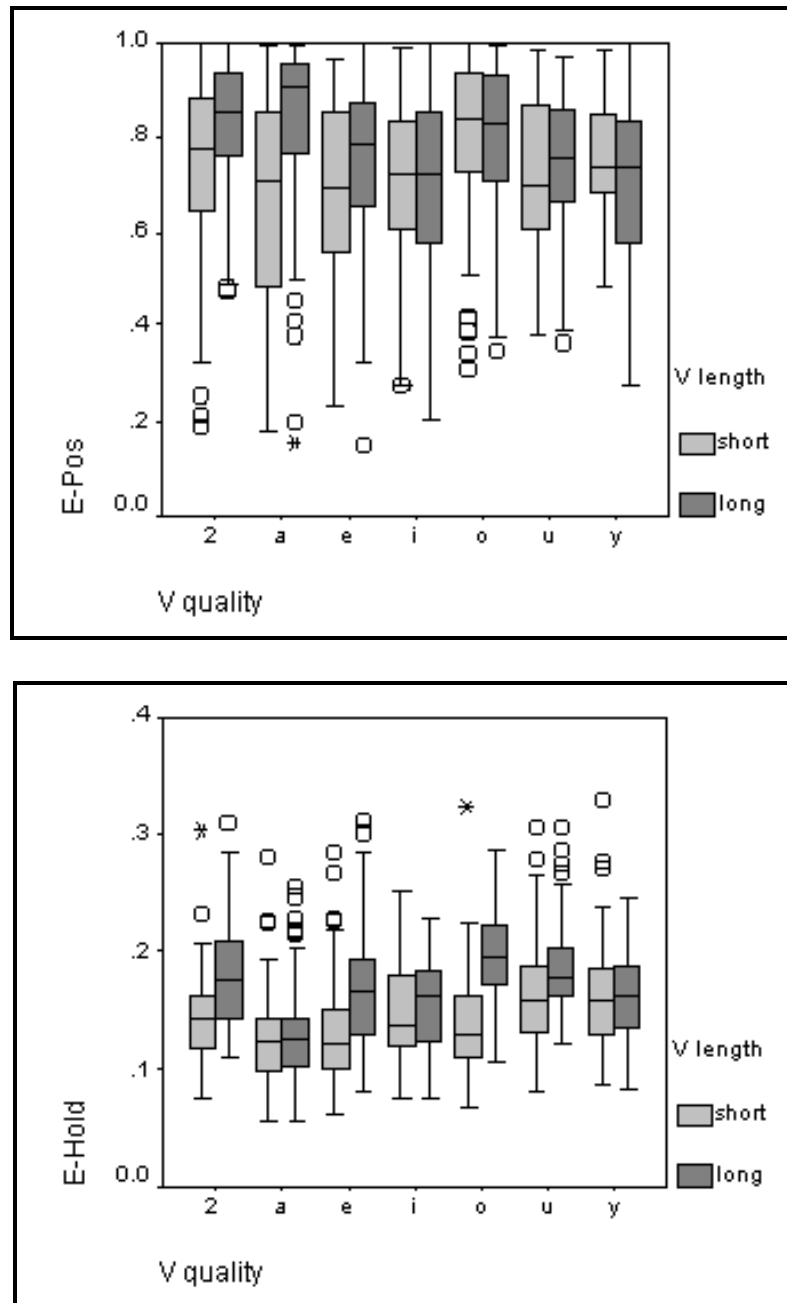


**Figure 6:** Vowel durations in each German vowel class (vowel quality is given in SAMPA).

In order to test the energy parameters along vowel quality, a t-test was performed for each vowel pair (long and short, approximately 80 realisations per each). While the examined parameter differed in almost all vowel pairs according the pattern described above, the difference was not significant at the 5% level for any of the vowels except for /ø/.

All long vowels had a larger number of energy maxima. Most vowels (except for /a/ and /u/) had a higher value for E-Norm.

The least reliable parameter was E-Pos. Three vowels did not show a difference at all (/i/, /o/, /u/), and in /e/, the difference was not significant. The tendency in E-Hold was not much clearer: 4 vowels matched the overall pattern, while three (/i, y, a/) did not (Figure 7). If E-Pos and E-Hold were calculated according to Spiekermann's method, the pattern was even less clear. No interaction with vowel height can be seen along the parameters.

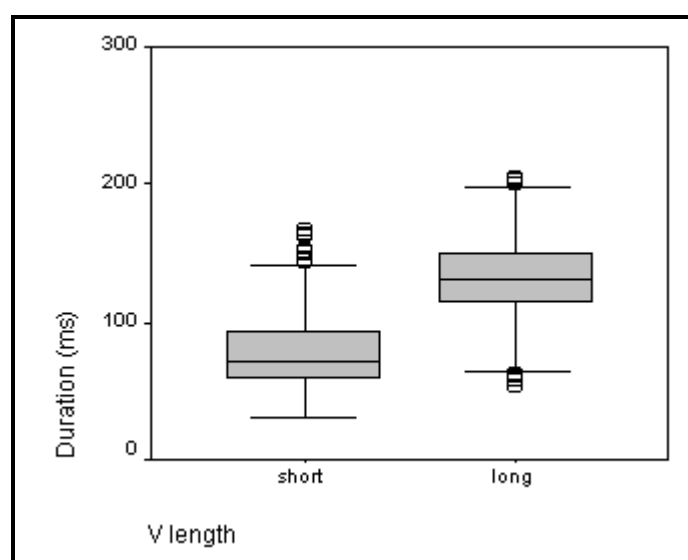


**Figure 7a,b:** E-Pos and E-Hold in short and long vowels for short and long German vowels (vowel quality is given in SAMPA). Circles stand for outliers (1.5–3 box lengths, i.e. interquartile distances from the upper/lower end of box), asterisks for extreme values ( $> 3$  interquartile distances).

## 3.2 *Hungarian vowels*

### 3.2.1 *Vowel length*

Hungarian long and short vowels differed significantly for duration. While Hungarian long vowels were somewhat shorter than German ones, the standard deviation is the same, while short vowels have the same mean but a clearly larger standard deviation. In other words, the difference between short and long vowels was less clear-cut than in German.



**Figure 8:** Duration for short (s) and long (l) Hungarian vowels. Mean(s) = 78 ms, SD(s) = 24 ms, mean(l) = 132 ms, SD(l) = 27 ms.

Correlations between duration and the energy parameters were approximately identical with those in German.

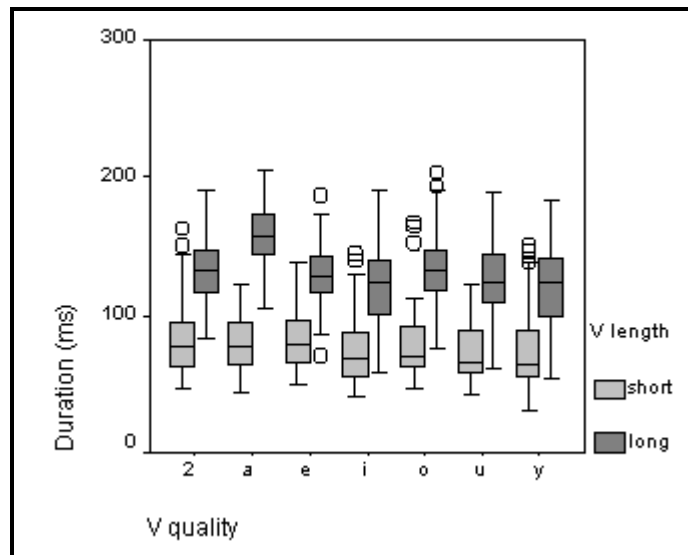
A comparison revealed the same tendencies between long and short vowels regarding E-Max and E-Norm: a higher number of energy peaks in long vowels, but their higher density in short vowels. E-Pos was located further back in long vowels that also had a larger maximum-minimum difference of intensity. E-Norm and E-Hold resembled the pattern described for German.

### 3.2.2 *Vowel classes*

All vowel classes differed significantly for duration. Short vowels showed, as seen in Figures 6 and 9, little variation of duration. The difference between the long vowel and its short counterpart was largest in /a/ and smallest in /y/. This finding is interesting in the context that long and short /a/ in Hungarian clearly

differ for quality ([a:] vs. [ɒ]<sup>5</sup>), while high vowels do only to a small extent (Kassai 1998). Figure 9 reveals a large overlap between short and long /y/. It is interesting, as Vicsi & Szaszák found that the shortest vowel in their corpus (BABEL) was the long vowel /y:/. It seems, that the duration distinction plays only a marginal role for this sound.<sup>6</sup>

All vowel classes match the results found for the entire set of data, but none of the classes reveals a significant difference for all three parameters.



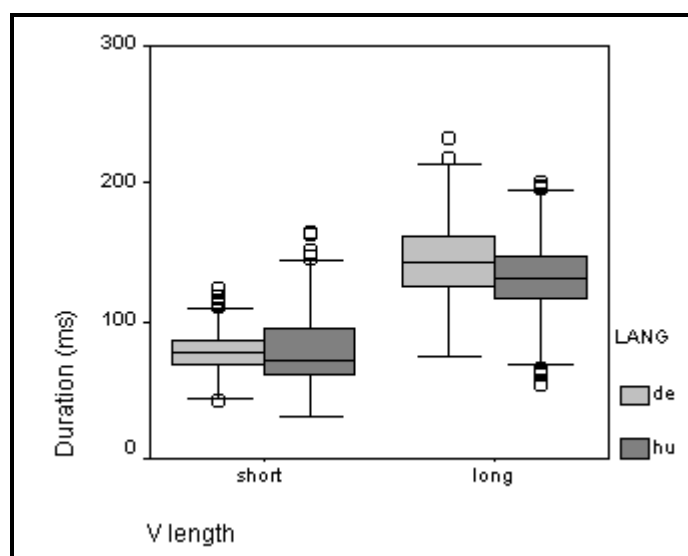
**Figure 9:** Vowel durations in each Hungarian vowel class (vowel quality is given in SAMPA).

### 3.3 *Contrasting German and Hungarian vowels*

As said in 2.1, both the German and the Hungarian corpora involved slightly different consonant contexts. Although both consonant PoA and voicing might have an impact on vowel duration, this could be ignored in the previous sections because each corpus was balanced for these factors. However, a comparison between German and Hungarian vowels required a corpus where consonantal environment was identical for both languages. Therefore, only stimuli with labial and velar consonants were considered for the contrastive corpus (n = 1747).

