ZASPiL Nr. 49 – Januar 2008

Papers in Phonetics and Phonology

Marzena Zygis & Susanne Fuchs
(Eds.)

Table of contents

Addresses of contributors

Anna Bloch-Rozmej
Department of Celtic, Institute of English
John Paul II Catholic University of Lublin
Aleje Raclawickie 14
20-950 Lublin
Poland
Email: abloch@kul.lublin.pl

Claire Brutel-Vuilmet
GIPSA-lab (ICP)
46 Avenue Felix Viallet
38031 Grenoble Cedex 01
France
Email: brutel@icp.inpg.fr

Laura Downing
ZAS
Schützenstr. 18
10117 Berlin
Germany
Email: downing@zas.gwz-berlin.de

Susanne Fuchs
ZAS
Schützenstr. 18
10117 Berlin
Germany
Email: fuchs@zas.gwz-berlin.de

Silke Hamann
Utrecht Institute for Linguistics OTS
Janskerkhof 13a
3512 BL Utrecht
The Netherlands
Email: silke.hamann@let.uu.nl

Elke Kasimir
ZAS
Schützenstr. 18
10117 Berlin
Germany
Email: kasimir@zas.gwz-berlin.de

Grzegorz Nawrocki
State Higher Vocational School (SHVS) in Tarnow
Institute of the Humanities
Mickiewicza 8
33-100 Tarnów
Poland
Email: annan1@wp.pl

Cédric Patin
Laboratoire de **P**honétique et **P**honologie
UMR 7018, CNRS/Sorbonne-Nouvelle
19 rue des Bernardins
75005 Paris
Email: cedric.patin@gmail.com

Daniel Recasens
Universitat Autònoma de Barcelona and Institut d'Estudis Catalans
Edifici B
Campus de la UAB
08193 Bellaterra
Spain
Email: daniel.recasens@uab.es

Marzena Zygis
ZAS
Schützenstr. 18
10117 Berlin
Germany
Email: zygis@zas.gwz-berlin.de

Rate effects on aerodynamics of intervocalic stops: Evidence from real speech data and model data

Claire Brutel-Vuilmet

GIPSA-Lab, Département Parole et Cognition (ICP), Grenoble, France CNRS - UMR 5216

Susanne Fuchs

Zentrum für Allgemeine Sprachwissenschaft, Berlin

This paper is a first attempt towards a better understanding of the aerodynamic properties during speech production and their potential control. In recent years, studies on intraoral pressure in speech have been rather rare, and more studies concern the air flow development. However, the intraoral pressure is a crucial factor for analysing the production of various sounds.

In this paper, we focus on the intraoral pressure development during the production of intervocalic stops.

Two experimental methodologies are presented and confronted with each other: real speech data recorded for four German native speakers, and model data, obtained by a mechanical replica which allows reproducing the main physical mechanisms occurring during phonation. The two methods are presented and applied to a study on the influence of speech rate on aerodynamic properties.

1   Introduction

Even if aerodynamics in speech production is getting a matter of particular interest, the development of the intraoral pressure is rarely considered in comparison to air flow variations. However, from a physical point of view, this parameter is crucial, particularly in the production of obstruents or for devoicing.

Indeed, a certain amount of intraoral pressure is a necessary requirement in the production of obstruents. If it is missing, as for instance in cleft palate speech (Gibbon & Lee, accepted), it can either cause a distortion in the production and perception of the intended sound or a total reorganization of the

production of the sound. For instance, cleft palate speakers often try to compensate for the lack of high intraoral pressure by means of a retracted place of articulation. Some authors even go so far to assume that during speech we could aim at aerodynamic goals, next to acoustic or articulatory ones (Warren et al. 1992, Huber et al. 2004).

Moreover, the intraoral pressure (or rather the transglottal pressure difference) is a crucial factor to understand the realization of vocal fold phonation (Hertegård et al. 1995). In obstruents, the increase of intraoral pressure associated with the occlusion of the vocal tract can entail the stop of the vocal folds vibration. This factor should also be dependent on the duration of the segment. The higher the intraoral pressure and the longer the segment, the more likely is the devoicing of sounds. On the contrary, oscillations can be maintained when the intraoral pressure does not increase to a large extent.

Maximal intraoral pressure in obstruent production can be reached when the glottis is wide open and intraoral pressure equals subglottal pressure (assuming an oral occlusion and the complete closure of the velar port). If the glottis is only spindle shaped or closed, intraoral pressure may not reach this maximum. Hence, the extent of intraoral pressure is not only a result of the moving supralaryngeal articulators, but also a consequence of the laryngeal-oral coordination.

In addition, the amount of peak pressure varies with manner of articulation as well as place of articulation (Fuchs & Koenig 2006). The more posterior the articulation of an obstruent the higher the pressure, and the wider the constriction in comparison to oral closure, the lower the pressure.

In summary, intraoral pressure is an important factor for the voicing or devoicing of a segment. It especially results from laryngeal-oral coordination, from the duration of a segment and from manner and place of articulation. These factors change under various speech conditions. Our study will concentrate on the intraoral pressure evolution in different speech rate conditions, i.e. on the impact of temporal parameters.

1.1 Speech rate

We do not intend to provide an extensive overview on investigations considering speech rate since a tremendous amount of work has already been published on this topic. We will only exemplify some ideas and results from the literature which are interesting for the topic of our study.

Weitkus 1931 (cited in Pfitzinger 2001, p. 131) analysed a corpus of spoken utterances with 3 levels of speech rate (slow, normal and fast) and separated all phonemes in two groups. First, all consonants except the voiceless fricatives, are *relatively shortened* with increasing rate and second, all vowels,

diphthongs, and voiceless fricatives are *relatively lengthened* in faster speech. Note that all segments are shorter in their absolute values at higher speech rate, but that is not the case if one considers the relative length of the relevant segment, normalized at the syllable or word level. Thus, speech rate has different effects on different segments which may cause a reorganisation of the segments in an uttered word.

By means of electropalatographic data Byrd and Tan (1996) investigated four levels of speech rate (which they called 'normal', 'medium', 'faster', 'fastest') in heterosyllabic sequences. They proposed different articulatory strategies which could underlie variations in speech rate. As one potential strategy they suppose that the faster the speaking rate, the stronger the *gestural overlap* of two adjacent segments (with no shortening of gestures). As another possibility they considered *gestural shortening* (without differences in the amount of gestural overlap) in faster speech. A combination of the two mechanisms was discussed too and additionally, a *reduction of the movement amplitude* (spatial reduction) was taken into consideration.

In addition to some speaker-specific variations, Byrd and Tan consistently found a temporal shortening and an increase in the amount of coarticulation in faster speech rates. More speaker-specific behaviour was reported with respect to spatial reduction of articulatory gestures at higher rates.

We suppose that there are aerodynamic consequences for the different strategies: The greater the gestural overlap and the spatial reduction of the closing gesture, the smaller the amount of intraoral pressure and the more likely the maintenance of vocal fold oscillation. In contrast, gestural shortening without an increase in gestural overlap of adjacent segments or a spatial reduction probably has less effect on the amount of intraoral pressure. There is no general agreement in the literature (see e.g. Pfitzinger, 2001) to explain the articulatory results under varying speech rate conditions (e.g. Tuller & Kelso 1984 proposing the invariance of the relative timing between articulators; Edwards, Beckman & Fletcher 1991 relating articulatory behaviour under varying speech rate (lengthening effects) to differences in the degree of stiffness; Lindblom (1963) assuming the so called 'target undershoot'). However, whatever explanation one might follow, one fact seems to be quite consistent – the articulatory strategies used are highly speaker specific and should have an effect on the development of intraoral pressure.

To summarise: (1) Speech rate can affect various segments in a different way. (2) Speech rate goes hand in hand with a temporal reduction of segments, but not necessarily with a spatial reduction. (3) Articulatory strategies used are highly speaker specific. (4) Speech rate should affect the aerodynamic properties during obstruent production. In the next section a review regarding previous investigations is provided.

1.2 Aerodynamics and varying speech rate

Arkebauer et al. (1967) reported averaged intraoral pressure peaks of selected consonants uttered under 3 rate conditions by 10 adults. They found higher intraoral pressure peaks in the faster rate condition in comparison to the slower condition when the data for all speakers were pooled together. They also mentioned that their results on speech rate may be confounded by changes in intensity.

Brown et al. (1969) investigated the relation of intraoral pressure to oral cavity size in various consonants and conditions. They did not only record intraoral pressure, but also the size of the vocal tract by means of x-ray. Among other factors, they asked their 15 adult subjects to vary speech rate (1 utterance per second & 3 utterances per second). Intraoral peak did not differ significantly between the two rate conditions.

Malécot (1969a,b) studied intraoral pressure under various speech conditions. He provided averaged data of 10 subjects uttering isolated bisyllabic words in a slow, average and fast condition. Malécot found only sporadic rate effects in intraoral pressure impulse (calculated as the area in mm between pressure curve and baseline) of the initial and final consonants, but significant effects for the consonants in intervocalic position. Pressure impulse decreased with increasing rate.

Similar to the different results found for the kinematics under varying speech rate conditions, aerodynamic results from previous investigations are rather mixed, but all studies pooled all the speakers together and did not take into account the inter-speaker variations.

1.3 Motivation for the current study

Our study is a first attempt towards a better understanding of the aerodynamic properties during speech production and their potential control. In recent years experimental studies on intraoral pressure have been rather rare although technology has been developed. Additionally, most studies concern air flow variation (Pelorson 1997). This was one reason to focus on intraoral pressure development. The major motivation for our study is to obtain relevant experimental data in order to test and to validate the physical modelling of speech production, which is not as advanced even for simple stop production (McGowan et al., 1995, Van Hirtum et al., 2004). To do so, two experimental methodologies were confronted: real speech data from four German native speakers and model data gathered by means of a simplified model of the phonation system which allows to reproducing the main physical mechanisms

occurring in speech production. In a first step we decided to test the influence of speech rate on aerodynamics results.

2 Rate effects on aerodynamics in real speech data

The following part describes the experimental procedure used to gather real speech data, their pre-processing, analyses, and the results.

2.1 Methods

A new experimental design has recently been developed (Fuchs & Koenig 2006) which allows to study aerodynamics in real speech in a relatively comfortable way for the subject. It consists of a piezoresistive pressure sensor (Endevco 8507C-2) of about 2.4 mm in diameter and 12 mm in length. It was glued mid-sagittaly approximately between the subject's hard and soft palate. The sensor measures the pressure difference between intraoral and atmospheric pressure. Atmospheric pressure was sensed via a very small plastic tube hanging outside the mouth. The set-up is easier to apply in comparison to methods involving tube insertion through the nose and it is not affected by saliva blocking the tube. Altogether 4 German native speakers (2 males: jd & rw, 2 females: sf & vh) were recorded at the phonetics laboratory at ZAS. For one speaker (jd), we had to stop recording after half of the session since the amplifier overheated and only half of the data are available.

Figure 1: Experimental set-up.

Acoustics were recorded simultaneously on Digital Audio Tape (DAT) with a sampling rate of 48 kHz. We also tested a new device measuring at what point

the upper and lower lip are in contact during the production of bilabials (simply by contact or no contact corresponding to 1 or 0). The device consisted of two very thin flexible copper plates glued mid-sagitally to the upper and lower lip. One of the subjects demonstrating the whole experimental set-up is shown in figure 1.

The subjects were instructed to speak the same sentence four times in a successive order, starting at a very slow speech rate and ending at a very fast one (total duration = 9 seconds). This method was chosen to get to the limits from very slow to very fast speech.

Our target words consisted of /CVCV/-sequences with C being one of the following consonants /p, b, t, v/ and V being /a, i, u/ (except for /v/ where we only used the /a/-context). Target words were embedded in the carrier phrase *Habe X gesehen.* (*Have X seen*) and repeated up to 7 times (x 4 speech rate conditions) in a randomized order. We will here concentrate on the target word /papa/.

2.2   Pre-processing and data analyses

2.2.1   *Intraoral pressure data*

The pressure and the lip contact data were acquired using PCQuirer version 5.0 at a sampling rate of 1375 Hz and subsequently imported into Matlab for processing. The pressure 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. Furthermore, the second derivate of the filtered pressure signal was calculated in order to label experimental data. This procedure was adopted from Koenig and Fuchs (2006) using the acceleration peak as a landmark for defining closure onset (CLOSon). The beginning of pressure drop (closure offset CLOSoff) was associated with the deceleration peak at the end of the pressure plateau, and voicing offset (VOICoff) was labeled too. Details of the labeling procedure are given in figure 2. From the temporal landmarks we calculated the difference between CLOSon and CLOSoff (tClosure) and the relative duration of voicing during closure (tVoicing/tClosure).

In addition, a script was used searching automatically for the pressure maximum in the interval between CLOSon and CLOSoff. Then, it calculates the pressure difference (DeltaIOP) between the pressure maximum and the pressure minimum at CLOSon. For the model data we will use similar labels, but with the extension *_mod*.
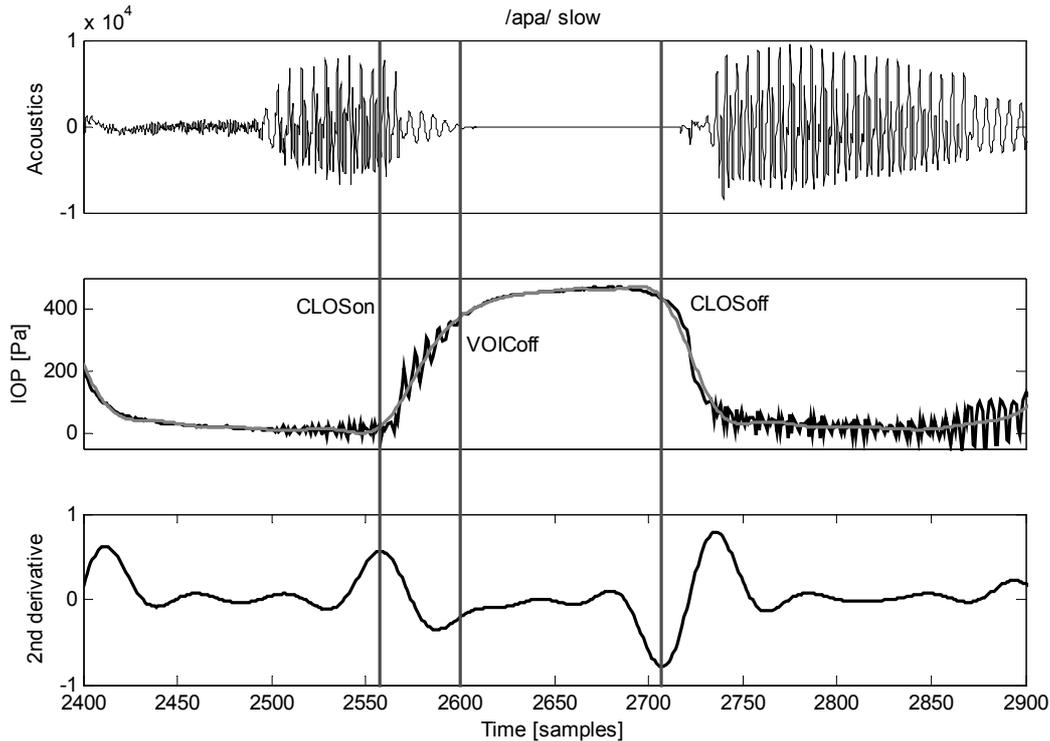
Figure 2: Temporal landmarks for closure onset (CLOSon), voicing offset (VOICoff) and closure offset (CLOSoff); 1st track acoustic signal, 2nd intraoral pressure signal (in black raw data and in gray filtered data), 3rd track second derivative of the pressure signal with acceleration and deceleration peaks.

2.2.2 Acoustic analyses

The acoustic analysis was carried out using PRAAT, version 4.4.20 (Boersma & Weenink 2006). In some speakers we had the perceptual impression that they did not only increased their speech rate, but also their loudness, starting from slow to fast speech. In order to check the potential mixture of effects, we labeled the intensity peaks in the surrounding vowels for each repetition.

Since temporal parameters could reflect a potential increase in loudness too (vowel lengthening coincides with louder speech), we additionally labeled the beginning and end of the second formant to measure the duration of the preceding and following vowels. In cases where the intervocalic /p/ became voiced, we considered a decrease (an increase) in the amplitude envelope as the relevant vowel offset (vowel onset).

2.3 Results

2.3.1 Potential confounding of faster rate with intensity

An increase in loudness affects the properties of vowels more consistently in comparison to consonants (Mooshammer et al., 2006). Generally, an increase in subglottal pressure, intensity, f_0, vowel lengthening was reported previously (most of these parameters go also hand in hand when a particular phrase is under a prosodic focus condition).

We will consider the vowel in the second syllable, the one we are interested in. Figure 3 displays the relation between intensity and vowel duration. For speakers jd and sf a strong negative correlation was found. The shorter the vowel and the higher the speech rate, the louder the maximal intensity of the sound. For rw and vh an increase in speech rate coincides with shorter vowel duration, but not necessarily with a higher intensity (for vh). For rw an increase in intensity from the first to the last repetition can also be found on average, but this speaker also varies his intensity during the whole experiment so that the scatterplots do not show a strong correlation.
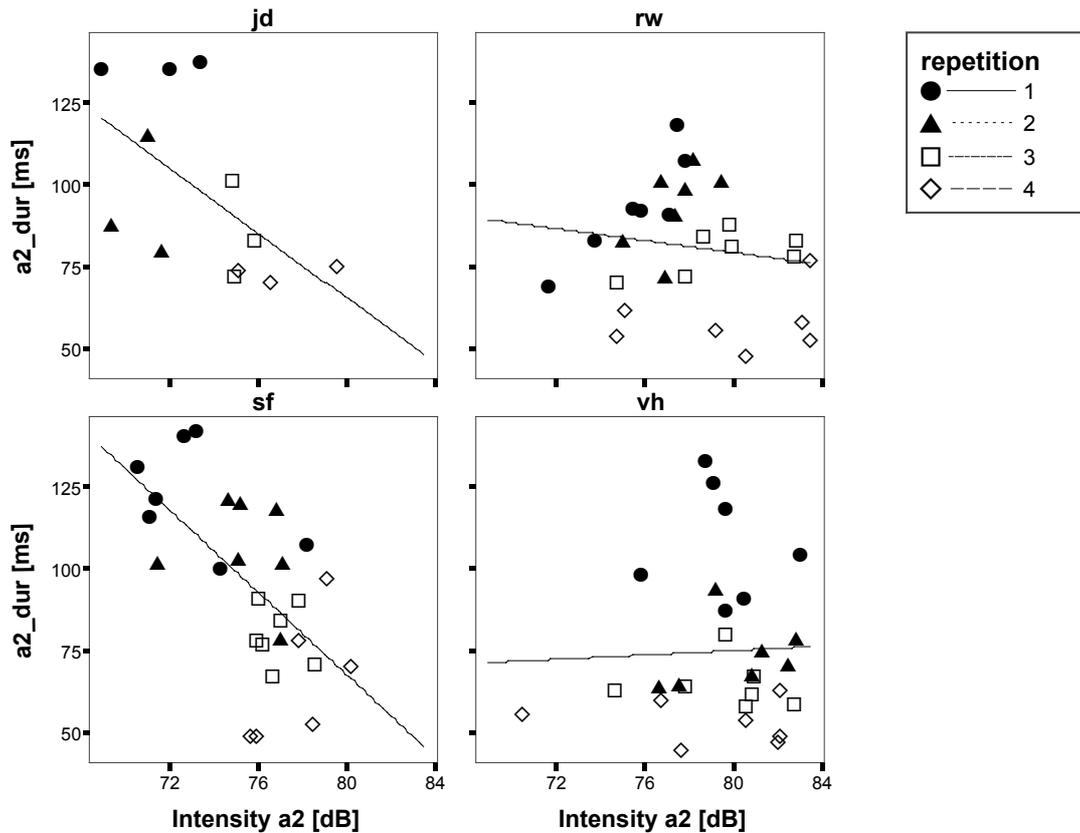
Figure 3: Scatterplots for duration (y-axis) and maximum intensity (x-axis) of the second vowel, speakers correspond to subplots. Different marker symbols = different repetitions (from 1: slow to 4: fast).

Hence, for most of the speakers (except vh) faster speech goes hand in hand with an increase in loudness. So far we are unable to tease the different effects in the data apart. However, physical model may be a good solution to control for these effects.

2.3.2 Relation between temporal behavior and intraoral pressure

The duration of voicing during oral closure should be dependent on the intraoral pressure rise and the duration of the oral closure. Everything else being equal, we expect that the higher the intraoral pressure and the longer the closure, the shorter the voicing duration. If faster speech rate goes hand in hand with shorter closure duration and less intraoral pressure increase, a relatively long voicing duration can be assumed. If faster speech rate coincides with shorter closure duration, but a high pressure, a relatively short voicing duration may occur. The latter could also correspond to the mixture of speech rate and loudness effects. Figure 4 shows different trends in a speaker-specific manner. Here the pressure difference (DeltaIOP) is plotted against the ratio between voicing duration and closure duration (tVoic/tClos). The smaller the difference between the two values of the ratio, the longer the voicing during closure and the closer the ratio value to 1.
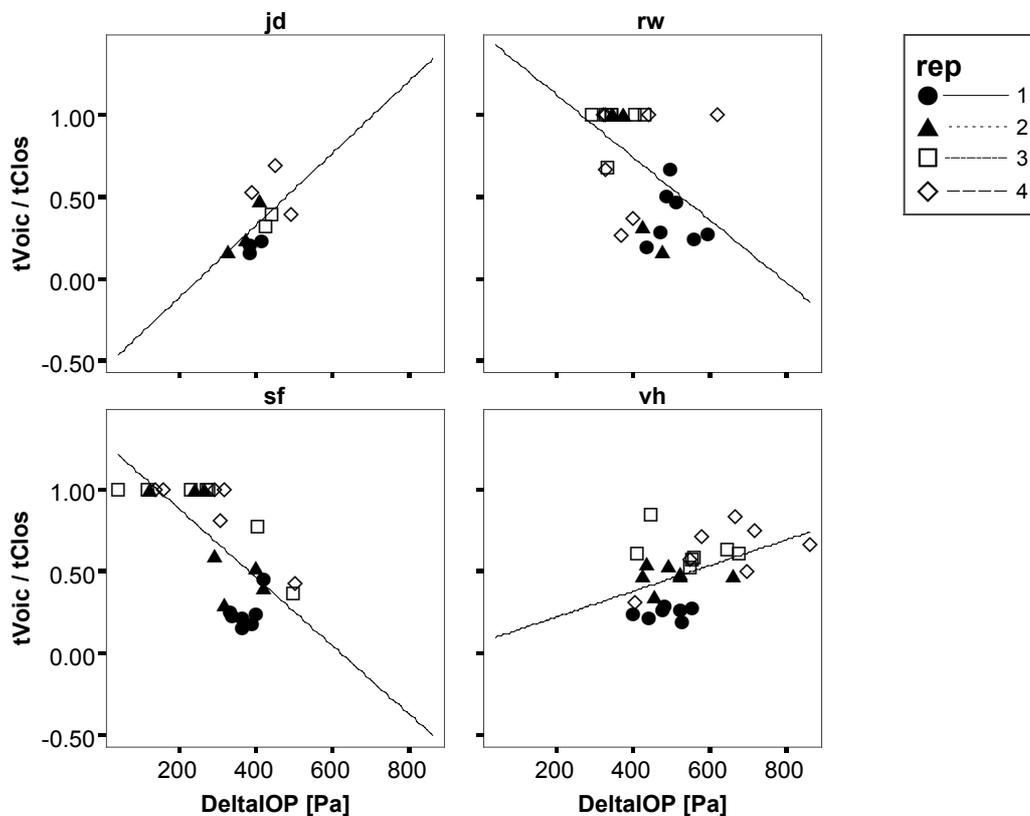
Figure 4: Scatterplots for tVoicing/tClosure (y-axis) and DeltaIOP (x-axis), speakers correspond to subplots. Different marker symbols = different repetitions (from 1: slow to 4: fast).

For all speakers it can be seen that the faster the speech rate, the closer the ratio gets to 1, i.e. the greater the voicing proportion to the closure. The slower the speech rate, the greater the temporal difference between the two parameters. Considering the temporal behavior in relation to the intraoral pressure development under various rates, speaker-specific strategies are found. For the speakers jd and vh an increasing intraoral pressure goes hand in hand with an increase in speech rate whereas for speakers rw and sf this relation is complementary.

Spearman Rho correlation coefficients for the different temporal and pressure relations are depicted in the table below.

Table 1: Spearman Rho correlation coefficients, P-values (* P<0.05, **P<0.001, ***P<0.001), number of samples included (N) for DeltaIOP with voicing duration, closure duration and the ratio tVoicing/tClosure, split by speaker.

speaker			tVoicing	tClosure	tVoicing/tClosure
Jd	DeltaIOP	r (Spearman Rho)	0.57	-0.75	0.65
		P-value	0.067	0.008**	0.029*
		N	11	11	11
Rw	DeltaIOP	r (Spearman Rho)	-0.08	0.63	-0.51
		P-value	0.789	0.000***	0.005**
		N	13	28	28
Sf	DeltaIOP	r (Spearman Rho)	0.02	0.75	-0.74
		P-value	0.950	0.000***	0.000***
		N	15	27	27
Vh	DeltaIOP	r (Spearman Rho)	0.46	-0.47	0.50
		P-value	0.014*	0.012*	0.006**
		N	28	28	28

Table 1 clearly shows that the development of intraoral pressure is closely linked to the duration of the oral closure. Additionally, the ratio between voicing and closure duration correlates with the pressure development. However, these correlations are sometimes positive and sometimes negative and we assume that they correspond to the following two strategies:

1. Speakers with a positive correlation between closure duration and intraoral pressure are those ones who either increase the amount of coarticulation between adjacent phonemes or reduce their spatial target.
2. Speakers with a negative correlation are those ones who speed up their articulation without a spatial reduction or an increase in coarticulation.

The second strategy can be tested by means of the set-up we used for modeling the aerodynamics of intervocalic stops.

3   Rate effects on aerodynamics in model data

In this part we describe the system used to gather the model data, obtained with a simplified model of the phonation system. We provide the details of the experimental set-up, the analysis procedure, and the results on the influence of speech rate.

3.1 Method

The set-up which has been developed by GISPA-lab for a few years presented in Figure 5 is composed of (Ruty et al., 2007):

1. air compressor and air reservoir (model of the lungs, not seen on the picture),
2. a replica of self-oscillating vocal folds (latex tubes vibrating due to air flow),
3. a resonator (rigid tube, oversimplified vocal tract)
4. a constriction having an adjustable height (to simulate oral or lip constriction)

This set-up does not pretend to be a realistic model of the phonation system, but is used to reproduce different physical mechanisms occurring during speech production, in simplified and controlled configurations.
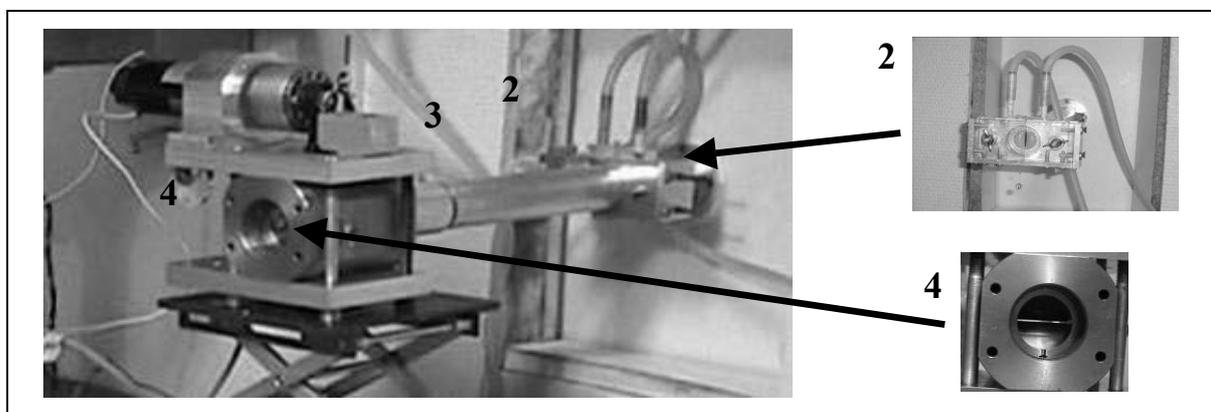
Figure 5: Experimental setup: 2- vocal folds replica; 3- tract; 4- constriction.

Measurements
The air pressure can be measured in different positions. In this paper only two points are considered: P0 is the alimentation pressure measured 2cm before the vocal folds replica (related to the subglottal pressure) and P2 is the pressure measured in the tube, a few centimeters before the constriction (related to the intraoral pressure). An optic sensor is used to measure the height of the

constriction h3 (the width of the constriction is fixed). Pacou is the acoustical pressure measured at the output of the tube (see Figures 5 and 6).
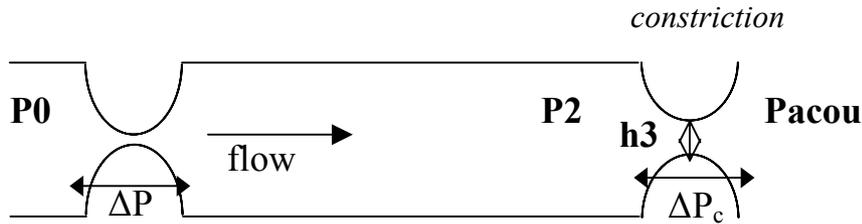
Figure 6: Definition of the studied physical parameters.

The motion of the constriction can be controlled by a motorized system or just by hand. The use of the motorized system imposes a sinusoidal evolution of h3 (in this case the duration of closure is not adjustable). By hand, both speed and time of closure can be controlled, but less accurately.

In this study, several values of the alimentation pressure P0 have been chosen ranging from P0=800Pa to P0=1400Pa.

The constriction is placed at the end of the tube, which simulates thus a configuration near to the production of the bilabial stop /p/. The resonator is a simple cylindrical tube, but more realistic geometries could be considered.

Some specificities of the experimental setup have to be taken into account to understand the differences to real speech data: 1. The boundaries of the tract are rigid, so that the volume of the tract remains constant, even during closure. 2. The air supply is continuous. It means that even if the tube is completely closed, the incident air flow is not stopped, it results in an increase of the pressure in the air reservoir (see paragraph 3.2.3). 3. The magnitude of the constriction motion can not be larger than about 2mm, i.e. the tract is always constricted (P2 never reaches 0).

3.2 Data analysis

3.2.1 Definition of the main parameters

The different parameters used to characterize the data are presented in Figure 7.

The closure duration is directly obtained from the signal h3, representing the opening height of h3 at the constriction. It is defined by $tClosure_{mod}$= $CLOSoff_{mod}$-$CLOSon_{mod}$, where $CLOSon_{mod}$ is the beginning of the decrease of h3 (beginning of the closing gesture), and $CLOSoff_{mod}$ is the beginning of the increase of h3 (end of closure and beginning of opening).

The duration of voicing is defined by $tVoicing_{mod}=VOICoff_{mod}-CLOSon_{mod}$, where $VOICoff_{mod}$ is determined from the pressure signals P0 or P2 as the offset of voicing.

$DeltaIOP_{mod}$ should be the maximum value of the pressure P2 during the closure. In fact, since the air supply is continuous, P0 and P2 increase slightly during the closure. So we chose $DeltaIOP_{mod}$ as the value of P2 observed when P2 begins to behave similar to P0.
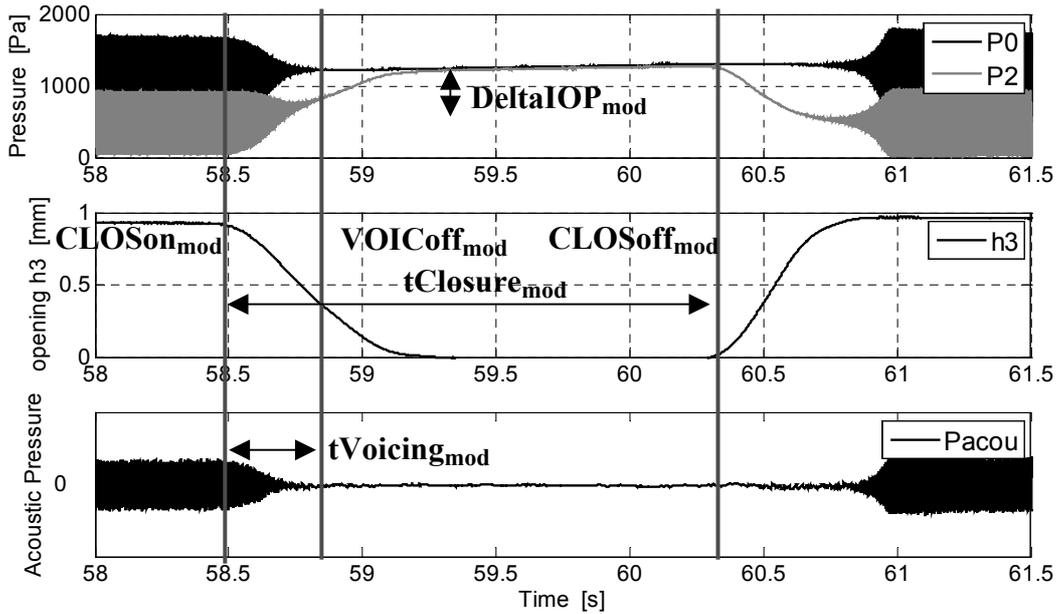
Figure 7: Temporal landmarks for closure onset ($CLOSon_{mod}$), closure offset ($CLOSoff_{mod}$), voicing offset ($VOICoff_{mod}$) and definition of $DeltaIOP_{mod}$; 1st track: pressures P0 and P2, 2nd track: height of the constriction h3, 3rd track: acoustic signal.

3.2.2 The relation between subglottal and intraoral pressure

A minimum value of P0 (or DeltaIOP) is needed to allow the self oscillations of the vocal folds replica. Figure 8 presents two recordings measured for two different values of the alimentation pressure P0, allowing or preventing the vibration of the vocal folds replica. (With the constriction, the pressure P0 should be higher than about 800Pa to obtain oscillations. Without any constriction, 300Pa is enough).
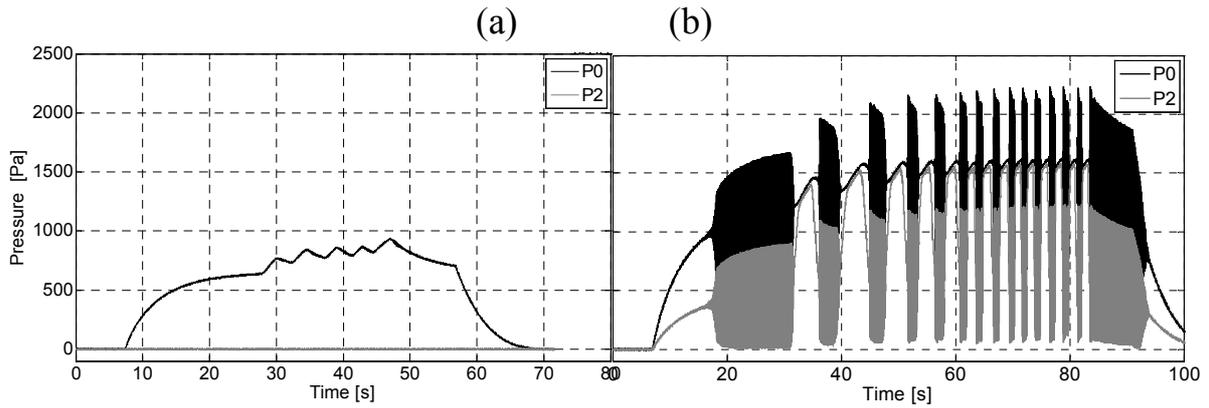
Figure 8: Pressure measurements obtained with (a) P0<1000Pa: no oscillations of the vocal folds replica (b) P0>1000Pa: self-oscillations of the vocal folds replica.

