

On the articulation, aerodynamics, and acoustics of voiceless consonant clusters

Susanne Fuchs^{1*}, Laura L. Koenig^{2,3}

¹Zentrum für Allgemeine Sprachwissenschaft (ZAS), Jägerstr. 10-11, 10117 Berlin
Germany

²Haskins Laboratories, 300 George Street, New Haven CT 06511 U.S.A.

³Long Island University, Department of Communication Sciences and Disorders, 1
University Plaza, Brooklyn, New York 11201-5372 U.S.A

fuchs@zas.gwz-berlin.de, koenig@haskins.yale.edu

Abstract. *This investigation aims towards a better understanding of the relations between intraoral pressure variations, articulation (tongue palatal contact changes), and acoustic events in long voiceless consonant clusters. We are in general interested to describe these relations globally, but also with respect to speaker specific realizations and their potential causes. To do so, a new experimental set-up has been developed combining Electropalatography and a piezoresistive pressure transducer. It guarantees a more comfortable recording in comparison to previous experimental designs. We recorded five German native speakers with a corpus of homorganic clusters. Here we present a subset including the following target clusters: /ft#t/, /ft#f/, /ft#tʃ/, /ft#ft/, /ft#st/, /ft#ts/, /st#ft/, /ts#ft/ and /tʃ#ft/. Results so far provide evidence that: (1) there is a global trend that intraoral pressure correlates more strong with tongue-palatal contact patterns in the front-back dimension (Centre of Gravity index for EPG) than with overall amount of contact (PC). In particular rising pressure peaks in the cluster are tightly linked with front articulation and oral closure for the stop, but not with more back articulation and a smaller oral cavity. (2) Individual realizations of the clusters vary to a great extent which is observed in the articulatory and aerodynamic data.*

1. Introduction

Current knowledge on the relation between articulation, aerodynamics, and acoustics is limited, in part, by a lack of sufficient experimental evidence: Often, data are acquired in one or two domains and assumptions are made about the others. By means of a new experimental set-up this study does a first attempt to record speech articulation, aerodynamics (intraoral pressure) and acoustics in a relatively comfortable way.

*Supported by a grant from the German Research Council (DFG) GWZ 4/8-1 P1.

It is well-known that in voiceless consonant clusters laryngeal events have to be well synchronized with supralaryngeal ones to meet the aerodynamic requirements. Variation in glottal opening of voiceless consonant clusters has been extensively studied by Yoshioka et al. (1981) for American English, by Löfqvist & Yoshioka (1980) for Swedish and by Ridouane et al. (2006) for Berber. All of these studies found that, for longer clusters, peak glottal opening varies in accordance with the relevant segment. Namely, the peak glottal openings occur in the fricative segment of the cluster. Ridouane et al. (2006) found that a cluster with n fricative segments are most often produced with n glottal opening peaks, unless the fricatives are adjacent. One of the questions they raised was whether all glottal opening peaks in a cluster are controlled by laryngeal muscle activity or whether some of the peaks could be a result of supralaryngeal articulation and intraoral pressure variation. Our study is designed to provide detailed information on how supraglottal articulation and intraoral pressure are related, and the results could be combined with modelling work in the future to explore such an issue.

The aims of this study are: (1) to investigate the relation between changes in supralaryngeal articulation (by means of tongue-palate contacts), intraoral pressure and acoustic landmarks in long consonant clusters, (2) to discuss individual control strategies in voiceless cluster articulation and aerodynamics.

2. Methods

2.1. Speech material

In order to investigate the interplay between articulation, aerodynamics and acoustics in long voiceless consonant clusters we constructed a corpus of 42 target words with frequent consonant clusters and affricates of German.