<sup>5</sup> Short /a/ is normally given as [ɒ] in Hungarian phonetics, but the vowel quality is better expressed by the 13. cardinal vowel [ɒ].

<sup>6</sup> These findings are unpublished and rely on personal communication with György Szaszák, Department of Telecommunication and Media Informatics, Budapest University of Technology and Economics.



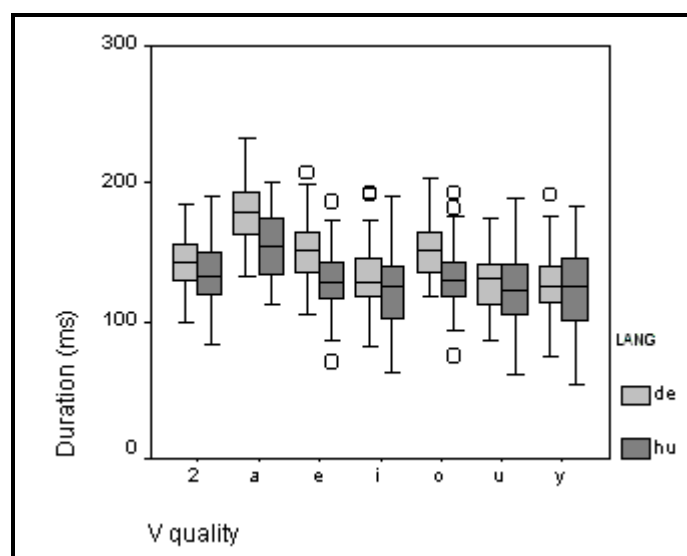
**Figure 10:** Duration for short and long German (de) and Hungarian (hu) vowels. Mean(Germ) = 144 ms, SD(Germ) = 27 ms, mean(Hung) = 131 ms, SD(Hung) = 28 ms.

As shown in Figure 10, long vowels in German and Hungarian differ significantly for duration. Therefore the energy parameters were again compared with regard to German and Hungarian long vowels.

The whole set of long German and Hungarian vowels differed significantly for E-Norm, and E-Pos, but not for E-Hold. In short vowels (that were not significantly shorter in Hungarian) the same tendency was observed. The energy parameters were related to duration in the same way as for each language: German vowels, that were significantly longer than Hungarian ones (see Figure 10), had a higher number of absolute but a lower number of relative maxima, which were located further back in a vowel segment, and the overall energy level was lower.

Finally, German and Hungarian vowels were compared class-wise. While the German vowels were longer in each case, the difference was not significant for the high vowels /i/, /y/, and /u/. These vowels had shorter durations in both languages than mid-high and low vowels, and the duration was less specific in Hungarian (expressed by a high SD, especially for /y/, see Figure 11).

German vowels, that were longer than their Hungarian counterparts, had significantly lower E-Norm and higher E-Pos values, but there was no clear tendency regarding E-Hold. While the German vowels were expected to have higher E-Hold, i.e. a larger intensity range within the segment, it was only true in four cases: for the middle vowels /e/, /ø/, /o/, and for the high back /u/.



**Figure 11:** Duration for long German and Hungarian vowels for each of the seven vowel classes.

#### 4 Discussion in a phonological framework

The theory of syllable cut is based on languages in which (1) long vowels are associated with tenseness and (2) stressed short vowels always require a consonantal syllable coda. In these languages, vowel length is only distinctive in stressed syllables. The main idea is that smoothly cut (long and tense) and abruptly cut (short and lax) vowels do not only differ in length and tenseness, but there is a hidden, more general category that governs the other two features. According to Spiekermann (2000), this feature can be detected in the different energy contours of the vowels.

After several decades of unsuccessful search for reliable acoustic correlates of the hypothetical syllable cut phenomenon, Spiekermann (2000) proposed that the difference relied on the different energy contours in vowels with smooth and abrupt syllable cut. He also proposed that non-syllable-cut languages like Finnish and Czech did not show the regularities that were found for German. The same tendency was expected for Hungarian which is a non-syllable-cut language (for arguments, see 1).

In the present study, duration and energy parameters proposed by Spiekermann were re-tested on a larger German and Hungarian speech corpus. The results of the investigation of vowel duration showed a significant difference between long and short vowels for both languages – as was expected. The investigation of the separate vowel classes revealed that German short vowels varied less than the Hungarian ones, while the variation of short and long vowels in Hungarian overlapped in many cases. In one case (/y/), the short vowel was often realised with a duration that reached beyond the longest



articulation of its long equivalent. The relative constancy of the German short vowels confirms Trubetzkoy's theory, according to which smoothly cut vowels are expandable while abruptly cut ones are not (see also Hoole & Mooshammer, 2002, for articulatory data). Becker (1998) sets up a theoretical framework for this phenomenon and suggests that if a syllable in a smoothly cut vowel is expanded, it is always the vowel that is lengthened, while in a syllable with an abruptly cut vowel it is rather the following consonant. The fact that German short and long vowels do not or only slightly overlap shows the stability of the opposition in question: there seems to be a strong tendency for the double distinction (in duration **and** quality), as syllable cut is not only theoretically distinctive in German, but there are lots of cases (minimal pairs) in which it does in fact differentiate between meanings. In other words, this distinction is very often used in German.

On the other hand, there is no clear duration distinction in Hungarian for some of the vowel classes. Thus, we may assume that the phonological opposition that is manifested phonetically by means of durational differences is not a stable one: an overlap of the short segment far into the central part of its long counterpart signals that it is not so important to make a clear-cut difference between a short and long vowel. This finding matches well the investigations of Kassai (1979) who pointed out that for some of the vowel pairs of Hungarian, only few minimal pairs exist, thus, she argued that the quantity opposition in this language was relatively instable.

Although our results regarding duration were in accordance with those of Spiekermann (2000), the energy contours we found were often slightly different. In our data, long vowels had more energy maxima than short ones, just as Spiekermann describes. On the other hand, if the number of the maxima was normalised by vowel duration, long (smoothly cut) German vowels had in fact less energy maxima than short vowels, and the energy contours diverged more than in short (abruptly cut) vowels. In other words, while the number of energy maxima in long vowels varied from one to ten, short vowels included three energy peaks at most. Thus, the idea that tense vowels are characterised by a constant and high energy level did not prove to be appropriate. On the contrary, short (lax) vowels seem to have a more compact energy distribution (relatively more energy maxima and a smaller decrease of the intensity level during the vowel segment) according to their higher E-Norm and lower E-Hold values.

These findings are somewhat surprising, as they do not fit the tentative descriptions given by some phonologists (i.e. Becker), who see a direct relationship between syllable cut, the amount of energy, and phonetic manifestation (i.e. duration and tenseness/centralisation of a vowel). They argue that in the case of abrupt cut, there is always a postvocalic segment (typically a

consonant) cutting the ballistic production of the preceding vowel – the result is a smaller energy content resulting in a short and lax, i.e. more centralised articulation of the vowel in question. In case of smooth cut, there is no postvocalic segment at all or if any, it does not cut the vowel before its energy climax, thus, the vowel is articulated long and tense, i.e. not centralised. On the other hand, the idea of syllable cut could be maintained despite our findings, if one would consider it from the other ‘side’, i.e. from the perspective of the expandibility concept. In this case, one could argue as follows: in case of a lacking cutting effect (i.e. if there is an unrelated or no postvocalic segment at all in the syllable) a vowel will be expanded, and expansion of a vowel can lead to a nonlinear increase of the number of intensity maxima (i.e. the number of maxima is not directly related to the duration difference), while it is not possible for abruptly cut vowels. This has already been stated regarding speech rate in Hoole & Mooshammer (2002) from an articulatory point of view. On the other hand, all speculations on the relationship between expandibility, energy content and syllable cut will be superfluous if one considers our results regarding Hungarian, where similar data were found for E-Norm dividing long and short vowels, although no difference was hypothesised on the basis of the syllable cut theory. Hence, the only possible conclusion is that the absolute and relative number of energy maxima in a vowel is rather related to differences in duration than to syllable cut.

As already said previously, one of the main ideas of syllable cut is that abruptly cut vowels are cut by the following segment while smoothly cut segments are not. Spiekermann’s (2000) results regarding E-Pos were therefore somewhat unexpected, as he found exactly the opposite tendency as had been supposed by phoneticians and phonologists since Eduard Sievers: the intensity maximum appeared further back in smoothly cut vowels than in abruptly cut ones. At the same time, he found similar patterns in non-syllable cut languages – thus, the only possible interpretation was that also E-Pos is related to duration and not to syllable cut. Our measurements could confirm Spiekermann’s data in one aspect: the energy maximum lied further back in long vowels than in short ones, both in German and in Hungarian. On the other hand, we could measure only small differences for E-Pos in both languages (approximately 6% of the vowel duration), therefore it is questionable whether E-Pos has any phonological relevance at all.

One of the fundamental findings in Spiekermann’s investigations are undoubtedly his results for E-Hold. Since he found a relationship between syllable cut and the difference of energy minimum and maximum in a vowel, but he could not find it in non-syllable cut languages, he concluded that E-Hold was to be seen as a stable acoustic correlate of syllable cut in German. Our measurements in German confirmed Spiekermann’s results: there are indeed

differences in the energy range between smoothly and abruptly cut vowels. On the other hand, our results for Hungarian do not support this conclusion, as we found similar data also for Hungarian, where syllable cut does not play a role. Thus, we conclude that also E-Hold is probably rather related to duration than to syllable cut, as it is also present in Hungarian and between German and Hungarian vowels. The situation is more confusing, if one considers the seven vowel pairs in the two languages. A comparison of some German vowel pairs does show a relation between language type and intensity differences found by Spiekermann (mid-high vowels and /u/). So we cannot exclude the possibility in general that E-Hold plays any role in distinguishing between the two language type.