3.2.3 The influence of the speed of closure

Figure 9 shows an example of a recording in which the motion of the constriction is controlled by hand, to slowly shorten the closure duration.
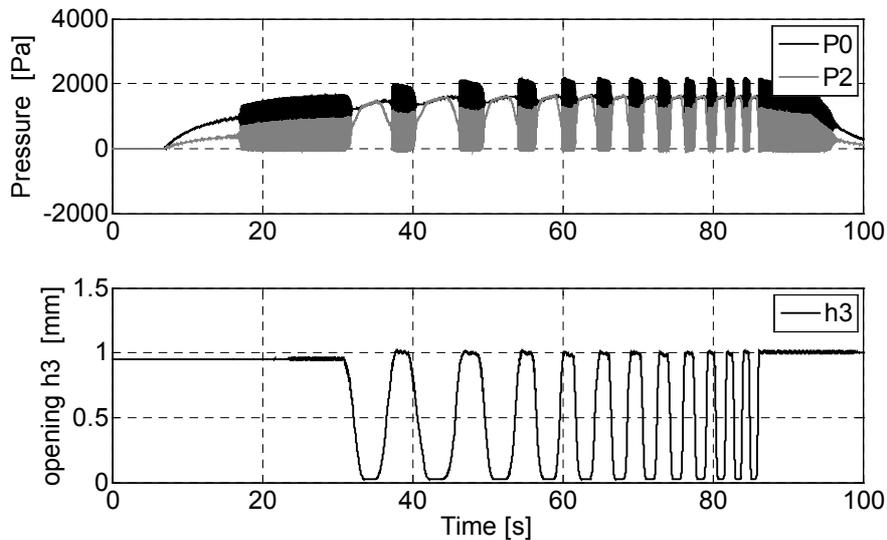
Figure 9: Model measurements with variations of the closure speed – Complete closure, 1st track: pressures P0 and P2, 2nd track height of the constriction h3.

It can be seen that each closure is associated with an increase of the pressure P2 followed by the stop of the oscillations. Each opening phase is associated with a decrease of P2 and the resumption of the oscillations.

14

For a complete closure (h3→0, Figure 9), P2 increases until it reaches the alimentation pressure P0. Note that the slow general increase of both P0 and P2 during the closure phases is due to the continuous air supply.
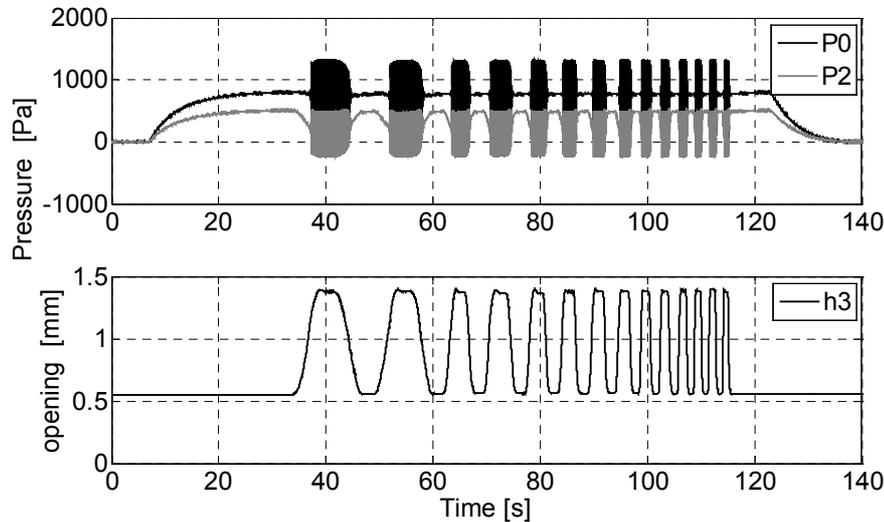
Figure 10: Model measurements with variations of the closure speed − No complete closure, 1st track: pressures P0 and P2, 2nd track height of the constriction h3.

When the closure is not complete (h3→0.5mm, Figure 10), the maximum value of P2 during the closure stays lower than the alimentation pressure P0. Moreover, the pressure P0 and the maximum of P2 are constant during the closure phases. Even if the air supply is continuous, the opening height at the constriction is enough to avoid an increase of P0 and P2 during the closure phase.

When the motion of the constriction becomes too fast in comparison to the response time of the vocal folds replica, the self sustained motion of the vocal folds replica becomes rather difficult. Figure 11 displays measurements obtained for two different speeds of closure (complete closure, sinusoidal motorized motion). It is clearly shown that the oscillations are very weak for the highest speed (graph (b)), which is a limit of the setup.

(a)

(b)
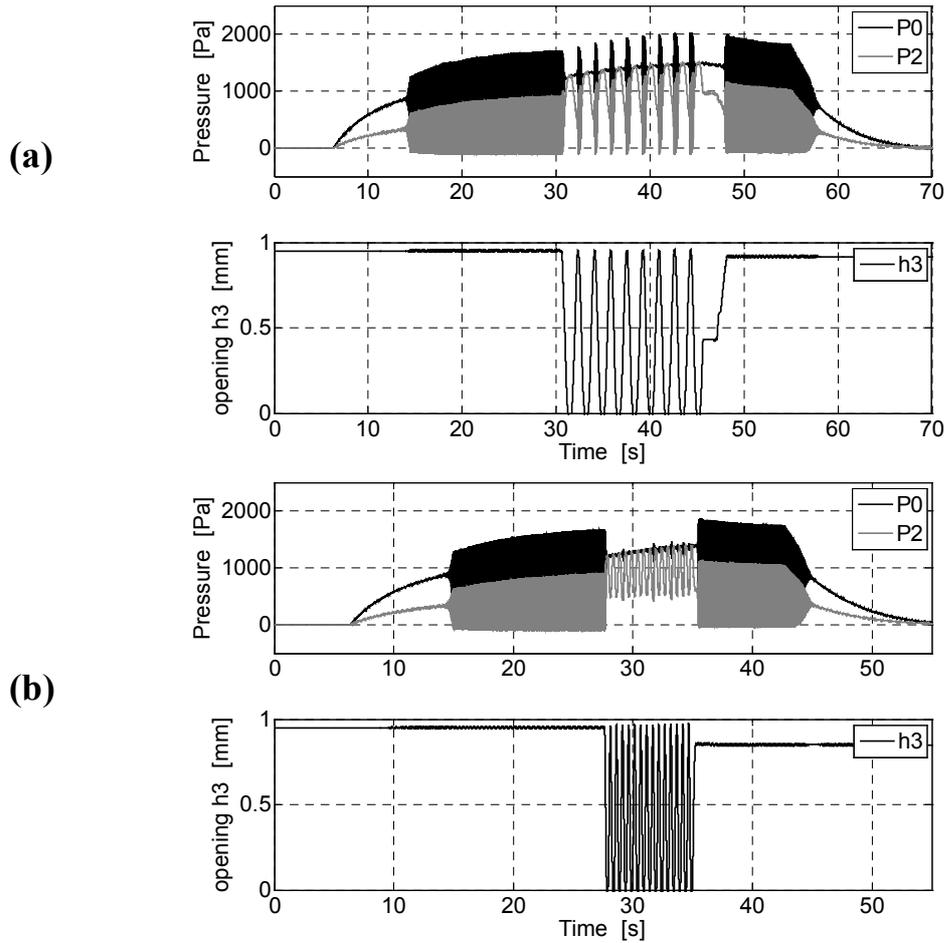
Figure 11: Model measurements with 2 different speeds of closure (a) self oscillations of vocal folds replica, (b) very weak motion of vocal folds replica, 1st track: pressures P0 and P2, 2nd track height of the constriction h3.

3.2.4 Comparison with the previous definition of the closure duration

In Figure 12, CLOSon$_{mod}$ and CLOSoff$_{mod}$ defined from the signal h3 are compared to CLOSon and CLOSoff defined from the second derivative of the filtered pressure signal P2 (definition used in the analysis of real speech data, see 2.2, and applied on P2 measurements). Even if some discrepancies are observed and should be considered, it is confirmed that the extrema of the second derivative of the filtered intraoral pressure signal are relevant indicators for determining the closure duration, defined as the time between the beginning of the closure gesture and the beginning of the opening gesture.

However, we also noticed that the h3 signal contains more information than the second derivative of P2. It clearly shows beginning *and* end of the closure and opening gestures.
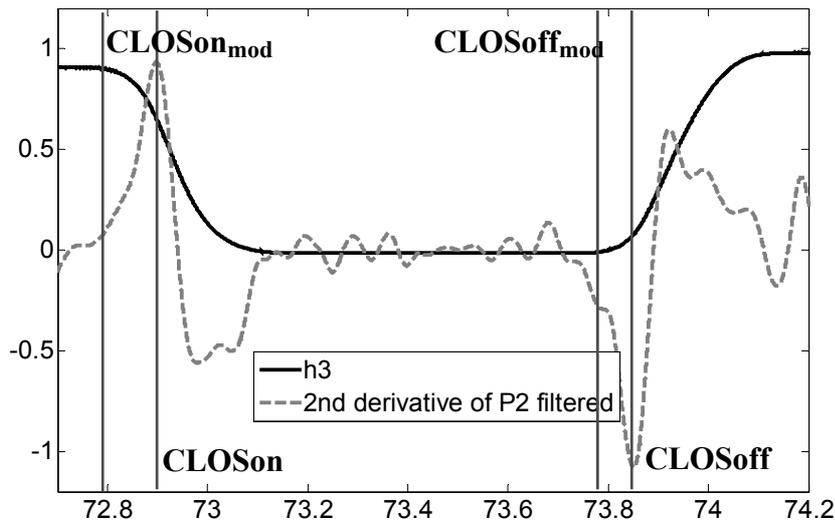
Figure 12: Comparison of CLOSon, CLOSoff (from the 2nd derivative of the filtered P2 signal), CLOSon$_{mod}$ and CLOSoff$_{mod}$ (from h3 signal).

3.3 Results

3.3.1 General considerations

Even if our experimental setup is a very simplified model used to reproduce the main phonation mechanisms, it mimics some interesting features. The evolutions of the acoustical pressure and the air pressure in the tract observed during a closure of the tube (Figure 7) are quite comparable to real speech during the production of intervocalic stops (Figure 2).

The closure of the tract results in an increase in the intraoral pressure P2 up to a threshold value. This increase results in the stop of the oscillations of the vocal folds replica. Then, the reopening is followed by a decrease of the intraoral pressure, which allows the resumption of the oscillations.

For a complete closure, the threshold value of P2 is equal to the alimentation pressure P0 (normal stop production). For an incomplete closure, the threshold value is lower than P0, depending on the minimum height of the constriction (leaky production).

However, some characteristics of the experimental setup should be considered to explain some observations: The continuous air supply results in a slow increase of P0 and P2 during closure phases in case of complete closure, which is of course not observed in real speech data. Moreover, due to the restricted maximum height of the constriction (h3max = about 2mm), the pressure P2 does not decrease until P2=0Pa during the opening phase because

the tract always presents an obstruction. For instance, in Figure 7 the mean value of P2 during the opening phase is about 500Pa.

3.3.2 Influence of speed of closure

Several measurements have been carried out: 2 tube lengths (L=20cm, 30cm), different threshold values of P0, and different speed of closure controlled by hand.

For each recording and for each closure, the parameters $CLOSon_{mod}$, $CLOSoff_{mod}'$, $VOICoff_{mod}$ and $DeltaIOP_{mod}$ are computed.

Figure 13 represents the ratio between the voicing duration $tVoic_{mod}$ and the closure duration $tClos_{mod}$ as a function of $DeltaIOP_{mod}$ with P0 ranging between 800 Pa (p1) and 1400 Pa (p4).
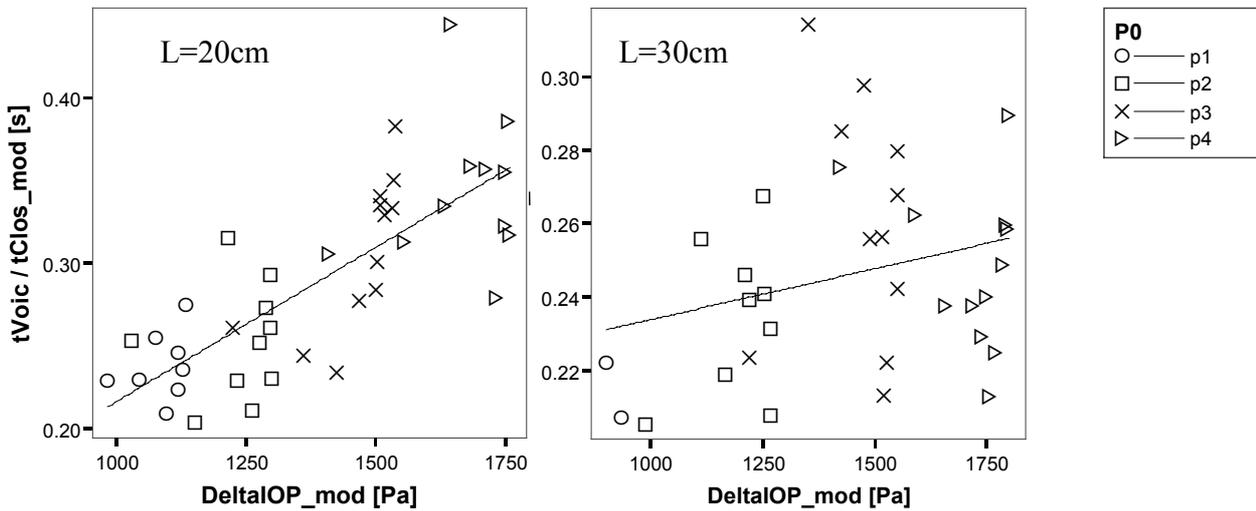
Figure 13: tVoic_mod/tClos_mod (y-axis) as function of DeltaIOP_mod (x-axis) for L=20cm (on the right) and L=30cm (on the left), with P0= 800Pa (p1), 1000Pa (p2), 1200Pa (p3), 1400Pa (p4).

These data show that the relative voicing duration globally increases with DeltaIOP, which is an expected physical result. Indeed, an increase of DeltaIOP corresponding to an increase of the alimentation pressure P0, results in a longer time to reach the threshold pressure drop $(\Delta P = P0 - P2)_{stop}$ associated to the resumption of the oscillations of the vocal folds replica.

For these model data, the maximum value of P0 and the closure duration are independently controlled (independent control of the alimentation pressure and of the motion of the constriction). The geometry of the tract is also fixed (rigid tract). Consequently, we are closer to the second strategy (see p.10), without spatial reduction and without increase of coarticulation.

The results are consistent with results obtained for speakers jd and vh (positive slopes in Figure 4), which seems to confirm that these speakers do not use spatial reduction at faster speech rate (second strategy).

Unfortunately, we never obtain model measurements with a ratio tVoic_mod/ tClos_mod = 1 (complete voicing during closure), because we can not get oscillations of the vocal folds replica for fast constriction motion, so we can not analyze cases without devoicing.

4   Conclusions

Two experimental methods have been developed and compared to gain a better understanding of aerodynamic properties in speech, especially during the production of intervocalic stops, focusing on the development of the intraoral pressure and the influence of speech rate.

Real speech data have been recorded of 4 native speakers of German. The analysis of intraoral pressure, voicing during closure phase, and closure duration shows that speech rate is heavily correlated to intensity (and hence, with the maximum of intraoral pressure). It confirms also that results are highly speaker specific and that different articulatory strategies should be considered.

Model data obtained from a mechanical replica with realistic physical behavior show a similar global intraoral pressure development as were found in real speech data. The analysis of the ratio between voicing during closure and closure duration confirmed our hypothesis that voicing is maintained longer when speech rate variations go hand in hand with loudness changes.

We conclude that an increase in speech rate coincides with a longer voicing during closure in the intervocalic stop. On the one hand this strategy can be attributed to a reduced spatial target and an increase in the amount of coarticulation, lowering intraoral pressure and sustaining voicing during closure. On the other hand, the data for 2 speakers of German and the model data can be attributed to a larger intraoral pressure going hand in hand with an increase in loudness (Note that this is only true for a certain vocal tract length).

Finally, this work has also shown that even if the experimental model is very simplified, it can be used to test various hypotheses on the aerodynamic properties concerning speech in controlled conditions.

The next step will be to compare these experimental data to theoretical data, in order to validate or improve new theoretical physical modeling.

5   Acknowledgements

6   References

Arkebauer, H.J., Hixon, T.J. & Hardy, J.C. (1967). Peak intraoral air pressures during speech. *Journal of Speech and hearing Research* 10: 196-208.

Bickley, C. (1991). Vocal fold vibration in a computer model of a larynx. In Gauffin, J. & Hammarberg, B. (eds.) *Vocal Fold Physiology : Acoustic, Perceptual and Physiology Aspects of Voice Mechanisms*. 37-46.

Boersma, P. & Weenink, D. (2006). Praat: doing phonetics by computer (Version 4.4.20) [Computer program]. http://www.praat.org/.

Brown, W.S. & McGlone, R.E. (1969). Relation of intraoral air pressure to oral cavity size. *Phonetica* 21: 321-331.

Byrd, D.; Tan, C. C. (1996). Saying consonant clusters quickly. *Journal of Phonetics* 24: 263-282.

Edwards, J., Beckman, M.E. & Fletcher, J. (1991) The articulatory kinematics of final lengthening *Journal of the Acoustical Society of America* 89(1): 369-382.

Fuchs, S. & Koenig, L. (2006). On the articulation, aerodynamics, and acoustics of voiceless consonant clusters. *ISSP Ubatuba Brazil:* 19-26.

Gibbon, F. E. & Lee, A. (accpeted) Producing turbulent speech sounds in the context of cleft palate. In Fuchs, S., Toda, M., Zygis, M. & Shadle, C. (eds.) *On the Phonetics and Phonology of Turbulent Sounds*. Mouton de Gryuter.

Hertegård, S., Gauffin, J. & Lindestad, P.-Å. (1995) A comparison of subglottal and intraoral pressure measurements during phonation. *Journal of Voice* 9(2): 149-155.

Huber, J.E., Stathopoulos, E.T. & Sussman, J.E. (2004) The control of aerodynamics, acoustics, and perceptual characteristics during speech production. *Journal of the Acoustical Society of America* 116(4): 2345-2353.

Malécot, A. (1969a). The lenis-fortis opposition: Its physiological parameters. *Journal of the Acoustical Society of America* 47(6/2): 1588-1592.

Malécot, A. (1969b). The effect of syllabic rate and loudness on the force of articulation of American stops and fricatives. *Phonetica* 19: 205-216.

McGowan, R.S., Koenig, L.L. & Löfqvist, A. (1995). Vocal tract aerodynamics in /aCa/ utterances: Simulations. *Speech Communication* 16: 67-88.

Mooshammer, C., Hoole, P. & Geumann, A. (2006). Interarticulator cohesion within coronal consonant production. *Journal of the Acoustical Society of America* 120(2): 1028-1039.

Koenig, L. & Fuchs, S. (2006). The time course of intraoral pressure changes during the formation and release of consonants and consonant clusters of German. *ISSP Ubatuba Brazil,* 349-356.

Lindblom, B.E.F. (1963) Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35(11): 1773-1781.

Löfqvist, A., Koenig, L.L. & McGowan, R. (1995). Vocal tract aerodynamics in /aCa/ utterances: Measurements. Speech Communication 16: 49-66.

Pelorson, X. Hofmans, G. Ranucci, M. & Bosch, R. (1997) On the fluid mechanics of bilabial plosives. Speech Communication 22: 155-172

Pfitzinger, H. (2001). Phonetische Analyse der Sprechgeschwindigkeit. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 38, 117-264.

Ruty, N. Pelorson, X. & Van Hirtum, A. (2007) An in vitro setup to test the relevance and the accuracy of low-order vocal folds models. *Journal of the Acoustical Society of America* 121(1): 479-490.

Tuller, B. & Kelso, J.A.S. (1984) The timing of articulatory gestures: Evidence for relational invariants. *Journal of the Acoustical Society of America* 76(4): 1030-1036.

Van Hirtum, A., Ruty, N., Pelorson, X., Fuchs, S. & Perrier, P. (2004) Laryngeal adjustment in the production of voiceless consonants: II. Physical modelling. *Proc. of the ICVPB Marseille.*

Weitkus, K. (1931). *Experimentelle Untersuchung der Laut- und Silbendauer im deutschen Satz*. PhD Dissertation at the University of Bonn.

Warren, D.W., Putnam Rochet, A., Dalston, R.M. & Mayo, R. (1992) Controlling changes in vocal tract resistance. *Journal of the Acoustical Society of America* 91(5): 2947-2953.

On the avoidance of voiced sibilant affricates

Marzena Żygis
Zentrum für Allgemeine Sprachwissenschaft, Berlin

In this paper it is argued that several typologically unrelated languages share the tendency to avoid voiced sibilant affricates. This tendency is explained by appealing to the phonetic properties of the sounds, and in particular to their aerodynamic characteristics. On the basis of experimental evidence it is shown that conflicting air pressure requirements for maintaining voicing and frication are responsible for the avoidance of voiced affricates. In particular, the air pressure released from the stop phase of the affricate is too high to maintain voicing which in consequence leads to a devoicing of the frication part.

1   Introduction

Phonemic inventories of the world's languages show various types of gaps which are accounted for by appealing to the phonetic properties of the sounds under question. These properties are in turn inevitably connected with the anatomy and functioning of the human vocal tract during speech production. For example, it has been widely shown that the voiced velar stop [g] does not occur frequently in the phonemic inventories of the world's languages in comparison to their voiced counterpart [k] (see Maddieson 1984, Boersma 1988). A possible explanation of this gap pertains to the difficulty in maintaining optimal conditions for voicing due to a short cavity behind the constriction.[1]

The present investigation expands the list of sounds which are avoided cross-linguistically. It shows, namely, that voiced sibilant affricates ([d͡z], [d͡ʒ]) tend to be eliminated from phonemic inventories. At the same time, their voiceless counterparts [t͡s], [t͡ʃ] show a different behavior: the affricates are stable and they create a part of many phonemic inventories. This observation also holds for languages with a complete voicing contrast in coronal stops and fricatives; see Slavic inventories discussed in section 2.

---

[1]   For other gaps in phonemic inventories see Maddieson (1984).

That voiced sibilant affricates are avoided is also confirmed by phonological processes which convert these sounds into others. In some languages we observe a productive devoicing of affricates even in a word-initial position (see section 3 for examples).

Explaining the avoidance of voiced affricates is a challenge for any phonological theory. Several phonological approaches to voicing contrasts, including those dealing with features, are not able to account for this gap; see e.g. Lombardi (1994, 1999), Iverson & Salmons (1995, 2003), Steriade (1997), Avery & Idsardi (2001), Wetzels & Mascaro (2001), Kehrein (2002).

How can we explain this cross-linguistic avoidance of voiced sibilant affricates? In this paper, phonetic evidence is taken into consideration. It is argued that voiced affricates are exposed to an aerodynamic conflict: the air pressure required for maintaining voiced frication is often too high due to the high pressure conditions in the stop phase. This conclusion is evidenced by experimental results presented in section 5.

The paper is organized as follows. In section 2 several inventories from various language families are provided illustrating the gap of voiced affricates in phonemic systems. Further evidence for the cross-linguistic tendency to avoid voiced sibilant affricates is provided in section 3 where examples of phonological processes converting voiced sibilants to other sounds are discussed. Section 4 deals with a typology of voicing based on 451 language inventories according to which voiced affricates are the least frequent phonemes among coronal obstruents. The remaining part of the study is devoted to a presentation and discussion of a piece of experimental evidence showing that aerodynamic relations or even aerodynamic conflicts in the production of voiced affricates are responsible for their avoidance. Section 6 concludes.

2   Phonemic inventories

The investigation of several typologically unrelated languages reveals a systematic gap in the phonemic inventories of the world's languages: voiced sibilant affricates are not attested.

The presentation starts with Slavic languages, followed by a discussion of languages from Romanic and Germanic language families. Examples from Asiatic and Arabic languages are provided as well.

Slavic languages undoubtedly serve to advance the hypothesis that voiced affricates are avoided. In almost all Slavic inventories these phonemes are either not present or they occur in a very limited number of (foreign) words. Phonemic inventories of languages representing West, East and South Slavic family are presented in Table 1, 2, and 3 respectively.

Table 1: Czech phonemic system.

	dental/alveolar		palatoalveolar		palatal	
fricative	s	z	ʃ	ʒ		
affricate	t͡s		t͡ʃ			
stop	t	d			c	ɟ

Table 2: Russian phonemic system.

	dental				retroflex		palatoalveolar	
fricative	s̪	z̪	s̪ʲ	z̪ʲ	ʂ	ʐ	ʃʲː	(ʒʲː)
affricate	t͡s̪						t͡ʃʲ	
stop	t̪	d̪	t̪ʲ	d̪ʲ				

Table 3: Bulgarian phonemic system.

	dental				palatoalveolar	
fricative	s̪	z̪	s̪ʲ	z̪ʲ	ʃ	ʒ
affricate	t͡s̪		t͡s̪ʲ		t͡ʃ	(d͡ʒ)
stop	t̪	d̪	t̪ʲ	d̪ʲ		

A common denominator of Czech and Russian phonemic inventories is the absence of voiced affricates. In Bulgarian, the voiced affricate /d͡ʒ/ is found in a few foreign words. At the same time, in all three languages coronal fricatives and stops display a voicing contrast.

An analysis of other Slavic inventories leads to similar results. Slovene displays a voicing contrast in fricatives and stops, but not in affricates. As far as the latter are concerned, /d͡z̪/ is not attested. The palato-alveolar /d͡ʒ/ is only found in words of foreign origin. For this reason its phonemic status is rejected by some scholars; see e.g. Dalewska-Greń (2002). A similar scenario is repeated in Czech. A voicing contrast is shown by stops and fricatives, while the affricate phonemes include only the voiceless segments /t͡s/ and /t͡ʃ/, see Kučera (1961:24).

In several Serbian and Croatian dialects, it is the voiced affricate /d͡z/ that is often missing; see, for example, the Jekavian dialect of Eastern Herzegovina, the Young Ikavian dialect, the Istrian Ikavian dialect, and a number of Slavonian dialects, see Ivić (1958).

In Upper Sorbian, the voiced counterparts of /t͡s̪/ and /t͡s̪ʲ/ are not attested, whereas the contrast between /t͡ʃ/ and /d͡ʒ/ is phonemically represented. By

contrast, in Lower Sorbian, no voiced affricates are attested despite the occurrence of a complex spectrum of voiceless affricates: /t͡s̪ t͡ʂ t͡ɕ/.

Although the Macedonian inventory seems to be complete, the symmetry between /d͡z̪/ and /t͡s̪/ as well as /d͡ʒ/ and /t͡ʃ/ is questionable. The voiced affricate /d͡ʒ/ is mainly found in words of foreign origin, particularly in those borrowed from Turkish. Dalewska-Greń (2002:93) observes that /d͡z̪/ occurs in a few native morphemes. Hence, Macedonian also shows a voicing contrast with respect to affricates, but the presence of voiced affricates, especially /d͡z̪/, is rather marginal.

Prima facie, the inventories of Belorussian and Ukrainian (especially its western dialects) show a voicing contrast in sibilant affricates. However, Zilyns'kyj (1979:114) points out that voiced affricates, especially /d͡z̪/, /d͡z̪ʲ/ and /d͡ʐ/, occur only marginally in western Ukrainian. They are present in a few words, primarily in those of foreign origin, e.g. kukuru[d͡z̪]a 'corn' nom.sg. In western dialects, the alveolo-palatal /d͡ʑ/ appears in loanwords from Polish, e.g. [d͡ʑ]ura 'hole.' In Belorussian, voiced affricates are less frequent than their voiceless counterparts. This observation pertains especially to /d͡z̪/, which is found especially in onomatopoetic words and loanwords, e.g. ks'on[d͡z̪] 'priest.'

According to Dalewska-Greń (2002), in two Slavic languages, Polish and Slovak, voiceless and voiced affricates are attested and, moreover, well represented in the vocabulary. However, a closer inspection of the vocabulary reveals that the occurrence of voiced affricates is very limited. In Polish the affricate /d͡z/ occurs in a very few proper names in word-initial position. Similarly, the occurrence of [d͡ʐ] is limited to a few foreign words, again in word-initial position. The two affricates are well represented in word-medial positions when they appear as outputs of morpho-phonological processes. A very similar conclusion pertains to Slovak voiced affricates which are found in word-medial positions, and their presence is mostly morphologically conditioned.

Finally, it has to be stressed that the voiced affricates did occur in all Slavic languages but they either completely disappeared or converted to other sounds; see Zygis (2006) for a detailed discussion.

Romanic languages do not prefer voiced affricates either. The only difference between Slavic and Romance languages is that the former also display voicing contrast in stops and fricatives, whereas the latter show voicing symmetries in stops only. For example, the Romanian coronal inventory contains the following obstruents: /t d s ʃ t͡s t͡ʃ/. Similarly, in the inventory of Galician, voiced fricatives and voiced affricates are not attested. The systems consists of /t d s ʃ t͡ʃ/.

There are at least two Germanic languages relevant for the present discussion. In German, /t d s ʃ z t͡s t͡ʃ/ are part of the phonemic inventory

while the voiced affricate /d͡ʒ/ only occurs in words of foreign origin (see Hall 1992). But even there they tend to be devoiced, see section 3 for examples. In Yiddish, the voicing opposition in coronal obstruents is symmetrical except for affricates which are limited to the voiceless /t͡ʃ/.

In African languages, it can be observed that voiceless segments are generally predominant. In Northern Soto, voiced affricates create a gap. In addition, a voiced /d/ and /z/ are not attested. Consider the tableau in Table 4.

Table 4: Northern Soto phonemic system.

	dental/alveolar		palatoalveolar	
fricative	s		ʃ	ʒ
affricate	t͡s	t͡sʰ	t͡ʃ	t͡ʃʰ
stop	t	tʰ		

In Ekegusii, a language spoken in Kenya, the coronal phonemic inventory is constituted by voiceless segments /t s t͡ɕ/ only. In Chingoni (southern Tanzania), the voicing system is symmetrical in fricatives, but not in stops and affricates: /t t͡ʃ s z/. In summary, the investigation of sibilant inventories of several languages shows that affricates show an asymmetry in voicing: voiced affricates are considerably less frequent than their voiceless counterparts. In many languages their status is either marginal, i.e. they occur in foreign words only, as for example in Slovene or Bulgarian, or they do not occur at all, as in Russian or Yiddish.

3   Phonological processes

The avoidance of voiced affricates is also evidenced by phonological processes attested cross-linguistically. In many languages the voiced affricates convert to other sounds, most frequently to fricatives and glides. Furthermore, it is also observed that the affricates devoice not only domain-finally but also in other, e.g. word-initial or intervocalic positions. Of special importance for the present study are (i) changes not conditioned by context which could possibly blur inherent properties of affricates and (ii) changes which are not motivated by internal language-specific factors as e.g. improving phonemic contrast. Most processes presented below are in accordance with these two conditions.

In Standard Bulgarian [d͡z] and [d͡ʒ] is pronounced as [z] and [ʒ] in words adopted from Bulgarian dialects, respectively. For example, /d͡z/vezda is pronounced as [z]vezda 'star' in Standard Bulgarian. Dalewska-Greń (2002:96) stresses that /d͡ʒ/ is found in foreign words and that it changes to [ʒ] in colloquial speech.

In Czech, loan words containing the affricates [d͡z] and [d͡ʒ] are converted to a sequence of a stop followed by a fricative, i.e. to [dz] and [dʒ]. For example, [d͡ʒ]udo 'judo' is pronounced as [dʒ]udo, see Kučera (1961:33), Short (1993:457). But this sequence is also found in native words [dʒ]ban 'jug' and [dʒ]ber 'tub' and at some morpheme boundaries, e.g. in po[d͡z]emi 'underground' a sequence [dz] rather than the affricate [d͡z] is pronounced (Short 1993:457).

In literary Upper Sorbian, /d͡z̦/ it is replaced by [z], e.g. in the declension of substantives ending in –ga. For example Jadwi[g]a 'proper name, nom.sg' appears as Jadwi[dz]e 'acc.sg.' The latter form converts to Jadwi[z]e; see Schuster-Šewc (1999:38).

In Romance languages, voiced sibilant affricates frequently convert to other sounds. In Standard Romanian as well as its southern dialects, [d͡z] changes to [z], e.g. lucre[d͡z]i 'you work' appears as lucre[z]i. The palato-alveolar [d͡ʒ] converts to [j] if the affricate is followed by back vowels: /d͡ʒ/ok 'game' is pronounced as [j]ok. In Florentine Italian, both voiced and voiceless affricates convert to fricatives in the word-initial position (d͡ʒ, t͡ʃ →ʒ, ʃ/#_). For example /d͡ʒ/orni 'days' is realised as [ʒ]orni; see Gianneli & Savoia (1979).

In the Graulhet dialect of Occitan (Western Lengadocian, near Tolosa and Albi), [t͡ʃ] and [d͡ʒ] from Standard Occitan merge into a single affricate [t͡s]. Thus, Standard Occitan [d͡ʒ]orn 'day' is pronounced as [t͡sun] in Graulhet dialect (Lieutard 2004).

In German, the voiced affricate [d͡ʒ] is often devoiced, especially if it occurs in the word-initial position (and not intervocalically). The foreign words as e.g. [d͡ʒ]ungle 'jungle' or [d͡ʒ]ob 'job' are frequently pronounced as [t͡ʃ]ungle and [t͡ʃ]ob by both Southern and Northern German speakers. The initial devoicing pertains, however, not only to affricates. Stops tend to be devoiced as well and sibilant fricatives are always voiced in word-initial position.

In Gulf Arabic, the voiced affricate [d͡ʒ] alternates with a palatalized [gʲ] and [j] depending on the context (but also on the dialect). The alternation [d͡ʒ] ~ [gʲ] is found in words which have a historical [g] usually followed by a front close vowel: mi[gæː]bil ~ mi[d͡ʒæː]bil 'opposite'. In other words [d͡ʒ] alternates with [j], e.g. [d͡ʒæː]r ~ [jæː]r 'neighbour'; see Holes (1990:262).

In Chitwan Tharu, the lamino-alveolar /d͡z/ converts to the lamino-alveolar [z] intervocalically: hai/d͡z/a 'malaria' is pronounced as hai[z]a and so/d͡z/a 'straight' as so[z]a (Leal 1972:20). In Haitian Creol /d͡ʒ/ changes to [j] in the word-final position (Tinelli 1981).

Finally, it should be stressed that there are also processes attested in which voiced affricates emerge. However, these occur less frequent in comparison to the processes converting affricates into other sounds. They seem to be exceptions rather than a tendency. For example, in Montenegrin we observe the

rise of the phoneme /d͡z/ from /z/: /z/eleno → /d͡z/eleno 'green' or /z/ubi → /d͡z/ubi 'teeth' (Greenberg 2000:298).

In summary, the processes presented above confirm certain instability of sibilant voiced affricates. The spectrum of the segments to which the affricates convert varies from devoiced affricates, sequences of stops and fricatives, voiced fricatives up to glides. Several of the changes presented are context-free, thereby highlighting marked properties of the voiced affricates.