Table 1: Speech material

| | Transcription | Phonological terms | Fricative=F, Stop=S, Word boundary # |
|-----------------------------|----------------------|---------------------------|---|
| <i>Verb#Compound</i> | | | |
| Nascht Tschadende | ʃt#tʃ | Cluster#affricate | FS#SF |
| Nascht Zanderende | ʃt#ts | Cluster#affricate | FS#SF |
| Nascht Stalinende | ʃt#st | Cluster#cluster | FS#FS |
| Nascht Stachelende | ʃt#ʃt | Cluster#cluster | FS#FS |
| Nascht Taschenende | ʃt#t | Cluster#consonant | FS#S |
| Nascht Schafende | ʃt#ʃ | Cluster#consonant | FS#F |
| <i>Compound</i> | | | |
| Maststelle | ʃt#ʃt | Cluster#cluster | FS#FS |
| Matzstelle | ts#ʃt | Affricate#cluster | SF#FS |
| Matschstelle | tʃ#ʃt | Affricate#cluster | SF#FS |

We focused on alveolars and postalveolars since we used Electropalatography (EPG) as a technique to study supralaryngeal articulation. We avoided bilabials and velars since they do not provide tongue-palate contact patterns in the EPG signal. In the current

work we will only present a subset of the original corpus as depicted in table 1. All target words were embedded in a carrier sentence and repeated 10 times by each subject. The target words are all real words of German, but most of the compounds we constructed do not exist in the German vocabulary. We also note that in German word initial $>st<$ is commonly pronounced as $/ʃt/$. There are a few exceptions grounded in non-native vocabulary. We included the word ‘Stalinende’ and instructed the subjects to realize it with initial $/st/$ which is a possible pronunciation in German, and the original pronunciation in Russian. In the compounds the first word is always stressed. However, both words have lexical stress too. Final clusters occur in coda position and initial clusters in onset position.

2.2. Experimental set-up

Tongue-palate contact patterns were recorded by means of EPG (Reading system, EPG 3) with a sampling rate of 100 Hz. Intraoral pressure changes were recorded synchronously by means of a piezoresistive pressure transducer (Endevco 8507C-2) which was fitted to the posterior end of the artificial palate via a flexible plastic tube (see Figure 1). The pressure sensor is about 2.4 mm in diameter and has a length of 12 mm. The sensor measures a pressure difference between intraoral and atmospheric pressure. Atmospheric pressure was sensed via a small plastic tube smaller in diameter than the bundled wires of the EPG palate. This new experimental set-up allows one to monitor tongue-palate contacts simultaneously with pressure variations. It is easier to apply in comparison to set-ups like tube insertion through the nose and it is not affected by saliva blocking the tube. The pressure data were acquired using PCQuirer version 5.0 at a sampling rate of 1859 Hz and subsequently imported into Matlab for processing. The data were smoothed with a 6th-order Butterworth filter using a 43 Hz cut-off so that low-frequency changes in the pressure could be monitored.

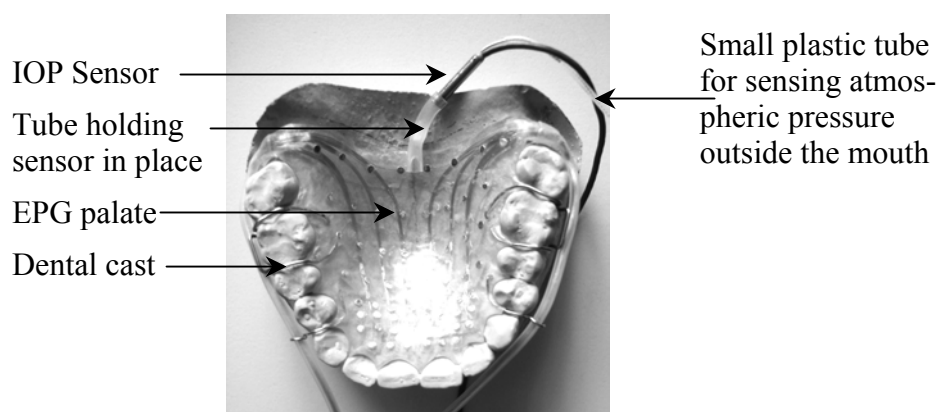


Figure 1. EPG palate with the pressure sensor in a plastic tube fixed at the posterior end of the palate.