Finally, we want to report about some problematic points of Spiekermann's parameters. It is not quite clear what E-Hold as a central category in Spiekermann's theory really means. We are probably not wrong in the assumption that E-Hold is a kind of acoustic implementation of what Maas (1999) metaphorically called "austrudeln" ('to fade out'). The fact that smoothly cut vowels "fade out" while abruptly cut ones do not, is interpreted such, that smoothly cut vowels have a constant energy contour, while abruptly cut ones show an abrupt change. Spiekermann intends to apprehend this by measuring the difference between the minimum and the maximum value of the energy contour in a vowel. However, there are at least two crucial problems connected to this analysis method. Spiekermann measured – as we did when re-testing his hypotheses – an intensity minimum regardless of its position in the global energy contour of the vowel, i.e. regardless of the fact whether it is at the beginning or at the end of the contour. We think that it is probably not unimportant to reduce the investigation only to a certain part of the intensity contour. If we speak about the virtual cutting effect of a postvocalic segment, this points to the VC transition. If one on the other hand takes the findings for E-Pos into account, one could argue for measuring the CV transition of the energy contour of the vowel in question. There are tentative arguments against the latter approach that are based on accidental findings. The German data we originally intended to use for our analysis had a small echo, so we set the boundary mark before the echo was visible in the oscillogram and the spectrogram. This led to a general shortening of all German vowel segments with the conceivable effect that a possible difference in E-Hold disappeared. As the shortening always took place at the end of the vowel, what might have led to a neutralisation of the duration difference concerning E-Hold, we assume that acoustic cues of syllable cut are probably rather to be found at the VC boundary.

Besides, E-Hold is a problematic parameter from the acoustic and perceptual point of view. Due to the nonlinearity of the decibel scale, the

quotient (energy maximum – energy minimum)/energy maximum still depends on the respective energy level and this lack of normalisation prohibits a comparison of E-Holds derived from different energy levels. Furthermore, E-Hold does not reflect perceived energy differences, that is, an E-Hold value of 0.1 does not refer to the same perceptive difference if the energy maximum is located at 80 or at 65 dB in the given segment.

For a better understanding of the phenomenon of syllable cut, several further aspects should be taken into account: what pattern can be found in (1) unstressed vowels, (2) diphthongs, and (3) reduced vowels? Do unstressed vowels that are long and tense phonologically but short in their phonetic realisation like /e/ in *Metall* show the characteristics of smooth or abrupt cut? Is duration a crucial factor for reduced vowels that are not marked for tenseness? How would the tenseness difference of Hungarian mid-high vowels influence the energy parameters in a larger corpus? The answers to these question may bring us closer to the question whether syllable cut is a relevant feature in the German vowel system.

## Acknowledgements

We would like to thank Christian Geng, Centre for General Linguistics, Typology and Universals Research (ZAS), Berlin, for providing us with the German and Hungarian data.

## References

- Becker, T. (1998) *Das Vokalsystem der deutschen Standardsprache*. Frankfurt: Lang.
- Hoole, P. & Mooshammer, C. (2002) Articulatory analysis of the German vowel system. In: P. Auer, P. Gilles & H. Spiekermann (eds.) *Silbenschnitt und Tonakzente*. Tübingen: Niemeyer, 129–152.
- Kassai, I. (1979) *Időtartam és kvantitás a magyar nyelvben*. Budapest: Akadémiai Kiadó.
- Kassai, I. (1998) *Fonetika*. Budapest: Nemzeti Tankönyvkiadó.
- Maas, U. (1999) *Phonologie*. Upladen: Westdeutscher Verlag.
- Mády, K. (2001) Kontrastive Phonetik Deutsch-Ungarisch in Hinblick auf zu erwartende Interferenzphänomene. *Linguistische Beiträge Pasmaniensis*, 1: 29–51.
- Siptár, P. & Törkenczy, M. (2000) *The Phonology of Hungarian*. Oxford: University Press.
- Spiekermann, H. (2000) *Silbenschnitt in deutschen Dialekten*. Tübingen: Niemeyer.
- Tronka, K. (2005) Die Vokallänge im Deutschen und Ungarischen. In: I. Szigeti (ed.) *Junge Germanisten stellen sich vor*. Frankfurt: Lang, 177–196.

- Tronka, K., Mády, K. & Reichel, U. D. (2006): A contrastive study of syllable cut in German and Hungarian: vowel length and energy. *3<sup>rd</sup> Old World Phonology Conference*, Budapest, 17–19. January 2006, 57–58.

# Some comments on the reliability of three-index factor analysis models in speech research

**Christian Geng**

*Zentrum für Allgemeine Sprachwissenschaft, Berlin*

**Phil Hoole**

*Institut für Phonetik und sprachliche Kommunikation der LMU, München*

---

Low- dimensional and speaker-independent linear vocal tract parametrizations can be obtained using the 3-mode PARAFAC factor analysis procedure first introduced by Harshman et al. (1977) and discussed in a series of subsequent papers in the Journal of the Acoustical Society of America (Jackson (1988), Nix et al. (1996), Hoole (1999), Zheng et al. (2003)). Nevertheless, some questions of importance have been left unanswered, e.g. none of the papers using this method has provided a consistent interpretation of the terms usually referred to as “speaker weights”. This study attempts an exploration of what influences their reliability as a first step towards their consistent interpretation. With this in mind, we undertook a systematic comparison of the classical PARAFAC1 algorithm with a relaxed version, of it, PARAFAC2. This comparison was carried out on two different corpora acquired by the articulograph, which varied in vowel qualities, consonantal contexts, and the paralinguistic features accent and speech rate. The difference between these statistical approaches can grossly be described as follows: In PARAFAC1, observation units pertain to the same set of variables and the observation units are comparable. In PARAFAC2, observations pertain to the same set of variables, but observation units are *not* comparable. Such a situation can be easily conceived in a situation such as we are describing: The operationalization we took relies on the comparability of fleshpoint data acquired from different speakers, which need not be a good assumption due to influences like sensor placement and morphological conditions.

In particular, the comparison between the two different approaches is carried out by means of so-called “leverages” on different component matrices originating in regression analysis, calculated as  $v = \text{diag}(A(A'A)^{-1}A')$  and delivering information on how “influential” a particular loading matrix is for the model. This analysis could potentially be carried out component by component, but we confined ourselves to effects on the global factor structure. For vowels, the most influential loadings are those for the tense cognates of non-palatal vowels. For speakers, the most prominent result is the relative absence of effects of the paralinguistic variables. Results generally indicate that there is quite little influence of the model specification (i.e. PARAFAC1 or PARAFAC2) on vowel and subject components. The patterns for the

articulators indicate that there are strong differences between speakers with respect to the most influential measurement as revealed by PARAFAC2: In particular, the most influential  $y$ -contribution is the tongue-back for some talkers and the tongue-dorsum for other speakers. With respect to the speaker weights, again, the leverage patterns are very similar for both PARAFAC-versions. These patterns converge with the results of the loading plots, where the articulator profiles seem to be most altered by the use of PARAFAC2. These findings, in general, are interpreted as evidence for the reliability of the PARAFAC1 speaker weights.

---

## **1 Introduction**

One broad research area aiming at a deeper understanding of the motor implementation of linguistic contrasts has been the search for efficient characterizations of vocal tract shapes by factor analytic methods. Nevertheless, the exact purpose of their application is not as homogeneous as it might seem at first glance: It has been suggested to evaluate statistical articulatory models in terms of their potential to mimic articulatory behavior expressed in terms of articulatory degrees of freedom as in the tradition of Maeda (1979a,1979b,1990). But also, a second tradition has focused its attention on these methods' ability to generalize over several speakers. Likewise, the statistical procedures are tuned to different rationales: The first tradition leads to intraindividually fitted models advantageous for control purposes, as applied in articulatory control models for speech synthesis (Badin et al., 2002) or facial animation (Maeda, 2005). Multispeaker solutions are characterized by the attempt to reveal latent building blocks underlying articulatory organization. In this work, we will concentrate on the latter approach, i.e. the "PARAFAC-tradition".

### **1.1 Classical PARAFAC1**

PARAFAC is a type of multi-mode analysis procedure and therefore contrasting with Principal Component Analysis (PCA) or factor analysis, which are two mode representations. PARAFAC requires an at least three-dimensional data structure with the third dimension usually being represented by different speakers, i.e. if all speaker weights are fixed to be one, then PARAFAC reduces to PCA. The advantage of PARAFAC is that there is no rotational indeterminacy as in PCA, in other words, PARAFAC gives unique results. The PARAFAC (in accordance with literature from now on called PARAFAC1) model can be written as (following Kiers et al., 1999, alternative notations are given in Harshman

et al., 1977 or Nix et al., 1996)

$$X_k = AS_kV^T \quad (1)$$

where  $X_k$  is the  $k$ th “slab” of the input data matrix, with  $k$  the number of speakers,  $A$  is the matrix of articulator loadings,  $S$  is the diagonalized matrix of speaker loadings for speaker  $k$  and  $V$  the loading matrix for vowels. The matrix of articulator weights is held constant for each slab of the data cube, i. e. for all  $k$  speakers. This addresses Cattell’s notion of parallel proportional profiles: “The basic assumption is that, if a factor corresponds to some real organic unity, then from one study to another it will retain its pattern, simultaneously raising or lowering all its loadings according to the magnitude of the role of that factor under the different experimental conditions of the second study.” (Cattell and Cattell, 1955, citing Harshman and Lundy, 1984, p. 151). Another way to put it is this (Harshman 1977, p. 609): “Thus if speaker A uses more of factor 1 than does speaker B for a particular vowel, then speaker A must use more of factor 1 than speaker B in all other vowels. The ratio of any two speakers’ usage of a given factor must be the same for all vowels.” Fitting the PARAFAC1 to the data in the least squares sense amounts to minimizing

$$\sigma_1(A, V, S_1, \dots, S_k) = \sum_{k=1}^k ||AS_kV^T||^2 \quad (2)$$

There is a unique solution minimizing (2) up to scaling and permutation. Cattell’s proportionality does not always have to be a plausible assumption though; it can also turn out to be too restrictive in some cases. For illustration, the other extreme would be to put no structure at all onto  $A$  -which is equal to reducing the PARAFAC model to a PCA and loosing the desirable uniqueness properties.

Before turning to the constraints that define PARAFAC1 and describing less restrictive alternatives, we give a brief review of the studies using this method.

## 1.2 Survey of studies using PARAFAC1

The presumably largest focus of interest in the late 80’s to the mid 90’s by researchers from the speech production area using multimode Data Analysis techniques has been an issue raised in a paper by Jackson (1988) concerning the number of factors that are reliably extractable by means of the PARAFAC1 algorithm. Jackson claimed to have extracted a three-factor solution from a



corpus of Icelandic data. This claim was rejected later by Nix et al. (1996) highlighting the importance of diagnostic measures for the assessment of reliability of a PARAFAC solution: Harshman & Lundy (1984) suggested to use the triple product over the three modes of the correlations between corresponding sets of weights for each pair of factors. This triple product, also referred to as “congruence” coefficient can, in the case of PARAFAC1, be calculated as the triple Hadamard product of the products of the component matrices with their transposes:

$$TC = (A^T A) \circ (B^T B) \circ (C^T C) \quad (3)$$

Harshman & Lundy (1984) suggested triple products more negative than -0.3 between a pair of factors indicating a degenerate solution since in this case both factors are attempting to capture similar portions of the total variance, resulting in a second factor being simply a degenerate version of the first. The reanalysis of the data published by Jackson (1988) carried out by Nix et al. (1996) indicated that the third factor in Jackson’s solution was not reliable which lead to disenchantment about the explicative claim made by this kind of modeling.