4   Typology of voicing

The conclusions presented above prompt the question to what extent we can talk about cross-linguistic generalizations, and how the conclusions relate to other coronal sound classes such as coronal stops and fricatives.

Figure 1 presents frequency of occurrence of voiced and voiceless coronal obstruents. It is based on the UCLA Phonological Segment Inventory Database (Maddieson & Precoda 1992) which includes the phonemic inventories of 451 languages.
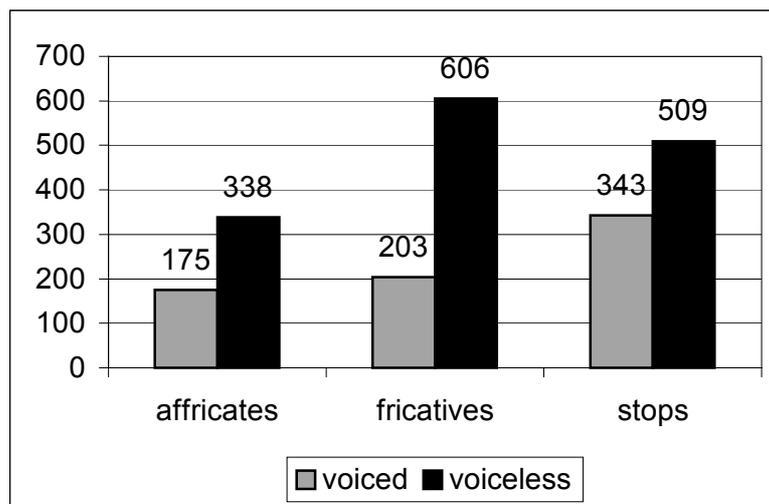
Figure 1: Frequency of occurrence of voiced and voiceless obstruents based on the UPSID database.

In all three classes, i.e. affricates, fricatives and stops, it is the voiceless segment that occurs more frequently than the voiced one. Thus, the conclusion can be drawn that the inherent properties of voicing including the articulation, aerodynamic conditions and laryngeal-oral coordination are responsible for the asymmetry.

How do these properties look in particular? Starting with the general aerodynamic properties of voicing, it is well known that in order to maintain

vocal fold vibration, the supraglottal pressure must be lower than the subglottal pressure. It is assumed that this pressure drop should be greater than the threshold of voicing, i.e. 2000 dyn/cm2 (200 Pa), in order to sustain vocal fold vibration; see Ladefoged (1964). The subglottal pressure building up below the glottis affects the lower parts of the adduced vocal folds initiating the vibration. Maintaining the pressure difference is especially difficult in stops due to the closure leading to an increase in the supraglottal pressure. To secure voicing throughout the closed phase or the final part of it, the volume of the vocal tract must expand so that the pressure will not build up too quickly. There are different strategies which enable the cavity above the glottis to be expanded. These include tissue compliance, muscularly actuated enlargement of the supraglottal cavity, the opening of the velopharyngeal port as well as jaw movements (see e.g Westbury 1983; for jaw movements see Mooshammer et al. 2007). All these manoeuvres are used to different extents by speakers in the production of voiced plosives. By contrast, voiceless plosives do not need an enlargement of the oral cavity as the vocal folds – being open – do not vibrate during their production.

In voiced sibilant fricatives, as observed by Ohala (1983), the problem of cumulating the air during the constriction phase with the risk of quenching voicing should disappear as the air continuously flows during the fricative articulation. However, this is not the case – as stressed by Ohala – because the turbulence characteristic of fricatives requires a high oral pressure. In other words, during the production of fricatives two conflicting conditions have to be fulfilled at the same time: a low oral pressure for maintaining voicing and a high oral pressure ensuring sufficient air velocity for creating frication. In addition, it is also observed that the opening of the glottis is coordinated with the supraglottal constriction so that the latter has to be smaller than the former to assure the production of the turbulence, see e.g. Scully et al. (1992).

Considering this issue in more detail, it should be noted that the emergence of the frication (turbulence) is generally determined by the flow rate, which depends on (i) the area of the constriction (A) and (ii) the difference in pressure between the two cavities on both sides of the constriction (ΔP_c). This means that the increasing pressure difference and the shrinking of the area of constriction leads to increased airflow rate, and consequently more turbulent airflow ($U = A (\Delta P)^{0.5}$) (Solé 2003:2761). If the fricative is voiced, a pressure drop across the glottis is additionally required (ΔP_g).

Figure 2 shows the allowable range of aerodynamic variation for voiced fricatives as estimated from aerodynamic data (adapted from Solé 2002:370).

glottis　　　supraglottal constriction

$$P_g=7.6 \quad P_o=5.6-3 \quad P_a=0$$

$$\Delta P_g \approx 2\text{-}3 \qquad \Delta P_c \geq 3$$
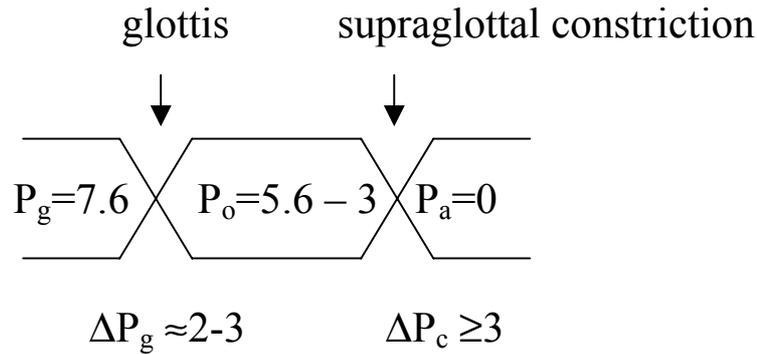
Figure 2: Estimated aerodynamic range for voiced fricatives.

The model presented in Figure 2 illustrates how voiced sibilants are constrained by aerodynamic conditions. Solé (1998) estimates a subglottal pressure (P_s) of 7.6 cm H_2O in the production of a voiced sibilant. The pressure drop across the glottis should be of at least 2–3 cm H_2O, which leaves an oral (supraglottal) pressure of at most 5.6 cm H_2O ($\Delta P_g = P_g - P_o$). The airflow then passes a supraglottal constriction, where turbulence is created. At this point, a pressure drop of approximately 3 cm H_2O creates frication. If this requirement is not met, the production of turbulence ceases; see Catford (1977), Ohala, Solé & Ying (1998), Stevens (1998).

Let us proceed to affricates, the main subject of interest. Their aerodynamics and articulation are even more complex than those of voiced plosives. This follows from the fact that the plosive is released into a voiced fricative, which in turn requires additional effort to maintain the voicing as described above. Importantly, we could hypothesize that the precise aerodynamic conditions required for the voiced fricative are partly hampered by the increasing air pressure in the stop part (see the discussion in section 5).

What does the articulatory complexity of affricates in fact mean? The main distinction between an affricate and a stop is that the release of the former is prolonged in the form of friction. A stop is inherently fricated, especially if it is followed by a high vocoid, but the frication length is too short to be perceived as a fricative component. In affricates, on the other hand, the frication is considerably longer, constituting the most prominent part of the affricates from the perceptual point of view.

Wierzchowska (1980:100) presents cinetracings of the Polish affricate [t͡ɕ] which clearly shows its two phases: the closure, and the (fricative) release. The x-ray tracings in Figure 4 illustrate that the tongue blade touches a postalveolar place of articulation (tracings 2, 3, and 4), and then, during the release, a fricative is created (tracings 5 and 6).
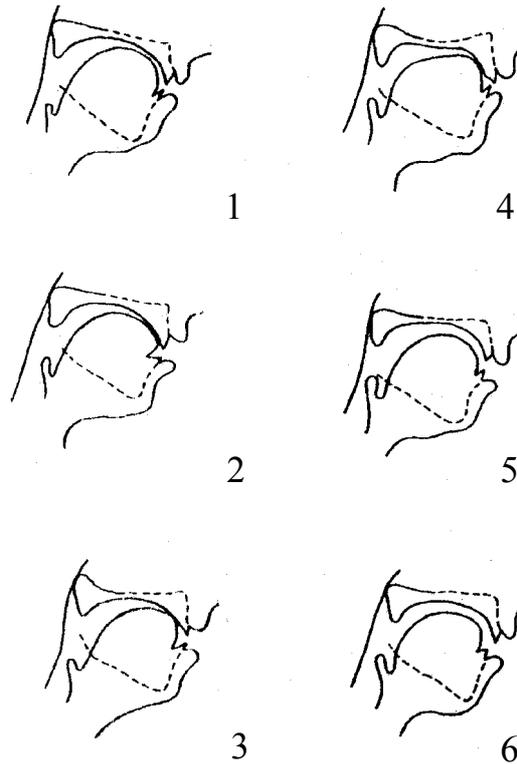
Figure 3: Cinetracings of the Polish affricate [t͡ɕ] (Wierzchowska 1980:100).

Stevens (1993a) notices that the release mechanism for an affricate differs from that for a stop consonant. This is due to the complexity of the affricate constriction, which consists of two parts: an anterior (A_1) and a posterior section (A_2). See the schematic model of the vocal tract for the English affricate [t͡ʃ] presented in Figure 4. (Stevens 1993a:33).
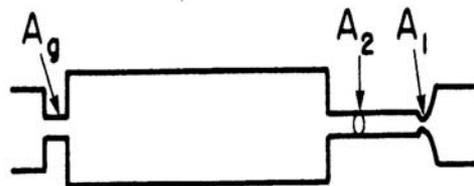
A_g: the cross-sectional area of the glottal opening

A_2: frication part

A_1: plosive constriction

Figure 4: A schematic model of the vocal tract for the English affricate [t͡ʃ].

Both components A$_1$ and A$_2$ can be manipulated independently. If a consonant is released, then the anterior section increases, whereas the posterior part remains unchanged for a few milliseconds. After this, a slow increase in the area A$_2$ takes place. An important difference between the two sections is that the initial release is rapid while the posterior part is maintained longer, which is reflected in a longer fricative component of the affricate.

Stevens (1993b) also notices that the tongue shape is different in both components. This is shown in Figure 5, where the midsagittal section of the vocal tract is illustrated prior to the initial release of the stop component [t͡ʃ] (dashed line) and 20–30 ms following the release (solid line) (Stevens 1993b:35).
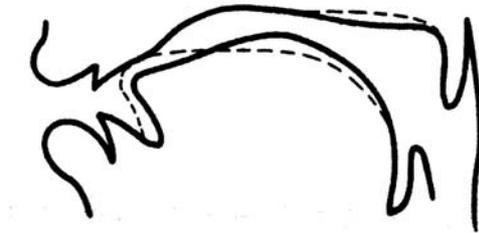
Figure 5: The midsagittal section of the vocal tract during the articulation of [t͡ʃ].

As can be inferred from Figure 5, the tongue blade and the tongue body behind the closure is depressed in comparison to its position for the fricative component. This can be explained by aerodynamic conditions, due to which the downward displacement of the tongue blade and the tongue body by the increased air pressure on the tongue surface during the closed phase is caused. At the moment when the closure is released, the air pressure decreases and the tongue moves upward; see Svirsky et al. (1992). As a consequence, the following fricative component should be articulated at a different place of articulation than the stop part.

However, it seems that the fast change from a closure where the tongue tip/tongue blade touches the palate to a constriction formed by the lateral parts of the tongue poses a challenge as the different articulatory gestures should be executed very quickly and precise in order to form an appropriate supralaryngeal constriction.

Ladefoged & Wu (1984:272) notice some small but systematic differences between the affricates and the fricatives when investigating Pekingese sibilants and affricates, [s], [ɕ], [ʂ] versus [t͡s], [t͡ɕ], [t͡ʂ]. Apart from the change in the tongue tip or blade required for making the stop preceding the fricative, the body of the tongue changes its position. It is often slightly higher during the stop than

during the fricative component.[2] Apart from these small disparities, the affricates and the corresponding fricatives share the same place of articulation.

In summary, voiced affricates are less preferred than voiceless ones because two different conditions must be maintained in their production. These are listed in (1).

(1) (i)   enlargement of the vocal tract cavity for maintaining voicing of the plosive;

(ii)   fulfilment of two conflicting conditions simultaneously in the fricative release: a low oral pressure for maintaining voicing and a high oral pressure ensuring sufficient air velocity for creating frication.

Taking into consideration all facts listed above, it is hypothesized that the aerodynamic conditions attested in voiced affricates are more complex than in the production of voiced stops and voiced fricatives (see also Dixit & Brown 1985 and König & Fuchs 2007 for voiceless affricates). On the basis of experimental data consisting of Polish fully voiced affricates it will be shown that the fricative component of affricates is often exposed to devoicing due to the high pressure peak in the closure phase; see the presentation and discussion of the results in the next section.

5   Experimental evidence

5.1   Experimental design

The goal of the experiment is not only to get a better insight into the aerodynamics of voiced affricates in contrast to voiceless ones, but also to compare aerodynamic conditions in the production of stops, fricatives and affricates. For this purpose, three simultaneous recordings were obtained: (i) oropharyngeal pressure changes, (ii) airflow at the mouth, and (iii) audio-signal.

All recordings were obtained by using PCquirer (version 8.9.8.6.). The audio-signal was recorded at a sampling rate of 22500 Hz for the oropharyngeal pressure and of 2750 Hz for the airflow. The data were analysed by PCquirer.[3]

Four native speakers of Polish (two female and two male) took part in the experiment. Each subject was equipped with a Rothenberg mask and

---

[2]   See Figure 5 for a different positioning of the tongue body during the production of English [t͡ʃ].

[3]   One of the disadvantages of using a Rothenberg mask is the rather weak acoustic quality.

additionally a piezoresistive pressure transducer was glued on the back part of the palate (Endevco 8507C-2) to measure intraoral pressure differences.

The material consisted of words containing coronal voiced and voiceless obstruents: stops /t d/, fricatives /s z ɕ ʑ/ and affricates /t͡s d͡z t͡ɕ d͡ʑ/[4], see Table 5. The obstruents appeared either in the intervocalic a_a context or in the word-initial prevocalic _a context. All words were bisyllabic with a stress falling on the first syllable. The subjects were asked to read all presented words embedded in a frame sentence *Powiedziała … do niego* 'She said … to him' and to repeat each sentence five times. The sentences were given in a randomized order.

Table 5: Polish words used in the experiment.

	initial	medial
/d/	/d/ata 'date'	ra/t/a 'installment'
/t/	/t/aca 'tray'	ra/d/a 'advice'
/z/	/z/araz 'in a moment'	ka/s/a 'cash'
/s/	/s/arna 'roe dear'	ga/z/a 'gauze'
/ɕ/	/ɕ/ara 'sulfur'	Ka/ɕ/a proper name
/ʑ/	/ʑ/arnka 'grain' pl.	Ka/ʑ/a proper name
/d͡z/	/d͡z/aga 'Dzag' name	sa/d͡z/a 'soot'
/t͡s/	/t͡s/ara 'tsar' gen.	ta/t͡s/a 'tray'
/d͡ʑ/	/d͡ʑ/ało 'cannon'	na/d͡ʑ/ać 'to stuff'
/t͡ɕ/	/t͡ɕ/ało 'body'	bra/t͡ɕ/a 'brother' pl.

The data were normalized for baseline shifts by setting the zero line in accordance with the closure phase in the airflow data and for the air pressure measurements in accordance with the pressure in a vowel [a]. Three parameters were measured in the segments:

(i) airflow peak
(ii) air pressure peak
(iii) duration of the components (stop and fricative phase in the case of affricates)

In Figure 6 the measurement points are shown for /ata/ excerpted from *rata* 'installment'.

---

[4] Polish retroflex affricates /t͡ʂ d͡ʐ/ were not considered due to the lack of the appropriate test words.

closure   release

a   **t**   **a**
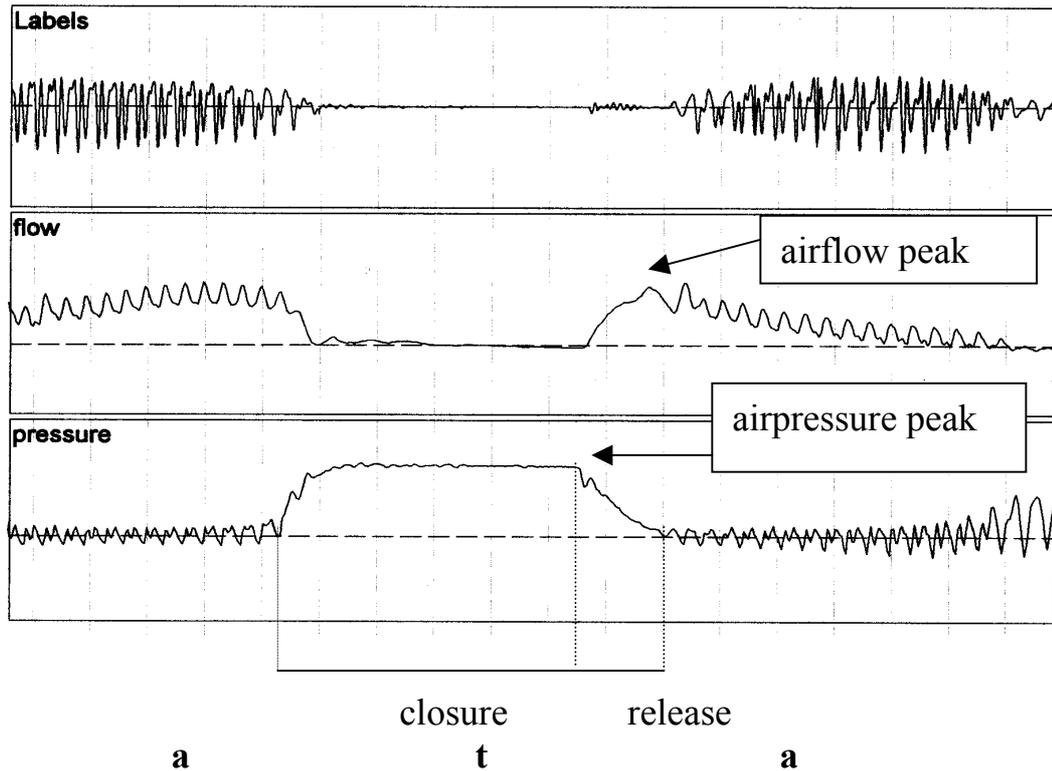
Figure 6: Measurement points in [t] excerpted from *rata*.

5.2 Results

The following discussion is focussed on the results of air pressure measurements which appear as the most relevant for the purposes of the present study. The comparison between voiced and voiceless affricates is of particular importance. Before presenting the results some discussion of the data is in order. Figure 7 illustrates how voiceless affricates are typically realized in the intervocalic position by Polish speakers.

In the realization of the closure phase of [t͡s] there is no airflow in contrast to the frication phase where the airflow is rising. As far as the air pressure is concerned, it is very high in the closure phase and diminishes in the frication phase.

The aerodynamics of the voiced affricates is notably different from their voiceless counterparts, as illustrated by Figure 8. The voiced affricate [d͡z] shows voicing throughout the whole segment. Therefore, it is difficult to determine where the boundary between closure and frication phase is. The airflow is nearly 0 and it slightly rises in the frication phase. Regarding the air pressure, it is higher in the closure phase and it decreases in the frication phase. The main difference between voiceless and voiced affricates is not only the amount and level of air pressure (see below) but the fact that air pressure varies

throughout both the closure phase and the frication phase, indicating the vibration of the vocal folds.
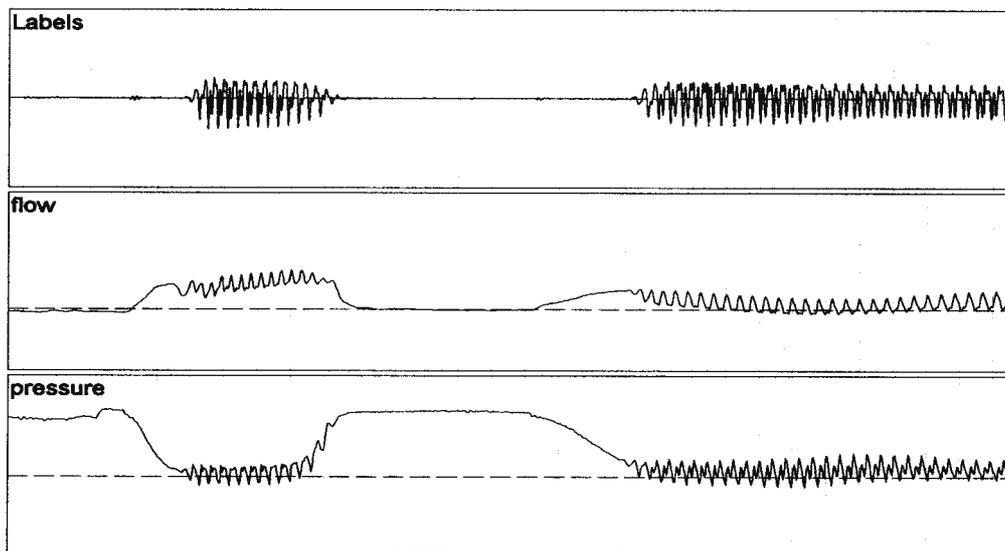
Figure 7: [at͡sa] excerpted from *cara*.
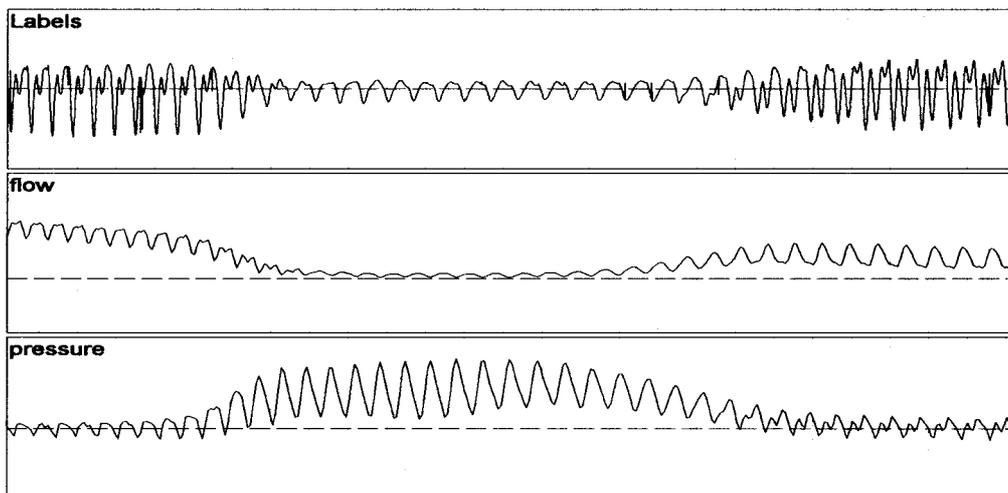
Figure 8: [ad͡za] excerpted from *sadza*.

Let us proceed to the results. Figure 9 presents average air pressure peak values as obtained for individual speakers.
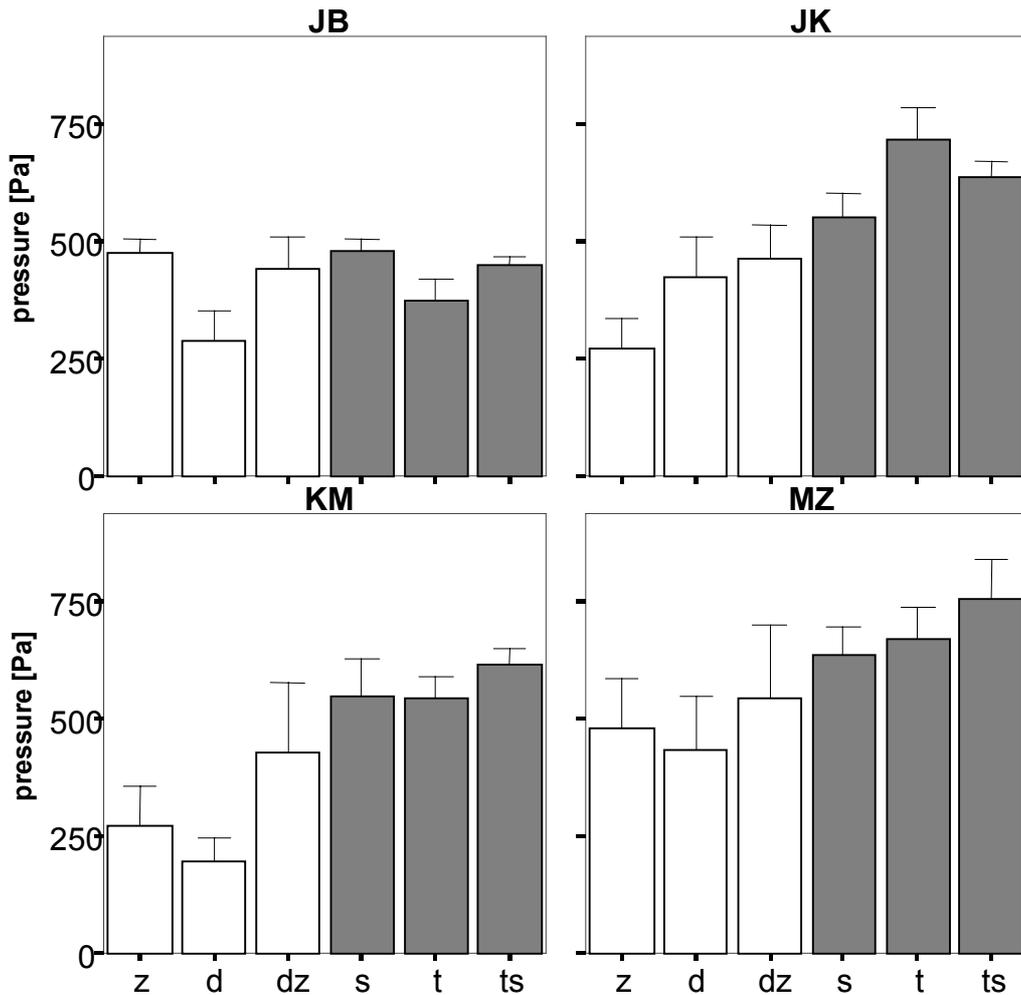
Figure 9: Bar plots showing average air pressure peak values with standard error as obtained for individual speakers (subplots); grey bars = voiceless phonemes, white bars =voiced phonemes.

The results show that for three speakers, JK, KM and MZ, the peaks values for voiceless segments are significantly higher than for the corresponding voiced affricates. The only exception is speaker JB in whose pronunciation the air pressure peaks in voiced items are as high as in voiceless items. A closer inspection of the data reveals, however, that in some cases the affricate – especially its fricative part – is devoiced, i.e. it does not show voicing. Such realizations appeared in the intervocalic position, but especially in the word-initial position, as exemplified in Figure 10. They have also been found occasionally in the pronunciation of other speakers.
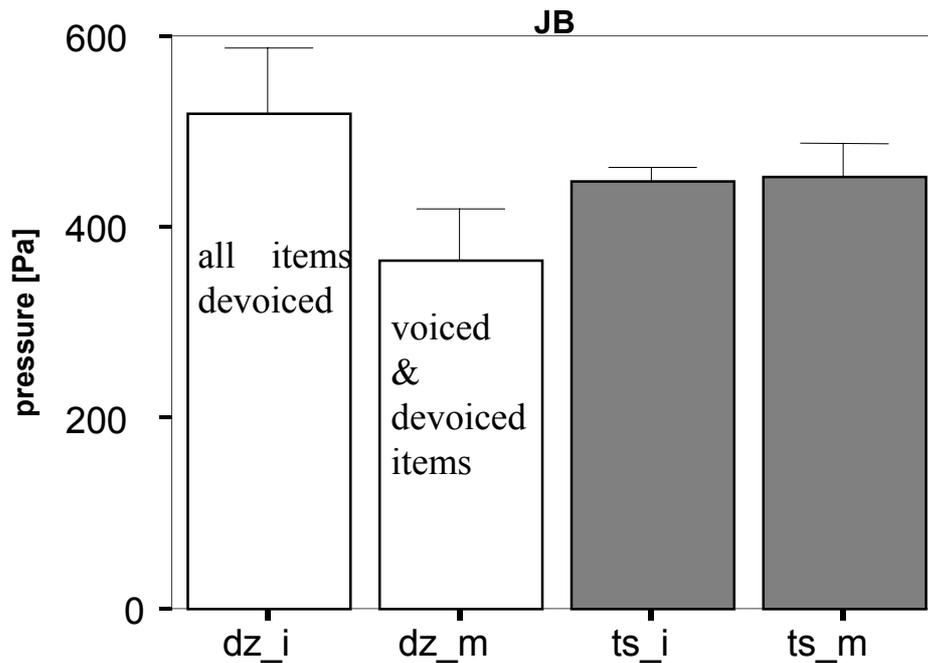
Figure 10: Bar plots showing average air pressure peak values and standard error as obtained for affricates [d͡z] (white bars) and [t͡s] (grey bars) in word-initial (_i) and word-medial position (_m) by speaker JB.

The air pressure profile of [d͡z] frequently found in the pronunciation of speaker JB (see Figure 101) and occasionally in the pronunciation of other speakers provides a possible answer as to why voiced affricates are devoiced. It seems that in the production of affricates an air pressure conflict occurs: on the one hand, the pressure naturally rises in the closure phase, and on the other hand, it has to be lower for maintaining voicing in the closure and in the frication phase. But, at some point the air pressure in the closure reaches the point where the voicing ceases. It appears to be difficult to balance the pressures in both phases. The difficulty is increased by the fact that the frication phase also needs a higher pressure for creating turbulence (Ohala 1983). Hence, in the production of voiced affricates a balance has to be reached between (i) raising air pressure in the closure, (ii) lower air pressure in the frication phase which is required for maintaining voicing and (iii) high air pressure in the frication phase required for creating turbulence. It also appears that the word-initial position is of special difficulty for producing voiced affricates which could be attributed to observation (i) and to the fact that the initial position is always long which favours devoicing. In other contexts, e.g. in the intervocalic position, the air pressure does not rise so significantly. An indirect piece of evidence in favour of

this proposal is the fact that the intervocalic context (but not the word-initial one) favors voicing cross-linguistically.
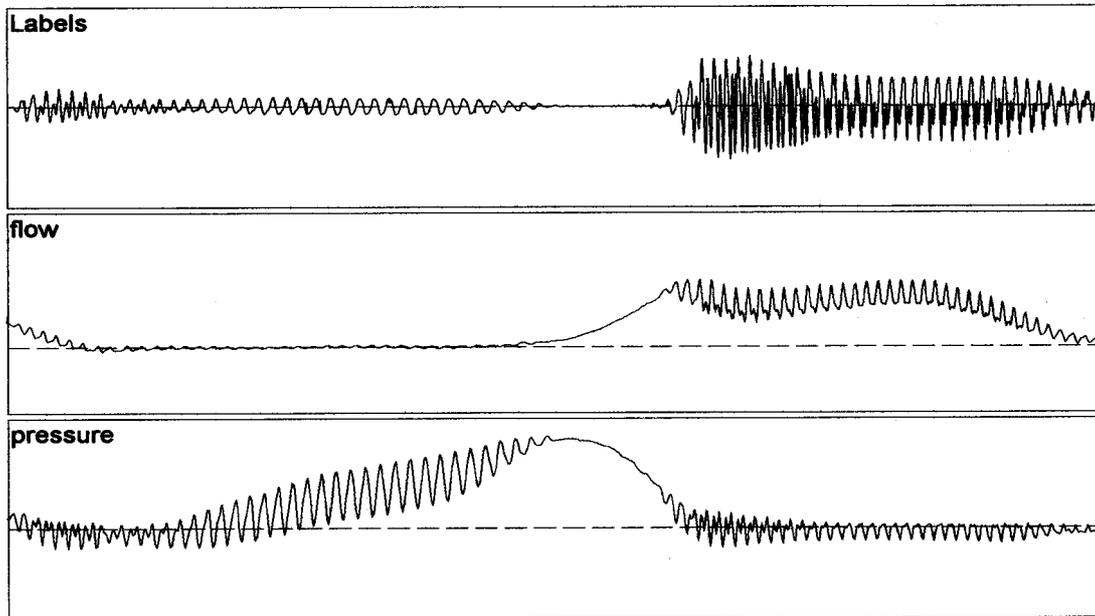
Figure 11: Devoicing in [d͡z] in word-initial position.

Other results are in line with studies examining the difference between voiced and voiceless items, namely that the airflow peak in voiced items is consistently lower than the corresponding peak in voiceless items. This is due to the wider glottal opening in the voiceless items; see Figure 12.