For comparison with the EPG data, the region around the cluster was also extracted from the larger files. High quality acoustic data were simultaneously recorded on DAT with a sampling frequency of 48 kHz. Five German native speakers, three males (DP, JD, RW) and two females (SF and SK) participated in the experiment. None of them had any known history of speech, language or hearing disorders. They are all colleagues from the ZAS lab.

2.3. Acoustic labeling procedure and calculation of EPG parameters

Since the acoustic signal has the highest sampling frequency and therefore the greatest temporal precision it served as the basis for defining landmarks. An interval of the whole cluster was defined from the first consonant to the last. Additionally, major acoustic events like bursts for the stops and frication on- and offsets for the fricatives were labeled.

For the EPG data we calculated 2 different parameters, first the overall percentage of contact (hereafter PC) and second, the Centre of Gravity index (hereafter COG) as a weighted index in the front-back dimension (Hardcastle et al. 1991). Higher COG values correspond to a more front articulation.

3. Results

The results section is structured as follows: We will first provide a brief summary, in particular of the reduction phenomena (3.1.), since the actual realization of the speech material varied between the subjects. In section 3.2. articulatory, aerodynamic, and acoustic results for two subjects will be presented in detail. In section 3.3. we will focus on articulatory and acoustic data of SK, since the results of this subject differ from the others due to her careful pronunciation. SF and RW are not included due to space limits, but they behave similar to JD.

3.1. Actual realization of consonant clusters

In general, reduction phenomena occur most frequently in coda position, i.e. in many cases the burst in final /t/ in /ʃt/ is not visible in the acoustics.

(1) Identical segments that are separated by a word boundary are most frequently realized as one, comparable with geminates. If the segment is a stop as in /ʃt#tʃ/, /ʃt#t/, and /ʃt#ts/ the acoustic signals show a relatively long closure duration and a burst in the word initial segment. If the segment is a fricative as in /tʃ#ʃt/ frication continues from one to the next without any obvious amplitude reduction or weakening. If the final stop was deleted as in some cases for /ʃt#ʃ/, /ʃt#st/, /st#ʃt/, and /ʃt#ʃt/ frication continued from the fricative in the coda cluster to the initial one. Very often it could not be separated clearly, but a weakening in amplitude was visible in the acoustic signals.

(2) Concerning inter-individual variation, the most careful pronunciation was produced by subject SK. All other subjects show reduction phenomena quite frequently, in particular JD and SF.

(3) The most variable pronunciation in general was found in the cluster /ʃt#ʃt/ which was realized as /ʃ#ʃt/ (deletion of 1st final /t/), /#ʃt/ (deletion of final /ʃt/), /ʃt#ʃ/ (deletion of 2nd final /t/). This extreme variation could be interesting within the discussion of the realization of speech errors. Pouplier (submitted) pointed to the role of shared gestures triggering speech errors. Thus, there may be two explanations for cluster reduction in /ʃt#ʃt/. First, it may be perceptually tolerable since the same cluster occurs on both sides of the word boundary, similar to the cases where /t/ is on both sides

of the word boundary and gets reduced to one. Second, however, it could also be a mechanism to avoid speech errors which are very likely when the same gestures are realized in adjacent words

3.2. Tongue-palate contacts, intraoral pressure changes, and acoustic landmarks

In a first step we analyzed the EPG parameters PC and COG within the acoustically defined interval in order to check the intra-individual articulatory variation from one repetition to the next. Except for subject SK, all speakers showed similar patterns in their EPG data for the different repetitions. Therefore, the following graphs are based on ensemble averages of the EPG parameters and the intraoral pressure changes. For subject SK repetitions could be divided categorically into those produced with careful speech and those produced with more casual speech. Within the two categories the articulation did not vary to a great extent, so that ensemble averages were calculated separately for SK's careful and casual speech (see also Figure 4).