A second major result of the discussion of the 80’s and 90’s was the format of input data for applications of the PARAFAC1 algorithm to articulatory problems: “Although measuring the shape of the tongue with respect to anatomically normalized vocal diameter gridlines<sup>1</sup> does reduce the initial representational dimension, this measurement scheme needlessly loses information such as the positions of the tip of the tongue in the horizontal dimension. More importantly, the range of possible solutions is artificially constrained by the orientation of the grid lines. For example, a factor representing protrusion and/or retraction of the tongue tip is not possible because no grid line is oriented in this direction.” Thus it is not too surprising that both of our factors contain a quite strong horizontal component, as our data are “fleshpoint data” (Nix et al., 1996, p. 3708). In other words, the quality of the data seen by the algorithm determines the solution obtained by fitting PARAFAC, and therefore also the interpretation of the factors. This in particular can become a hot topic concerning the relevance of analyzes obtained by this method considering the advent of three-dimensional acquisition techniques in speech production research.

The first application of the PARAFAC algorithm in a reviewed journal contribution to three-dimensional tongue configurations was published in a paper by Zheng et al. (2003). The essential novelties apart from the three spatial dimensions of the input data consist in (a) a more thorough discussion of rea-

---

<sup>1</sup>The original PARAFAC work was based on the measurement of distances along anatomically defined reference lines forming a “measurement grid”, which was calculated for each speaker

sonable preprocessing strategies for the application of the algorithm to tongue configurations and (b) the assessment of the solutions obtained by PARAFAC1 by more recent diagnostics of model degeneracy. With respect to the first point, the authors apply additional scaling subsequent to centering as applied in previous studies. The purpose of the scaling procedure is to normalize each speaker's data to unit sum-squared variation, "so that talkers with greater variability and/or larger vocal tracts do not dominate the PARAFAC fitting process" (Zheng et al., 2003, p. 482).

With respect to the second point in the preceding paragraph, i.e. the application of more recent diagnostics of the reliability of model fits, it is useful to have a closer look at family relationships between N-way methods. Here, PARAFAC1 can be considered as a special case of a more general method of three-way factor analysis, Tucker3 (Tucker, 1966). The structural model of Tucker3 is given in formula (4):

$$\underline{X} = VG(S \otimes A)^T. \quad (4)$$

Here,  $\underline{X}$  denotes the higher-way array to be modeled,  $V$ ,  $S$  and  $A$  are the component matrices ( $S$  the speaker weights,  $A$  the articulator weights and  $V$  the vowel weights).  $G$  denotes the so-called "Tucker core" matrix.  $|\otimes|$  denotes the Kronecker tensor product. Now, PARAFAC1's structural model implies a hypercube as shaping of the core array, e.g. for a 2-factor solution the core array has the dimension  $2 \times 2 \times 2$ . Furthermore, all elements off the hyperdiagonal of the core array are required to be zero for a valid PARAFAC solution, i.e. the core array is required to exhibit superidentity and therefore cancels in the following representation of PARAFAC1:

$$\underline{X} = V(S| \otimes |A)^T. \quad (5)$$

Here,  $S| \otimes |A$  denotes the Khatri-Rao product. This conceptualization of PARAFAC is used in Bro (1998) for the development of an alternative criterion of the number of factors and the detection of model degeneracies in PARAFAC1 models. It measures the percentage of the variation in the Tucker core matrix  $G$  consistent with PARAFAC1's requirement of core hyperdiagonality. Bro & Kiers (2003) suggest that a core consistency of at least 90% is a good indicator of a valid model.

### 1.3 PARAFAC2

Above, we have mentioned Cattell’s notion of “parallel proportional profiles”, which does not always have to be a valid assumption; it can also turn out to be too restrictive in some cases, and, as we have shown elsewhere (Geng & Mooshammer, 2000), a less restricted algorithm, PARAFAC2, offers an attractive alternative. Referring to the notion we have used in equation (1) for PARAFAC1 for a single “slab” of the multiclassified array, PARAFAC2 can be expressed as

$$X_k = A_k S_k V^T \quad (6)$$

Within PARAFAC2, each loading matrix for the articulators,  $A_k$ , is expressed as  $A_k = P_k A$ .  $P_k$  is an  $I * R$  matrix, where  $R$  denotes the number of factors and  $I$  the number of measurements in the articulator domain.  $A$  is constant over all these individual profiles and of size  $R * R$ . The rotational freedom provided by the PARAFAC2 model is adequate for approximating certain deviations from the strict linearity required in PARAFAC1. PARAFAC2 incorporates an invariance constraint on the factor scores as a milder version of factorial invariance: The cross-product matrix  $A_k^T A_k$  is constrained to be constant over  $k$  speakers. The model structure is determined by the choice of the structure of  $A_k$ . Bro (1998) compares PARAFAC2’s flexibility in this respect to Procrustes analysis. In Geng & Mooshammer (2000) we have shown that the strict assumptions required in the classical PARAFAC1 model were too strong to capture stress-specific variation in full detail. In contrast, PARAFAC2 allowed to account for systematic variation produced by word stress by imposing this weaker structure on the data. In particular, PARAFAC2 modeled the physical properties of the vocal tract shape in a more realistic and plausible way with respect to the description of mean factor shapes.

## 2 Method

### 2.1 The Corpora

In this study, we will reanalyze two distinct corpora. Both of them sample vowel nuclei acquired with fleshpoint tracking methods. The first corpus, which we will term the “stress corpus” was described in Geng & Mooshammer (2000), the second corpus, which we will refer to as the “speech rate corpus” was published in a paper by Hoole (1999). We will reiterate the description of these in order to pinpoint the differences between them, which could potentially endanger our interpretations concerning the method comparison we wish to achieve.

### 2.1.1 The Stress Corpus

Six native speakers of German (4 males, JD, PJ, CG and DF and 2 females, SF and CM) were recorded by means of an electromagnetic midsagittal articulo-graphic device. The speech material consisted of words containing /tVt/ syllables with nuclei (V= /i,ɪ,y,ʏ,e,ɛ,ɛɪ,ø,œ,a,ɑ,o,ɔ,u,ʊ/) in stressed and unstressed positions. Stress alternations were fixed by morphologically conditioned word stress and contrastive stress. So each symmetric /CVC/-sequence was embedded in the carrier phrase *Ich habe tVte, nicht tVtal gesagt.* (*I said –, not –*) with the first test syllable /tVt/ always stressed and the second always unstressed. For each of the 15 vowels, between six and ten repetitions of these vowels were recorded. Tongue, lower lip and jaw movements were monitored by EMMA (AG100, Carstens Medizinelektronik). Four sensors were attached to the tongue, one to the lower incisors and one to the upper lip. Two sensors on the nasion and the upper incisors served as reference coils to compensate for head movements under the helmet during the recording session. Jaw and lower lip movements will not be included in the analysis.

### 2.1.2 The Speech Rate Corpus

This corpus consists of seven adults, six males and one female. The experimental conditions were similar with respect to apparatus, tongue sensor placement, vowel environment and preprocessing to the stress corpus described in the previous section. The test utterances were formed by inserting the vowels into three different consonant contexts /p\_p/, /t\_t/ and /k\_k/. Each symmetric CVC sequence was embedded in a carrier phrase with the structure *Ich habe geCVC gesagt* (*I said –*). The subjects were tested in two separate recording sessions, usually a few days apart, which lasted about one hour each. In the first recording session the speakers produced the utterances at normal speech rate, in the second recording at a fast speech rate.

### 2.1.3 Potential Problems

- The data of the speech rate corpus, in contrast to the stress corpus, were recorded on two different occasions. Therefore, the sensors had to be attached twice, potentially resulting in artifacts of sensor placement.
- The material of the stress corpus contained two test words per item. It does not seem implausible to assume that the amplitudes of articulatory movements reduce over the course of the intonation phrase.

### 3 Results

For both corpora, we performed two analyzes of the data, the first using PARAFAC1 and the second PARAFAC2. The analysis of the rate corpus concerning PARAFAC1 is a partial reanalysis of results published in Hoole (1999) and therein referred to as the model for “multiple consonantal contexts”. Therefore, reprinting the displays already published in the paper mentioned would be redundant and is skipped with reference to the original publication. To stay in line with the results published in this paper, we also used the same preprocessing strategy as in Hoole (1999): The data delivered to the algorithm consisted of displacements from the average articulatory configuration of each subject. This amounts, in “standard terminology” (Harshman & Lundy (1984)) to centering across the vowel mode. This does not necessarily have to be the optimal preprocessing strategy, as elaborated in Zheng et al. (2003)<sup>2</sup>, but was adopted here for optimal comparability. The same strategy was applied to the stress corpus as well, for comparability purposes. Note that beforehand, in Geng & Mooshammer (2000), we had applied centering across vowel and speaker mode<sup>3</sup>, so that the solutions are not directly comparable to these results. Furthermore, as mentioned in Geng & Mooshammer (2000), we had to constrain some modes in some models to orthogonality in order to obtain non-degenerate solutions.

The results section is organized as follows: In the first part, we will have a look at global fit measures like the percentage of variance explained in order to achieve some basic insight into the structure of the models and to substantiate our solutions as valid. In the second, descriptive part, we display the conventional results on the solutions obtained, i.e. loading plots of extracted factors. In the third part, we proceed with analyzing the leverages, i.e. the influences that determine the exact solutions and the differences in fit between them.

#### 3.1 Global fit

In the first step, we will have a look at the global measures for the different solutions. As mentioned above, some of the models were constrained to orthogonality in the vowel mode in order to prevent strongly correlated factors and degeneracy. This holds for both solutions analyzing the stress corpus, and

---

<sup>2</sup>We crosschecked the congruence between differently preprocessed factor solutions, more precisely between the strategy adopted here and the strategy recommended by Zheng et al. (2003) with additional scaling in the speaker mode. This measure resulted to 0.99 and evidences an almost identical solution.

<sup>3</sup>contrary to the citation in Zheng et al. (2003).

for the PARAFAC2 model of the rate corpus<sup>4</sup>.