The duration measurements show that (i) the closure duration is significantly longer in voiced and voiceless affricates than the frication duration and (ii) more specifically, the closure of the voiceless affricates is significantly longer than the closure of voiced affricates (see Figure 13). The latter conclusion is drawn for all speakers apart from JB, whose affricates are devoiced, as discussed above. Regarding the frication duration, it holds for all four speakers that the frication lasts longer in voiceless affricates than in voiced ones. The results are consistent with those obtained by Dogil & Jessen (1989) and Zygis (2006).
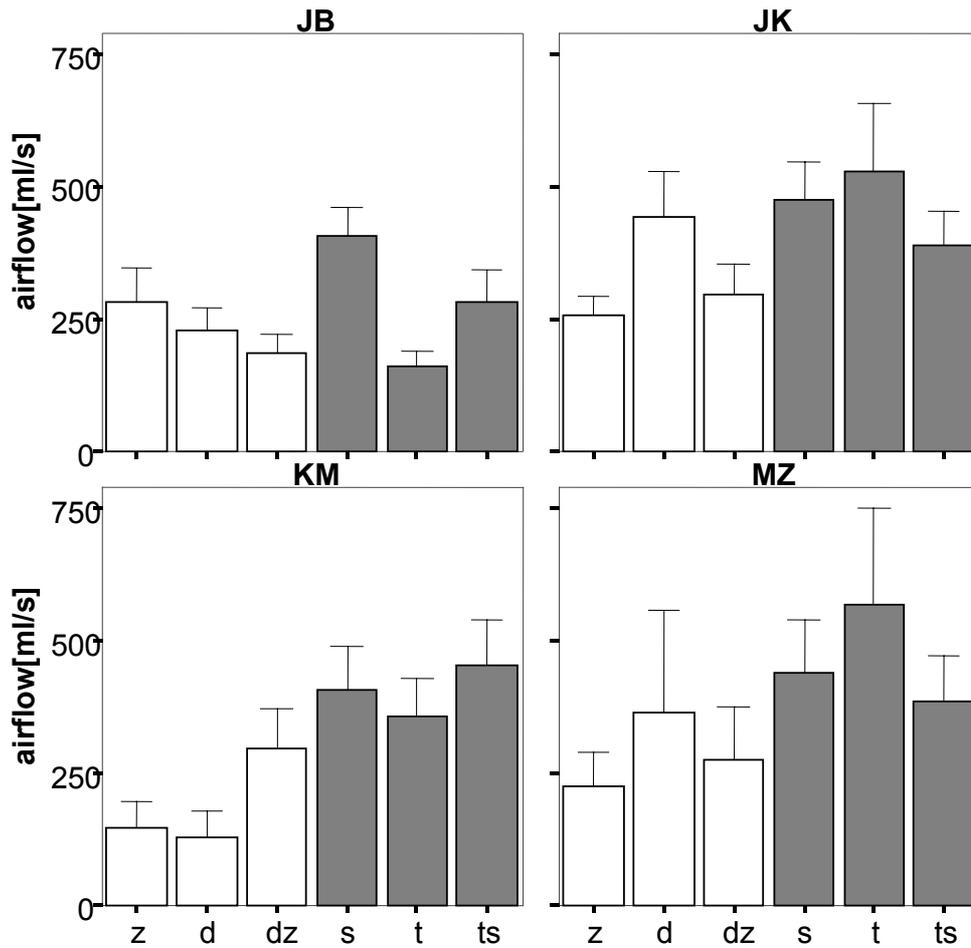
Figure 12: Bar plots showing average airflow peak values and standard error as obtained by individual speakers (subplots); grey bars = voiceless phonemes, white bars = voiced phonemes.
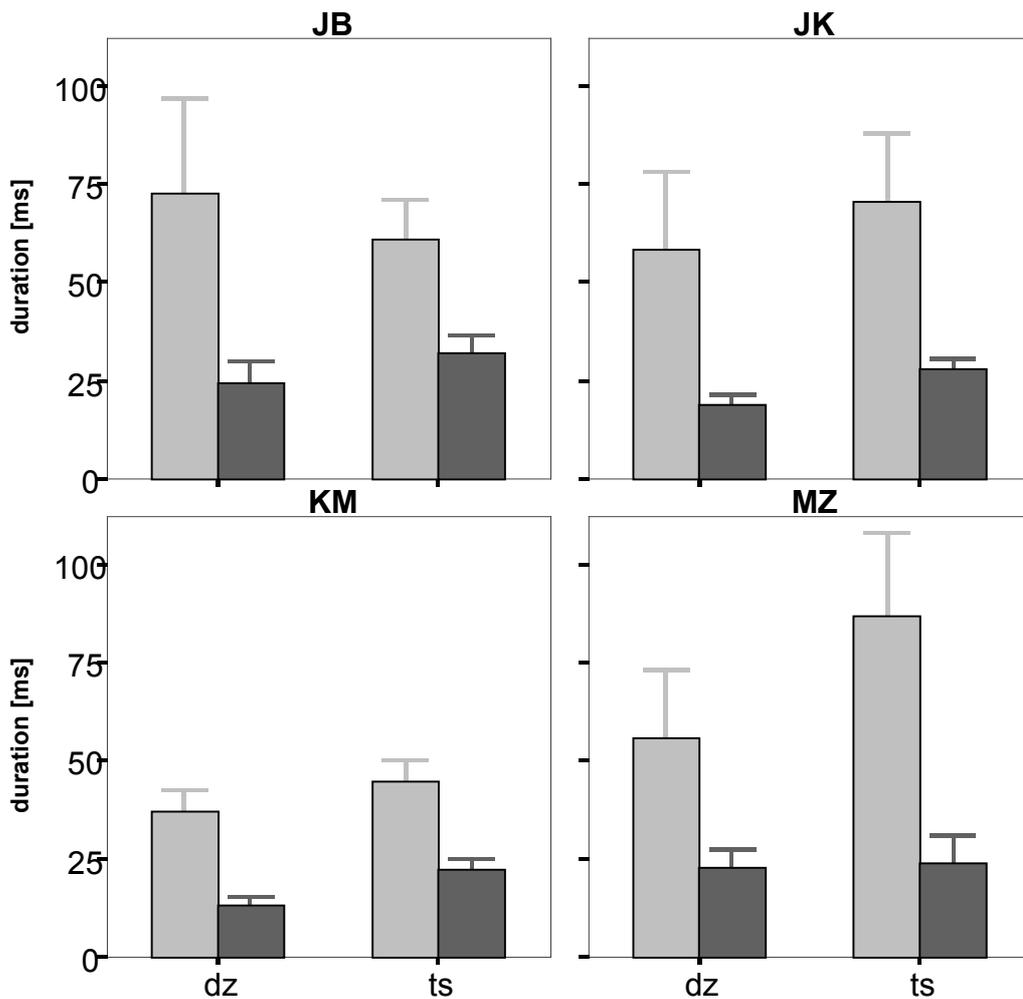
Figure 13: Bar plots with standard error showing average values of closure and frication duration for individual speakers (subplots); grey bars= closure duration, black bars = frication duration.

In summary, the results presented in this section, especially those related to air pressure measurements, provide a possible answer as to why voiced affricates are avoided. It is namely the complexity of air pressure relations which are inherently antagonistic in the frication phase and aggravated by the raising air pressure in the closure phase, especially in the word-initial position.

6   Conclusion

This paper reveals that several typologically unrelated languages share the tendency to avoid voiced sibilant affricates. In the phonemic inventories of Slavic, Romance, Germanic and other languages, a phonemic gap is attested: voiced coronal affricates do not occur while their voiceless counterparts do. In several of these inventories, stops and fricatives create a contrast with respect to

voicing as well. The avoidance of voiced affricates is supported by examples from phonological processes in which the existing voiced affricates change to other sounds.

It has been argued that this cross-linguistic tendency is to be attributed to the articulatory and aerodynamic complexity of affricates. As evidenced by the experimental results, the frication component of affricates is subject to devoicing. This is due to the air pressure rise in the stop part which is too high to maintain optimal conditions for fricative voicing.

This first attempt at explaining the avoidance of voiced affricate should be deepened by EPG and transillumination studies in order to gain more insight into the laryngeal-oral coordination. Despite the well-known fact that the glottis is more open in voiced fricatives than in voiced stops it remains to be seen how the opening of the vocal folds is coordinated with the supralaryngeal articulation in voiced affricates. Finally, this study poses a question as to why languages choose different sounds when dispensing with voiced affricates. I leave this question open for further study.

ACKNOWLEGMENTS

References

Avery, P. & W. Idsardi (2001). Laryngeal dimensions, completion and enhancement. In: Hall, T. A. (ed.) *Distinctive Feature Theory.* Berlin: de Gruyter. 4-70.

Boersma, P. (1998). *Functional Phonology.* Den Haag: Holland Academic Press.

Catford, J. C. (1977). *Fundamental Problems in Phonetics.* Edinburgh: Edinburgh University Press.

Dalewska-Greń, H. (2002). *Języki słowiańskie.* Warszawa: Wydawnictwo Naukowe PWN.

Dixit, R.P. & W.S. Brown (1985). Peak magnitudes of oral air flow during Hindi stops (plosives and affricates). *Journal of Phonetics* 13, 219-234.

Dogil, G. & M. Jessen (1989). Phonologie in der Nähe Phonetik. Die Affrikaten im Polnischen und Deutschen. In: Prinzhorn, M. (ed.) *Phonologie.* Opladen: Westdeutscher Verlag. 223-279.

Gianneli, L. & L. Savoia (1979). Indebolimento Consonantico in Toscana. *Revista Italiana di Diallettologia* 2, 23-58.

Greenberg, R. (2000). The dialects of Macedonia and Montenegro: Random Linguistic Paralles or Evidence of a Sprachbund? *Južnoslovenski filolog* 56, 295-300.

Hall, T. A. (1992). *Syllable Structure and Syllable-Related Processes in German.* Tübingen: Max Niemeyer Verlag.

Holes, C. (1990). G*ulf Arabic.* Routledge: London.

Iverson, G. & J. C. Salmons (1995). Aspiration and laryngeal representation in Germanic. *Phonology* 12, 369-396.

Iverson, G. & J. C. Salmons (2003). Legacy specification in the laryngeal phonology of Dutch. *Journal of Germanic Linguistics* 15, 1-26.

Ivić, P. (1958). *Die serbokroatischen Dialekte. Ihre Struktur und Entwicklung.* Vol. 1: Allgemeines und die štokavische Dialektgruppe. 'S-Gravenhage: Mouton.

Kehrein, W. (2002). *Phonological Representation and Phonetic Parsing: Affricates and Laryngeals.* Tübingen: Max Niemeyer Verlag.

König, L.L. & S. Fuchs (2007). The sensivity of intraoral pressure in consonants and consonant clusters to following vowel context in German. ICPhS 16, 641 - 644.

Kučera, H. (1961). *The Phonology of Czech.* The Hague: Mouton & Co.

Ladefoged, P. (1964). Comment on 'Evaluation of methods of estimating subglottal air pressure'. *Journal of Speech and Hearing Research* 7, 291-292.

Ladefoged, P. & W. Z. Wu (1984). Places of articulation: An investigation of Pekingese fricatives and affricates. *Journal of Phonetics* 12, 267-278.

Leal, D. (1972). *Chitwan Tharu. Phonemic Summary.* Kirtipur: Summer Institute of Linguistics and Institute of Nepal Studies Tribhuvan University.

Lieutard, H. (2004). *Phonologie et morphologie du parler occitan de Graulhet (Tarn). Structure, contenu et rôle de la syllabe.* Montpelhièr: CEO.

Lombardi, L. (1994). Laryngeal features and laryngeal neutralization. *Lingua* 98, 46-53.

Lombardi, L. (1999). Positional faithfulness and voicing assimilation in Optimality Theory. *NLLT* 17, 276-302.

Maddieson, I. (1984). *Patterns of Sounds.* Cambridge: Cambridge University Press.

Maddieson, I. & K. Precoda (1992). UPSID and Phoneme. UCLA Phonological Segment Inventory DatabaseVersion 1.1.

Mooshammer, C., Geumann, A., Hoole, P., Alfonso, P., van Lieshout, P. & S. Fuchs (2003). Coordination of lingual and mandibular gestures for different manners of articulation. In: Solé, M. J., Recasens, D. & J. Romero (eds.) *Proceedings of the 15th ICPhS, Barcelona.* 81-84.

Mooshammer, C., Hoole, P. & A. Geumann (2007). Jaw and order. Language and Speech 50, 145-176.

Ohala, J. J. (1983). The origin of sound patterns in vocal tract constraints. In: MacNeilage, P. F. (ed.) *The Production of Speech.* New York: Springer Verlag. 189-216.

Ohala, J. J., Solé, M. J. & G. Ying (1998). The controversy of nasalized fricatives. *Proceedings of the 135th Meeting of the International Congress on Acoustics of Acoustical Society of America. Seattle, Washington.* 2921-2922.

Schuster-Šewc, H. (1996 [²1999]). *Grammar of the Upper Sorbian Language.* München: Lincom Europa.

Scully, C. et al. (1992). Analysis and simulation of speaker's aerodynamic and acoustic patterns for fricatives. *Journal of Phonetics* 20, 39-51.

Short, D. (1993). Czech. In: Comrie, B. & G. G. Corbett (eds.) *The Slavonic Languages.* London: Routledge. 455-532.

Solé, M. (1998). Phonological universals: Trilling, voicing, and frication. *Proceedings of the Twenty-Fourth Annual Meeting of the Berkeley Linguistics Society.* 403-416.

Solé, M. (2002). Assimilatory processes and aerodynamic factors. In: Gussenhoven, C. & N. Werner (eds.) *Papers in Laboratory Phonology.* Berlin: de Gruyter. 351-386.

Solé, M.-J. (2003). Aerodynamic characteristics of onset and coda fricatives. *Proceedings of the 15th ICPhS, Barcelona.* 2761-2764.

Steriade, D. (1997). Phonetics in Phonology: the Case of Laryngeal Neutralisation. Ms. UCLA.

Stevens, K. N. (1993a). Modelling affricate consonants. *Speech Communication* 13, 33-43.

Stevens, K. N. (1993b). Models for the production and acoustics of stop consonants. *Speech Communication* 13, 367-375.

Stevens, K. N. (1998). *Acoustic Phonetics.* Cambridge, MA: MIT Press.

Svirsky M. A., Stevens, K. N., Matthies, M. L. & J. S. Perkell (1992). Tongue surface deformation during obstruent stop consonants. *JASA* 92, 2390.

Tinelli, H. (1981). *Creole Phonology.* The Hague: Mouton.

Westbury, J. R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *JASA* 73, 1322-1336.

Wetzels, W. L. & J. Mascaro (2001). The typology of voicing and devoicing. *Language* 77, 207-244.

Wierzchowska, B. (1980). *Fonetyka i fonologia języka polskiego.* Wrocław: Zakład Narodowy im. Ossolińskich: Wydawnictwo Polskiej Akademii Nauk.

Zilyns'kyj, I. (1979). *A Phonetic Transcription of the Ukrainian Language.* Cambridge, MA: Harvard University Press.

Zygis, M. (2006). Contrast Optimisation in Slavic Sibilant Systems. Habilitationsschrift. Humboldt-Universität zu Berlin.

Focus and prominence in Chichewa, Chitumbuka and Durban Zulu[*]

Laura J. Downing
Zentrum für Allgemeine Sprachwissenschaft, Berlin

Much work on the interaction of prosody and focus assumes that, cross-linguistically, there is a necessary correlation between the position of main sentence stress (or accent) and focus, and that an intonational pitch change on the focused element is a primary correlate of focus. In this paper, I discuss primary data from three Bantu languages – Chichewa, Durban Zulu and Chitumbuka – and show that in all three languages phonological re-phrasing, not stress, is the main prosodic correlate of focus and that lengthening, not pitch movement, is the main prosodic correlate of phrasing. This result is of interest for the typology of intonation in illustrating languages where intonation has limited use and where, notably, intonation does not highlight focused information in the way we might expect from European stress languages.

1   Introduction

In this paper, I investigate the role of intonation – defined as phrase or sentence level prominence-related prosody – in three Bantu languages. Following work like Cruttenden (1997), Jun (2005), Ladd (1996), Gussenhoven (2004), I assume that length, loudness or intensity, and/or pitch are the primary prosodic

---

correlates of intonational prominence. Whereas most work on intonation has been concerned with pitch (movement), I show that in these Bantu languages another prosody - culminative vowel lengthening – fulfills the demarcative function of providing a prosodic cue to syntactic boundaries (or other groupings at the phrase or sentence level) that pitch accent provides in more familiar European languages.

In many languages, sentence level prosody is conditioned not only by syntactic factors but also by semantic ones, primarily focus and utterance type, like question vs. declarative (see Bruce 1977; Gussenhoven 1984, 1996, 2004; Ladd 1996 and many others). That is, the same type of prosody often can serve two functions, signaling both constituent edges and information structure. Indeed, much work on the interaction of prosody and focus assumes that, cross-linguistically, there is a necessary correlation between the position of sentential prominence (or accent) and focus. The formulation below comes from Samek-Lodovici (2005: 697), and very similar proposals are found in: Reinhart 1995; Selkirk 1984, 1995, 2004; Roberts 1998; Rooth 1992, 1996; Szendröi 2003; Truckenbrodt 1995:

(1) PROMINENCE-FOCUS
     For any XP_f and YP in the focus domain of XP_f, XP_f is prosodically more prominent than YP,

In other words, a focused element (XP_f) is required to have the culminative (i.e., highest) level of some suprasegmental feature which correlates with prominence – pitch, duration and/or amplitude or intensity – in its prosodic domain, normally, the Intonation Phrase. For example, in English, 'normal' sentence stress is sentence-final. However, a sentence-initial focused subject must receive sentence stress. That is, sentence stress must 'move' in English to satisfy the Prominence-Focus correlation, as shown in (2), where small caps indicate the word with sentence stress:

(2) Broad vs. narrow focus and stress in English (Samek-Lodovici 2005: 688)
(a) [John has LAUGHED.]$_f$ Context: What happened?
(b) [JOHN]$_f$ has laughed.   Context: Who has laughed?
(c) *[John]$_f$ has LAUGHED.   Context: Who has laughed?

However, other work on intonation, like Ladd (1996) and Hayes & Lahiri (1991), has pointed out that the PROMINENCE-FOCUS correlation in (1) is mainly supported by European word stress languages like English where cues for sentence level prominence co-occur on the head syllable of focused constituents. A more universal correlate of focus, they argue, is prosodic (re-)phrasing:

narrow focused constituents trigger different prosodic phrasing from broad focused constituents. Sentence level prominence is a potential – not universal – cue to prosodic phrasing, not directly to focus. I show that sentential or phrasal prominence is not a correlate of focus in these three Bantu languages, as we find numerous mismatches between prominence and focus. These languages support the proposal that prominence is, indeed, a cue to phrasing and that phrasing, rather than prominence, is the more universal correlate of focus.

2   Prosodic phrasing and stress in Durban Zulu, Chichewa and Chitumbuka

Chichewa and Chitumbuka are two of the three major languages of Malawi (Yao is the third). Durban Zulu is a dialect of one of South Africa's official languages. This section presents the prosodic phrasing algorithms for the three languages, showing the role of syntax and focus in conditioning prosodic phrasing. The data and phonological analysis presented in this section come from my own fieldwork on these languages, except where noted otherwise.

2.1   Phonological phrasing algorithms

In all three languages, both syntax and focus play a role in determining the phonological phrasing. However, as we shall see, different syntactic constituents define neutral phonological phrasing in the three languages. Focus also plays a different role in each language. In all three languages, lengthening of the phrase penult syllable is the easiest to identify – and most consistent – correlate of prosodic phrasing. (Parentheses in the data indicate prosodic phrasing.) Although this cannot be effectively transcribed, it should be noted that in all three languages, the penult of the sentence-final prosodic phrase is noticeably longer than the penults of sentence-medial prosodic phrases. That is, sentence penult vowles have culminative lengthening at the sentence level. (We return to this point in section 3.1, below.)

2.1.1   Durban Zulu prosodic phrasing

Cheng & Downing (2007a, b) show that prosodic phrasing in Durban Zulu is almost identical to that of Xhosa, a closely related Nguni Bantu language analyzed by Jokweni (1995). Under broad focus, there is a prosodic phrase break at the right edge of CP (roughly, a clause), as shown below:[1]

---

[1]   In the morpheme-by-morpheme glosses in all of the data, the numbers indicate the agreement class (or gender) of the noun and the verbs and modifiers which are required to

(3)　Durban Zulu neutral phrasing

(a)　The teacher read to the parents a letter.

　　　[_{CP}(úm-fúndísi　　ú-fúndel-ê:　　ábá-zal'　　ín-cwa:di)].
　　　　　1-teacher　　　　1-read to-TAM　　2-parent　　9-letter

(b)　We believe that the children are playing outside.

[_{CP}(Si-khólwa　　[_{CP} úkúth'　　ábá-ntwána　　bá-dlalá　　phá:ndle)]].
　　we-believe　　　　that　　　2-child　　　2-play　　　outside

(c)　The man who is wearing a hat saw the visitors.

[_{CP} [_{DP} [_{CP} (Ín-dod'　　é-gqoke　　ísí-gqo:ko)]]　í-bon-é　　ízi-vaká:shi)].
　　　　9-man　　　REL9-wear　　7-hat　　　9-see-TAM　　　8-visitor

(d)　We like the hat the man is wearing.

[_{CP} (si-thánd'　　[_{DP} [_{CP} ísí-gqok'　　ín-dod'　　é-si-gqok-ilê:-yo)]]].
　　we-like　　　　　　6-hat　　　9-man　　REL9-OM6-wear-TAM-Rel

Focus only indirectly conditions prosodic phrasing in Durban Zulu. Focused verb complements must occur in Immediately After the Verb (IAV) position, and a prosodic phrase break separates them from any other postverbal complements. This is illustrated by the Wh-question and answer pairs in (4a, b), below. The answer to a Wh-question on the subject is clefted, and clefts are obligatorily set off by a prosodic phrase break, as shown in (4ci). Cheng & Downing (2007a, b) argue that these phrasings are also consistent with proposing a prosodic phrase break at the right edge of CP, as the right edge of a cleft and the right edge of IAV position correspond to the right edge of CP.

(4)　Durban Zulu focused postverbal complements and clefts

　　　Right dislocations – WH particle and answer–both focused– are in IAV

(a)　Q-What did the visitors buy for their families?

　　　(ízí-vaká:shi)　zí-**yí**-thengel-ê:-ni)　　(ímí-ndeni　　yâ:zo)?
　　　10-visitors　　　10SM-OM4-buy for-TAM-what　4-families　　4.their

　　A-The visitors bought clothing for their families.

　　　(ízí-vakáshí　bá-**yí**-thèngel-é　　ízí-ngu:bo)　　(ímí-ndeni　yâ:bo).
　　　10-visitors　　2SM-OM4-buy for-TAM　10-clothes　　　4-families　2.their

---

agree with the noun. For example, in (3a), the word for 'teacher' is in agreement class 1. As it is the subject, the verb's subject prefix is also from agreement class 1. (See Doke 1954 and Mchombo 2004 for more discussion of Bantu agreement class systems.) Other frequently used abbreviations in the morpheme-by-morpheme glosses are: REL 'relativizer'; Cop 'copula'; TAM 'tense/aspect marker'; OM 'object agreement prefix'; Loc 'locative'.

(b)   Q   What did the teacher give to the winner?
      (úm-fúndís'   ú-**m**-nik-ê:-ni)   (ó-wín-i:le)?
      1-teacher   1-OM1-give-TAM-Q   Rel1-win-TAM
      A   The teacher gave a medal to the winner.
      (úm-fúndí:sí)   ú-**m**-nikez-é: í-méndlè:la)   (ó-wín-i:le).
      1-teacher   1-OM1-give-TAM   5-medal   Rel1-win-TAM

Clefts - note distinction in tone on the first syllable of the head of cleft vs. non cleft

(ci)   *clefted subject*, Answers, 'Who is playing at school?'

The children are playing at school. = It is the children who are playing at school.

    (<u>Abá-ntwa:n'</u>)   (abá-dlal'   é-sí-kóle:-ni).
     Cop2-child   Rel2-play   Loc-7- school-Loc

(cii) *subject relative*

The children who are playing at the school live near the school.

(<u>Ábá-ntwán'</u>   <u>ábá-dlal'</u>   é-sí-kóle:-ni)   (bá-hlál'   édúzáne   nésí-kó:le).
2-child   Rel2-play   Loc-7-school-Loc   2-live   near   to.7-school

To sum up this section, in Zulu, prosodic phrasing is syntactically conditioned by the right edge of CP. Focus plays an indirect role in conditioning prosodic phrasing. Clefts and the IAV focus position arguably are each at the right edge of a CP, and so also condition a prosodic phrase break by the usual phrasing algorithm.

2.1.2   Chichewa prosodic phrasing

As Kanerva's (1990) detailed study of prosodic phrasing in Chichewa shows, under neutral phrasing a smaller syntactic constituent (roughly, XP) conditions prosodic phrase breaks in this language. The subject NP, VP (verb and all its complements) and a Topic NP are the three potential syntactic subconstituents of the clause in Kanerva's analysis.[2] Each of these is parsed into its own prosodic phrase. Note that prosodic phrases, like the syntactic phrases they parse, can contain a single prosodic word:

(5)   Chichewa neutral phrasing (Kanerva 1990)
(a)   (fíisi)   (a-na-gúlá   chi-péwá ku-San Francíscó   dzuulo)
      1.hyena   1-TAM-buy   7-hat   Loc-San Francisco   yesterday
      The hyena bought a hat in San Francisco yesterday.

---

[2]   See Bresnan & Mchombo (1987), Bresnan & Kanerva (1989) and Mchombo (2004) for detailed discussion of this syntactic approach to Chichewa sentence structure.

(b) (aána) (a-na-góná m-nyumbá yá mávúuto)
       2.child 2-TAM-sleep Loc-9.house 9.of Mavuto
       The children slept at Mavuto's house.
(c) (a-leenje) (zi-ná-wá-luuma) (njúuchi)
       2-hunter 10-TAM-2.OM-bite 10-bee
       The hunters, they bit them, the bees [did].

Unlike Durban Zulu, Chichewa allows in situ focus of verb complements. As Kanerva's (1990) work shows, we find the following pattern of phrasing under focus. A prosodic phrase boundary follows a constituent in narrow focus, and non-focused VP complements are each parsed into a separate prosodic phrase. Downing et al's (2004) work on a different variety of Chichewa shows that VP-final focused constituents are preceded by a prosodic phrase boundary in some varieties of Chichewa. This study also confirms Myers' (1996) observation that the phrase containing the focused element is raised in pitch, indicated with an upward arrow, in some varieties. (Focus raising is discussed briefly in section 4.2, below.) The data in (6) and (7) are from Downing et al.'s (2004) study of Ntcheu Chichewa:

(6) S/he hit the house with a rock.
(a) /A-ná-meny-a nyumbá ndí mwalá/
       s/he-SIMPLE PAST-hit-FV house with rock
(b) broad VP focus
       (A-ná-mény-a nyumbá ndí mwáálá).
(c) Oblique PP focus
       Q (A-ná-ménya nyuúmbá) (ndí mwáálá) (kapéná ndí ndoodo)?
          'Did s/he hit the house with a rock or with a stick?'
       A (A-ná-mény-a nyuúmbá) (ndí mwáálá) ↑.
(d) Object NP focus
       Q (A-ná-ménya chiyáani) (ndi mwáálá)?
          'What did s/he hit with the rock?'
       A (A-ná-mény-a nyuúmbá) ↑ ‖ (ndí mwáálá).
(e) V focus
       Q (Nyuúmba) (i-ná-táá-ní)?
          'What happened to the house?'
       A (A-ná-méeny-a) ↑ (nyuúmbá) (ndí mwáálá).

(7)   The chief gave the child clothes.
(a)   broad VP focus

   (M-fúumu)   (i-ná-páts-a   mw-aná   zóóváala).
   9-chief   9-Simple Past-give-FV   1-child   10.clothes

(b)   Direct Object NP focus

   Q   (A-ná-páts-a mw-aáná) (chiyáani)?
       What did he give to the child?
   A   (A-ná-páts-a mw-aáná) (zóóváala) ↑.

(c)   Indirect Object NP focus

   Q   (A-ná-pátsa ndáání) (zóóváala)?
       'Who did he give clothes to?'
   A   (A-ná-páts-a mw-aáná) ↑ ‖ (zóóváala).

To sum up this section, in Chichewa, prosodic phrase breaks are syntactically conditioned by the major subconstituents of the clause: Subject, VP and Topic. Focus plays a direct role in conditioning prosodic phrasing, as constituents within the VP must be set off by a prosodic phrase break when they are focused in situ.

2.1.3   *Chitumbuka prosodic phrasing*

Downing (2006) shows that neutral prosodic phrasing in Chitumbuka is conditioned by the right edge of NP.[3] As in Chichewa, this means that Subject NPs and Topics are phrased separately from the rest of the clause. In contrast to Chichewa, the entire VP does not form a single prosodic phrase unless the VP is very short. Instead, a verb plus its first complement form a single phrase, and following complements are generally phrased separately. As a result, the neutral phrasing of VPs in Chitumbuka is essentially identical to the focus-induced phrasings of VPs in Chichewa which were illustrated in (6) and (7). Compare in particular (8e), below, with (6c, d):

(8)   Chitumbuka neutral phrasing
(a)   We are cooking porridge.

   (ti-ku-phika   sî:ma)
   we-TAM-cook   9.porridge

(b)   The children help the friends.

   ([β]-â:na)   ([β]a-ku-[β]a-vwira   [β]a-bwê:zi)
   2-child   2-TAM-2.OM-help   2-friend

---

3   Chitumbuka is the least well studied of these three languages. There is no thesis length work on prosodic phrasing or even a grammar of the language. See Downing (2006) for a preliminary sketch of the syntax and prosody of focus in Chitumbuka.

(c)   We saw a thief at the market.
    (ti-ka-wona   mu-nkhúngu   ku-msî:ka).
    we-TAM-see   1-thief   Loc-3.market

(d)   The women sewed clothes for the bride.
    ([β]-anakâ:zi)   ([β]a-ka-sona   vy-akuvwara   vya   mu-kwâ:ti.)
    2-woman   2-TAM-sew   8-clothes   8.of   1-bride

(e)   The boy hit a house with a rock.
    (m-nyamâ:ta)   (wa-ka-timba   nyû:mba)   (na   lî:bwe).
    1-boy   1-TAM-hit   9.house   with   5.rock

Focus regularly conditions prosodic rephrasing in two main contexts in Chitumbuka. The answer to a Wh-question and a Wh-question particle are followed by a prosodic phrase break, as shown in (9a) vs (9b) and in (10).

(9)   Wh-Qs on verb complements and Answers

(a)   The woman washes clothes for the children.   [neutral reading]
    ([β]a-mâ:ma)   ([β]a-ku-chápa   vy-akuvwára   vya   [β]-â:na).
    2P-woman   2P-TAM-wash   8-clothes   8.of   2-child

(b)
Q-   Who does the woman wash clothes for?
    (Kâ:si),   ([β]a-mâ:ma   [β]a-ku-chapíra   njâ:ni)   (vy-akuvwâ:ra)?
    Q   2P-woman   2P-TAM-wash for   1.who   8-clothes

A-   The woman washes clothes for the children.
    ([β]a-mâ:ma)   ([β]a-ku-chapíra   [β]-â:na)   (vy-akuvwâ:ra).
    2P-woman   2P-TAM-wash for   2-child   8-clothes

(10)
Q-   Who did you buy the green mangoes for at the shop?
    (U-ka-mu-gulira   njâ:ni)   (mango   ya   [β]î:si)   (ku-gorosâ:ri)?
    1-TAM-1.OM-buy for   1.who   6.mango 6.of   unripe   Loc-grocery

A-   I bought green mangoes for my friend at the shop.
    (N-kha-mu-gulira   mu-nyâ:ne)   (mango   ya   [β]î:si)   (ku-gorosâ:ri).
    I-TAM-1.OM-buy for   1-my friend   6.mango 6.of   unripe   Loc-grocery

Association with focus morphemes – *pera* 'only'; *-so* 'also'; *yaye* 'no; not' – are also followed by an obligatory prosodic phrase break, as shown in (11):

(11)   Association with focus (Downing 2006)

(a)   *pera* 'only'
    They showed their homes only to the visitors.
    ([β]a-léndo   *pê:ra*)   ([β]a-ka-[β]onésya   pamúzi   pâ:wo)
    2-visitor   only   2-TAM-show   homes   their

(b)   *-so* 'also'
        Are you also weeding the *maize*?
        (Ku-limilirâ:-*so*)   (ngô:mâ:)?
        You/TAM-weed-also   9.maize

(c)   *yaye* 'no; not'
        The monkey did not make the child *cry*.
        (m-bwê:ngu)   (wa-ka- lísya   *yâ:ye*)   (mw-â:na).
         1-monkey   1-TAM-make cry   not   1-child

To sum up, in Chitumbuka, prosodic phrase breaks are syntactically conditioned by noun phrase edges, though an entire VP can be parsed into a single prosodic phrase if it is short. Focus plays a direct role in conditioning prosodic phrasing. Constituents within the VP which are focused in situ (i.e., WH-questions and answers) must be followed by a prosodic phrase break. Association with focus particles must also be followed by a prosodic phrase break.

2.2   Phrasal stress and focus

As we have seen, in all three of these Bantu languages, the prosodic phrase is the domain for the lengthening of phrase-penult syllables. (See Downing (to appear) for a recent survey of penult stress in Bantu languages.) Indeed, duration is a common cross-linguistic correlate of stress, as noted in work like Hyman (1977) and Odden (1999). We have also seen that, in all three languages, prosodic phrasing is conditioned by focus, at least indirectly. As a result, focused constituents often have a prosodic phrase boundary at their right edge, and so receive phrasal stress. We noted, though this cannot be effectively transcribed, that the penult of the sentence-final prosodic phrase is noticeably longer than sentence-medial lengthened penults in all three languages. This gives the sentence-final penult culminative prominence in the sentence. With this background in mind, we are prepared to consider whether these languages support the PROMINENCE-FOCUS correlation in (1). This question is taken up in the next section.

3   Mismatches between stress and focus

In this section I argue that the following mismatches between stress and focus in Chichewa, Durban Zulu and Chitumbuka show that there is no correlation between the position of prosodic prominence and focus. First, sentence stress – realized as extra penult lengthening – remains fixed on the final word of an utterance; it is not attracted to focus. Phrasal stress is also fixed on the last word of the phrase, not necessarily the one in narrow focus. Finally, we shall see that

in Chitumbuka, the productive focus particle, *-so* – the equivalent of English *also* introduced briefly in (11), above – attracts phrasal stress to its verbal host, not to its focused argument.

3.1   Sentence stress remains fixed in utterance final position

The PROMINENCE-FOCUS correlation in (1) requires focused constituents, as heads of the Intonation Phrase (roughly equivalent to a clause), to have the culminative prosodic prominence within their domain – that is, they should have the highest level of some measurable prosodic property associated with prominence, such as duration, pitch or intensity. This correlation is satisfied if stress is flexible, as in English and other Germanic languages, and moves to the stressed position, as shown in (2), above. It is also satisfied if word order is flexible, as in Italian and Hungarian (Samek-Lodovici 2005, Szendrŏı 2003), and focused words move to the stressed position. It can also be satisfied, as in Danish (Grønnum 1998) or French (Beyssade et al. 2004, Féry 2001), if pitch is compressed in non-focal constituents, lending focused constituents passive prominence. In all three of these Bantu languages, though, the highest degree of stress (defined as vowel lengthening) in the sentence is fixed on the penult syllable of the final prosodic phrase in the sentence. Downing et al.'s (2007) exploratory phonetic study shows that penults in the sentence-final prosodic phrase are significantly lengthened in Chichewa compared to sentence-medial (non-prepausal) phrase penult vowels. The penult vowels of phrases containing focused elements are never longer than penults in sentence-final phrases. My informal phonetic analysis of Durban Zulu and Chitumbuka shows a similar pattern for these languages.

   In short, the final constituent in a sentence is always made the most prosodically prominent in these three languages through penult lengthening, whether it contains the focused constituent or not.[4] This clearly violates the PROMINENCE-FOCUS correlation.

3.2   Phrasal stress remains fixed in phrase final position

In the data presented so far, entire XPs were in focus, and often these XPs (especially the noun phrases and prepositional phrases) and the prosodic phrases which parse them consist of just one prosodic word. In all of these cases, phrasal stress occurs in a position that is consistent with scope of focus as defined in Selkirk (1995), either the phrase-final word or the entire XP. A look at data illustrating contrastive focus in more complex XPs in the three languages shows

---

[4]   See Zerbian (2006) for detailed discussion of a similar finding for Northern Sotho.

that prosodic phrase boundaries fall at the right edge of the XP containing the focused word, and phrase stress falls consistently on the phrase penult syllable. This leads to mismatches between the position of phrasal stress and the position of focus, if the focused word is not at the right edge of its XP.

These points are illustrated by the Chitumbuka data in (12). As we can see, contrastive focus is clearly on the word towards the left edge of the prosodic phrase, but phrasal stress is assigned to the non-focused word which occurs at the right edge of the prosodic and syntactic phrase, leading to a mismatch between phrasal stress and focus:

(12) Chitumbuka (Downing field notes)

(a)

Q- Did the child carry the basket for an old *man* or an old *woman*?

(Mw-â:ná) (wa-ka-yeyera   chi-tê:te) (dada   mu-chekû:rû:)
1-child   1-TAM-carry for   7-basket   1.man   1-old

(panyákhe   mw-anakazi   mu-chekû:ru)?
or   1-woman   1-old

A- The child carried the basket for an old *man*.

(Mw-â:na) (wa-ka-mu-yeyera   chi-tê:té) (dada   mu-chekû:ru).
1-child   1-TAM-1.OM-carry for 7-basket   1.man   1-old

(b)

Q- Is he building the new houses *in* the village or *outside* the village?

(Kâ:si, wa-ku-zenga   nyumba   zî:-pyá) (mu-kati   mwa-mû:zî:)
Q   1-TAM-build   10.house   10-new   Loc-in   Loc-village

(pa-nyákhe   ku-walo   kwa-mû:zi)?
or   Loc-outside Loc-village

A- He is building *some* new houses *in* the village (and) *some outside*.

(Wa-ku-zenga   nyumba   zi-nyákhe   mu-kati   mwa-mû:zí)
1-TAM-build   10.house   10-some   Loc-in   Loc-village

(zi-nyákhe   kuwâ:lo).
10-some   Loc-outside

Similar phrasing and prominence assignment for similar data are found in Chichewa and Durban Zulu (Downing field notes), as shown by the data in (13):[5]

---

[5] In (13c,d) we see that a left-dislocated NP can be parsed into a separate prosodic phrase from what follows in Zulu. This is surprising, given the generalization from section 2.1.1 that prosodic phrases align with the right edge of a clause. See Cheng & Downing (2007b) for an analysis of this phrasing pattern.