Subjects DP and JD

Figure 2 shows the time normalized, ensemble averaged pressure values for tongue palatal contact changes and intraoral pressure of subject DP. For visualization purposes the EPG-COG values were multiplied by 15 and the IOP values by 0.1. The figure also shows the relative temporal locations of the burst and frication on- or offset. These represent means of at least 6 tokens (excluding cases where some acoustic landmarks were missing or could not be measured reliably). For each token, the landmarks were converted into a location as a percentage of 100% of the cluster duration. The figures show the averaged relative landmarks for all available tokens for the respective utterance.

A pearson correlation coefficient was calculated with SPSS (version 11.5) for each speaker and utterance type pooling the 100 data points of each of the 9 target words together ($n=900$). We found that intraoral pressure correlates to a greater extent with the front versus back articulation (IOP with COG: $R=0.696$) than with the overall amount of contact (IOP with PC: $R=0.599$). Results of subject JD shown in figure 3 show a similar picture with $R=0.768$ for a correlation between IOP with PC and $R=0.834$ for IOP with COG. In order to test whether PC or COG are more closely related to intraoral pressure we additionally carried out a Wilcoxon-test for DP and JD with the two correlation coefficients for each cluster. It turned out that although the trend goes towards a closer relation between IOP and COG than IOP and PC, differences are not significant.

The temporal location of the pressure peak coincides with oral closure followed by oral release (the burst of the stop) in the acoustics and with a maximum of the EPG-COG value, corresponding to the most front articulation. That means that for casual (not careful) speech n pressure peaks coincide with n stops unless they are adjacent. Examples of 2 pressure peaks can be clearly seen in the clusters /st#ʃt/, /tʃ#ʃt/, /ts#ʃt/, and /st#ʃt/. The realization of the compounds differs individually. In DP intraoral pressure stays relatively stable for the consonants on both sides of the word boundary whereas in JD's data pressure clearly rises towards the cluster in onset position.

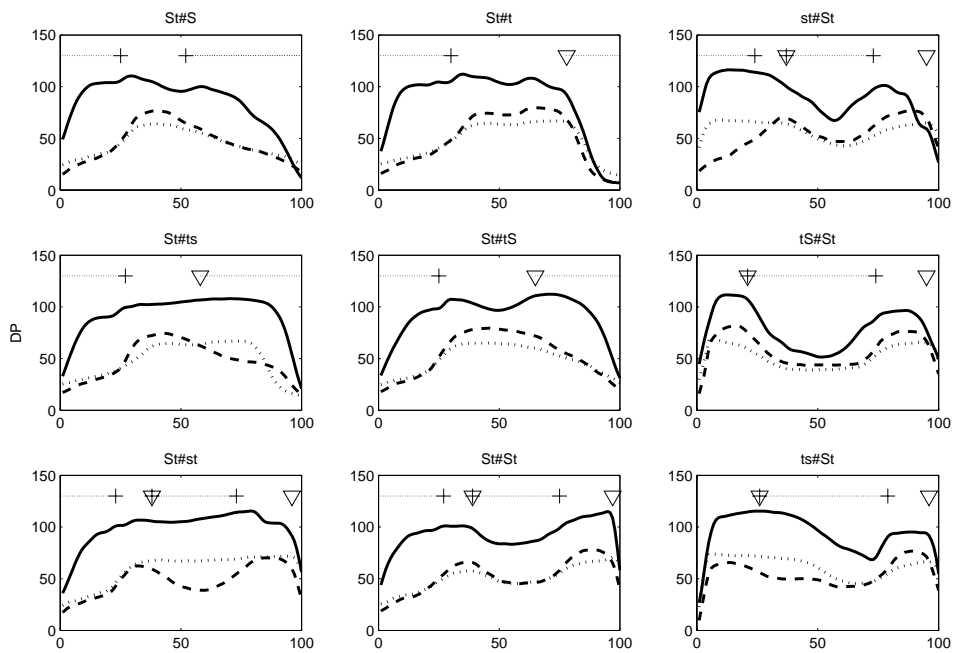


Figure 2. Time normalized, ensemble averaged IOP values (bold line), EPG-COG values (dotted line), EPG-PC values (dashed line) for DP; averaged acoustic landmarks '+' = frication on- or offset, 'v' = burst.