Note that, unlike in principal component analysis, the sum of the variances explained by single factors does not necessarily have to sum up to the total percentage of variance explained by the whole model. Table 1 summarizes these statistics. The percentage of variance explained for the PARAFAC1 in the speech rate corpus was around 80% , as already published in Hoole (1999). The amounts explained for the first and second factors amount to 61% and 24% . The fit of PARAFAC2 with respect to this dataset is slightly better. For the whole model this amounts to 82% and for the single factors to 21% and 61% respectively.

Concerning the stress corpus, we observed 86% variation explained for the total PARAFAC1 solution and 69% and 17% for the two factors separately. For PARAFAC2, the same indicators amount to 90% ,18% and 72% . Taken together, for this corpus, the benefit in explained variances by using PARAFAC2 was substantial in contrast to the speech rate corpus. The core consistency diagnostic can only be calculated for PARAFAC1 model and can take values less than or equal to 100. According to (Bro & Kiers, 2003, p. 276), a core consistency close to 100% implies an appropriate model, and, as a rule of thumb, a core consistency above 90% can be interpreted as ‘very trilinear’. Accordingly, the consistency for both solutions reported here can be seen to almost perfectly conform to the PARAFAC1 model. In other words, valid solutions seem to be warranted and we can turn to the display of conventional loading plots.

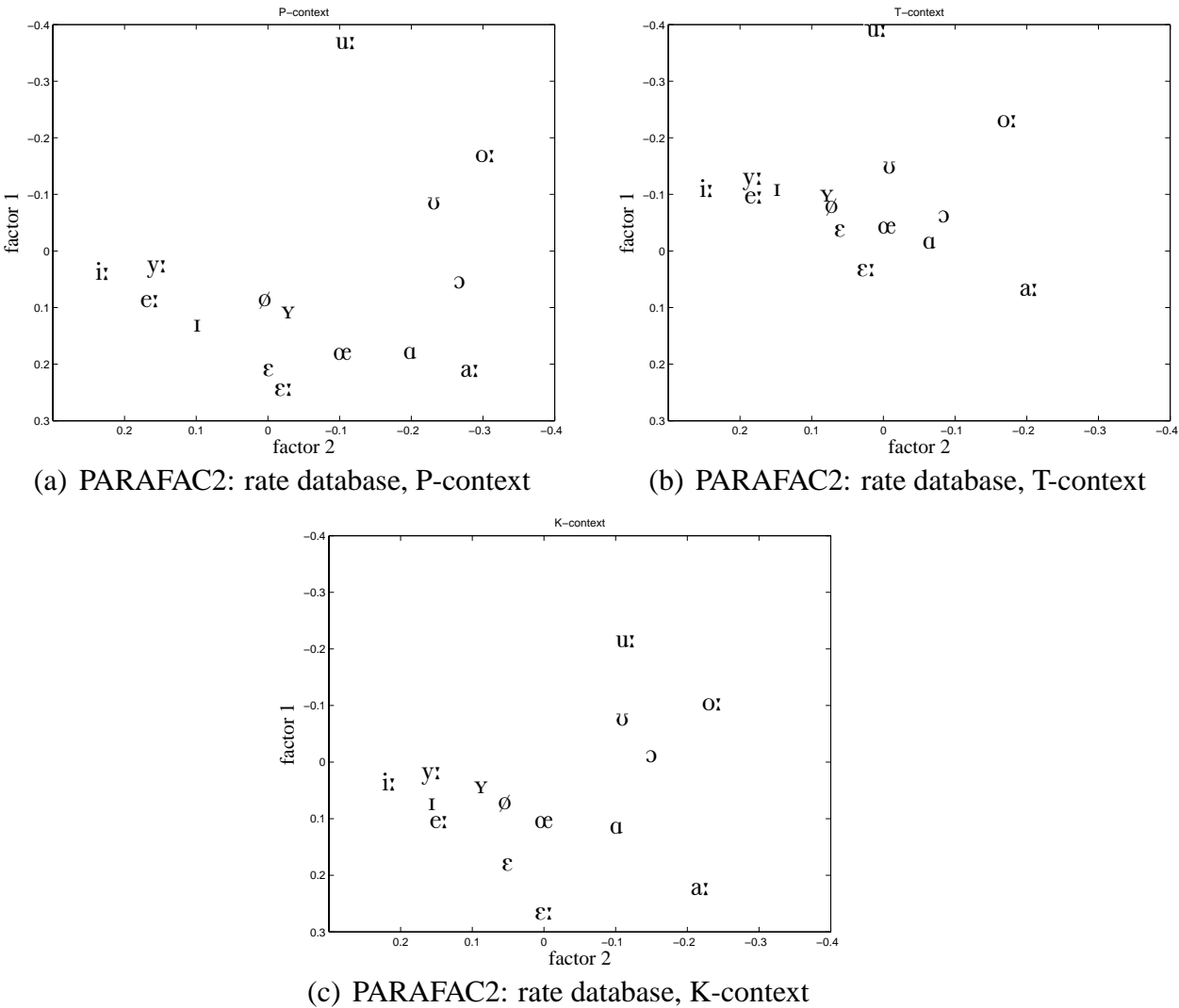
**Table 1.** Summary statistics for fitted models

	<b>Perc.expl.tot</b>	<b>Perc.expl.F1</b>	<b>Perc.expl.F2</b>	<b>CoreCond</b>	<b>Congr. tot</b>
<b>P1 rate</b>	80%	61%	24%	100	-0.05058
<b>P2 rate</b>	82%	21%	61%	-	1
<b>P1stress</b>	86%	69%	17%	98.3	0.00008
<b>P2 stress</b>	90%	18%	72%	-	1

### 3.2 Loading Plots

As can already been seen from table 1, the ordering of the factors is reversed in the PARAFAC2 solutions resulting in a second factor with a higher percentage of variance explained than the first factor. For the plots of vowel loadings, the axes were changed according to convention, i.e. with high front vowels in

<sup>4</sup>Note that constraining the first (vowel) mode to orthogonality implies constraining the second (articulator) mode. Nevertheless, congruences of around .95 for the unconstrained speaker modes were indicating non-degenerate models

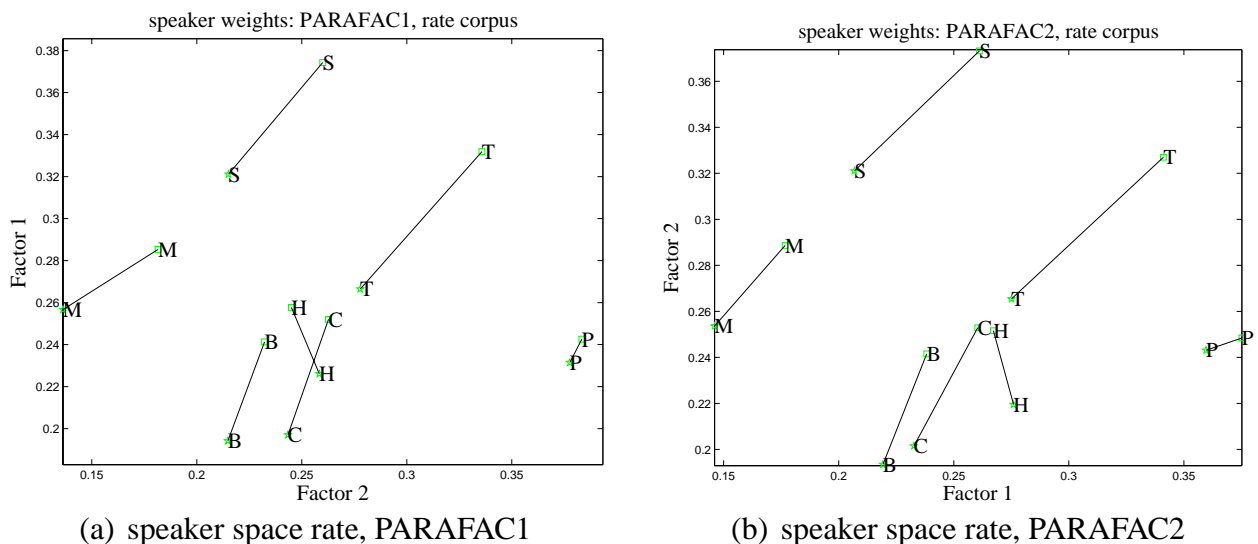


**Figure 1.** Results for the *SPEECH RATE* database for whole corpus, vowel loads

the top left corner. This is not in general the case for the plots of the speaker weights.

### 3.2.1 Speech Rate Corpus

Figure 1 shows the loading plots for the vowel mode split by consonantal contexts for the speech rate corpus. These plots can be directly compared to figure 4 in Hoole (1999). The PARAFAC2 solution can be seen as a rotated version of the vowel space as described in Hoole (1999), i.e. the topological information is retained. This implies that we do not have to discuss this aspect in further detail. Similarly for the speaker weights shown in figure 2. Here, the results for the PARAFAC1 solution are identical with the results of figure 6 in Hoole (1999). Both solutions conform to a scaling down of articulation for subjects B, C, M, S, and T in the fast rate condition, and a different behavior for speakers H and P, conforming to the fact that an increase in speech rate can be achieved by either downscaling the amplitudes of articulatory movements or by increasing movement velocities. For the current study the absence of a substantial and interpretable topological change between the two solutions is the interesting aspect.