(13)
Chichewa

(a) Q   Did the child carry the basket for the old *man* or the old *woman*?
     (Mw-aáná)   (a-ná-nyámulira   dengu   bambo   wókálaamba)
     1-child   1-TAM-carry for   5.basket   1.man   1.old
     (kapená   máí   wókálaamba)?
     or   1.woman   1.old

(b) A   She carried the basket for the old *man*, not the old *woman*.
     (A-ná-nyámulira   dengu   bambo   wókálaamba)
     1-TAM-carry for   5.basket   1.man   1.old
     (ósatí   máí   wókálaamba).
      not   1.woman   1.old

Zulu

(c) Q   Did the chief build houses *inside* our village or *outside*?
     (ín-du:na)   (izí:-ndlu)   (í-z-akh-é: é-sí-gódi-ni   se:thu)
     9-chief   10-house   9-10.OM.build-TAM   Loc-7-village-Loc   7.our
     (nomá   nga-phá:ndle)?
     or   Loc-outside

(d) A   The chief built houses *inside* our village, not *outside*.
     (ín-dú:na)   (izí:-ndlu)   (í-z-akh-é: é-sí-gódi-ni   se:thu),
     9-chief   10-house   9-10.OM.build-TAM   Loc-7-village-Loc   7.our
     (hháyí   nga-phá:ndle).
     not   Loc-outside

This pattern of focus-prominence mismatch within NPs and/or Prepositional Phrases has parallels in Swahili (Geitlinger & Waldburger 1999), as well as in non-Bantu languages like Italian (Ladd 1996, Swerts et al. 2002) and Egyptian Arabic (Hellmuth 2005). In all these languages, the generalization is the same: prosodic phrasing respects XP constituent edges, and phrasal prominence remains fixed on the phrase-final word. Prominence does not highlight a pre-final focused element within the phrase.

3.3 No sentence level stress on arguments of focus particles

In English, sentence-level stress marks all types of focus, including focus on the italicized argument of association with focus particles like 'also' in (14c):

(14)

(a) Where are you going to eat dinner on Friday?
      We are going to *an Italian restaurant* for dinner on Friday.

(b) We are going to an *Italian* restaurant, not a *Thai* restaurant.

(c) We are <u>also</u> going to an Italian restaurant on *Saturday* night.

However, the analogous focus particle in Chitumbuka – the association-with-focus verbal enclitic, *-so* 'also; again'– does not lend culminative prosodic prominence to its focused argument. As shown in (15) - (18), below, it attaches only to verbs and is itself followed by a prosodic phrase boundary (indicated with parentheses). As a result, both the verb host and the word the enclitic puts in focus realize the same prominence-related prosody – penult lengthening and contour tone – which correlates with prosodic phrase boundaries. This leads to potential ambiguity about what is in focus. For example, in (17b), the subject, the verb, the verb phrase or the object could be interpreted as the argument of *-so* without the context in (17a) to disambiguate:

(15)

(a) (n-khu-limilíra   ma-pû:no).
      I-TAM-weed   6- tomatoes
      'I am weeding tomatoes.'

(b) (Ku-limilirâ:-*so*)   (ngô:mâ:)?
      You/TAM-weed-also   9. maize
      'Are you also weeding the maize?'

(16) The friend who killed the snake also brought their father to the hospital.
(Mu-nya[β]o uyo   wa-ka-yi-koma n-jô:ka)   (ndiyo   wa-k-izâ:-*so*)
 1-friend   1.REL   1-TAM-9.OM-kill 9-snake   is.who   1-TAM-bring-also
(na   [β]a-dada   [β]-â:[β]o)   (ku-chi-patâ:la).
with   2P-father   2P-their   Loc-7-hospital

(17)

(a) Q- Is it only the doctor who helps the teacher?
      (Ni   [β]a-dokotala pê:ra) (a[β]o   [β]a-ku-vwíra [β]a-sambî:zî:)?
      COP   2P-doctor   only   2P.REL   2P-TAM-help   2P-teacher

(b) A- No, the chief also helps the teacher.
      (Yâ:yí), ([β]a-fúmu   [β]a-ku-vwirâ:-*so*)   ([β]a-sambî:zi).
      no   2P-chief   2P-TAM-help-also   2P-teacher

(18)

(a) Q-   Are you going to Lilongwe today?

   (Kâ:si),   (mu-ku-luta   ku-Lilô:ngwe)   (mw-ahû:nô:)?
   Q   you-TAM-go   Loc-Lilongwe   today

(b) A-   Yes, and I am also going to Salima.

   (Ê:nya),   (n-khu-lutâ:-*so*)   (ku-Salî:ma).
   yes   I-TAM-go-also   Loc-Salima

Work by Rooth (1992) on focus-related morphemes has argued that focus particles like -*so* should be morphologically and phonologically uninteresting. The focused argument of these morphemes should be made prominent either phonologically, by having the same focus prosody as other focus constructions, like Q/A pairs and in situ contrastive focus, or morphologically, by adjacency of the focusing morpheme and its argument. However, in Chitumbuka, the focused argument of the focus particle is not always made prominent by either phonology or morphology. Data like (17b) shows that -*so* is not always morpho-syntactically adjacent to its argument. And while the focus particle triggers prosodic rephrasing, its focused argument is not unambiguously highlighted by any culminative sentence-level prosody.

4   Intonation and information structure

Intonation in the sense of pitch manipulation does not play a consistent and striking role in signaling focus or other aspects of information structure in these languages. However, it does play some role, and this will be briefly sketched in this section.

4.1 Question intonation

In Zulu, there is no special prosody associated with Wh-questions. In both Chichewa and Chitumbuka, the overall pitch of yes-no questions is higher than statements, and there is an Intonational rise-fall (or fall-fall) melody over the final two syllables of the question, as shown by the data below:[6]

---

[6]   See Myers (1996, 1999) for a detailed phonetic study of question intonation in Chichewa, and see Downing et al. (2004, 2007) for more detailed discussion of focus raising in one variety of Chichewa.

(19) Chitumbuka (Downing field notes)
(a)   Did the goats jump over the wall?

Kási,	mbû:zi	zi-ka-duka	pa-chi-phû:phâ:?
Q	10.goats	10-TAM-jump	Loc-7- wall

(b)   Did the monkey make the child cry?

Kási,	mbwê:ngu	[β]a-ka-lilisya	mw-â:nâ:?
Q,	1.monkey	1-TAM-make cry	1-child

(20) Chichewa (Downing field notes)
(a)   Did the dog make the child laugh?

Kódí,	gaálu	a-ná-seketsa	mw-àánâ:?
Q	1.dog	1-TAM-make laugh	1-child?

(b)   Are the boys feeding the pigs?

Kódí,	a-nyamáàta	a-ku-dyétsa	nkhùúmbâ:?
Q	2-boys	2-TAM-feed	10.pigs

These intonation patterns are found elsewhere in Bantu languages. Ashton (1947) shows that Swahili yes-no questions are marked by a rise-fall melody over the last two syllables. An overall raised pitch for yes/no questions is found in Northern Sotho (Zerbian 2004, 2006) and Jita (Downing 1996). Indeed, as Yip (2002) shows, it is fairly common for tone languages to use boundary tones or an overall raising of pitch register to mark questions. Lexical tone does not preclude the intonational use of pitch. It is rather surprising, then, that these languages do not indicate focus by use of sentential pitch prominence, as intonational melody (and register) are manipulated to distinguish statements from questions.

4.2   Focus register raising in Chichewa

Although Chichewa does not use sentential stress or accent to indicate focus, Downing et al. (2004, 2007) have shown that in some varieties of Chichewa, phrasal register raising accompanies focus. (Myers (1996: fn 8) also mentions the occurrence of focus register raising in Chichewa, but unfortunately provides no phonetic details.) Downing et al.'s main finding – briefly summarized in table 1 – is that focus leads to systematic raising of f_0 within the Phonological Phrase containing the focused element. The table 1 provides the mean maximal pitch values for the Phonological Words in (6b, c, d, e), above, repeated below the table for convenience:

Table 1: Bolding highlights the significantly raised f_0-values (in Hz; and their sd in parentheses) of the Prosodic Words under focus (underlined) as revealed by Scheffé post hoc tests for an ANOVA over pitch maxima split by Phonological Word (**: $p < .01$; *: $p < .05$). Bold cell borders indicate Phonological Phrasing. The duration in ms of pauses is given in column 'p:'; the 'N' column indicates the number of repetitions out of 5 containing a pause at that position.

	a-ná-mé(e)ny-a	nyu(ú)mbá	p: [ms]	N	ndí-mwáálá
(6b)	147.6 (3.96)	113.7 (2.96)	-	-	110.9 (3.18)
(6c)	144.0 (7.29)	115.2 (5.07)	193.2 (32,23)	2	**120.0 (6.02)***
(6d)	154.1 (8.71)	**134.4 (15.22)****	252.7 (52.43)	5	109.6 (3.63)
(6e)	**179.0 (11.9)****	109.9 (4.21)	-	-	101.2 (1.37)

(6)   S/he hit the house with a rock.
(a)   /A-ná-meny-a   nyumbá   ndí   mwalá/
       S/he-SIMPLE PAST-hit-FV   house   with   rock
(b)   (A-ná-mény-a nyumbá ndí mwáálá).
(c)   (A-ná-mény-a nyuúmbá) (ndí mwáálá) ↑.
(d)   (A-ná-mény-a nyuúmbá) ↑ || (ndí mwáálá).
(e)   (A-ná-méeny-a) ↑ (nyuúmbá) (ndí mwáálá).

The mean maximal pitch values in table 1 show that the pitch of High tone sequences in the Phonological Phrase containing the narrowly focused element (underlined) is significantly higher (bolded) than when the same constituent is not focused. Note that downstep is not reset by focus raising. High tones undergo declination across the sentence in all the data. We also find a pattern of declination between High-toned sequences within the Phonological Phrase (set off by bolded vertical lines). As a result, the focused element does not have the highest pitch in its phrase (or in the sentence), unless it is also sentence or phrase initial. In short, focus raising does not lead to culminative pitch prominence for focused elements at either the phrase or the sentence level.

5   Conclusion

To conclude, let us return to the questions that we started off with. First, do Chichewa, Durban Zulu and Chitumbuka have sentence-level prominence? As we have seen, yes, they do all have culminative sentential prominence, realized as significant lengthening of the sentence-penult syllable. One contribution of the paper, then, is to point out that pitch is not the only form of prosody that can serve intonation's role of marking sentence and phrase edges. Secondly, does sentence prominence correlate with focus? No, as we have seen, in these languages sentence prominence is fixed at the end of the sentence, but focus can

occur in an earlier prosodic phrase. Focus register raising, found in some varieties of Chichewa, also does not give culminative pitch prominence to the focused element. Finally, does phrasal prominence correlate with focus? No, as we have seen, phrasal prominence is fixed on the phrase-penult syllable. The focused word need not be in a position in the phrase where it can receive phrasal prominence. Further, focus particles in Chitumbuka highlight their host, not necessarily their arguments.

In short, these are languages where re-phrasing provides a prosodic correlate of focus but sentence-level prominence does not. This study adds to the body of work showing that re-phrasing is an important cross-linguistic correlate of focus (Hyman 1999, Ladd 1996, Jun 1996, Hayes & Lahiri 1991). Sentence level prominence is conditioned only by syntax, and plays the important demarcative function of identifying sentence edges. Another contribution of the paper is, then, to emphasize that the demarcative and focus functions of sentence- and phrase-level prosody are separable. The same prosody is not recruited to fill both functions in all languages the way it is, for example, in English.

6   References

Ashton, E.O. 1947. *Swahili Grammar (including Intonation)*. 2d edition. London: Longmans.

Beyssade, C., E. Delais-Roussarie, J. Doetjes, J.-M. Marandin & A. Rialland. 2004. Prosody and information in French. In F. Corblin & H. de Swart (eds.). *Handbook of French Semantics.* Stanford: CSLI, 477-499.

Bresnan, J. & J. Kanerva. 1989. Locative inversion in Chichewa: a case study of factorization in grammar. *Linguistic Inquiry* 20, 1-50.

Bresnan, J. & S. Mchombo. 1987. Topic ,pronoun and agreement in Chichewa. Language 63, 741-782.

Bruce, G. 1977. Swedish Word Accents in Sentence Perspective. *Travaux de l'Institut de Linguistique de Lund* XII.

Cheng, L. & L. Downing. 2007a. The prosody and syntax of Zulu relative clauses. *Bantu in Bloomsbury: Special Issue on Bantu Linguistics. SOAS WPL* 15, 51-63.

Cheng, L. & L. Downing. 2007b. Where's the topic in Zulu? ms. University of Leiden & ZAS. [under review]

Cruttenden, Alan. 1997. *Intonation*. 2nd edition. Cambridge: Cambridge University Press.

Doke, C. M. 1954. *The Southern Bantu Languages*. London: Oxford University Press.

Downing, L. J. 1996. *The Tonal Phonology of Jita*. Munich: LINCOM Europa.

Downing, L.J. 2006. The prosody and syntax of focus in Chitumbuka. In L.J. Downing, L. Marten & S. Zerbian (eds.), *Papers in Bantu Grammar and Description*. *ZASPiL* 43, 55-79.

Downing, L. J. to appear. Accent in African languages. In R.W.N. Goedemans, H.G. van der Hulst & E.A. van Zanten (eds.), *Word Prosodic Systems in the Languages of the World*. Berlin: Mouton.

Downing, L. J., A. Mtenje & B. Pompino-Marschall. 2004. Prosody and information structure in Chichewa. *ZASPiL* 37, 167-186.

Downing, L. J., A. Mtenje & B. Pompino-Marschall. 2007. The focus prosody of Chichewa and the Stress-Focus constraint: a response to Samek-Lodovici (2005). Under review.

Féry, C. 2001. Focus and phrasing in French. In C. Féry & W. Sternefeld (eds.), *Audiatur Vox Sapientiae: A Festschrift for A. v. Stechow*. Berlin: Akademie Verlag, 153-181.

Geitlinger, K. & D. Waldburger. 1999. Intonation in Swahili. *Brücken und Grenzen: Werkschau Afrikastudien* 2. Münster: LIT Verlag, 419-434.

Grønnum, N. 1998. Intonation in Danish. In Daniel Hirst & Albert Di Cristo (eds.), *Intonation Systems: A Survey of Twenty Languages*. Cambridge University Press, 131-151.

Gussenhoven, C. 1984. *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.

Gussenhoven, C. 1996. Sentence accents and argument structure. In I.M. Roca (ed.), *Thematic Structure: Its Role in Grammar*. Berlin: Foris, 79-106.

Gussenhoven, C. 2004. *The Phonology of Tone and Intonation*. Cambridge University Press.

Hayes, B. & A. Lahiri. 1991. Bengali Intonational Phonology. *NLLT* 9: 47-96.

Hellmuth, S. 2005. No de-accenting in (or of) phrases: Evidence from Arabic for cross-linguistic and cross-dialectal prosodic variation. In S. Frota, M. Vigário & M. J. Freitas (eds.), *Prosodies*. Berlin: Mouton de Gruyter, 99-121.

Hyman, L. M. 1977. On the nature of linguistic stress. In Larry Hyman (ed.), *Studies in Stress and Accent*. *SCOPIL* 4. Los Angeles: University of Southern California, 37-82.

Hyman, L. M. 1999. The interaction between focus and tone in Bantu. In G. Rebuschi & L. Tuller (eds.), *The Grammar of Focus*. Amsterdam: John Benjamins, 151-177.

Jokweni, M. W. 1995. *Aspects of IsiXhosa Phrasal Phonology*. Ph.D. dissertation, University of Illinois at Urbana-Champaign.

Jun, S.-A. 1996. *The Phonetics and Phonology of Korean Prosody*. New York: Garland.

Jun, S.-A. 2005. Prosodic Typology. In Sun-Ah Jun (ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 430-458.

Kanerva, J. 1990. *Focus and Phrasing in Chichewa Phonology*. New York: Garland.

Ladd, D. R. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.

Mchombo, S. 2004. *The Syntax of Chichewa*. Cambridge: Cambridge University Press.

Myers, S. 1996. Boundary tones and the phonetic implementation of tone in Chichewa. *SAL* 25, 29-60.

Myers, S. 1999. Downdrift and pitch range in Chichewa intonation. *Proceedings of ICPhS99*, 1981-1984.

Odden, D. 1999. Typological issues in tone and stress in Bantu. In S. Kaji (ed.), *Cross-linguistic Studies of Tonal Phenomena: Tonogenesis, Typology, and Related Topics*. Tokyo: ILCAA, 187-215.

Reinhart, T. 1995. Interface strategies. *OTS Working Papers in Theoretical Linguistics* 02-001, Utrecht: OTS, Utrecht University.

Roberts, C. 1998. Focus, the flow of information, and universal grammar. In *The Limits of Syntax, Syntax and Semantics* 29, 109-160. Academic Press.

Rooth, M. 1992. A theory of focus interpretation. *Natural Language Semantics*: 75-116.

Rooth, M. 1996. Focus. In S. Lappin (ed.), *Handbook of Contemporary Semantic Theory*. Oxford: Blackwell, 271-297.

Samek-Lodovici, V. 2005. Prosody-syntax interaction in the expression of focus. *NLLT* 23: 687-755.

Selkirk, E. O. 1984. *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge, Mass.: MIT Press.

Selkirk, E. O. 1995. Sentence prosody: intonation, stress and phrasing. In J. A. Goldsmith, (ed.), *Handbook of Phonological Theory*. Cambridge, Mass.: Blackwell, 550-569.

Selkirk, E. O. 2004. Bengali intonation revisited. In C. Lee, M. Gordon & D. Büring (eds.), *Topic and Focus: A Cross-Linguistic Perspective*. Dordrecht: Kluwer, 217-246.

Swerts, M., E. Krahmer & C. Avesani. 2002. Prosodic marking of information status in Dutch and Italian: a comparative analysis. *Journal of Phonetics* 30, 629-654.

Szendrői, K. 2003. A stress-based approach to the syntax of Hungarian focus. *The Linguistic Review* 20: 37-78.

Truckenbrodt, H. 1995. *Phonological Phrases: Their Relation to Syntax, Focus, and Prominence*. Ph.D. dissertation, MIT.

Yip, M. 2002. *Tone*. Cambridge: Cambridge University Press.

Zerbian, S. 2004. Intonation in Northern Sotho. Poster presented at the TIE Conference, Santorini, 9-11 September 2004.

Zerbian, S. 2006. *Expression of Information Structure in the Bantu Language Northern Sotho*. Ph.D. dissertation, Humboldt-University Berlin. [published as *ZASPiL* 45].

Prosodic correlates of subclausal quotation marks

Elke Kasimir
Zentrum für Allgemeine Sprachwissenschaft, Berlin

We present the results of an experimental study which targets prosodic correlates of subclausal quotation marks. We found that written sentences containing passages enclosed by quotation marks were read aloud in a manner that significantly differs in prosody from spoken realizations of corresponding disquoted counterparts. However, we also observed that such prosodic marking of subclausal quotation wasn't strong enough to survive subsequent back-translation into written language: there was no correlation between the presence/absence of quotation marks in the original written examples, and the presence/absence of quotation marks in corresponding back-translations from oral renditions. We investigated three different kinds of uses of quotation marks and found no systematic difference between them with respect to prosodic marking.

1   Introduction

This paper presents an experimental study undertaken at the University of Potsdam in winter 2005/2006 which targets prosodic correlates of quotation marks. Given that written quotation marks in languages like English or German usually have no morphological correlate in spoken language[1], the question has often been raised whether written quotation marks have a counterpart in spoken language in the form of special prosodic marking. The experiments were sought to investigate the extent to which this question can be given a positive answer: to which extent are written quotation marks translated into the spoken idiom by using a recognizably marked prosody, and to which extent does such a marking - if it in fact exists - help to communicate the interpretive differences conveyed by quotation marks? We also asked whether distinctive *uses* of quotation mark behave differently with respect to prosodic marking.

---

[1]   The English form "quote-unquote" and its German counterpart "in Anführungsstrichen" or "in Gänsefüßchen" do not generally replace written quotation marks (see also footnote 7).

The results can be summarized as follows: the presence or absence of subclausal quotation marks in a sentence significantly influences its prosodic realization. However, such prosodic marking fails to be a real spoken substitute for written quotation marks. We conclude this from the fact that quotation marks did not survive translations into the spoken idiom followed by back-translations into the written idiom. Furthermore, the prosodic means were apparently rather diverse and not subject to any specific phonological strategy. Finally, the three different usage patterns of subclausal quotation marks acknowledged in the experimental design (partial reported speech, names or titles, modal uses) were treated alike when being translated into the spoken idiom.

After a review of previous studies on the prosody of quotation and reported speech, the rest of this paper will give an overview of the experimental design, describe the material and methods being used and finally present and discuss the results. A thorough discussion of possible theoretical consequences of the findings reported here will be provided in a separate paper.

2 Previous works on the prosody of quotation

This study is to our knowledge the first investigation ever of the prosody of *subclausal* quotation; it is also the first study which explicitly targets not only reported speech but other uses of quotation marks as well. The prosody of reported speech in spontaneous German and English speech data has been investigated in Brünner 1991, Günthner 1997, 1998, Couper-Kuhlen 1998, and Klewitz & Couper-Kuhlen 1999. There is agreement that both direct and indirect reported speech may receive a special prosodic marking. According to Klewitz & Couper-Kuhlen 1999, prosodic means include shifts in global pitch (register), loudness, speech rate or isochronous timing and sometimes also drastic changes in voice quality, for instance whispering. Brünner 1991 observes that the prosodic marking of reported speech is used to characterize the speaker at hand of stereotypes, often in derogatory or mocking ways (see also Buttny 1997). This suggests that at least certain more articulate forms of prosodic marking are neither a grammatical marker of reported speech, nor, as has often been claimed in the linguistic and philosophical literature, a reproduction of phonological properties of some antecedent utterance, but rather an independent comment on the purported speaker of the utterance. According to Klewitz & Couper-Kuhlen 1999, this kind of mocking or derogatory comment not only occurs with direct reported speech, but also with utterances just describing the quotee, and spoken counterparts of *free indirect style* (*style indirect libre, erlebte Rede*) as known from literary language. In the case of direct quotation, prosodic marking doesn't exactly line up with quotation boundaries but is somewhat flexible with regards to onset and offset.

A phonetically oriented study of reported speech based on elicited spontaneous speech data is presented in Bertrand & Espesser 2002. The authors found significant differences in pitch range between direct self-reports on the one hand, direct reports of quotees other than the source speaker on the other hand. They didn't find significant differences between the latter type of direct reported speech and ordinary discourse. Phonetically oriented studies based on corpus data are reported in Grosz & Hirschberg 1992, Hirschberg & Grosz 1992 and Jansen et. al. 2001. Grosz and Hirschberg used spoken material from a professional radio news speaker, together with the original written texts. They investigated how annotators assigned quotational readings when provided either just the sound data (first group), or both sound & text (second group). They found that both groups correlate direct reported speech with greater pitch range and a sentence-final pronounced drop in intensity. Notice that such prosodic marking was produced by a trained speaker and is yet merely a statistical tendency. Jansen et. al. 2001 found that direct reported speech is positively correlated with higher pitch range and the likelihood of the alignment of the report sentence with an intonation phrase boundary.

3 Overview of the experimental procedure

The initial step of our study consisted of a production task where participants translated sentences containing subclausal quotation marks, as well as corresponding sentences lacking the quotation marks, into spoken language. This production task was followed by two classification tasks where participants were asked to match the acoustic tokens from the production task with the corresponding written tokens. In a final reproduction task participants were asked to back-translate items that had been particularly "successful" in the classification tasks back into the written idiom. Figure 1 gives an overview of the overall experimental procedure.

In the classification tasks, things were made as easy as possible for the participants: the latter had just to decide whether a certain passage being embedded in a whole sentence was surrounded by quotation marks, or not. The primary goal of this classification task was to measure the degree to which the presence or absence of quotation marks was encoded in the spoken signal. The participants therefore had a role similar to that of annotators in corpus-based studies. But since this procedure provided us with a measurement of the degree to which particular acoustic tokens were correctly classified, it also helped us to identify particularly "successful" token pairs, that is, tokens pairs where the speakers had encoded the presence or absence of quotation marks in a way that allowed most or all of the listeners to recover this information from the acoustic signal.

```
┌─────────────────────────────────────────────────────┐
│  original written item                              │
│           ↓ production task                         │
│   acoustic token                                    │
│           ↓ classification tasks                    │
│   signal quality  ⟶  "promising" acoustic token     │
│                          ↓ reproduction task        │
│                       written token                 │
└─────────────────────────────────────────────────────┘
```
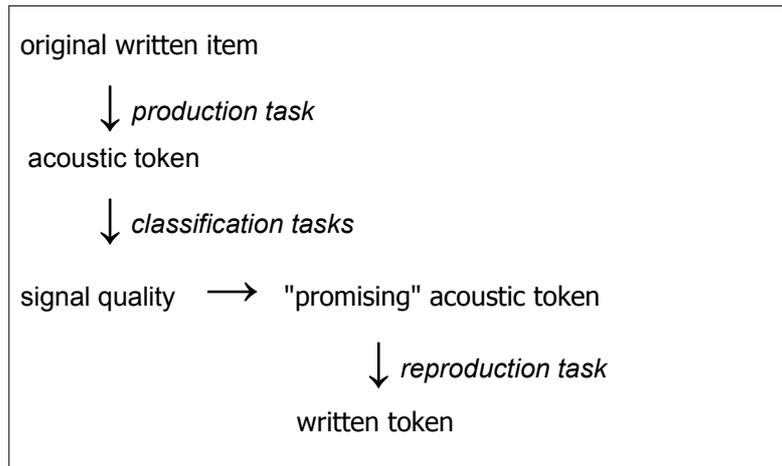
Figure 1: Experimental procedure

Some few of these "successful" items were used for the final reproduction task. This time, participants were asked to listen to whole texts containing either the quote or its disquoted counterpart, and typewrite them into a computer text file. By comparing the results to the original written texts, we could determine to what degree written quotation marks had survived a translation into spoken language and a subsequent back-translation into written language. This in turn seemed to us to be a good indicator of the degree to which the encoding of quotation in spoken language is actually communicatively effective.

4   Material

The material consisted of a collection of subclausal quotation items in Standard German of which the following is an example:

> Die Grüne Bundestagsabgeordnete Monika Knoche kritisiert, dass die Verordnung damit einem "faktischen Verbot" von Codein gleichkommt.

> *The Green deputy Monika Knoche criticises that therewith this bye-law amounts to a "factual verdict" on codeine.*

Each of these items were paired with a disquoted variant:

> Die Grüne Bundestagsabgeordnete Monika Knoche kritisiert, dass die Verordnung damit einem faktischen Verbot von Codein gleichkommt.

> *The Green deputy Monika Knoche criticises that therewith this bye-law amounts to a factual verdict on codeine.*

For every such sentence pair, a short embedding text was construed such that both variants have a plausible reading in that context.

We used three kinds of subclausal quotation: (I) subclausal direct reported speech (*partial quotes*, *mixed quotation*), (II) names or titles which were derived from descriptive terms by quotation marks, (III) modal uses of quotation marks, including scare quotes (*modalisierende Verwendungen* in the sense of Klockow 1980):[2]

> The byelaw amounts to a "factual verdict" on codeine.
>> *(= (I) reported speech)*

> Lord Anson's "Journeys round the world" made it into the headlines.[3]
>> *(= (II) names or titles)*

> The anticlimax was credited to his "political fanaticism".
>> *(= (III) other / modal uses)*

Six items of each kind were chosen, where each item came in two variants, the first containing, the second lacking the quotation marks in question, yielding a total of 36 texts. All subclausal quotations were basically of the category *noun phrase*. However, in order to facilitate prosodic marking and its subsequent recognition, we let the actual quotation exclude a determiner or adjective whenever this was possible. Furthermore, we took care that the right edge of the quotation didn't coincide with a clause boundary.[4] All items were somehow derived from naturally occurring examples, mostly newspaper articles and literary texts that were selected from the text corpus database of the Institut für Deutsche Sprache, Mannheim. The surrounding texts were generally taken from the original context but were always simplified both syntactically and with respect to content in order to facilitate loud reading. Some texts had to be invented in order to obtain a context which was compatible with both the quoted and the disquoted version of the expression we were interested in. We took care

---

[2]   We didn't include metalinguistic quotation (*pure quotation*) as in
    (i) *'Boston' is disyllabic.*
  because of the obvious difficulties involved in the construction of contexts compatible with both the quoted and the disquoted variant.

[3]   Notice that in German, names do not receive a special treatment with respect to uppercase-lowercase, so that the German version of the example is disambiguated by quotation marks alone.

[4]   In 14 of 18 items (78%), either a leading determiner or a leading adjective was excluded from the quoted expression. In 17 of 18 items (94%), the right edge of the quoted expression didn't coincide with a phrasal boundary.

that the embedding texts contained further quotation marks in order to conceal the specific differences between the two versions of the texts. The embedding short texts were 65 words on average. The presence/absence of quotation marks generally led to differences in the meaning of the texts, although these differences where in some cases rather subtle.

5   Methods

For the reading task we selected eight speakers of Standard German, four male and four female, who had experience in reading aloud, mostly due to the fact that they regularly read to little children. The participants read all 36 texts first silently and then aloud in four sessions per participant. One session took about 20 minutes. Readings of different versions of the same text were generally at least two weeks apart in order to conceal the difference between the versions. Texts were read again at most one time per session if there were speaking errors in sentences that contained the relevant quotation. Whenever the second reading led to another error, the text was repeated in some subsequent session. The order in which the texts were presented was randomized. The participants didn't become aware of the fact that the experiments had to do with quotation, as was verified by a questionnaire that was presented to the participants at the end of the last sessions.

In the first classification task, 4 participants were asked to rate all 143[5] acoustic sentence token pairs (a pair consisting of the quoted and the disquoted variant). The audio tracks were presented together with the written counterparts (both in random order) in an Internet browser window, and the participants were asked to determine which track belongs to which sentence version. The participants could listen to the tracks as often as they wanted and had no time limit. Each token pair was presented twice.

In the second classification task, 2 educated participants were asked to rate all 286 acoustic sentence tokens. The audio track was presented with the corresponding written versions (quoted and disquoted) in an Internet browser window and the participants were asked to determine the sentence version to which the audio track belongs. The participants could listen to the track as often as they wanted and had no time limit. Each token was presented once. One listener was an advanced student of phonology, the other one was a researcher in a phonetics scientific project. That way, the increased difficulty of the task, when compared to the first classification, was partially compensated by choosing trained listeners. The goal of the second experiment was to verify that

---

[5]   One recorded token pair had to be discarded because of a speech error in one of the tokens.

speakers do not just make a relative difference between quoted and disquoted realization, but instead render quotation marks in some absolute, non-relational sense.

As was mentioned earlier, the results of the classification task were also used to select successful items to be used in the subsequent reproduction task. In the above experiment, each token pair can be said to correspond to 12 classifications (two presentations of the pair to 4 participants in the first task, and one classification of one token of the pair to 2 participants in the second task). We randomly selected 6 token pairs which were rated correctly in at least 11 of these 12 classifications for use in the reproduction task.

For the reproduction task, 6 participants were provided with 6 text tokens, each being drawn from a different token pair, in the form of audio files, and were instructed to typewrite the text into a computer file by using the full set of punctuation marks as well as boldface and italics. The listeners were not aware of the goal of the experiment.

6   Results

In both classification tasks, there were more correct than incorrect classifications, indicating that the presence/absence of quotation marks was in fact (prosodically) encoded in the spoken signal. The results are highly significant[6] (Table 1):

Table 1: Results of the two classification tasks

classifications:	correct	total
1st task (four participants)	763 (67%)	1144
2nd task (two participants)	337 (59%)	572

Furthermore, it turned out that all three kinds of subclausal quotation roughly behaved identical in the classification task (Table 2).

---

[6]   The post-hoc probability that the high number of correct classifications is accidental is $p < 2*10^{-30}$, $p < 8*10^{-6}$ respectively, according to the one-sided Sign Test.

Table 2: Kinds of quotation in the 1st classification task

classifications:	correct	total
(I) reported speech	239 (63%)	383
(II) names or titles	267 (70%)	384
(III) other	257 (69%)	377

We finally observed a considerable variability in the performance of "teams" consisting of a speaker and a listener (Figure 2).
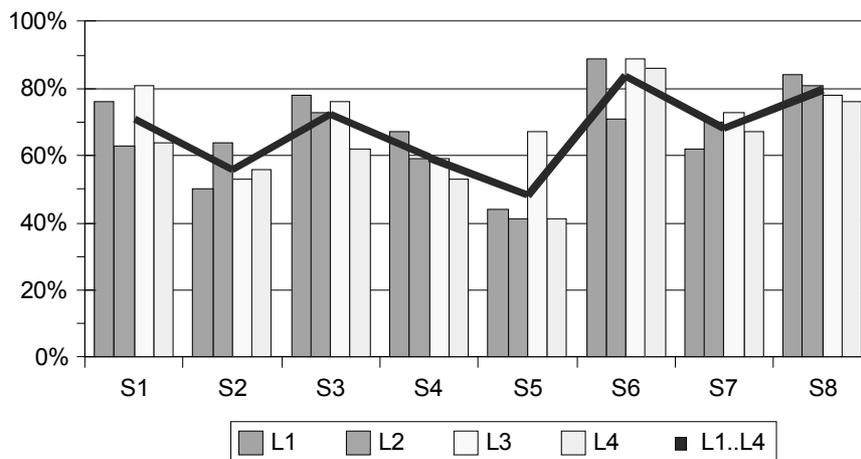
Figure 2: 1st classification task: speakers (S) and listeners (L)

In particular it appears as if the tokens produced by speaker 5 have been merely confusing to listeners 1, 2 and 4, whereas speaker 3 is moderately successful in grasping the quoted-disquoted distinction (see again Figure 2).

In the reproduction experiment we found that even though the tokens being used had been successfully classified in the foregoing classification tasks, not even a tendency towards a reproduction of the quoted-disquoted distinction could be observed (Table 3).

Table 3: Reproduction task

Versions	translations into quoted versions	expected
non-quoted	3	0
quoted	3	18

We were somewhat liberal in what we regarded as a correct reproduction by allowing both quotation marks and italics and by accepting cases where the marking appeared in approximately the same region which was quoted in the original items (see Table 4).

Table 4: List of reproductions that were classified as *quotational* reproductions in the reproduction task. Columns (left to right): the version underlying the acoustic token, the expression which was quoted in the quoted version of the original item, the unedited (!) reproduction as produced by the listener-writer.

Version	Quote	Reproduction
disquoted	Katharina Ehler	An einem Mittwoch verlässt die „Katharina Ela" *Marseille* und begibt sich auf die Reise richtung Lateinamerika.
disquoted	Katharina Ehler	An einem Mittwoch verlässt die *Katharina Ela* Marseille und begibt sich auf die Reise Richtung Lateinamerika.
quoted	Katharina Ehler	An einem Mittwoch verlässt die *Katharina-Ela* Marseille und begibt sich auf die Reise Richtung Lateinamerika.
quoted	Katharina Ehler	An einem Mittwoch verlässt die „Katharina Ehler" Marseille und begibt sich auf die Reise Richtung Latein Amerika.
disquoted	Faktischen Verbot	Die Grüne Bundestagsabgeordnete Monika Knochel kritisiert, dass die Verordnung damit „einem praktischen Verbot von Codein gleichkommt".
quoted	Reise um die Welt	Lord Ansons *Reise um die Welt* machte im damaligen London Schlagzeilen.