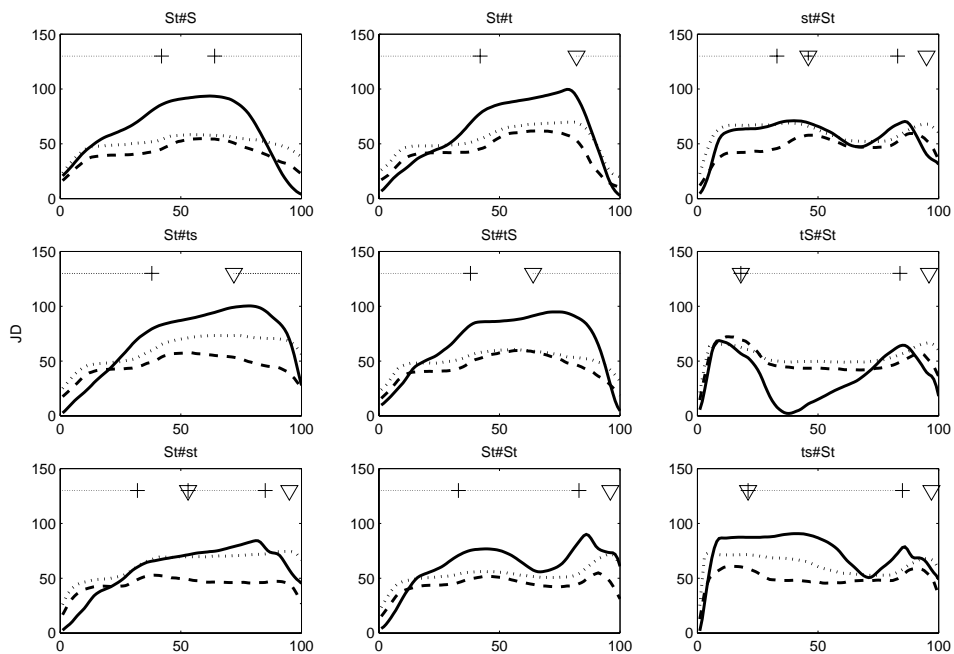


Figure 3. Time normalized, ensemble averaged IOP values (bold line), EPG-COG values (dotted line), EPG-PC values (dashed line) for JD; averaged acoustic landmarks '+' = frication on- or offset, 'v' = burst.

These effects are less obvious in the EPG data (see also Figure 4), which may be explained due to the fact that /s/ and /ʃ/ are known to be relatively resistant to coarticulation or boundary effects. However, a more detailed analysis is necessary to derive firm conclusions on this point.