**Figure 2.** Results for the *SPEECH RATE* database, speaker weights

In contrast to Hoole (1999), we show the loading plots for the articulator weights split by paralinguistic conditions. These plots can be seen as the effects of the factors on the tongue configurations of an average speaker. For the speech rate corpus, this information is given in figure 3. In general, both solutions cohere with the interpretation of the factors published in Hoole (1999). The



most striking result in the PARAFAC1 solution appears to be the absence of a strong difference in tongue shape between the projections at normal and fast rate, the projection at fast rate being a somewhat downscaled version of the projection at normal rate. This is slightly different for the PARAFAC2

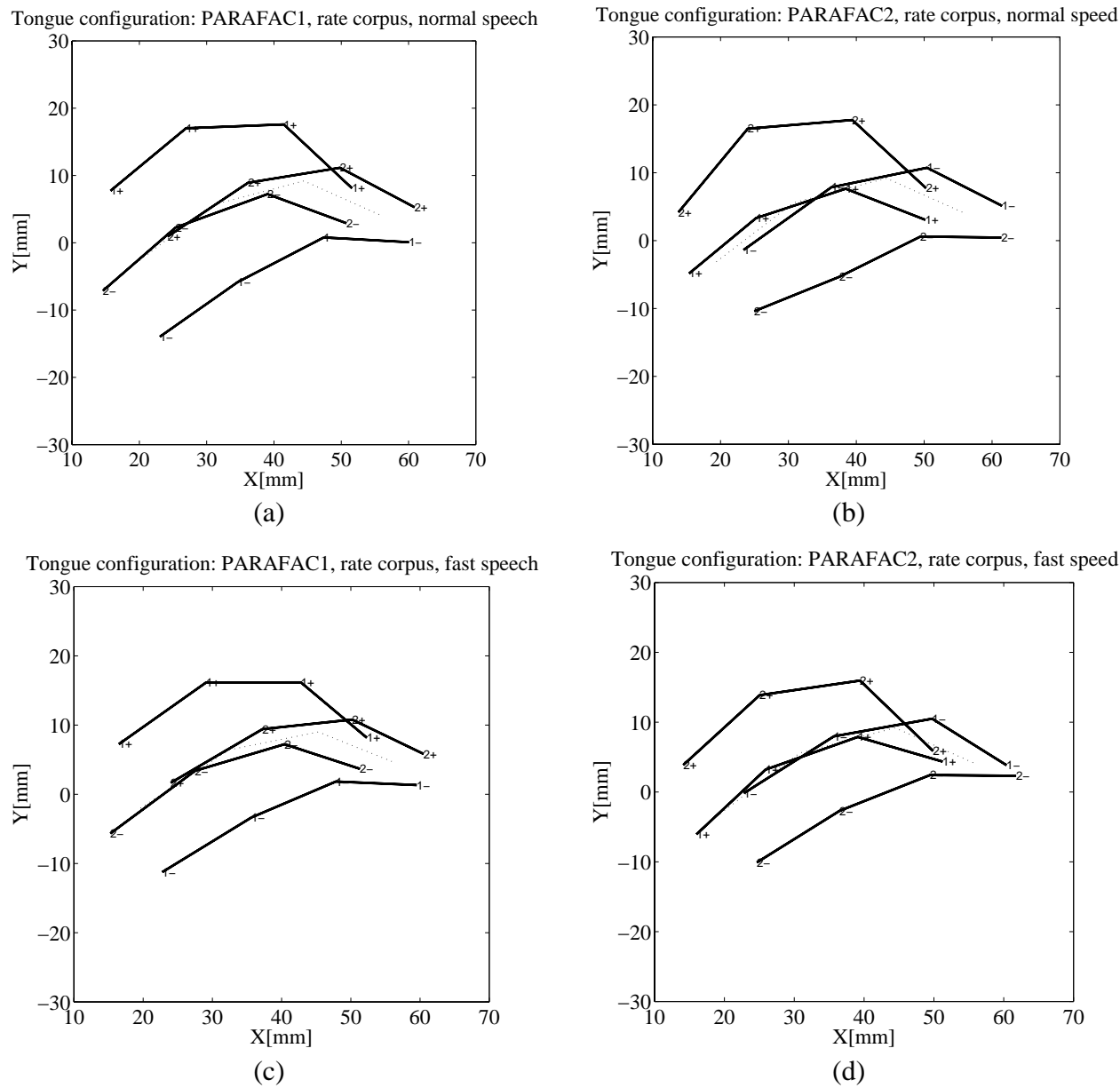
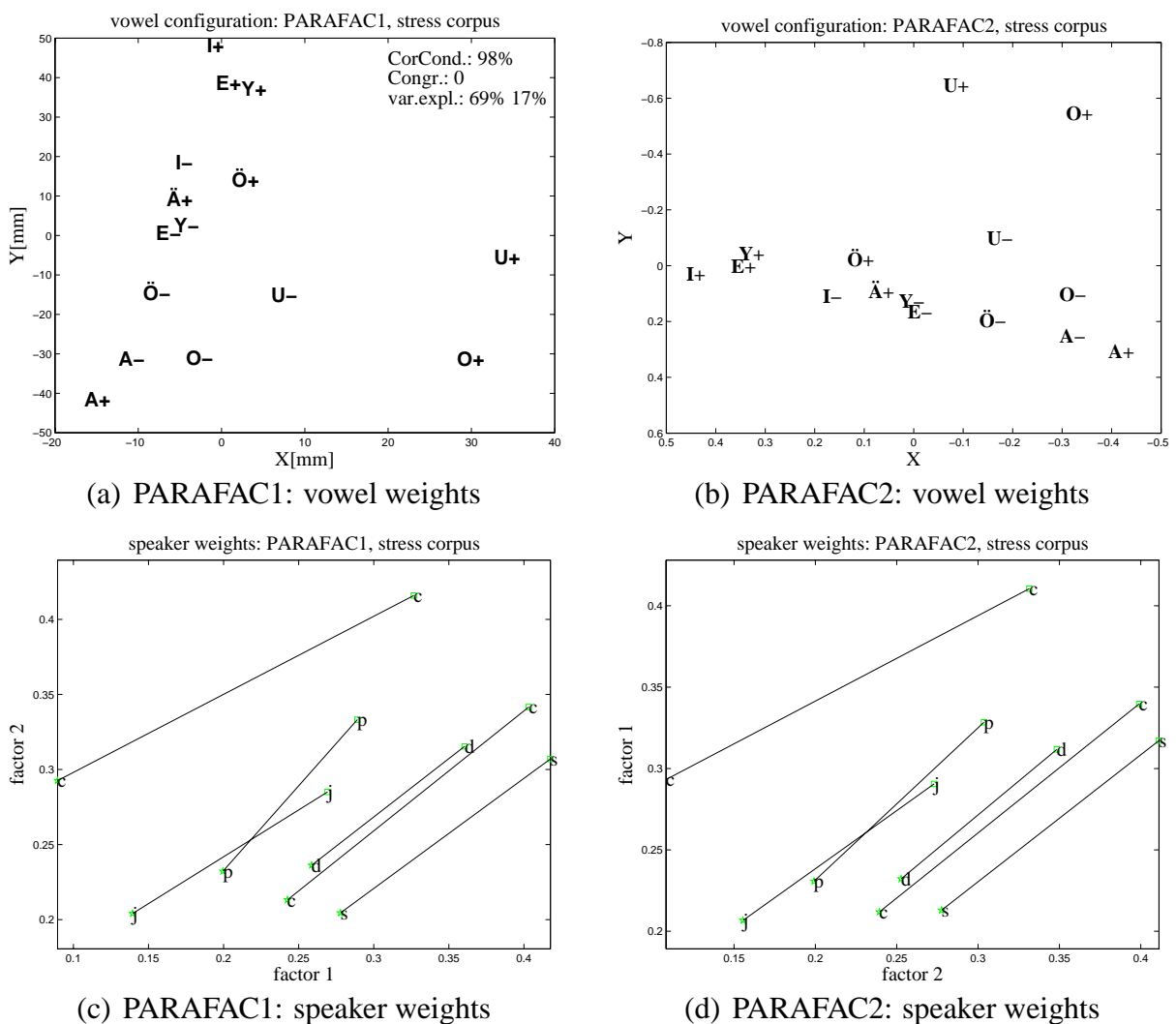


Figure 3. *speech rate* database, articulator loadings

solution. A downscaling of amplitudes is indeed observed, but additionally, there are some shape-relevant aspects in this solution worth mentioning: First, the negative shape of factor2 -corresponding to factor 1 in PARAFAC1, front-raising- in the normal-rate condition is characterized by a lower tongue blade sensor in comparison to the surrounding tongue tip and tongue dorsum sensors. If this factor is assumed to encode a movement from an /a/-like to an

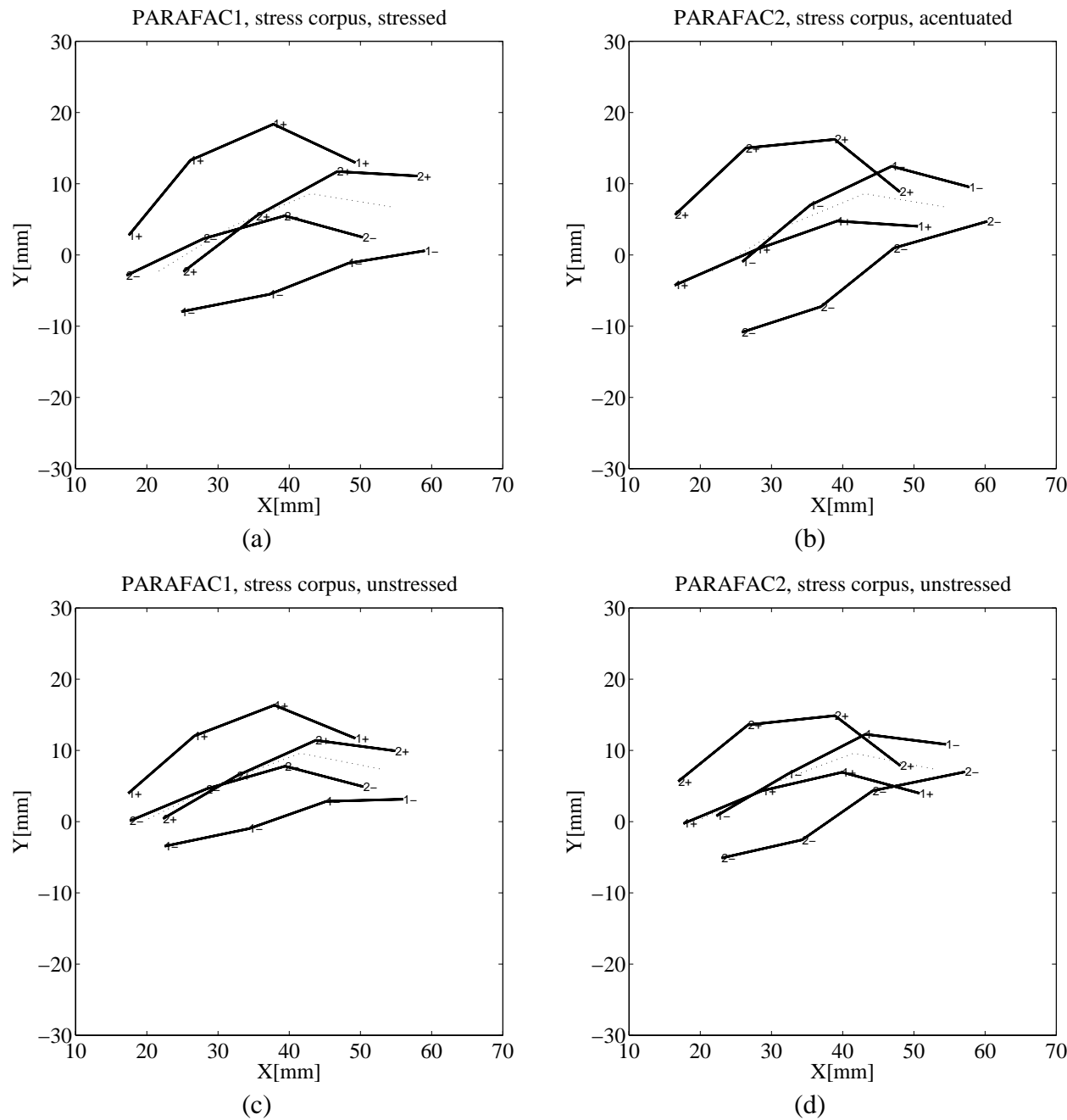
/i/-like shape, the /a/-pole of this factor appears to be a more reasonable configuration than the first factor of the PARAFAC1 solution. Subsidiary evidence comes from the comparison of the /i/-like pole of the front-raising factor at normal speech rate: The tongue tip appears to be “more down” in the PARAFAC2 solution, which as well seem to be more reasonable. Interestingly, the informative patterns with respect to factor 2 arise in the fast-rate condition. Hoole (1999, p.1026) had noted that his second factor shares with the solution found in Harshman et al. (1977) “ the responsibility for forming a constriction in the velar region, but our factor 2 shows above all a pattern of advancement and retraction, which is hardly the case for the “back raising” factor.” This tendency