One might finally wonder which particular prosodic means were used in those pairs that were correctly classified in the classification tasks.[7] Although a phonetic analysis is still missing, our naive auditory impression suggests that at least four different strategies can be distinguished: an audible leading pause, an audible trailing pause, a change in the voice quality (i.e. lengthening of stressed

---

[7]   None of the speakers used any morphological, lexical or syntactic means as a translation of quotation marks.

vowels) and shifting of the location of pitch accents. The latter strategy can be illustrated at hand of the following item:

> Der Ford Scorpio ist ein PKW des "oberen Segments" (obere Mittelklasse oder Oberklasse - je nach Motorisierung und Ausstattung).

> *The Ford Scorpio is an automobile of the "upper segment" (upper middle class or upper class - depending of motor equipment and configuration).*

The main accent of the phrase *Der Ford Scorpio ist ein PKW des oberen Segments* (*The Ford Scorpio is an automobile of the upper segment*) was realized in the quoted version on the adjectival modifier *oberen* (*upper*) by one speaker, whereas the same speaker realized the accent in the default position, i.e. on the modified noun *Segment* (*segment*), in the disquoted version.

7   Discussion

Our results suggest that although speakers prosodically mark renderings of quotation marks, they do not do this in a fashion that effectively supports communication. The speakers also do not seem to rely on some unique strategy for the prosodic marking, and do not differentiate between different usages of quotation marks.

How is this result to be interpreted? One natural explanation is that speakers recognized the need to convey the information contributed by quotation marks when reading aloud the example items. However, spoken language doesn't supplied them with a suitable conventional device to do so; and as a consequence, speakers rely on paralinguistic and non-conventional or at least not fully conventionalized devices. These devices may have been derived from experience with stylized spoken language from mass media, or may have been spontaneously drawn from a set of rather unspecific means to highlight parts of a linguistic utterance or mark boundaries. Alternatively, the failure to communicate subclausal quotation marks could just be a performance error that is induced by the somewhat artificial situation of reading a text aloud instead of producing speech spontaneously. This seems to us however rather unlikely given that the experiment relied on previously selected "successful" encodings and given that other features of the texts were conveyed rather accurately.

The results are consistent with previous results for reported speech as summarized above where it was observed that the quoted-disquoted distinction is marked in a merely statistically significant way by various prosodic means. As expected, our speakers didn't use the more articulate prosodic means which Brünner (1991) and others observed in spontaneous spoken language (see

section 2 above); if these means are in fact used to characterizing the cited speaker at hand of prototypes, they would have been inappropriate for the newspaper and literary texts that were used in our experiment.

8   References

Bertrand, R. & R. Espesser (2002). Voice diversity in conversation: a case study. In: *Proceedings of the 1st International Conference on Speech Prosody. Aix-en-Provence, France,* 171-174.

Brünner, G. (1991). Redewiedergabe in Gesprächen. *Deutsche Sprache* 1, 1-16.

Buttny, R. (1997). Reported speech in talking race on campus. *Human Communication Research* 23, 477-506.

Couper-Kuhlen, E. (1998). Coherent voicing. On prosody in conversational reported speech. *Interaction and Linguistic Structures* 1, 1-28.

Grosz, B. & J. Hirschberg (1992). Some intonational characteristics of discourse structure. In *Proceedings of the International Conference on Spoken Language Processing, Banff*, 429-432.

Günthner, S. (1999). Polyphony and the "layering of voices" in reported dialogues: An analysis of the use of prosodic devices in everyday reported speech. *Journal of Pragmatics* 31, 685-708.

Günthner, S. (1997). Direkte und indirekte Rede in Alltagsgesprächen - zur Interaktion von Syntax und Prosodie in der Redewiedergabe. In: P. Schlobinski (ed.). *Zur Syntax des gesprochenen Deutsch*. Westdeutscher Verlag, 227-262.

Hirschberg, J. & B. Grosz (1992). Intonational features of local and global discourse structure. In: *Speech and Natural Language: Proceedings of a workshop held at Hariman, New York, February 23-26*.

Jansen, W., J. M. Brenier et. al. (2001). Prosodic correlates of directly reported speech: Evidence from conversational speech. Unpublished paper.

Klewitz, G. & E. Couper-Kuhlen (1999). 'Quote - Unquote? The role of prosody in the contextualization of reported speech sequences. *Interaction and Linguistic Structure*, 1-34.

Klockow, R. (1980). *Linguistik der Gänsefüßchen. Untersuchungen zum Gebrauch der Anführungszeichen im gegenwärtigen Deutsch*. Frankfurt a.M.

Blending in heterosyllabic consonant clusters in three Catalan dialects

Daniel Recasens

Universitat Autònoma de Barcelona and Institut d'Estudis Catalans, Barcelona

This study is an electropalatographic investigation of clusters composed of /n/ or /l/ followed by the (alveolo)palatal consonants /ʎ, ɲ/ or by dental /t/ in three Catalan dialects, i.e., Majorcan, Valencian and Eastern. Data show that articulatory blending through superposition occurs in the palatalizing environment except when C1 is highly constrained (e.g., dark /l/) or C2 is purely palatal and therefore, produced at a distant articulatory location from C1. Contrary to previous descriptions in the literature, data for /nt, lt/ reveal that blending through superposition rather than assimilation is at work. The implications of these data for theories of speech production are discussed.

1   Introduction

This study seeks to explore whether the DAC model of coarticulation may account satisfactorily for blending processes in clusters composed of front lingual consonants produced with the tongue tip, blade and predorsum, i.e., dentals or dentoalveolars, alveolars and alveolopalatals. It presents experimental data on C-to-C adaptation mechanisms in place of articulation for thirteen speakers of three Catalan dialects, i.e., Majorcan, Valencian and Eastern, and has a much larger scope than previous studies dealing with data for fewer speakers of the Eastern dialect only (Recasens and Pallarès, 2001, Recasens, 2006).

According to the DAC model (Recasens, Pallarès and Fontdevila, 1997, Recasens and Pallarès, 2001), the articulatory realization of consonant clusters is conditioned by the degree of articulatory constraint for the consonants in the cluster, i.e., by the requirements that speakers impose on the articulatory structures for the formation of a consonant closure or constriction. The alveolar trill /r/, the alveolar and alveolopalatal fricatives /s/ and /ʃ/ and dark /l/ appear to be more constrained than oral and nasal stops, i.e., the dental /t/, the alveolar /n/

and the alveolopalatals /ɲ/ and /ʎ/. Indeed, in comparison to the latter consonants, the former ones are more resistant to change and exert more prominent effects on other consonants in line with the precise aerodynamic and articulatory mechanisms involved (Solé, 2002). The alveolar lateral /l/ is not equally dark in all Catalan dialects: it is strongly dark in Majorcan, moderately dark in Eastern and clear in Valencian. The DAC model postulates that the degree of articulatory constraint for /l/ should increase with darkness degree in view of the strict demands involved in performing a postdorsal approximation in addition to an apical closure for the dark but not for the clear variety.

Differences in degree of constraint among Catalan consonants keep a certain relationship with place of articulation. Thus, in Eastern Catalan, the two highly constrained lingual fricatives /s, ʃ/ and the trill /r/ are invariably centroalveolar or postalveolar, while the less constrained consonants /t, n/ are more anterior, i.e., /t/ is fixedly dental and /n/ may be articulated at the front alveolar zone. Closure for the alveolopalatals /ʎ/ and /ɲ/ may occur simultaneously at the back alveolar zone and at the prepalate, but also just at the alveolar or dentoalveolar zone in the case of /ʎ/ (Recasens and Espinosa, 2006). The two varieties of /l/ are anterior, more so in Majorcan (dental or dentoalveolar) than in Eastern and Valencian (front alveolar).

A basic prediction of the DAC model regarding C-to-C adaptation mechanisms in consonant clusters is that highly constrained consonants should prevail upon less constrained ones through an assimilatory action or through robust coarticulatory effects (see also Fowler and Saltzman, 1993). On the other hand, two relatively unconstrained consonants ought to undergo blending and thus, result into a compromise articulatory outcome between the two original consonants in the cluster.

The goal of this study is to verify the validity of the prediction that consecutive consonants which share a similar degree of articulatory constraint and are not highly contrained are realized through a blending mechanism. The articulatory implementation of gestural blending will be investigated in two different scenarios, i.e., in a palatalizing environment in the case of sequences composed of the alveolars /n, l/ followed by the (alveolo)palatals /ʎ, ɲ/, and in a dentalizing environment in the case of clusters with /n, l/ followed by dental /t/. Different articulatory outcomes are expected to occur in clusters with /l/ depending on whether the consonant is dark or clear (see section 1.1). Several sequences will be excluded from analysis, namely, consonant combinations with C1=/t/ since this consonant often assimilates to a following nasal or lateral in Catalan, and the sequence /nɲ/ which is realized [ɲ:] in all three dialects.

1.1 Palatalizing environment

Two possible blending mechanisms may apply to consonant clusters. A first mechanism occurs when the closure area for the outcoming articulation results from superimposing the closure areas of the two adjacent consonants taken independently. This blending mechanism has been found to operate in clusters with C1=/n/ and, less so, C1=/l/ and C2=/ʎ, ɲ/ in Eastern Catalan (Recasens, 2006), and will be referred to as "blending through superposition". In case that clusters composed of C1=/n, l/ and C2=/ʎ, ɲ/ exhibit blending through superposition, we expect C1 to undergo some contact increase at the back closure border since C1 is dental, dentoalvolar or alveolar while C2 is alveolopalatal. Moreover, the front closure border should reach a fairly anterior location, mostly due to C1 in clusters with C2=/ɲ/, and to C1 and/or C2 in clusters with C2=/ʎ/.

According to a second production strategy, the blending outcome is realized at an intermediate zone between the places of articulation of C1 and C2, and will be referred to as "blending through intermediation". This instance of blending has been reported to take place in the CV sequence /ki/ or in consonant clusters such as /tθ/ and /sj/ (Browman and Goldstein, 1989, 1992).

While all clusters just referred to undergo blending, its articulatory manifestation may differ depending on the case. More evidence needs to be gathered in order to find out what factors determine the implementation of the two blending types referred to. Two plausible factors are closure or constriction extent for the two adjacent segments and the spatial distance between their places of articulation. Thus, blending through superposition is expected to apply in consonant clusters if the two adjacent consonants are articulated nearby (as for /n/ and /ɲ/), and at least one of them involves a large degree of contact at closure or constriction location (as for alveolopalatals).

A related issue is the temporal manifestation of blending, i.e., whether the compromise articulation between C1 and C2 occurs already at C1 onset or else at a later point in time during C1. If blending through palatalization starts at C1 midpoint or at C1 offset, the front and back closure location at the onset of C1=/n, l/ should be more anterior than the closure location for the alveolopalatals C2=/ʎ, ɲ/. Therefore, we would be facing a palatalized alveolar rather than an alveolopalatal realization at this temporal period (e.g., [nʲ] in the case of the cluster /nʎ/).

The conditions preventing blending from taking place completely or partially will also be paid attention to. In principle, blending may fail to apply if the two adjacent consonants in the cluster involve distant articulatory targets, as when C2 is purely palatal rather than alveolopalatal (e.g., /ɲ/ for some Majorcan speakers). It could also be delayed or absent in clusters with C1=/l/ due to

requirements on laterality and on darkness, since in both cases the tongue predorsum is relatively lowered in order to allow the passage of airflow through lateral oral channels (for laterality) and to facilitate the achievement of a /w/-like percept (for darkness).

1.2   Dentalizing environment

The articulatory outcome of clusters composed of C1=/n, l/ and C2=/t/ is hard to ascertain. The traditional view is that place assimilation is at work here and therefore, alveolars should acquire the dental place for /t/ throughout the entire closure period (see, for example, Navarro Tomás, 1972 for Spanish). Within the DAC model framework, the fact that all those consonants are relatively unconstrained and produced at the same or at a close articulatory zone renders the implementation of blending possible as well.

2   Methodology

2.1   Recording procedure

Linguopalatal contact configurations and acoustic data were gathered synchronously with the Reading electropalatography (EPG) system every 10 ms using artificial palates equipped with 62 electrodes. Acoustic data were digitized at 10 kHz.

Recordings included the clusters /nt, nʎ, lt, lʎ, lɲ/ occurring across a word boundary embedded in meaningful sentences four to six syllables long, e.g., /nt/ in the sentence "*d'anys en té set*" ("he/she is seven years old"), /nʎ/ in the sentence "*això no enllaça*" ("this does not work out well"). These clusters were recorded seven times by the five male Majorcan speakers AR, BM, MJ, ND and CA, seven times by the five male Valencian speakers JM, VB, MS, VG and AV, and a variable number of times by the three male Eastern speakers DR (five times) and JP and JS (three times). The cluster /lʎ/ was not recorded by the Valencian subjects.

Data were also processed for single /t, n, l, ʎ, ɲ/ in (quasi-)symmetrical VCV sequences in the context of front, low and back rounded vowels in order to determine the extent to which closure location in clusters differs from closure location in intervocalic position. VCV recordings were only carried out for /t, n, l/ in Majorcan and Valencian, and for /t/ in Eastern, e.g., /iti/ in "*aprèn l'hitita*" ("learn the Hittite language"), /əta/ in "*beu-te la tassa*" ("drink what is in the cup"), /otu/ in "*el llapis rotula*" ("the pencil works"). Data for intervocalic /n, l/ in Eastern and for intervocalic /ʎ, ɲ/ in the three dialects were taken from mean

contact patterns across the three contextual vowels published elsewhere (see Recasens and Pallarès, 2001).

The EPG data could not be processed for /lʎ/ for any of the Valencian speakers and for the Eastern speaker JP because /l/ assimilated to following /ʎ/ in this case, and for /lt, lʎ, lɲ/ for speaker AR since he vocalized preconsonantal /l/ almost systematically.

2.2 Measurement criteria

All cluster tokens were submitted to articulatory analysis. Closure or constriction location was measured on linguopalatal contact patterns placing a cursor on simultaneous EPG, spectrographic and waveform displays using the MultiSpeech 3700 program of Kay Pentax.

The linguopalatal contact patterns under analysis show 62 electrodes distributed into eight horizontal rows and four vertical columns at each side of the palate surface (see Figure 1). Filled electrodes have been contacted by the tongue while empty ones have not. Electrodes are grouped into two major articulatory zones, alveolar at rows 1 through 4, and palatal at rows 5 through 8. Several articulatory subzones were identified for better data interpretation, namely, dental (exclusively at row 1, just behind the upper teeth), dentoalveolar (at rows 1-2, 1-3 or 1-4), centroalveolar and postalveolar (at rows 2-3 and 3-4, respectively), alveolopalatal (at a continuous contact area encompassing the alveolar rows 1 through 4 and the palatal rows 5 through 8), prepalatal (at row 5), mediopalatal (at rows 6-7) and postpalatal (at row 8).
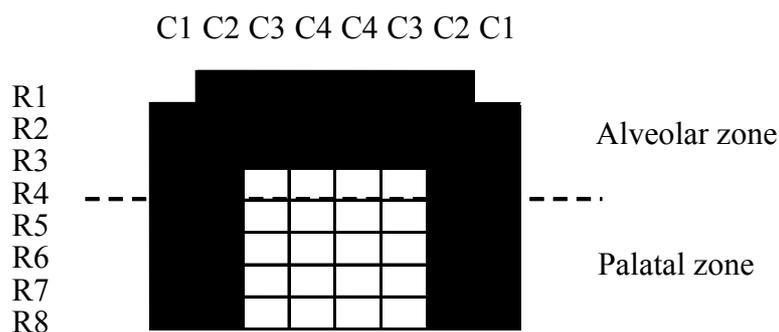
Figure 1. Linguopalatal contact configuration with 62 electrodes distributed into rows (R1-R8), symmetrical columns (C1-C4) and articulatory zones (alveolar, palatal).

For /t, n, ʎ, ɲ/, the consonant was taken to last from the onset of a complete closure until the release of any of the central electrodes. In the case of /l/, consonant onset and offset were determined by the presence or absence,

respectively, of at least one of the two centralmost alveolar electrodes. The application of this articulatory criterion was assisted by inspection of spectrographic displays.

Place of articulation was measured at several temporal points during C1 and C2 so as to explore the temporal extent of the C-to-C adaptation effects: at C1 onset, midpoint and offset, and at C2 onset and midpoint. Place of articulation for the intervocalic consonants was measured at consonant midpoint.

Two values corresponding to maximal contact activation at the front and back closure borders were taken at each point in time, e.g., 1-2 or 2-4 whenever full electrode activation occurred at rows 1 and 2 (in the former case) or at rows 2, 3 and 4 (in the latter case). For a given row, maximal activation was considered to apply whenever lingual contact occurred at all its electrodes, at all its electrodes except for one of the two central ones, or just at its two central electrodes. Both borders were assigned the same value if maximal activation turned out to be present at a single row. Closure fronting values for the intervocalic consonants were determined according to the same criterion and averaged across vowel contexts.

Front and back closure location values for consonant clusters were compared with those for the two consonants in intervocalic position in order to find out whether the final articulatory outcome was arrived at through blending through superposition or through blending through intermediation in the palatalizing environment, and through assimilation or through any of the two blending processes in the dentalizing environment. Initially, we carried out ANOVAs with repeated measures using the same conditions referred to below with one averaged score per condition for each speaker. Results were discouraging since significance was achieved in very few instances due perhaps to the small size of the data set and to the fact that the original data proceeded in one unit steps and therefore, were not actually continuous (see above). For this reason, we will provide results obtained from regular ANOVAs performed on data for all tokens and speakers. Statistical comparisons between the back closure location values for the clusters /nt, lt/ and for intervocalic /n, l, t/ were carried out with 'consonant condition' as the only factor and three levels of the independent variable, i.e., '/n/ in the cluster /nt/' or '/l/ in the cluster /lt/', 'intervocalic /n/' or 'intervocalic /l/', and 'intervocalic /t/'. Data for the intervocalic consonants were compared statistically with those for the clusters at C1 midpoint and at C1 offset but not with data at C1 onset since closure location at this temporal point was fairly anterior and did not reflect the implementation of a blending mechanism. Statistical comparisons for Eastern Catalan involved data for intervocalic /t/ but not data for intervocalic /n, l/ since there were not multiple tokens of the last two consonants in intervocalic position in this dialect.

Bonferroni post-hoc tests were run on the significant main effects whenever appropriate and the degree of significance was set at p < 0.05.

3 Results

3.1 Palatalizing environment

Figure 2 plots the front and back closure border trajectories averaged across tokens and speakers for clusters composed of C1=/n, l/ and C2= /ʎ, ɲ/ (thick lines), and for those same consonants in intervocalic position (discontinuous straight lines for C1, continuous straight lines for C2).
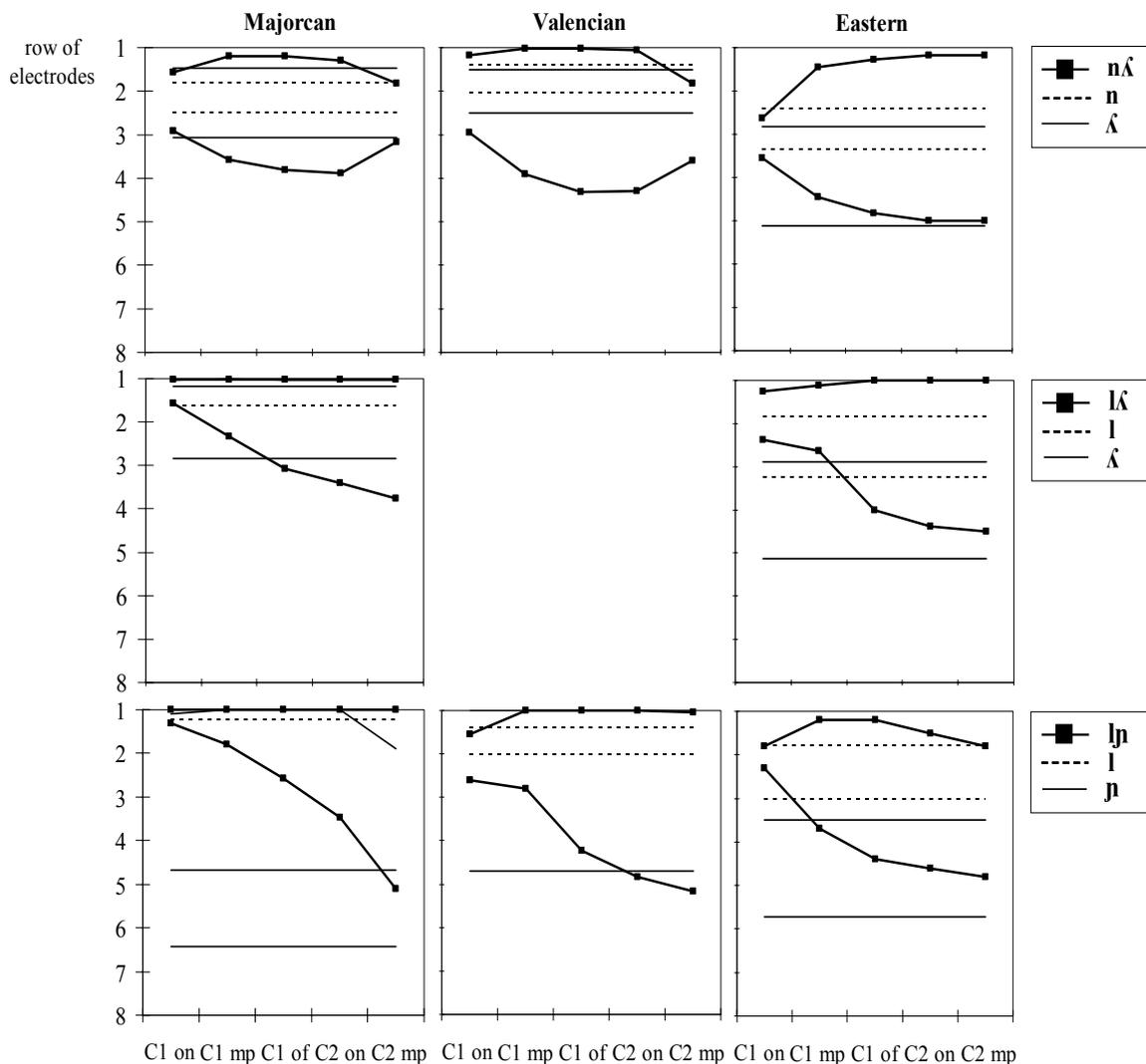
Figure 2. Front and back closure border location at C1 onset, midpoint and offset and at C2 onset and midpoint for the clusters /nʎ/ (top), /lʎ/ (center) and /lɲ/ (bottom). Closure location at both borders for intervocalic /n, l/ (discontinuous straight lines) and intervocalic /ʎ, ɲ/ (continuous straight lines) is also given. Data correspond to averages across tokens and speakers.

85

Straight lines for the intervocalic consonants show that /n, l/ are generally articulated somewhere at the central area of the alveolar zone except for strongly dark /l/ in Majorcan which is articulated at row 1 only and therefore, exhibits probably a dental articulation. Regarding intervocalic /ʎ, ɲ/, the lateral is basically alveolar in Majorcan and Valencian and alveolopalatal in Eastern, while the nasal is alveolopalatal in Valencian, alveolopalatal or prepalatal in Eastern and purely palatal in Majorcan.

For all clusters, thick lines reveal that closure extent is restricted to a relatively small contact area at C1 onset and expands posteriorly to the back border as C2 is approached. This backward expansion movement often yields a comparable closure area to that for the second consonant in the cluster, and it occurs earlier in clusters with C1=/n/ (at about C1 midpoint) than in those with C1=/l/ (at about C1 offset). It applies to a lesser extent to strongly dark /l/ in Majorcan than to clear /l/ in Valencian and to moderately dark /l/ in Eastern. Therefore, it appears that the laterality and darkness degree in /l/ delay and may block the palatalization effect. An increase in contact at the front border often reaches row 1 already during the first half of the alveolar nasal and the alveolar lateral in all consonant combinations.

As a general rule, closure location at C1 onset is fairly front and resembles that for /n, l/ rather than that for /ɲ, ʎ/. Accordingly, dialect-dependent differences in closure anteriority at C1 onset match those found for the same consonants in intervocalic position, i.e., fronting decreases in the progression Valencian, Majorcan > Eastern for /n/ and with darkness degree in the progression Majorcan > Valencian, Eastern for /l/.

As pointed out in the Introduction, an open issue is whether blending is implemented through the superposition of the closure areas for C1 and C2 such that the outcoming realization encompasses them, or else through the achievement of an intermediate closure location between the closure areas for the two consonants.

According to data for Majorcan and Valencian in the top graphs of Figure 2, contact degree at the front and back closure borders is greater at C1 midpoint and offset for the cluster /nʎ/ than for /n/ and /ʎ/ in intervocalic position. In Eastern Catalan, contact extent runs more anteriorly for /nʎ/ than for intervocalic /n/ and /ʎ/ at the front closure border; as for the back closure border, the contact trajectory for /nʎ/ runs more posteriorly than that for intervocalic /n/, and approaches but does not quite reach that for intervocalic /ʎ/ until at about C2. Inspection of the individual speakers' data (not shown) reveals that a closure area generally greater than or, less so, equal to that predicted by blending through superposition occurs at C1 midpoint for all Majorcans, all Valencians and the Eastern speaker DR. The Eastern speakers JP and JS, on the other hand,

show an intermediate closure location between that for C1 and C2 all throughout the cluster when the front and back closure borders are taken into consideration.

Regarding clusters with C1=/l/ in the graphs located at the center and bottom of the figure, data for the back closure border indicate that blending through superposition is not reached until about C1 offset in the case of /lʎ/ in Majorcan. This is so for most speakers (BM, MJ, CA), while others achieve blending already at C1 midpoint (ND). On the other hand, data for the back closure border for /lʎ/ in Eastern and for /lɲ/ in Valencian and Eastern reveal that the two adjacent consonants blend essentially into articulatory configurations which are intermediate between those for intervocalic C1 and for intervocalic C2. This strategy is found for most speakers (/lʎ/, DR, /lɲ/ JM, VB, MS, VG, DR, JP, JS), while others prefer blending through superposition either at C1 midpoint (/lɲ/, AV) or at C1 offset (/lʎ/, JS). In all cases, the front closure border for clusters /lʎ, lɲ/ reaches maximal fronting at row 1 early in the cluster.

A particular strategy applies to the cluster /lɲ/ in Majorcan (see bottom left graph of Figure 2), where the back closure border for C1= /l/ extends more posteriorly than that for intervocalic /l/ but falls short of the front closure border for intervocalic /ɲ/. Data for the individual speakers reveal that there is often a very large articulatory distance between the closure targets for the two consonants /l/ and /ɲ/ in intervocalic position, i.e., dark /l/ is essentially dental or front dentoalveolar, while closure for /ɲ/ may cover the prepalate and mediopalate (MJ), the entire palatal zone (CA) or just the back palate (BM, ND). This may account for why it may take so long for the articulatory target for C2=/ɲ/ in the cluster /lɲ/ to be reached and why, just as in Figure 2, C1=/l/ in this cluster exhibits an intermediate realization between intervocalic /l/ and /ɲ/ in the case of speakers BM, MJ and CA. Data for speaker CA, for whom /l/ is strongly dark and intervocalic /ɲ/ is alveolopalatal, reveal the presence of carryover effects involving depalatalization and thus, of palatal contact loss during C2. Speaker ND (whose data have not been averaged with those for the other three speakers in the figure) shows two separate targets and no gestural overlap for the two consonants in the cluster, i.e., the tongue travels fast from the alveolar zone at /l/ offset to the postpalatal zone at /ɲ/ onset.

3.2 Dentalizing environment

3.2.1 General trends

Data reported in Figure 3 allow investigating whether clusters with C1=/n, l/ and C2=/t/ undergo regressive assimilation or blending. Closure border location values for intervocalic /n, l/ are identical to those appearing in Figure 2. As for intervocalic /t/, the continuous straight lines in Figure 3 indicate that the front

closure border for this consonant reaches row 1 at the teeth, while the degree of alveolar fronting at the back closure border varies according to dialect in the progression Valencian > Majorcan > Eastern.

In view of the closure location characteristics for /t, n, l/ in intervocalic position and for /nt, lt/, the following processes of segmental adaptation may be predicted to occur in the two clusters:

(a) Regressive assimilation should apply whenever the back closure border is more anterior for intervocalic /t/ than for intervocalic /n/ or /l/, and that for the cluster is not more retracted than that for /t/. In this particular case it may be claimed that C1 exhibits the same closure location as C2 in the clusters /nt, lt/ because place regressive assimilation is at work. As mentioned in the Introduction, regressive place assimilation should involve the presence of the C2 place of articulation during the entire C1 closure period and thus, from C1 onset until C1 offset.

(b) Blending through intermediation between C1 and C2 should take place whenever the back closure border for the cluster occurs about halfway between the back closure border location for the two adjacent consonants in intervocalic position.

(c) Blending through superposition is expected to occur when the closure area for the cluster exceeds the closure area for its two consonants in intervocalic position, or equals that for intervocalic /n/ or /l/ if the back closure border for these two realizations extends more posteriorly than that for intervocalic /t/.

Data for the mean front closure border trajectories for clusters in Figure 3 (thick lines) are in agreement with those for the individual speakers in showing an anteriormost location at row 1 all along C1 and C2. This outcome could be indicative of regressive assimilation (maximal fronting is also found for C2=/t/ in intervocalic position), but is also compatible with a blending account (maximal fronting could result from the superposition of a maximally front dental C2 and a more retracted alveolar C1).

Relevant variations in contact placement over time for the clusters /nt, lt/ occur at the back closure border. As shown in Figure 3, there is some backward expansion from C1 onset to C1 midpoint and C1 offset which may be considered a general articulatory characteristic of stop closures towards the expansion of their area as consonant midpoint is approached.
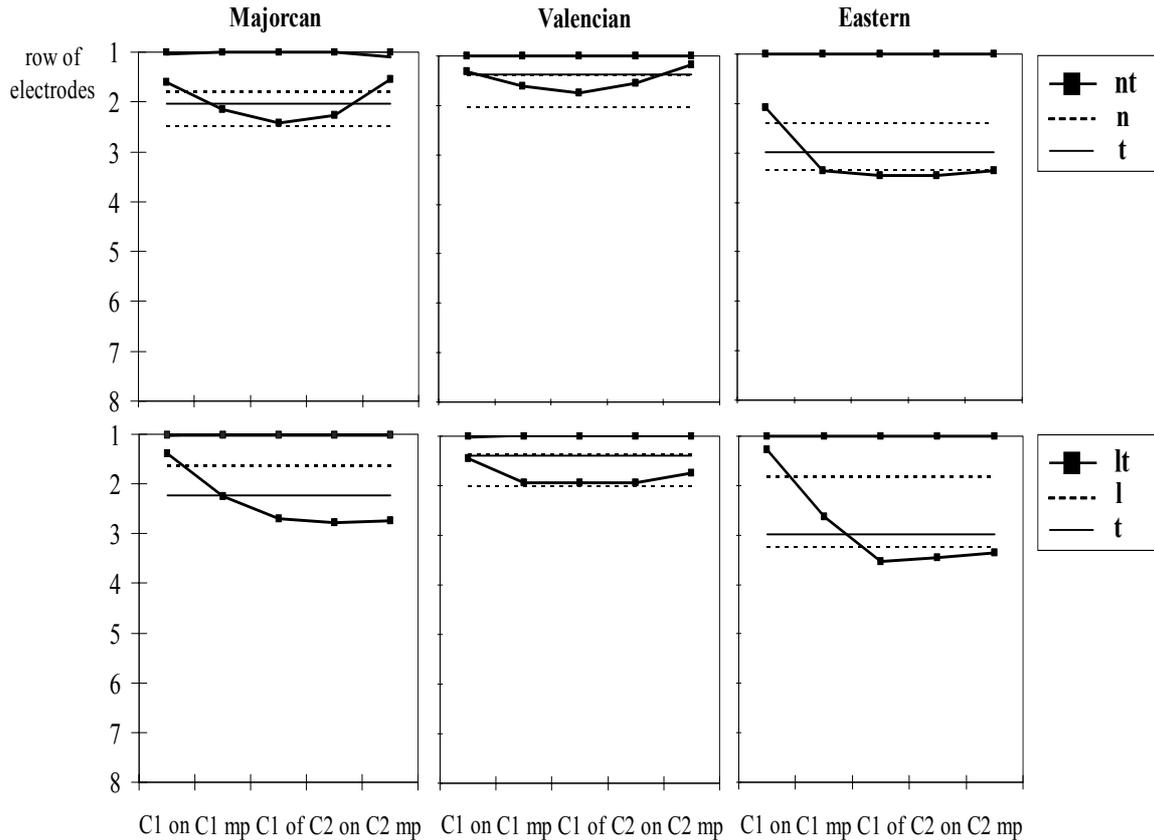
Figure 3. Front and back closure border location at C1 onset, midpoint and offset and at C2 onset and midpoint for the clusters /nt/ (top) and /lt/ (bottom). Closure location at both borders for intervocalic /n, l/ (discontinuous straight lines) and intervocalic /t/ (continuous straight lines) is also given. Data correspond to averages across tokens and speakers.

ANOVAs on the back closure border data for /nt/ yielded a significant main 'consonant condition' effect at C1 midpoint and offset in Majorcan ($F(2, 276) = 9.68$, p=0.000; $F(2, 276) = 9.78$, p=0.000) and Valencian ($F(2, 269) = 29.51$, p=0.000; $F(2, 269) = 25.99$, p=0.000). According to the results from post-hoc tests, Majorcan favors blending through superposition at C1 offset where closure is more retracted for /nt/ and for intervocalic /n/ than for intervocalic /t/, and /nt/ and intervocalic /n/ do not differ significantly in the degree of closure fronting. Valencian, on the other hand, favors blending through intermediation also at C1 offset where closure fronting decreases significantly in the progression /t/ > /nt/ > /n/. Contact trajectories in Figure 3 are in accordance with these statistical results. Regarding Eastern Catalan, ANOVAs yielded non-significant effects between /nt/ and intervocalic /t/ at C1 midpoint and at C1 offset. Data for this dialect in the figure suggest that the cluster /nt/ and the two intervocalic consonants /n/ and /t/ exhibit a similar degree of closure fronting which renders the cluster production strategy ambiguous: there could be regressive assimilation

since the back closure border location for C1=/n/ in the cluster is comparable to that for intervocalic /t/, but also blending through superposition since the closure area for C1=/n/ is equivalent to the closure areas for intervocalic /n/ and /t/.

ANOVAs on the back closure border data for /lt/ yielded a significant main effect of 'consonant condition' at cluster midpoint and offset in Majorcan ($F_{(2, 219)} = 10.66$, $p=0.000$; $F_{(2, 219)} = 19.09$, $p=0.000$) and Valencian ($F_{(2, 237)} = 29.30$, $p=0.000$; $F_{(2, 237)} = 28.66$, $p=0.000$), and at closure midpoint but not at closure offset in Eastern ($F_{(1, 84)} = 4.82$, $p=0.03$). Post-hoc tests reveal that Majorcan favors blending through superposition at closure offset where the back closure border is more posterior for /lt/ than for both intervocalic /l/ and /t/ (closure fronting values decrease in the progression /l/ > /t/ > /lt/ in this case). Valencian also favors blending through superposition at C1 midpoint and offset since the back closure border is more retracted for intervocalic /l/ and for the cluster /lt/ than for intervocalic /t/ at both temporal points. Statistical comparisons between data for /lt/ and for intervocalic /t/ suggest that Eastern Catalan favors blending through superposition at C1 offset since /lt/ becomes significantly more posterior than /t/ at this temporal point. Data on the back closure border trajectories for this dialect in Figure 3 show that /lt/ could also be more retracted than intervocalic /l/.