3.3. Tongue-palatal contacts and acoustic landmarks

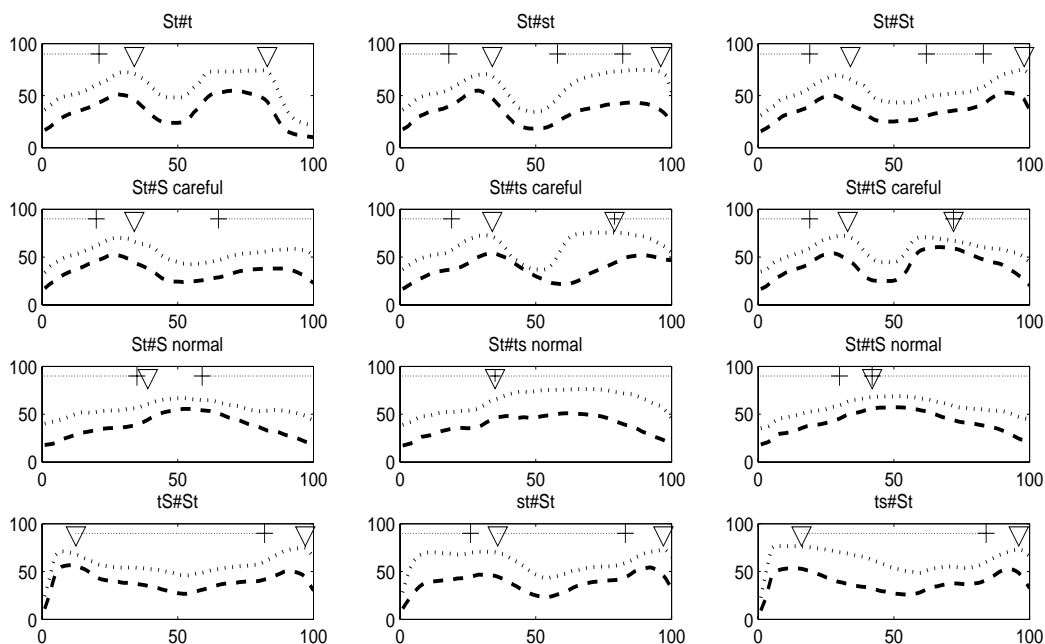


Figure 4. Time normalized, ensemble averaged EPG-COG values (dotted line), EPG-PC values (dashed line) for SK; averaged acoustic landmarks '+' = frication on- or offset, 'v' = burst.

As noted above, SK is the subject with a clear categorical distinction between careful and casual pronunciation. In her careful speech, SK does not reduce the clusters at all and realizes for instance both /t/'s in /st#tʃ/ resulting in 2 peaks (EPG-COG and PC) for the stops. Although less frequent, SK produces casual speech mostly in clusters with /t/'s on both sides of the boundary which get reduced to one /t/ and one peak. Similar to all other subjects, peaks in the observed EPG patterns can be associated with the alveolar stop in the cluster and both articulatory parameters are highly correlated and behave in parallel over the whole consonant cluster.

4. Summary and future work

This study does a first attempt to understand the relations between intraoral pressure variations, articulation (tongue palatal contact changes), and acoustic events in long voiceless consonant clusters. In general we found strong correlations between intraoral

pressure changes and the overall amount of tongue palate contact as well as with contacts in the front-back dimension. The latter (EPG-COG) showed however a stronger correlation. More front articulation coincides with greater intraoral pressure which may also be a result of our speech material, since we did not include velar stops. Pressure peaks occurred in stop production at the boundary between oral closure and oral release (burst in the acoustics). The actual realization of the clusters varied from one speaker to the next. Clusters in coda position are often reduced, in particular the alveolar stop. The greatest variation and reduction was found in the target word where the same cluster occurred on both sides of the boundary, in /st#ʃt/.

In a next step we will analyze the clusters in more detail by: (1) considering the realization of the elements of the cluster, (2) studying the effect of stress and boundaries, (3) comparing individual production and their potential causes.

5. Acknowledgements

We thank Jörg Dreyer for the development of the experimental set-up, Ralf Winkler & Mark Tiede for computational effort for PCquirer file conversion, Anke Busler for acoustic analysis, Phil Hoole and Jorge Lucero for information on filtering, and Tine Mooshammer and Jim Scobbie for comments of an earlier version of this paper. This work is dedicated to Dieter Fuchs.

References

- Hardcastle, W.J., Gibbon, F. & Nicolaidis, K. EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics* 19: 251-266, 1991.
- Löfqvist, A. and Yoshioka, H. Laryngeal activity in Swedish obstruent clusters. *JASA* 68(3): 792-799, 1980.
- Pouplier, M. Complex frequency relationships as error triggers. *Submitted*.
- Ridouane, R., Fuchs, S. & Hoole, P. Laryngeal adjustments in the production of voiceless obstruent clusters in Berber. In Harrington, J. & Tabain M. (Eds.): *Speech production: Models, Phonetic Processes and Techniques*. Psychology Press: New York, 275-297, 2006.
- Yoshioka, H., Löfqvist, A. & Hirose H. (1981) Laryngeal adjustments in the production of consonant clusters and geminates in American English. *JASA* 70(6): 1615-1623, 1981.