**Figure 4.** Results for the Stress database. Left Panels: PARAFAC1, right panel: PARAFAC2. Tense vowels(+), lax vowels(-)

appears to be even more prominent for the tongue back sensor in the fast rate condition for the PARAFAC2 solution, where no raising movement at all is ob-

servable.



**Figure 5.** Articulatory configurations for the stress database

3.2.2 Stress Corpus

As mentioned above, for the stress corpus, both models were constrained to orthogonality in the vowel mode. The PARAFAC1 model was quite close to degeneracy, but we have shown the core consistency above (table 1) indicating an acceptable coherence with trilinearity. In short, a pattern comparable to the one

for the speech rate corpus was observed concerning vowel and speaker weights shown in figure 4: There is no evident change in topology in vowel and speaker plots comparing PARAFAC1 and PARAFAC2. With respect to the articulatory configurations, the patterns are partly similar to the speech rate corpus: The “trough-like” -shape of the tongue blade is also evident for /a/-like configuration, but is visible in both stress conditions. Interestingly, both PARAFAC2 factors have quite a strong horizontal component except for the back raising factor in unstressed condition.

### *3.2.3 Preliminary Summary*

In this paragraph, the results obtained until so far will be summarized. PARAFAC1 and PARAFAC2 solutions give comparable results with respect to speaker weights and vowel weights. This kind of topological invariance could be substantiated in a more formal way by showing that e.g. the shape difference in the vowel spaces of PARAFAC1 and PARAFAC2 solutions is uniform, i.e. only trivial translation, scaling, and rotation operations are involved. This idea is not tracked further here, but could be performed by Generalized Procrustes analysis (Gower, 1975). The articulatory configurations for the “modal speaker” with respect to the paralinguistic features seem to show enhanced “flexibility” for PARAFAC2. Nevertheless, the gain in variance explained is only substantial in the stress corpus - the “/t/-only” data set. In the next paragraph, we will apply a method to identify the most influential observations shaping the particular solutions, particularly with regard to possible biases in the speaker weights caused by the compromise quality of the PARAFAC1 solution.

### *3.3 Leverages*

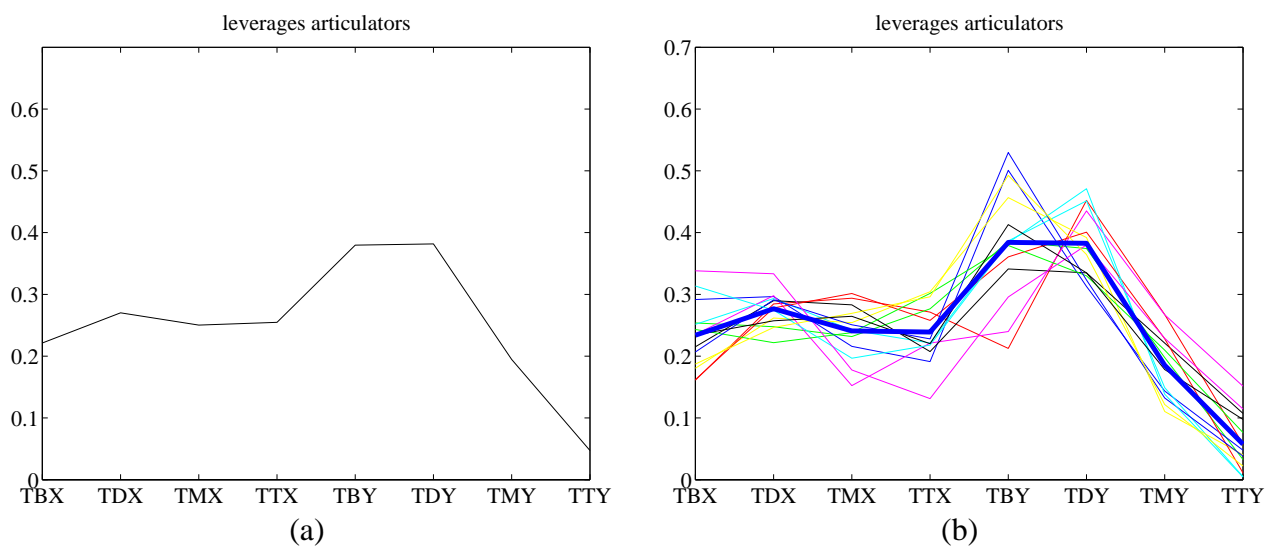
Leverages were originally developed for regression analysis as a tool for residual and influence analysis. For this reason, it might be more appropriate to speak of the “squared Mahalanobis distance” in the context of a factoring method like PARAFAC. Anyway, leverages are also widespread in two-way Principal Component Analysis, therefore the term “leverage” also seems appropriate (Bro, 1997). For a particular loading matrix, e.g. for first mode loadings, leverages can be calculated as

$$v = \text{diag}(A(A'A)^{-1}A') \quad (7)$$

Their possible range is between 0 and 1. A high value indicates that an observation is influential, a low value indicates the opposite. As mentioned already, leverages could have been calculated for each of the two factors separately for

every mode. Here, we will limit ourselves to their evaluation with respect to the factor structure as a whole. The basic results with respect to the leverages in the vowel mode point to a corpus effect: The least influential observations are the palatal vowels, the strongest contributions are made by long back vowels and /a/. This presumably is a corpus effect of the structure of the German vowel system with its numerous front vowels. Furthermore, lax vowels are generally less influential than their tense counterparts. The leverages in the speaker-mode show hardly any effect of the paralinguistic variables: Fast rate and unstressed shapes are generally less relevant for the total solution than normal rate and stressed shapes. This holds with the exception of speaker H, who is characterized by a deviant articulatory implementation of speech rate Hoole (1999).

The calculation of leverages in the articulator mode offers an additional inter-

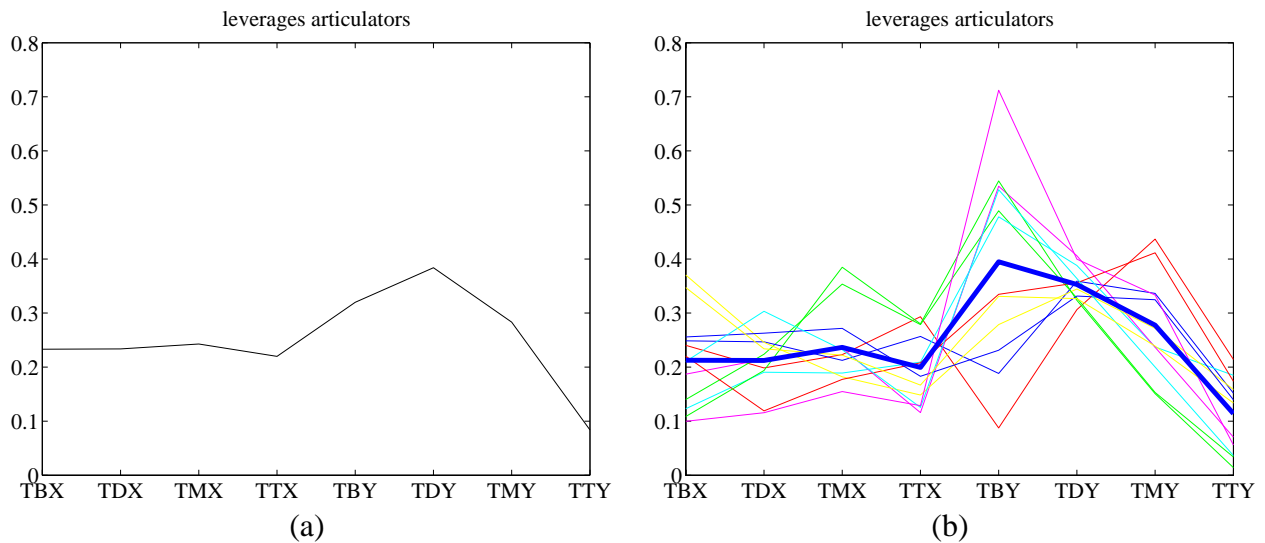


**Figure 6.** rate corpus: Leverages articulator mode, Sensor naming conventions: TB:tongue back, TD: tongue dorsum, TM: tongue mid, TT: tongue tip. x and y denote horizontal and vertical components.

esting property: Whereas for PARAFAC1, only one leverage profile can be calculated for the matrix of articulator weights, PARAFAC2 offers the possibility of calculating leverages for each speaker separately. Figure 6 and 7 show these plots for the two data sets used in this study. The left panels show the leverages for the articulator weights of the respective PARAFAC1 solution, the right panels show the leverages for the individual speakers as obtained by PARAFAC2. The bold line in the right panels depicts the average values of the single speaker's articulators for PARAFAC2.