In summary, results from ANOVAs and inspection of contact trajectories in Figure 3 reveal that the cluster /lt/ is implemented through blending through superposition in most cases. Blending appears to be the preferred production mechanism for the cluster /nt/ as well.

3.2.2 *Individual speakers*

A more thorough insight into the production mechanisms of the sequences /nt, lt/ may be gained from an analysis of the data for the individual speakers. Tables 1 and 2 present front and back closure border location values at C1 onset, midpoint and offset for /nt/ and /lt/ according to all Majorcan, Valencian and Eastern speakers (right panel), and at consonant midpoint for /n/ or /l/ and for /t/ in intervocalic position (left and central panels). Thus, for example, data for intervocalic /n/ and /t/ and for /n/ in the cluster /nt/ for speaker ND reveal that consonant closure in the cluster extends from row 1 (front border) to somewhere between rows 3 and 4 (3.36), while closure for the two intervocalic consonants spans from the same or a more retracted location at the front border (2.4 for /n/, 1 for /t/) to a slightly more anterior location at the back border (about 3.1 for both /n/ and /t/).

The closure fronting values for intervocalic /n, l, t/ will be compared with those at the midpoint and offset of C1 in the two clusters /nt, lt/ but not with those at C1 onset since blending does not occur at this moment in time yet.

Standard deviations for the mean values in the tables are generally small, i.e., zero for the front closure border data, and about a fifth of the corresponding mean for the back closure border data.

Table 1: Front border (FCB) and back border (BCB) closure placement values at the midpoint of intervocalic /n/ and /t/ (left and middle panels), and at onset, midpoint and offset of C1=/n/ in the cluster /nt/ (right panel). Data are presented for all speakers of Majorcan, Valencian and Eastern.

Dialect	Speaker	/VnV/		/VtV/		C1 ons (/nt/)		C1 mp (/nt/)		C1 of (/nt/)	
		FCB	**BCB**	**FCB**	**BCB**	**FCB**	**BCB**	**FCB**	**BCB**	**FCB**	**BCB**
Majorcan	AR	1.95	2.24	1.00	1.19	1.00	1.07	1.00	1.21	1.00	1.36
	BM	1.43	2.38	1.00	2.14	1.00	1.27	1.00	2.00	1.00	2.00
	MJ	1.76	2.38	1.00	1.33	1.00	1.15	1.00	1.62	1.00	1.92
	ND	2.40	3.10	1.00	3.05	1.21	2.29	1.00	3.36	1.00	3.71
	CA	1.52	2.33	1.00	2.43	1.00	2.21	1.00	2.50	1.00	3.07
Valencian	JM	1.12	1.62	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	VB	1.33	1.70	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MS	1.24	2.24	1.00	1.71	1.00	1.93	1.00	2.43	1.00	2.93
	VG	1.62	2.33	1.00	1.43	1.00	1.14	1.00	1.64	1.00	1.79
	AV	1.55	2.00	1.00	1.71	1.00	1.50	1.00	1.93	1.00	2.00
Eastern	DR	1.33	3.00	1.00	3.60	1.00	1.80	1.00	3.60	1.00	3.80
	JP	2.83	3.83	1.00	1.67	1.00	1.67	1.00	2.33	1.00	2.33
	JS	3.00	3.83	1.00	3.67	1.00	3.00	1.00	4.00	1.00	4.00

Table 2: Front border (FCB) and back border (BCB) closure placement values at the midpoint of intervocalic /l/ and /t/ (left and middle panels), and at the onset, midpoint and offset of C1=/l/ in the cluster /lt/ (right panel). Data are presented for all speakers of Majorcan, Valencian and Eastern.

Dialect	Speaker	/VlV/		/VtV/		C1 ons (/lt/)		C1 mp (/lt/)		C1 of (/lt/)	
		FCB	**BCB**	**FCB**	**BCB**	**FCB**	**BCB**	**FCB**	**BCB**	**FCB**	**BCB**
Majorcan	BM	1.00	1.24	1.00	2.14	1.00	1.00	1.00	2.43	1.00	2.86
	MJ	1.00	1.29	1.00	1.33	1.00	1.00	1.00	1.33	1.00	1.33
	ND	1.05	2.67	1.00	3.05	1.14	1.71	1.00	3.14	1.00	4.00
	CA	1.00	1.31	1.00	2.43	1.00	1.71	1.00	2.00	1.00	2.43
Valencian	JM	1.05	2.00	1.00	1.00	1.17	1.17	1.00	1.17	1.00	1.00
	VB	1.67	1.76	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	MS	1.70	2.60	1.00	1.71	1.00	1.83	1.00	2.33	1.00	2.33
	VG	1.19	1.71	1.00	1.43	1.00	1.14	1.00	2.00	1.00	2.00
	AV	1.38	1.95	1.00	1.71	1.00	2.00	1.00	2.86	1.00	3.00
Eastern	DR	1.00	2.50	1.00	3.60	1.00	1.20	1.00	2.80	1.00	3.80
	JP	1.67	2.50	1.00	1.67	1.00	1.00	1.00	2.00	1.00	2.67
	JS	2.67	4.00	1.00	3.67	1.00	1.67	1.00	3.00	1.00	4.00

According to the data in the tables, assimilation (scenario *(a)*; see section 3.2.1) appears to hold for /nt, lt/ for several speakers (/nt/, AR, BM, JM, VB; /lt/, JM,VB). Thus, for example, data for speaker JM show that the back closure border is more retracted for intervocalic /n/ (1.62) than for intervocalic /t/ (1), and equally anterior at the midpoint and offset of /n/ before /t/ in the cluster /nt/ (1) than at the midpoint of intervocalic /t/ (1).

Blending through intermediation (scenario *(b)*; see section 3.2.1) holds for /nt/ (MJ, VG, JP) and /lt/ (MS). Thus, according to speaker MJ, the back closure border values at C1 midpoint and offset in the cluster /nt/ (1.62, 1.92) are intermediate between those for intervocalic /t/ (1.33) and for intervocalic /n/ (2.38).

Blending through superposition (scenario *(c)*; see section 3.2.1) involving a larger closure for the cluster than for its two consonants in intervocalic position operates in the case of /nt/ for speakers ND, CA, MS and JS and in the case of /lt/ for speakers BM, ND, VG and AV. Thus, according to speaker ND, back closure border location values for /n/ before /t/ at closure midpoint and offset (3.36, 3.71) exceed those for /n/ and /t/ in intervocalic position (about 3.1). As pointed out in section 3.2.1, blending through superposition may be also said to apply whenever the closure extent for the cluster matches that for the most widespread intervocalic consonant if this consonant is not /t/. This condition is at work for /nt/ for speaker AV and for /lt/ for speakers JP and JS, since the closure area for C1=/n, l/ in the two clusters is analogous to that for the same consonants in intervocalic position, and that for intervocalic /n, l/ exceeds that for intervocalic /t/.

The remaining cases are ambiguous. This is so for the clusters /nt, lt/ for speaker DR and for the cluster /lt/ for speaker CA, where the closure area for C1=/n/ and C1=/l/ is close or analogous to that for intervocalic /t/, and that for intervocalic /t/ is more posterior than that for intervocalic /n/ and/or /l/. The production strategy for /lt/ for speaker MJ is also unclear since the closure location for C1=/l/ in the cluster is as anterior as that for intervocalic /l/ and /t/.

4   Discussion

Linguopalatal data on closure location and extent has allowed us to determine the production strategy of consonant clusters composed of dentals, alveolars and alveolopalatals in Majorcan, Valencian and Eastern.

When followed by an alveolopalatal consonant, the alveolar nasal /n/ undergoes frontward and backward closure expansion, and blending through superposition at about C1 midpoint, as a general rule. Regressive palatalization is less obvious for clusters with C1=/l/, mostly so if the alveolar lateral is dark: indeed, while the front closure border achieves maximal fronting early in the cluster, the back closure border may either fall somewhere between that for the two adjacent consonants or else may achieve a more retracted location not earlier than C1 offset. Blending through superposition may also fail to apply when the articulatory distance between C1 and C2 is too large; thus, in Majorcan clusters where /ɲ/ is palatal rather than alveolopalatal, the final articulatory outcome exhibits either an intermediate configuration between that of C1 and

C2 or else two separate targets. In all cases, closure location at C1 onset occurs at the alveolar zone and is analogous to that for intervocalic /n, l/. These data indicate that blending through superposition applies naturally when both adjacent consonants are considerably unconstrained and articulated close to each other, while a greater articulatory separation between those consonants or an increase in degree of constraint for one of them may either cause a delay in the blending implementation or else may yield an articulation which is intermediate between the two consonants in the cluster.

A comparison between the degree of back closure border placement for the cluster and for its two consonants reveals that /nt, lt/ may exhibit blending or assimilation depending on the speaker taken into consideration. The presence of a dental place of articulation at cluster onset could be taken as evidence for the two adaptation processes. Mean values across speakers indicate that blending through superposition is the favored strategy and statistical results allow concluding that this strategy applies mostly at C1 offset. In Valencian, however, blending through intermediation has also been found to occur. Speakers may prefer one strategy over another in both clusters or may split their preference according to cluster, and also favor blending mostly through superposition over assimilation. In view of the small data set submitted to statistical analysis in the present investigation, more experimental evidence will have to be collected in order to ascertain the validity of the production strategies just referred to.

The notion of articulatory blending formulated by Articulatory Phonology has proved useful for dealing with the articulatory outcome of clusters composed of relatively unconstrained front lingual consonants in Catalan. Blending does not necessarily result in the formation of an intermediate articulatory configuration between the two adjacent consonants in the cluster, however. Thus, sequences composed of an alveolar followed by an alveolopalatal consonant favor blending through superposition unless one of the two consonants is particularly constrained (e.g., dark /l/) or the two are articulated far away from each other (e.g., whenever /ɲ/ is palatal rather than alveolopalatal). Articulatory blending in these clusters is implemented gradually such that its temporal onset does not usually occur at the beginning of the cluster, i.e., closure location at C1 onset is basically the same as that for the consonant in intervocalic position. The production strategy involved in clusters with a dental C2 is less straightforward though blending through superposition appears to be preferred over assimilation by our Catalan speakers.

Data for other Majorcan clusters with /n, l/ followed by (alveolo)palatal consonants other than /ʎ, ɲ/ not submitted to analysis in the present investigation are in agreement with the notion that blending may fail to apply if C1 is especially constrained and/or if the two adjacent consonants are articulated too far away from each other. Thus, in addition to /lɲ/, other clusters starting with a

strongly dark variety of /l/ may involve two separate targets for C1 and C2. This is so for /lj/ for all speakers and for /lc/ for speakers BM, MJ and ND, where /j/ and /c/ exhibit a purely palatal realization (/c/ is an allophone of /k/ in Majorcan and therefore should be transcribed [c]). If C1 is unconstrained and C1 and C2 are fairly distant articulatorily, as for the cluster /nc/ for the same speakers BM, MJ and ND, segmental adaptation does not result in blending but in regressive assimilation. In this case, the phonetic outcome [ɲc] exhibits a palatal, not an alveolopalatal, nasal realization, just as /nk/ may be realized [ŋk] in many languages.

On the other hand, the final outcomes for the Majorcan clusters /nc/ and /lc/ with an alveolopalatal instead of a palatal realization of /c/, are basically the same as those for /nʎ/ (/nc/) and for the sequence /lɲ/ with a strongly dark variety of /l/ and an alveolopalatal variety of /ɲ/ (/lc/). Indeed, data for speaker CA indicate that the alveolopalatal variety of /c/ undergoes blending after the relatively unconstrained stop /n/, and carryover depalatalization involving dorsopalatal contact loss after highly constrained dark /l/.

An open issue is how large the articulatory distance between the two adjacent consonants in the cluster ought to be so that a given sequence composed of an alveolar C1 and an (alveolo)palatal C2 undergoes one adaptation process or another. Data reveal that two types of palatal consonants need to be differentiated in this respect, i.e., alveolopalatals which are articulated with the blade and the predorsum and whose place of articulation may coincide partly with that for alveolars, and palatals proper which are basically mediodorsal or medio-postdorsal and involve no alveolar central contact. Clusters with an alveolopalatal C2 may undergo blending (/nc/) or C2 depalatalization (/lc/), while those with a palatal C2 are implemented through regressive assimilation (/nc/) or two successive targets (/lc/).

Acknowledgments

5   References

Browman, C.P. & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology* 6, 201-251.

Browman, C.P. & Goldstein, L. (1992). Articulatory Phonology: an overview. *Phonetica* 49, 155-180.

Fowler, C.A. & Saltzman, E. (1993). Coordination and coarticulation in speech production. *Language and Speech* 36, 171-195.

Navarro Tomás, T. (1972). *Manual de pronunciación española*. Madrid: CSIC, 17th edition.

Recasens, D. (2006). Integrating coarticulation, blending and assimilation into a model of articulatory constraints. In: L. Goldstein, D. Whale & C. Best (eds.). *Laboratory Phonology 8*. Berlin-New York: Mouton de Gruyter, 611-634.

Recasens, D. & Espinosa, A. (2006). Articulatory, positional and contextual characteristics of palatal consonants. Evidence from Majorcan Catalan. *Journal of Phonetics* 34, 295-318.

Recasens, D. & Pallarès, M.D. (2001). Coarticulation, blending and assimilation in Catalan consonant clusters. *Journal of Phonetics* 29, 273-301.

Recasens, D., Pallarès, M.D. & Fontdevila, J. (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America* 102, 544-561.

Solé, M.J. (2002). Assimilatory processes and aerodynamic factors. In: C. Gussenhoven & N. Warner (eds.). *Laboratory Phonology 7*. New York: Mouton de Gruyter, 351-386.

How do voiced retroflex stops evolve?
Evidence from typology and an articulatory study

Silke Hamann
Utrecht University, Utrecht, The Netherlands

Susanne Fuchs
Zentrum für Allgemeine Sprachwissenschaft, Berlin

The present article illustrates that the specific articulatory and aerodynamic requirements for voiced but not voiceless alveolar or dental stops can cause tongue tip retraction and tongue mid lowering and thus retroflexion of front coronals. This retroflexion is shown to have occurred diachronically in the three typologically unrelated languages Dhao (Malayo-Polynesian), Thulung (Sino-Tibetan), and Afar (East-Cushitic). In addition to the diachronic cases, we provide synchronic data for retroflexion from an articulatory study with four speakers of German, a language usually described as having alveolar stops. With these combined data we supply evidence that voiced retroflex stops (as the only retroflex segments in a language) did not necessarily emerge from implosives, as argued by Haudricourt (1950), Greenberg (1970), Bhat (1973), and Ohala (1983). Instead, we propose that the voiced front coronal plosive /d/ is generally articulated in a way that favours retroflexion, that is, with a smaller and more retracted place of articulation and a lower tongue and jaw position than /t/.

1   Introduction

Retroflex segments are often understood as articulations that involve a bending backwards of the tongue tip (see, e.g., Trask 1996, p.308). This narrow definition excludes segments in a large number of languages that are traditionally described as retroflexes, such as the postalveolar fricative in Mandarin (see Ladefoged & Wu 1984). For this reason, the present study defines retroflexion as an articulation with the tongue tip (apical) or tongue underside (subapical or sublaminal) against the alveolar, postalveolar or palatal region, following Ladefoged & Maddieson (1996). This broader definition

includes segments such as the postalveolar fricatives in Russian and Polish, whose retroflex status is debatable (see the discussion in Hamann 2004). The tongue tip raising of retroflex articulations requires a flattening of the tongue middle, which co-occurs with a retraction of the tongue back (the retraction of the tongue back is argued to be a general property of retroflexes by Hamann 2002, 2003; but see Bhat 1974, and Flemming 2003).

The complexity of gestures involved in the articulation of retroflexes might be the reason why this segmental class occurs relatively seldom in the languages of the world. For instance, only an estimated 11 percent of all languages have a retroflex stop (Ladefoged & Bhaskararao 1983, p.292). Furthermore, retroflexes occur only in larger coronal inventories, no language is known to us in which retroflexes are the only coronals.[1]

In his thorough study on retroflexes, Bhat (1973) discusses several diachronic processes that introduced this articulatorily complex class into languages. He mentions assimilatory influences of adjacent back vowels, rhotics, and velar consonants, but also introduction of a single voiced retroflex /ɖ/ via voiced dental implosives (p.55). For the latter, Bhat refers to Greenberg (1970), though the explanation given by Greenberg (p.129) actually comes from Haudricourt (1950): Voiced dental implosives are quite often retracted, which can lead to a retroflex implosive and eventually to a pulmonic retroflex stop.

Ohala (1983, p.200) also describes a development of a voiced retroflex stop from a voiced apical implosive (also referring to Greenberg), and furthermore elaborates that this process has an aerodynamic cause: Voiced apical stops can maintain voicing longer if the tongue body is lowered during the closure, and since the tongue lowering goes together with a retraction of the tongue tip, it is argued to result in retroflexion. This aerodynamic explanation is obviously independent from implosion, though Ohala does not make this point explicit. We propose in the present article that *both* implosives and retroflexes can develop from voiced (not voiceless) front coronal stops, and argue that several factors (mostly also based in aerodynamics, but also articulatory and perceptual requirements) are responsible for this development.

The view taken by Haudricourt (1950), Greenberg (1970), Bhat (1973) and Ohala (1983) is depicted in (1), with implosion as the only possible development of retroflex voiced stops in a language with no other retroflexes (note that the intermediate step of a retroflex implosive is not mentioned by Bhat and Ohala).

---

[1] Maddieson (1984) lists Kota as having only one sibilant, namely a retroflex voiceless [ʂ], which can therefore be interpreted as a counterexample to the statement that retroflexes always occur with other coronals. Emeneau (1944), the original source for Maddieson's classification, however, describes this sound as [s], in free variation with [tʃ], which is realized as retroflex only adjacent to other retroflexes (see also Flemming 2003, p.354).

(1)   ɗ (> ɗ̢) > ɖ̢

Our alternative view is represented in (2), where both retroflexion (a) and implosion (b) can develop from voiced stops. This does not preclude the possibility that some retroflexes originated from apical implosives, as in (1).

(2) a)   d > ɖ̢
     b)   d > ɗ

Focus of the present study is the emergence of retroflex sounds from voiced stops proposed in (2a), though we come back to the development of implosives from plain voiced stops in sections 2.3 and 5 below. Evidence for our proposal in (2a) comes from diachronic developments of retroflexes in a number of languages where no intermediate stage of implosion is reported. Furthermore, we illustrate with articulatory data from German that there are general differences in place of articulation and tongue and jaw height between voiced and voiceless alveolar stops favouring retroflexion of /d/. Both the diachronic and the articulatory evidence support the phonetic naturalness of the process in (2a), which makes an intermediate stage of implosion redundant.

The present article is structured as follows. Section 2 elaborates on the articulatory and aerodynamic characteristics of voiced coronals, especially the similarities between plain stops, retroflexes and implosives. In section 3, we discus three typologically unrelated languages that have [ɖ] as only retroflex. Section 4 provides synchronic data from German, and section 5 is the conclusion.

2   Voiced coronal stops

To provide evidence for the claim that voiced but not voiceless front coronal stops are prone to develop into retroflexes and also into implosives, we first look at the articulatory differences between voiced and voiceless front coronal stops (§2.1), including possible explanations for this difference. We then compare the characteristics of voiced front coronals with those of retroflex stops (§2.2) and coronal implosives (§2.3). The last subsection (§2.4) discusses explanations and examples for developments of retroflexes via implosion.

In the following, we do not distinguish between dental and alveolar coronal stops but summarize them under the term 'front coronals'. Furthermore, we focus on segments in intervocalic position, for the following two reasons. First, we usually find fully voiced segments in this position (Keating 1984), which allows us to compare across languages without having to pay attention to the

actual realisation of the voicing contrast. And second, the intervocalic position is a location where all of the segmental types that we compare, that is, front coronals, retroflexes, and coronal implosives, can occur (note for instance that retroflex segments do not occur in initial position in a number of languages, see Steriade 2001; and Hamann, 2003, pp.114-118).

2.1 Front coronal voiced stops

Studies on a variety of languages have shown that there are systematic differences between the articulation of voiced and voiceless front coronal stops. /d/ is usually realized with a more posterior position of the tongue tip and thus a more backed place of articulation than its voiceless counterpart, see for instance the electropalatographic studies by Dixit (1990) on Hindi, Moen & Simonsen (1997) on English and Norwegian, and Farnetani (1989, 1990) on Italian. In all of these studies we can also observe a smaller amount of tongue palatal contact and more contextual variation for /d/ than for /t/. A further systematic difference lies in the active articulator: /t/ is often articulated with the tongue blade, whereas /d/ is usually produced with the tongue tip (see, e.g., the x-ray data by Dart 1991, 1998, on French and English), though this only holds for languages that have a single series of coronal stops. Some studies found a stronger tongue pressure against the palate during the closure of /t/ and deduce from this a higher tongue position for /t/ (e.g., Wakumoto, Masaki, Honda & Ohue 1998 and Fujimura, Tatsumi & Kagaya 1973 for Japanese). Others showed that /d/ is produced with a lower jaw position than /t/ (e.g., Fujimura & Miller, 1979 for American English; Dart, 1991, for French; and Mooshammer, Hoole & Geumann 2006, 2007 for German). A further observation is that voiced /d/ is usually shorter than its voiceless counterpart (Stevens, Keyser & Kawasaki 1986, p.432).

Several explanations have been proposed for the observed differences between voiced and voiceless front coronal stops. The first and most commonly given is the aerodynamic requirement for voicing. Vibration of the vocal folds is only possible when there is a pressure difference between the subglottal and the intraoral cavity. Such a transglottal pressure difference can easily be produced with an open vocal tract. However, during the production of plosives, the vocal tract is closed for a certain time, resulting automatically in an increase of intraoral pressure. In order to maintain voicing during oral closure, as required for thoroughly voiced stops, it is necessary to enlarge the oral cavity (either actively or passively). Mechanisms of cavity enlargement for /d/ are manifold and include for instance a change from tongue blade to tongue tip, a lowering of the tongue, the jaw or the larynx, and an extension of the cheeks (Perkell 1969; Bell-Berti 1975; Westbury 1983; for German see Fuchs, 2005). Recall from

section 1 that cavity enlargement was Ohala's explanation for diachronic changes from alveolar implosive to voiced retroflex stop.

A second explanation for the difference between /t/ and /d/ is also based on voicing requirements. Because the transglottal pressure difference can only be maintained for a certain time unless actively maintained (see the mechanisms of cavity enlargement discussed above), voiced stops have mostly a shorter duration than their voiceless counterparts, the latter having in principle no restriction on the length of their closure. The shorter duration of /d/ can then account for all other above-mentioned differences with /t/ in the following way. It has been argued that for coronal stops the tongue tip or blade is aiming at reaching a target somewhere above the constriction location (Fuchs, Perrier & Mooshammer 2001, 2006; Löfqvist & Gracco 2002), since no exact location is necessary compared to the precise positioning required for sibilants or trills. Voiced coronal stops cannot fully reach this target because they have only little time to do so, and this so-called target undershoot (Lindblom 1963) results in a lower tongue and jaw position and in a more variable articulation.

The third explanation discussed here is again grounded in aerodynamics. Voiceless stops have a greater oral pressure than voiced ones (both mean and peak pressure; see Ladefoged & Maddieson 1996, p.96), since the airflow is not arrested by the vibrating vocal folds. Consequently, they require a firmer closure at the place of articulation than voiced ones. Following Ladefoged & Maddieson (1996) we can argue that the articulatory characteristics of /t/ described above, which correlate with a more forceful articulation than for /d/, might be "an anticipation of this need to make a firmer seal" (p.96).

A last account for the articulatory difference between /t/ and /d/ proposed in the literature is that voiceless stops require a more salient burst than voiced ones. This prominent burst is an important perceptual cue to distinguish voiceless from voiced coronal stops (Lisker & Abramson 1964; Repp 1979). The higher intra-oral pressure required for such a salient burst can be achieved by a higher tongue and jaw position. Furthermore, the use of the lower teeth as a second noise source can enhance the strength of the burst and is also only possible with a high tongue and jaw position. With respect to the jaw, Mooshammer et al. (2006; 2007) found a high and stable jaw position for /t/ in German. For /d/, on the other hand, the jaw was positioned lower, which gives the tongue more freedom to move and to accommodate to the context.

Most of these explanations cannot be evaluated separately. Thus the less salient burst and the less forceful seal result both in a generally lower articulatory effort for /d/, and so does target undershoot. Only the mechanism of cavity enlargement predicts an additional active control of gestures for /d/. If the lowering of tongue and jaw were actively controlled then we would expect the voiced /d/ to show less contextual variation and to be more stable in its

articulation than /t/ (see Mooshammer et al., 2006, p.22, for a similar argumentation). This is, however, not what we find in the literature. Instead, we saw that /d/ shows a much higher variability, and hence the tongue and jaw position of /d/ are less tightly controlled than that of /t/. We can therefore exclude cavity enlargement as explanation for the difference between /t/ and /d/. The remaining three explanations can only indirectly account for the difference in place of articulation between /d/ and /t/, namely via the assumption that apical articulations are preferably alveolar and laminal ones preferably dental (Ladefoged & Maddieson 1996:20-21).

We will see in the following section that the difference between /t/ and /d/ in articulation and duration makes the voiced stop prone to change into a retroflex.

2.2 Retroflex voiced stops

Retroflexes are articulated with a raised and retracted tongue tip, that is, they are always apical or subapical, with a place of articulation between the alveolar and palatal region. The raising and retraction of the tongue tip requires a lowering of the tongue middle and a retraction of the tongue back (see introduction). Though tongue lowering usually goes together with jaw lowering, we could not find any explicit mentioning of a low jaw position for retroflexes in the literature. Retroflex segments seem also to be shorter than other consonants, see for instance Anderson & Maddieson's (1994) study on Tiwi coronal stops, where the closure duration of retroflex segments was the shortest of all coronal consonants.

Retroflex articulations in general are described as being strongly context-dependent and showing large variability due to vowel coarticulation (see Švarný´ & Zvelebil 1955; Ladefoged & Bhaskararao 1983; Dixit 1990; Dixit & Flege 1991; Krull, Lindblom, Shia & Fruchter 1995; Simonsen, Moen & Cowen 2000). Most of these studies show that retroflexes are articulated furthest back (and thus most retroflex-like) in /u/ context, and furthest front (i.e., most front coronal-like) in /i/ context. Phonological studies have shown that retroflexes often avoid /i/ context, since the two have antagonistic tongue gestures (Flemming, 2003; Hamann, 2003:94-107). The context of /u/, on the other hand, has been reported to cause retroflexion of front coronals (Bhat, 1973; Hamann, 2003:90-94), as /u/ has a similar lowered tongue middle and retracted (and raised) tongue back. The emergence of retroflexes in Australian languages is, for example, ascribed to backing of front coronals in /u/ context (Dixon 1980).

A difference between voiced and voiceless retroflex stop similar to that between voiced and voiceless front coronal stop discussed above is expected,

though we found only little work that was explicit on this point. Dixit (1990), for example, observed that the voiced retroflex stop has a narrower constriction than its voiceless counterpart, and a palatographic study by Khatiwada (2007) shows that the voiced retroflex stop in Nepalese is articulated further back and with more contextual variation than the voiceless one.

Apicality, lowered tongue middle, short duration, and strong contextual variation are characteristics that retroflex voiced stops share with the voiced front coronal stop /d/, see section 2.1 above. This strong similarity between a voiced front coronal stop and a voiced retroflex leads us to propose that the two can be considered endpoints on a continuum from no or few to a large amount of retroflex characteristics, supporting Dixit's (1990:190) observation that retroflexion is not so much a place of articulation than a descriptive term for a particular tongue shape. This tongue shape occurs sometimes together with a dental or alveolar place of articulation. This front-back retroflex continuum of voiced coronal stops supports our claim that a /d/ can develop into a /ɖ/ without an intermediate stage of implosion, by a slight shift along this continuum.

2.3 Implosives

Whereas voiced front coronal and voiced retroflex plosives differ from each other exclusively in their place of articulation, voiced coronal implosives are articulated quite differently. They can be defined by three successive articulatory stages, namely glottal closure (plus a closure along the supralaryngeal cavity), larynx lowering, which results in rarefaction of the air between the two closures, and an implosive release, where the pressure is equalized (Catford 1939). Implosive consonants are always stops and can be voiced and voiceless, but voiceless implosives are extremely rare in the languages of the world.

Though implosives are produced with an ingressive airstream, the voiced ones allow simultaneous pulmonic egressive airflow. According to Laver (1994:179), the egressive air is "not enough to overcome completely the rarefaction of the enclosed volume of air in the vocal tract caused by the descending larynx." Catford (1977a:75) proposes on the basis of cineradiographic films that there is no active pulmonic airflow in voiced implosives, and the airflow that causes the vocal fold vibration comes actually from the downwards movement of the larynx against a static pulmonic pressure.

Ladefoged (1964) describes three possibilities for producing implosive sounds, namely first the aforementioned larynx lowering with ingressive airflow at release, second a sound with laryngealized voicing, and third a preglottalized sound. These possibilities can be transcribed for instance for alveolars as [ɗ], [d̰]

and [ˀd], respectively. Ladefoged proposes that all three possibilities should be considered variants of one category, based on the following four arguments. First, the real implosive type of articulation often co-occurs with laryngealized voicing, as for instance in Hausa. Second, Ladefoged (1964:60) finds it difficult to consistently distinguish between the laryngealized and preglottalized variants. Third, some Mayan languages show positional variations of implosives, with the real implosive articulation in initial position, and preglottalized sounds intervocalically. And finally, no language has a phonemic contrast between any of these three, according to Ladefoged. This leads Ladefoged to summarize all three articulations under the category 'injective'. Clements & Osu (2003) use a similar cover-category, but employ the term 'nonexplosive stops'.

A summary of the three articulations as 'implosive' is questionable in the light of the fact that there are African languages contrasting two of the three articulatory possibilities for implosives listed by Ladefoged. Clements & Osu (2003) show in a phonetic study that the Niger-Congo language Ikwere (of the Igbo family) has a phonemic contrast between a bilabial voiced implosive and a bilabial voiced, glottalized implosive.[2] We therefore employ the term 'implosive' in the following to refer only to the real implosive articulation of this class, and not to preglottalized or laryngealized stops.

If we compare the characteristics of an implosive to that of a plain voiced stop articulated at the same place – coronal for our purposes – the two seem to differ in the movement of the larynx and the direction of the airflow, only: the implosive shows a lowering of the larynx and ingressive air at the release. Unfortunately, even the class of implosives that fall within the restricted definition employed here do not always display these two characteristics. Clements & Osu (2003) found that none of the Ikwere implosives is realized with larynx lowering, although these sounds show ingressive airflow. Similarly, Lex (2006) illustrates that the implosives in the Fouladou dialect of Fula, another branch of the Niger-Congo languages, do not always have an ingressive airflow (see also Ladefoged, 1964). Ordinary voiced stops, on the other hand, often can be accompanied by larynx lowering, for instance in English and French (Ewan & Krones 1974). These and similar findings lead Ladefoged (1964, 1971) and Ladefoged & Maddieson (1996, p.82) to suggest the difference between plain voiced stops and implosives is gradient, "lying primarily in the comparatively larger and more rapid descent of the glottis in implosive" (Ladefoged, 1971:27).

---

[2] Goyvaerts (1986) mentions a possible contrast between voiced implosives and preglottalized sounds in the East Nilo-Saharan language Lendu. Dimmendaal (1986) and Demolin (1988) argue against such a contrast since the phonetically preglottalized sounds in Lendu are phonemic sequences of glottal and plain stops.

From this we can conclude that larynx lowering and ingressive airflow are no reliable characteristics of implosives. Whether a sound in a language is categorized as (alveolar) implosive therefore depends very much on the definition of implosive employed by the linguist. For instance, all Chadic languages have the implosives /ɓ/ and /ɗ/ (see Schuh 2003). These are usually glottalized, which is the reason why they are often simply described as glottalized or laryngealized stops in the literature on Chadic, as pointed out by Clements & Rialland (2005:20).

Ladefoged's (1964 et seq.) idea that implosives without ingressive airflow form a gradient continuum with plain voiced stops, with no clear boundary between the two categories, is similar to the continuum we proposed for alveolar and retroflex articulations in section 2.2. Whereas the plain-retroflex continuum is one that differs in tongue shape, this plain-implosive continuum differs in amount and velocity of glottis lowering. The two are thus orthogonal to each other and create a two-dimensional space, including a gradient continuum from plain to implosive retroflex, but neglecting the dimension of ingressive airflow. We will come back to this proposed space in the general discussion in section 5.

2.4   Developments of retroflexes from implosives

On the affinity between retroflexes and implosives, Greenberg (1970:129) noted that an implosive corresponding to a non-implosive dental in a language is often "retracted to the alveolar or alveopalatal position and is consistently apical, often with accompanying retroflexion". Haudricourt (1950) explains that the negative air pressure (due to the larynx lowering) causes a vacuum which tends to suck in the mobile tongue tip. Such retracted implosives then tend to lose their glottalic feature, a development repeated in (3a).

(3)   a)   ɗ > ɗ̢ > ɖ̢   Haudricourt, Greenberg
        b)   ɗ > ɖ̢   Bhat, Ohala

The proposal by Bhat (1973) and Ohala (1983) in (3b) does not include an intermediate retroflex implosive, and Ohala's explanation for the development in (3b) does not refer to the negative air pressure of implosives. Instead he proposes that retroflexion is caused by cavity enlargement, where the tongue tip is retracted due to a lowering of the tongue. However, we saw in section 2.1 above that an account based on cavity enlargement does not require an implosive in the initial stage.

Let us look at languages supporting the two assumptions in (3). Greenberg bases his proposed development of voiced retroflex stops primarily on Tucker and Bryan's (1966) description of the retroflex implosives in Moru-Madi, a

branch of East Central Sudanic languages of the Nilo-Saharan family. For these sounds, "the retroflex tongue position is in fact a more distinguishing feature than the manner of articulation, which hardly seems implosive at all" (Tucker & Bryan 1966:102). This indicates a variation between retroflex implosive and voiced retroflex stop at the time of description. However, Moru-Madi languages have an additional phonemic retroflex voiced stop (see Watson 1991, in general; Demolin & Goyvaerts 1986, for Madi; Andersen 1987, for Lulubo; and Bender 1992, for a reconstruction of the contrast in Proto-Central-Sudanic), which makes a realisation of the voiced implosive as retroflex and thus a neutralisation between the two phonemes unlikely (though not impossible).

The development in (3b) is better documented. It occurred, for instance, in the Gbe languages (e.g., Fon, Ewe, Maxi) of the Niger-Congo Kwa family (Bantu), see the comparative study by Stewart (1995). Interestingly, the change in Gbe was preceded by a change in Bantu, were the coronal implosive is usually assumed to be a reflex of Proto-Bantu *d (Clements & Rialland 2005, p.21, Guthrie 1967-1971).

In the following section, we provide evidence for diachronic changes from front coronal to retroflex stops from three unrelated language families. Together with developments of implosives from plain stops as just elaborated for Bantu this illustrates that implosion and retroflexion can both be independent developments, supplementing the proposals by Haudricourt (1950), Greenberg (1970), Bhat (1973), and Ohala (1983).