The most evident patterns of figures 6 and 7 are (a) the general peak in influence observed for tongue back and tongue dorsum in y-direction (b) the su-

perfectionally close similarity between the average leverage profile of PARAFAC2 and the corresponding PARAFAC1 profile. This correspondence is almost perfect in the rate data set. Contrastingly, for the stress data set, a shift of the most important sensor for tongue dorsum to tongue back can be observed for PARAFAC2 in comparison to the PARAFAC1 profile.

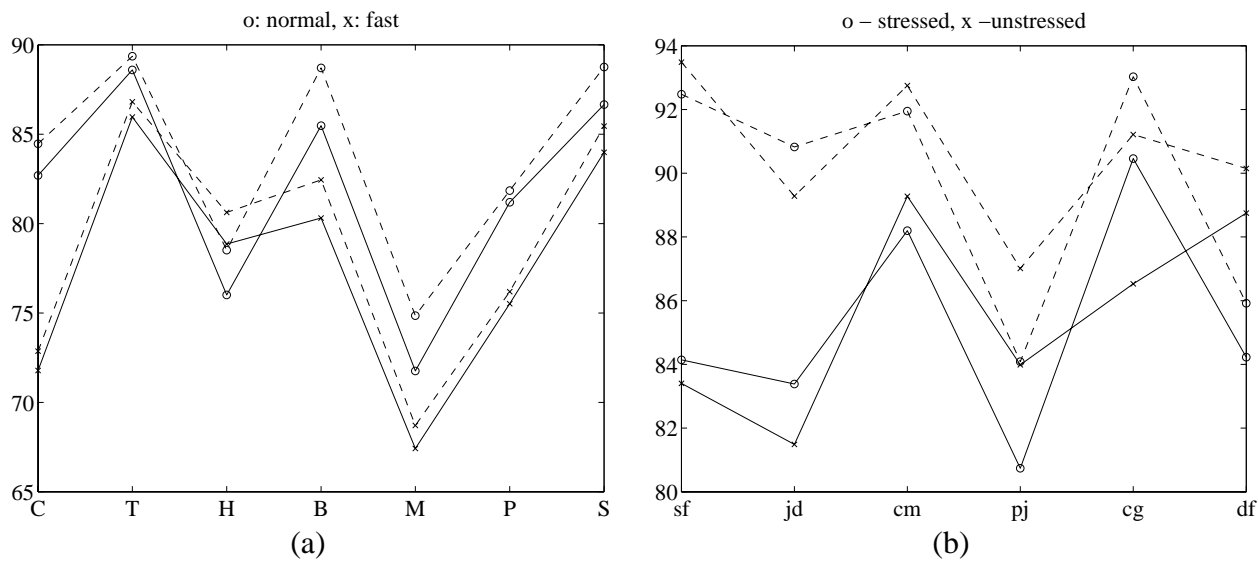


**Figure 7.** Stress corpus: Leverages for articulators. For explanation of the x-axis see figure 6.

Taken together, the patterning of the different modes in the leverage analyses is equivalent to the patterning in the loading analyses in previous sections with respect to the factoring method. This raises the following question: If the greater flexibility of PARAFAC2 to account for interindividual differences in the articulatory configurations by a separate subspace for each speaker does not relevantly influence speaker and vowel spaces, what is the origin of the increase in fit observed for the stress data set? In order to answer this question, we calculated explained percentages of variance for each speaker separately across data sets and factoring method. The result is shown in figure 8. The results show a general increase in explained variances for PARAFAC2 for both paralinguistic features in both data sets. Conforming with the results on the explained variances for the total samples, this increase in fit is less pronounced for the speech rate than for the stress corpus.

## 4 Discussion

The results of these analyses can be summarized as follows: When comparing two different PARAFAC-versions on two different corpora, findings with respect to the explained variances are better for the PARAFAC2 algorithm for



**Figure 8.** Percent explained per speaker for rate corpus (a) and for stress corpus (b). dashed lines: PARAFAC2, solid lines: PARAFAC1. left panel (a): circles indicate normal rate, crosses fast rate. right panel (b) circles indicate stressed, crosses unstressed position

both corpora. Note that the smaller impact of the choice of the modeling strategies on vowel and especially speaker weights is not precluded by the algorithmic decision made; rather it originates in the data analyzed. This is evidenced by the relatively large impact of interindividual differences on the leverages in the speaker modes in PARAFAC2 for both corpora.

We mentioned in section 2.1.3 that the data of the speech rate corpus were recorded on two different occasions implying that the sensors had to be attached twice. This does not appear to be problematic. The solution of the speech rate corpus is more stable than the solution of the stress data set: A scenario in which sensor placement differences between sessions would be recovered in individual articulator spaces would have been easily conceivable. However, this is not the case, as the speech rate data set benefits less from PARAFAC2 in terms of explained variances than does the word stress data set. Therefore, it seems to be conclusive that the greater impact of PARAFAC2 on the solution in the word stress data set can be traced to the coarticulatory influences of alveolar articulation present in these data. In this sense the word stress data set analysed here confirms the findings of Hoole (1999), where analyses of single consonant contexts ran into problems for the alveolar context. This is nicely illustrated by leverages of vowel mode loadings where the most influential observations stem from long back vowels. One need not go as far and claim that the canonical 2-

factor PARAFAC solution is largely incompatible with vowel data acquired in alveolar contexts, but often stabilizing orthogonality constraints have to be applied in order to end up with an interpretable model. In these cases, PARAFAC2 can capture speaker-specific variation more accurately than PARAFAC1.

This converges with the finding that the tongue back / tongue-dorsum y-components are the most influential observations in the data sets. If the vowel shapes are influenced by a surrounding /t/-gesture, the tongue-dorsum and tongue-back becomes articulatorily more constrained during the vowel. This leads to a decrease in the variance generated in the “backward-upward” direction and an increase in variance in “front raising”-conform direction. In other words, the “dominance” of the “front-raising” factor increases leading to a fragile second factor or even degeneracies. We think that the Procrustes-like relaxed version of the Parallel Proportional Profile as implemented in the PARAFAC2 algorithm allows for a speaker-specific definition of this backward-upward movement and prevents degeneracy as shown in the stress data set. There are other situations where this “backward-upward direction” of the tongue variance could be ill-defined across speakers and where PARAFAC2 could be a remedy against methodological pitfalls: Using “fleshpoint methods” like the magnetometer, the experimenter decides for particular landmark definitions while gluing the sensors on the tongue. Acquiring data using different methods like the three-dimensional reconstruction of MRI data as described in the paper by Zheng et al. (2003), the input data are the outputs of surface reconstruction algorithms, and the definition of landmarks is carried out at a later stage in the analysis. Here, the analysis could benefit from the fact that PARAFAC2 fits the data directly and not cross-products between column units. Therefore, it is easier to handle missing data, and in the case of three-dimensional data, the strong concept of the landmark is to some extent relaxed. It would be possible to let the dimension of the articulator mode differ from slab to slab, hence each slab  $k$  could have its own specific articulator mode dimension and thus the solution potentially depends less on a particular landmark definition.

Zheng et al. (2003) seem to have encountered morphological problems and solved them by applying advanced preprocessing strategies consisting in an implicit vocal tract length normalization (see section 1.2). We crosschecked the preprocessing recommendations published in Zheng et al. (2003) against the word stress corpus applying the vowel centering we were using in this study and in Hoole (1999). This led to practically identical solutions compared to the PARAFAC1 solution we reported in this study. For three-dimensional data however, such a preprocessing approach might be of greater benefit than for the analysis of EMA corpora, because static three-dimensional data might em-



phasize the importance of vocal tract morphology. But at the same time, in static MRI settings, the speaker weights might be more strongly biased by these methodological problems. At the moment, we only can state, that for our EMA analyses reported here, speaker weights were remarkably stable, although we still cannot offer a conclusive interpretation on what they measure.

## References

- Badin, P., Bailly, G., Rveret, L., Baciú, M., Segebarth, C., & Savariaux, C. (2002). Three-dimensional linear articulatory modeling of tongue, lips and face, based on MRI and video images. *Journal of Phonetics*, 30:533–553.
- Bro, R. (1997). PARAFAC. tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38:149–171.
- Bro, R. (1998). *Multi-way Analysis in the Food Industry Models, Algorithms, and Applications*. Ph.D. thesis, Department of Dairy and Food Science Royal Veterinary and Agricultural University Denmark.
- Bro, R. & Kiers, H. (2003). A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, 17:274–286.
- Geng, C. & Mooshammer, C. (2000). Modeling the german stress distinction using PARAFAC2. In: *Proc. 5th Speech Production Seminar*, 161–164.
- Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika*, 40:33–51.
- Harshman, R. & Lundy, M. (1984). Data preprocessing and the extended PARAFAC model. In: H. Law (ed.) *Research Methods for Multimode Data Analysis*, 216–284. New York: Prager.
- Harshman, R. A., Ladefoged, P., & Goldstein, L. (1977). Factor analysis of tongue shapes. *Journal of the Acoustical Society of America*, 62:693–707.
- Hoole, P. (1999). On the lingual organization of the german vowel system. *Journal of the Acoustical Society of America*, 106:1020–1032.
- Jackson, M. T. T. (1988). Analysis of tongue positions: Language-specific and cross-linguistic models. *Journal of the Acoustical Society of America*, 84:124–143.
- Maeda, S. (1979a). An articulator model based on a statistical analysis. In: J. Wolf & D. Klatt (eds.) *Speech Communication Papers*, 67–70. Acoustical Society of America, NY.

- Maeda, S. (1979b). Un model articulatoire de la langue avec des composantes lineaires. *10mes journées d'études sur la Parole, Groupe Communication Parle*, 152–164.
- Maeda, S. (1990). *Speech Production and Speech Modeling*, chapter Evidence from the Analysis and Synthesis of Vocal Tract Shapes using an articulatory Model, 131–149. Behavioural and Social Sciences. Hardcastle, J. and Marchal, A.
- Maeda, S. (2005). Face models based on a guided pca of motion-capture data: Speaker dependent variability in /s/-/?/ contrast production. *ZAS Papers in Linguistics*, 40:95–108.
- Nix, D. A., Papcun, M. G., Hodgen, J., & Zlokarnik, I. (1996). Two cross-linguistic factors underlying tongue shapes for vowels. *Journal of the Acoustical Society of America*, 99:3707–3718.
- Tucker, L. (1966). Some mathematical notes on three-way factor analysis. *Psychometrika*, 31:279–311.
- Zheng, Y., Hasegawa-Johnson, M., & Pizza, S. (2003). Analysis of the three-dimensional tongue shape using a three-index factor analysis model. *JASA*, 113:478–486.