3 Languages with retroflexed voiced stops only

The data for the diachronic development of retroflex voiced stops comes from three typologically unrelated languages or language groups, namely the Malayo-Polynesian Dhao (§3.1), the Sino-Tibetan language Thulung (§3.2), and the East Cushitic languages Afar, Somali and Rendille (§3.3).

3.1 Malayo-Polynesian: Dhao

Dhao, also called Ndao, Dao, Ndaonese or Ndaundau, is a Central Malayo-Polynesian language, subsumed under the Bima-Sumba subgroup (Gordon 2005).[3] It is spoken on Ndao, and partly on Rote and Timor; all three are islands in the Sabu Sea of Indonesia. Dhao has the coronal stops /t d ɖ ɗ/, where the

---

[3] The subgroup of Bima-Sumba languages is based on the classification by Jonker (1896; 1903; see also Esser, 1938) and has been criticised for its lack of evidence in terms of shared innovations (see Ross 1995:83). Fox (2004:7-8) argues for a more fine-grained distinctions between the languages of Sumba, those of Bima and Manggarai, and a separate subgroup of Sabu and Ndao.

retroflex is released with frication (Grimes, 2006:4)[4]. The closely-related Sabu (or Sawu(nese), Hawu, Havu) is spoken on the neighbour-islands of Sawu. Sabu has implosives, but no retroflexes, and its coronal stops are /t d ɗ/. Ngad'a, a further Bima-Sumba language, is spoken on Westflores and has like Sabu only implosives but no retroflexes (Arndt 1933, Klamer 1998), and the same holds for its neighbouring languages Lio and Kambera (Baird 2002).

The retroflex in Dhao corresponds to a plain stop in cognate words of Sabu, and the plain voiced stop to a palatal implosive; full correspondences between Dhao and Sabu voiced coronal stops are given in (4) (from Grimes, 2006:8).

(4)   Dhao   Sabu
    a)   ɖ   d
    b)   d   ʄ
    c)   ɗ   ɗ

The retroflex stop in Dhao and the Sabu alveolar stop (both 4a) are assumed by Grimes (2006) to stem from a voiceless alveolar, retroflex or palatal stop in Proto-Malayo-Polynesian (PMP). Evidence for the reconstruction of a voiceless segment for these voiced sounds comes from the fact that the sounds in (4c) correspond to voiceless stops in neighbouring languages (Jonker, 1903:86). The exact place of articulation of the PMP sound is difficult to determine and depends to a large extent on what has been reconstructed for Proto-Austronesian. For the purpose of the present article we can summarize Grimes' assumption that Dhao developed voiced retroflex stops from voiceless coronal stops, and not from implosives. Whether this development went via an intermediate stage of voiced front coronal stop is open to speculation.

Interestingly, the alveolar implosives in Dhao and the neighbouring Sabu in (4c) are assumed to have developed from a retroflex or palatal voiced stop in PMP (see Grimes 2006), as depicted in (5). Most authors (e.g., Dempwolff, 1934; Dyen 1971; Ross, 1992) assume a voiced and a voiceless retroflex stop in PMP, whereas others (such as Wolff 1974, 1991) propose palatal stops instead.

(5)   a)   * ɖ > ɗ
    b)   * c > ɗ

If the change did take place as in (5a), then we would have a reversal of the general development in (1) assumed by Haudricourt (1950), Greenberg (1970), Bhat (1973) and Ohala (1983).

---

[4]   This sound might be a retroflex affricate, though we found no further indication for this in the literature.

3.2 Sino-Tibetan: Thulung

Thulung, also called Thulung(e) Rai (e.g. Lahaussois 2003), is a Sino-Tibetan language and belongs to the subgroup of Western Kiranti languages. It is mainly spoken in Eastern Nepal. Thulung has an extensive coronal inventory, with four laryngeal settings for dental plosives and affricates: /t tʰ d dʰ ts tsʰ dz dzʰ/ (Ebert, 1997:14). According to Ebert (1994, 2003), Thulung is the only Kiranti language with retroflex stops in addition to this dental series.[5] The voiced retroflex /ɖ/ is phonemic, since it forms minimal pairs with initial /d/ in native words. The voiceless retroflexes [ʈ ʈʰ] are marginal and do not contrast with other coronals (Ebert, 1994; Lahaussois 2003:1).

If we compare Thulung words having a voiced retroflex to cognates in neighbouring languages, we can see that other Western Kiranti languages (such as Dumi, Khaling, Jero) have a voiceless stop /t/, and the Eastern Kiranti languages (such as Camling, Bantawa, Yamphu) have a voiced stop /d/ in its place (Michailovsky 1994), see (6a).[6]

(6)	Western Kiranti (except Thulung)	Thulung	Eastern Kiranti
a)	t	ɖ	d
b)	d	d	t
c)	t	t	t

For the voiced /d/ in Thulung, we find the same phoneme in the other Western Kiranti languages, but a voiceless /t/ in the Eastern Kiranti languages, see (6b). Of importance for a historical reconstruction of Proto-Kiranti is furthermore that

---

[5] The discussion on Kiranti is restricted to initial consonants. Other Kiranti languages like Limbu and Camling have retroflex consonants in this position, but almost only in loanwords from Nepali (Ebert 2003:14; Driem 1987:27). The Western Kiranti language Jero seems to be a case like Thulung because it has the phoneme /ɖ/ in native words. However, Opgenort (2005:59) describes somewhat uninterpretable that its use instead of /d/ "seems to be generally determined by personal style or preference" (ibid.). He goes on to say that the retroflex flap [ɽ] is an allophone of /ɖ/ in intervocalic position, and is a common sound in native Jero words, indicating again that the postulation of a phoneme /ɖ/ is justified.
According to Ebert (2003:14), the Eastern Thulung language Athpare has no dental coronals, but retroflex segments instead. No further information on this language could be obtained.
Michailovsky (1994:766) lists Sunwar as Kiranti language with dentals and retroflexes. However, Sunwar is usually not considered a Kiranti language, but as belonging to the Kham-Magar-Chepang-Sunwari languages, which form together with the Kiranti languages the Mahakiranti branch of Himalayish (Gordon 2005).

[6] The Eastern-Kiranti Limbu has no voiced stops.

Thulung /t/ in (6c) merged with the cognates of Thulung /ɖ/ in the other Western Kiranti languages, and with the cognates of Thulung /d/ in the Eastern Kiranti languages (Opgenort, 2005). This intricate relationship led several scholars to reconstruct three sounds in Proto-Kiranti corresponding to the ones in (6a) - (c), namely *t for the uniformly voiceless stops in (6c), *d for the sounds in (6b), and a preglottalized *ʔt for the sounds in (6c) (see Starostin 1994, and Opgenort, 2005; Michailovsky 1994, assumes a glottalized segment at a later stage). Michailovsky (1994:770) points out that the reconstruction of a preglottalized segment is somewhat speculative since there is no direct evidence for it. Opgenort (2005:14) agrees, but proposes that the preglottalized consonant might go back to the Tibeto-Burman prefix *ʔə. None of these authors accounts for the change in voicing that has to have taken place, if one assumes the development *ʔt > ɖ. In any case, there is no indication that the reconstructed segment was realized as an implosive, nor did it give rise to an implosive in any Kiranti language.[7] We can therefore take Thulung as evidence for a further language in which a voiced retroflex stop developed directly from a front coronal stop without an intermediate stage of implosion.

3.3 East Cushitic: Afar, Somali, Rendille

East Cushitic languages belong to the Afro-Asiatic family and are spoken in Somalia, Ethiopia, Eritrea, and Kenia. A number of East-Cushitic languages are reported to have a voiced retroflex stop /ɖ/, namely Afar (Bliese 1981), Somali and Rendille (Sasse, 1979:25; Lloret 1995:69). The related languages Boni, Arbore and Elmolo have instead an alveolar implosive usually transcribed as /d'/, see Sasse (1979:25). Sasse (ibid.) also mentions Dasenech in this context. Tosco (2001), however, describes the Dasenech sound as a retroflex implosive, realized as a plain retroflex stop [ɖ] or flap [ɽ] intervocalically (p. 21). In Oromo, a further East-Cushitic language, the cognate sound is also realized as retroflex implosive (see Gragg 1976, on the Western dialect Wellega, and Stroomer 1987, on the Southern dialects Boraana, Orma and Waata). A summary of the correspondences between these languages is given in (7).

(7) a) ɖ   Afar, Somali, Rendille
     b) ɗ   Boni, Arbore, Elmolo
     c) ɗ̢   Dasenech, Oromo (Western and Southern dialects)

---

[7]   Note that Opgenort (2005) proposes the existence of a preglottalized nasal *ʔn in Proto-Kiranti to account for the implosive /ɗ/ in Jero, which corresponds to plain nasals in all other Kiranti languages.

The sounds in (7) all stem from the same Proto-East Cushitic segment, which Sasse (1979:25) reconstructs as a voiced coronal stop *d' and describes as "glottalized or otherwise affected". Since this glottalized segment could be argued to have been an implosive (it resulted in implosives in neighbouring languages, and recall the discussion in section 2.3 on varied articulations and therefore inconsequent descriptions of implosives), the languages Afar, Somali and Rendille do not seem to provide strong evidence in favour of our argument that retroflexes did not necessarily develop from implosives.

It has to be mentioned, however, that Heine (1978) proposes a sub-classification of the Eastern Cushitic languages Somali, Rendille, and Boni as what he terms "Sam" languages (see also Tosco 2001), and reconstructs a Proto-Sam retroflex *ɖ which he assume to persisted into present-day Somali and Rendille but changed in Boni to an implosive /ɗ/. This reconstruction would, if correct, provide another example for the reverse development of a retroflex into an implosive, like the case of Dhao in (5).

The retroflex implosive in Dasenech (7c), which at present has a plain retroflex allophone [ɖ~ʈ] in intervocalic position (Tosco, 2001) that was not reported in earlier sources, provides an example for Haudricourt's (1950) assumption that plain retroflex voiced stops develop from retroflex implosives (3a).

To sum up, we illustrated with the examples of three typologically unrelated languages that diachronic developments of retroflex voiced stops do not necessarily proceed from alveolar or retroflex implosives. Furthermore, we saw two examples for a possible reverse development from a retroflex into an implosive, namely the change from Proto-Malayo-Polynese *ɖ to Dhao and Sabu /ɗ/, and from Proto-Sam *ɖ to Boni /ɗ/. For all the example developments discussed in this section, we have to keep in mind that we are dealing here with diachronic reconstructions of languages, which are usually based on the comparison of daughter languages and lack any kind of direct evidence. This is especially problematic for language families that exhibit a huge variety due to large areal spread and/or continuous contact with other language families (such as Austronesian).

While the languages presented up to now developed retroflex phonemes across several generations, the data on German in the following section differ in two ways: They are synchronic, and they illustrate allophonic retroflexion for one speaker (the other speakers in this study show allophonic backing). But again they provide evidence for the emergence of retroflexion from a voiced coronal stop without an intermediate stage of implosion.

4   A German case study

We chose German to provide us with synchronic data on the difference between front coronal voiced and voiceless stops and the affinity of /d/ to retroflexes for two reasons. First, it is a language without retroflex stops and where therefore the alveolars /t, d/ as the only coronal stop phonemes can considerably vary in their place of articulation (cf. the findings for French and English coronals by Dart 1998). Second, articulatory data on German in the form of Electromagnetic Articulography (EMA) and Electropalatography (EPG) was available from the study by Fuchs (2005), who looked at the realization of voicing in German obstruents. Data presented here are restricted to an intervocalic, unstressed position, because in this position a true voicing contrast is most likely for German. In initial position we find a contrast between plain and aspirated voiceless segments (Jessen 1998) and in final position a subtle contrast or none at all (due to final devoicing). The intervocalic position is also the one in which there seem to be no restrictions on the occurrence of plain stops, retroflexes and implosives.

We tested whether the voiced stop in German is realized in a way that favours retroflexion, that is, with a more retracted place of articulation, less palatal contacts, a lower tongue and lower jaw position, and with more contextual variation than /t/.

4.1 Methods

In order to test the above-mentioned differences we investigated tongue and jaw movements together with tongue-palate contact patterns by means of simultaneous EPG (Reading EPG3) and EMA recordings (AG 100, Carstens Medizinelektronik). Tongue tip (tt) movement was associated with the movement of a sensor placed midsagittally approximately 1 cm behind the tip. Tongue back (tb) movement was associated with a sensor that was placed at the posterior end of the tongue where it touches the soft palate. Since this sensor came loose during the recording session for 2 of the 4 subjects, we do not discuss it here. Two sensors, one for tongue mid (tm) and one for tongue dorsum (td), were placed in between and in equal distance to the tt and tb sensors. Jaw movement was associated with a sensor at the lower incisors. Two sensors served as reference points to compensate for helmet movements, one at the nasion and one at the upper incisors. Speech signals were recorded on Digital Audio Tape (DAT). Sampling frequencies were 16 kHz for the acoustic data, 100 Hz for EPG and 200 Hz for EMA data respectively.

Four German subjects were recorded, three male (Speakers 1-3) and one female (Speaker 4). The speech material consisted of nonsense words

"geC_1VC_2e" where C_1 and C_2 were either /t/ or /d/. The consonant C_2 occurred in an unstressed word medial position and the vowel preceding C_2 was always one of the stressed tense vowels /a, i, u/. We included different vowel context since we expected a retroflex-like articulation in /u/ context but not in the context of /i/ (recall the discussion in section 2.2). The target word was embedded in the carrier phrase *Ich habe geCVCe nicht Y erwähnt,* 'I said geCVCe not Y', with Y being another target word which is not the focus of this study here. Each sentence was repeated 10 times in a randomised order. We should point out that the measured tongue sensor signals are composed of both, the tongue and the jaw, since decomposition is not a straight forward process. For further details of the study, see Fuchs (2005).

On the basis of the EMA data we labelled the consonantal target, defined as the highest vertical position of the tongue tip sensor in correspondence with tongue palate contacts in the alveolar part of the palate. For this point, the following three measures were carried out:

(8)   a)   the horizontal (x) position of the tongue tip,
       b)   the vertical (y) position of the tongue dorsum and of the jaw, and
       c)   the frequency of tongue palatal contacts over all repetitions.

Although jaw lowering (in 8b) has not been mentioned as a potential characteristic of retroflexes before, we assume that it goes hand in hand with the tongue dorsum lowering to allow more flexibility for the apical articulation (it may also be a requirement for tongue tip curling). By contrast, a high jaw position makes a retroflex tongue configuration very unlikely.

In addition to the consonantal target we also labelled the vowel target, defined as the lowest vertical position (or most backward position in /u/-context), corresponding approximately to the turning point for all tongue sensors.

Both vowel and consonant target are used for the measure of the tongue tip angle. This is a measure introduced by Tiede, Gracco, Shiller, Espy-Wilson & Boyce (2005) in their study on variations of American /r/ to distinguish retroflex from bunched varieties. The tongue tip angle is calculated using three successive sensor coils on the tongue starting at the tip. It is the angle between the line connecting the first (tt) and second sensor (tm) and the line connecting the second (tm) and third sensor (td), depicted in figure 1 in the left graph with the dotted line. If this angle is greater than 180 degrees the tongue has a bunched shape, if the angle is 180 degrees or lower the tongue has a retroflex shape.
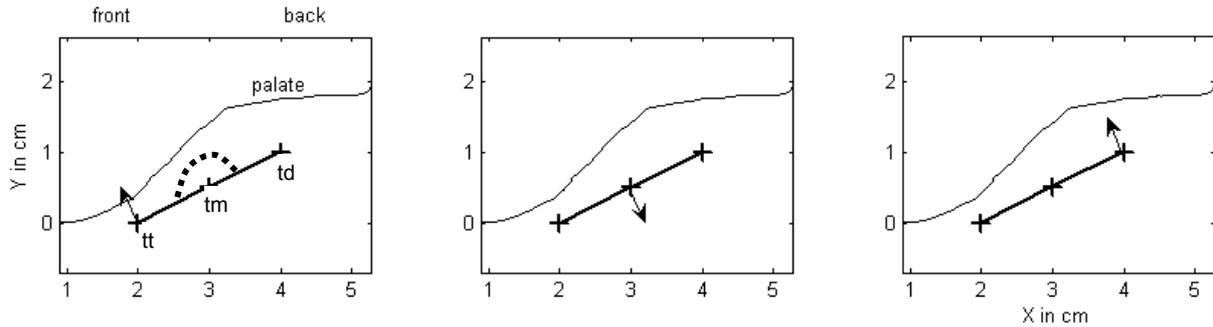
Figure 1: Schematic representation of Tiede et al.'s (2005) tongue tip angle for retroflex tongue configurations.

Figure 1 also shows that a retroflex tongue configuration according to this measure is not only possible with an upward movement of the tongue tip (graph on the left), but also with a lowering of the tongue mid (graph in the middle) and an upward movement of the tongue dorsum (graph on the right).

For the statistical analyses of the four measures we used SPSS (version 15.0).

4.2 Results

We discuss the results in the following order: horizontal position of the tongue tip in §4.2.1, frequency of tongue palatal contacts in §4.2.2, vertical position of the tongue dorsum and the jaw in §4.2.3, and the tongue tip angle in §4.2.4.

4.2.1 Retracted tongue tip position for /d/

Figure 2 displays the results for the horizontal position of the tongue tip at the consonantal target in the context of /a/ and /u/ based on EMA. It clearly shows that all speakers realize a significantly more retracted tongue tip position for /d/ in comparison to /t/ (for descriptive results and significance values see Appendix I). The differences are particularly pronounced for speaker 1 (up to 4 mm) and speaker 3 (up to 7 mm) whereas for speaker 2 and speaker 4 they are rather small (approximately 1 mm). The context of the front vowel /i/ was not included here, because in this context only speaker 3 had significant differences between /d/ and /t/.
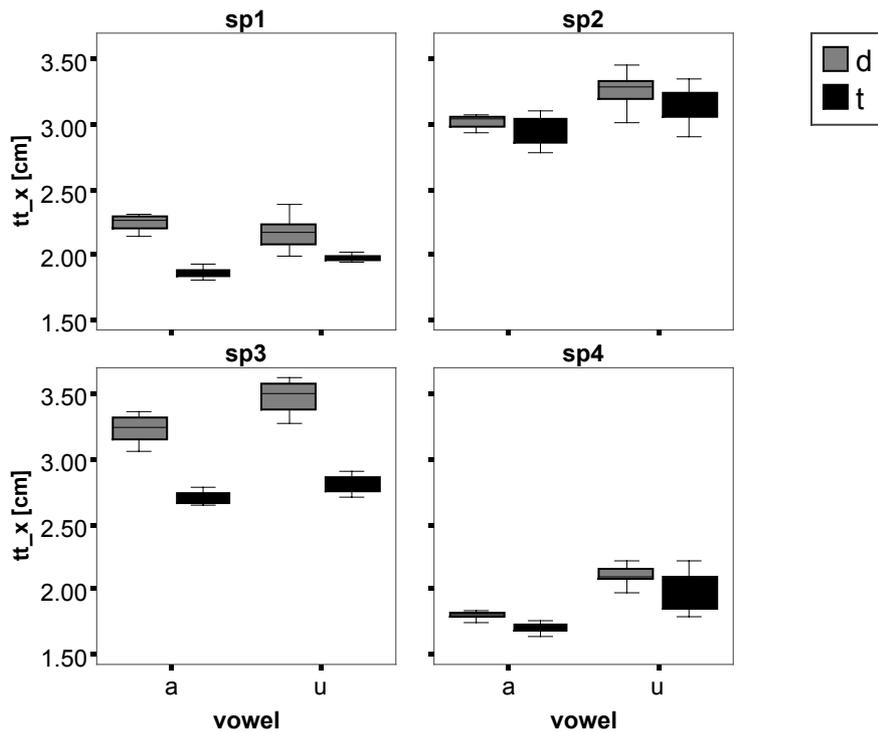
Figure 2: Boxplots with standard deviations for tongue tip horizontal target position for /d/ (grey) and /t/ (black) for the 4 speakers and /a, u/- contexts; lower values = more front articulation.

On the basis of the EMA data in figure 2, we can only specify the position of the tongue tip by its flesh point marker in the mid-sagittal plane, and conclude from it the place of articulation. Conclusions on the actual size of contact can only be gained from EPG data.

4.2.2   Area of contact for /d/

In figure 3, we see EPG frequency plots, which show the pattern of the tongue touching the palate at the consonantal target over all repetitions. The four columns in figure 3 below correspond to the four subjects, the four rows to /at/, /ad/, /ut/, and /ud/. The highest y-value corresponds to the most anterior row at the EPG palate and the lowest y-value to the most posterior row.
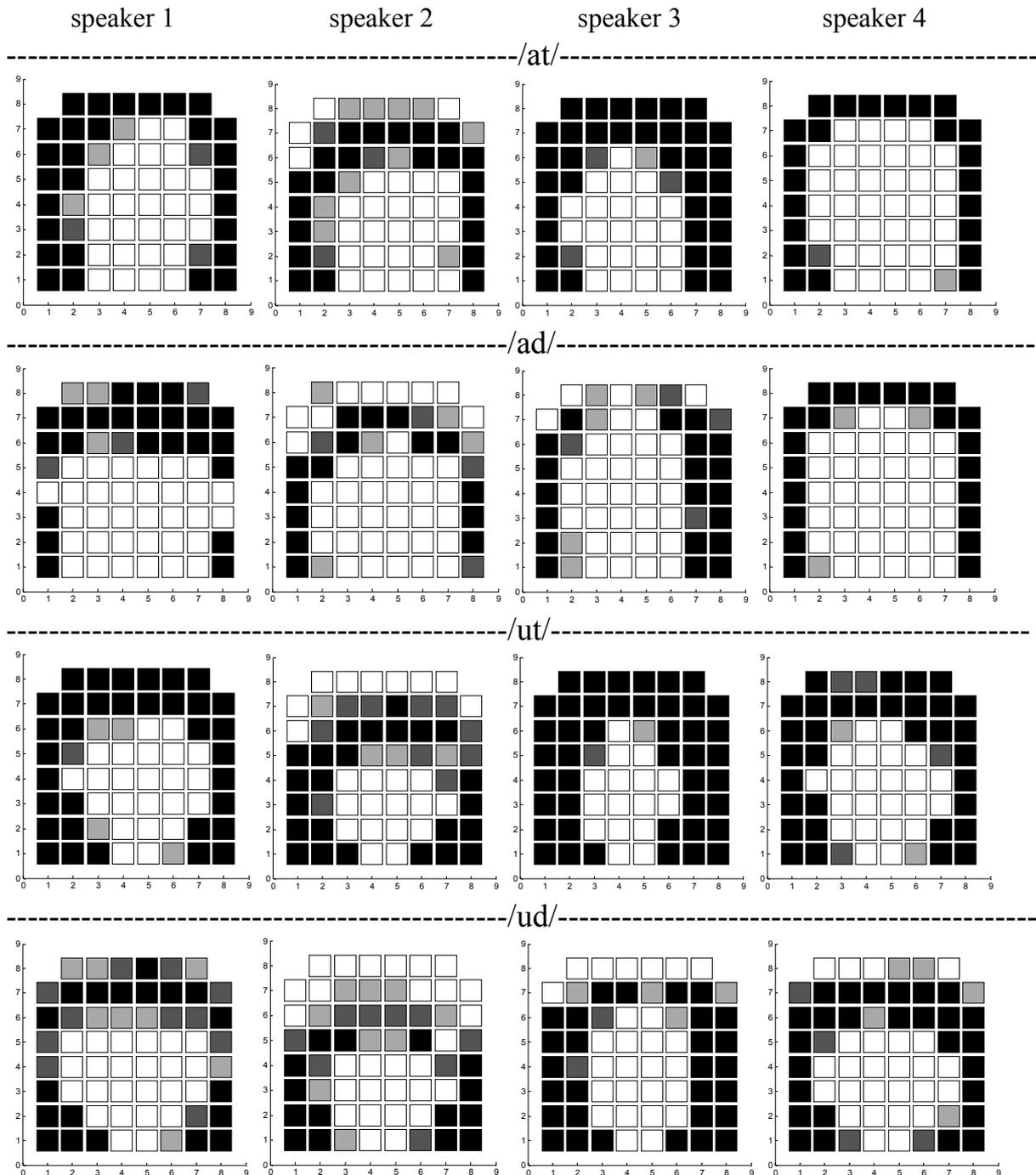
Figure 3: EPG frequency plots for all speakers (4 columns) in /a/-context (first 2 rows) and /u/-context (last 2 rows); /t/ = 1st and 3rd row, /d/ = 2nd and 4th row; black markers correspond to 76-100% tongue palatal contact with respect to all the subject's repetitions, dark grey markers to 51-75%, light grey to 26-50%, and white markers to 0-25%.

The EPG data in figure 3 shows that /d/ has generally a more retracted place of articulation than /t/, both in /a/- and /u/-context. The only exception is speaker 4, who shows equal amount of contact in /a/-context. This might be due to the fact that speaker 4 has a very fronted articulation in this context,

possibly with dental contact, which cannot be recorded with EPG. The percentage of contact over the whole palate for all speakers is significantly greater for /t/ than for /d/, and we can see that /d/ is often produced with less lateral contacts than /t/. Both findings can be interpreted as a more forceful articulation of /t/ (with a higher target position) and a difference in active articulator between the two (where the voiced stop is being articulated with the tongue tip and the voiceless one with the tongue blade). Furthermore, the contact for /d/ shows more variation than that for /t/ (see Appendix II for all descriptive statistics and significances).

Figure 4 displays the EPG frequency plots for the /i/-context. Although there are still subtle differences between /t/ (top row) and /d/ (bottom row), the overall amount of tongue palatal contact is very large for both, especially at the lateral margins of the palate.
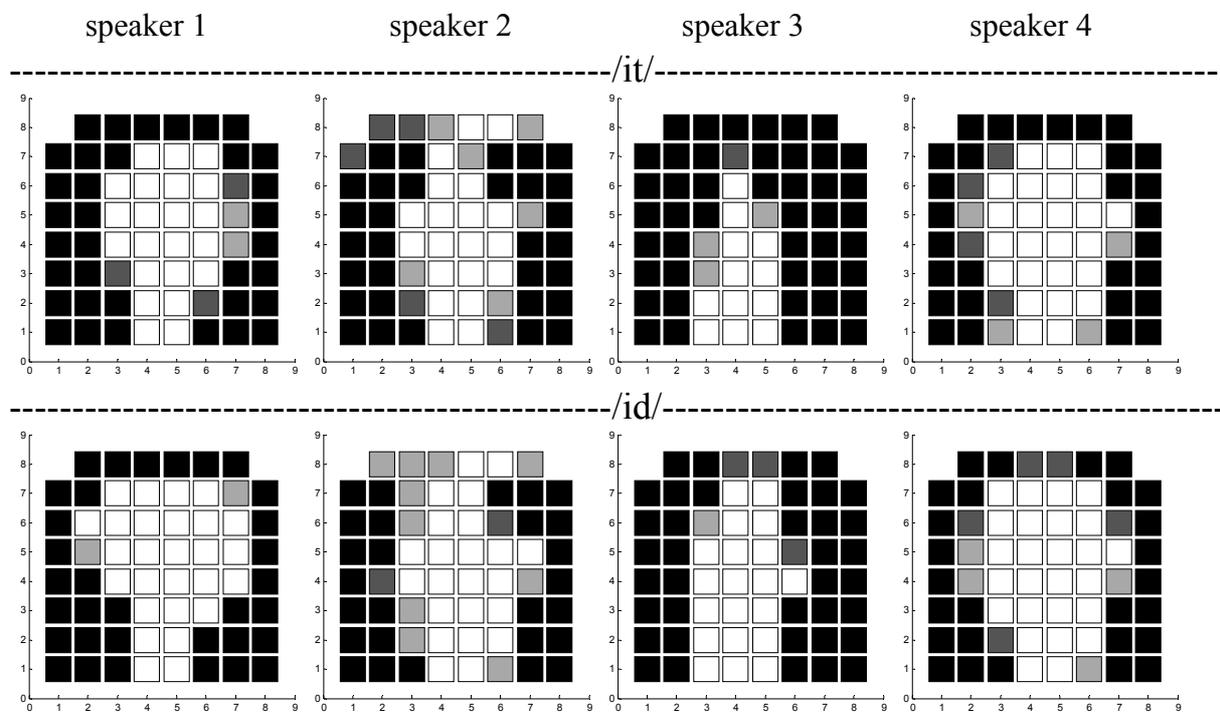
Figure 4: Same as previous figure, but for /i/-context. Note that for speaker 2 only 3 repetitions of /t/ could be included since this speaker realized the relevant target word in most cases with a /d/.

4.2.3   Lowering of tongue dorsum and jaw for /d/

A lowered tongue dorsum is a typical property of retroflex segments, and often goes together with a lowering of the jaw, as discussed in section 2.2. To what extent can these properties also be found in /d/ compared to /t/? In a univariate ANOVA we took tongue dorsum y position and jaw y position as

dependent variables and phoneme (/d/ versus /t/) and vowel context (/a, i, u/) as factors (data were split by speaker). The descriptive statistics and significances are given in Appendix III.

For the vertical tongue dorsum position we found main effects of vowel for all speakers, and a main effect of phoneme for speakers 1 and 2. All speakers but speaker 3 show an interaction between the two factors: In /a/-context, /d/ is realized with a lower tongue dorsum position. In /i/-context, both consonants have a similar tongue position (except for speaker 1 who shows a slightly higher dorsum for /d/). In /u/-context, results vary speaker-dependently: speakers 1 and 2 show similar results for /d/ and /t/, speaker 3 shows a higher /t/ than /d/ and speaker 4 the reverse. These findings show that vertical tongue movement is influenced to a large degree by the vowel context: For the articulation of the high back vowel /u/ a raising of the tongue dorsum is necessary, and for the high front vowel /i/ the dorsum is raised along with the necessary raising of the tongue blade.

For the vertical jaw position, we found main effects of vowel for all four speakers, a main effect of phoneme for all but speaker 2, and an interaction of the two only for speaker 1.
/d/ is articulated with a lower jaw than /t/ for three of the four speakers. Considering the actual values, it becomes evident that although significant, jaw differences are often very subtle. The most pronounced differences are consistently found for speaker 1.

From these findings we can conclude that there are obviously speaker-dependent strategies in the use of the tongue and the jaw. Speaker 3 was the only one who did not show a significant tongue lowering for /d/ (in /u/-context), and speaker 2 the only one who did not show a significant jaw lowering for /d/. Thus, whereas some speakers show tongue lowering for /d/, others show jaw lowering, and some show both.

4.2.4 Retroflex tongue configuration for /d/

The tongue tip angle (Tiede et al. 2005) is a measure to separate a retroflex tongue configuration from a bunched one, independent of the actual phonological retroflexivity of the sound. Although our dataset does not include retroflex phonemes, speaker 1 shows consistently a retroflex tongue configuration for /d/ in back vowel context. All other speakers do not show such configurations. The following figure is therefore restricted to an illustration of our findings for speaker 1.
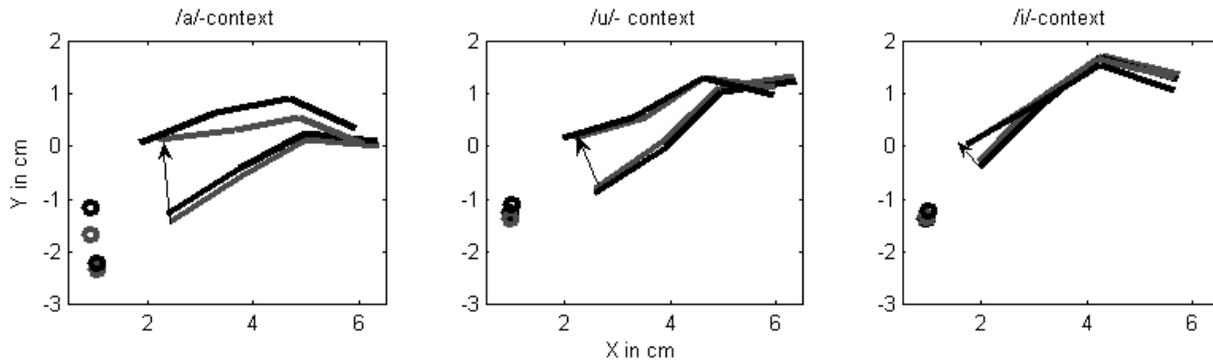
Figure 5: Averaged tongue contours for speaker 1 based on the connection of the 4 flesh point markers at the tongue; x-axis = horizontal movements in cm (left is anterior), y-axis = vertical movements in cm; lower lines = vowel targets, upper lines = consonantal target; little arrows mark the movement from vowel to consonantal targets; grey = /d/, black /t/; from left to right: /a, u, i/-context; empty circles correspond to jaw position.

In figure 5, an arrow indicates the movement from the vowel (lower line) to the consonant (upper line). In /a/-context (left plot) and /u/-context (middle plot), a retroflex tongue configuration can be seen for /d/. Here the tongue tip angle for the consonantal target is below 180 degrees. It is interesting to note that in these cases the tongue tip angle for the preceding vowel target is also below 180 degrees, thus retroflex, but to a far lesser extent (for the /a/-context this is not even visible). In /u/-context, the retroflex-like tongue configuration is also observable for the voiceless stop, and can thus be attributed to the vowel. For /a/-context, however, it is unique to the voiced stop.

In /i/-context (right plot) the tongue tip angle for /d/ is above 180 degrees, and we can see no retroflex tongue configuration in the average tongue contours. Here, the overall movement is very small and very close to the palate.

4.3 Discussion

Our data show that there is a systematic difference between /d/ and /t/ for all four speakers of German. This difference is mainly restricted to the context of /a/ and /u/, where /d/ has a smaller constriction and less lateral contacts (both can be interpreted as an apical articulation), shows more variation in the location of this constriction, and also has a more retracted place of articulation than /t/. Furthermore, three of the four speakers showed a small but significant difference in jaw position, with /d/ having a lower jaw, independent of context. These findings are in accordance with the literature on the difference between /d/ and /t/, recall the discussion in section 2.1, and indicate that the voiced alveolar stop in German is realized in a way that favours retroflexion.

In /i/-context, we could not observe significant differences between /d/ and /t/ in any of these parameters apart from jaw position. This influence of

vowel context coincides with previous observations that retroflex tongue configurations avoid /i/-context and that /u/ context leads to retroflexion, see the discussion in section 2.2.

The expected lowering of the tongue dorsum could only be found in the context of /a/ (for all but speaker 3). This is because the tongue dorsum plays an integral part in the articulation of non-low and back vowels, and if these vowels are adjacent to coronal consonants, they seem to influence the position of the dorsum to a large degree.

For one speaker (speaker 1) we actually found a tongue configuration for /d/ in /a/ and /u/-context that was distinctively retroflex, with a raised tongue tip, a lowered tongue mid, and a raised tongue back. We also found such a configuration for /t/, but only in /u/-context. We conclude from this that retroflexion in /u/-context is independent of the retroflexion of voiced coronal stops, since it influences both voiced and voiceless stop.

In sum, our data illustrate not only that German /d/ is articulated in a way that favours retroflexion, but that retroflexion is an acceptable articulation of /d/ in German, a language that does not have any other voiced coronal stops.

5   Conclusion

In this study we looked at the question whether voiced retroflex stops can develop from front coronal voiced stops. Our aim was to illustrate that this development is possible without an intermediate stage of implosion, contrary to what has been proposed by Haudricourt (1950), Greenberg (1970), Bhat (1973), and Ohala (1983). In addition, we proposed and tested a possible phonetic motivation for this process, namely the articulatory affinity between voiced front coronal stops and voiced retroflex stops.

For these purposes, we provided data from sound changes that introduced voiced retroflex stops as single retroflex in a language from a front coronal stop. Our diachronic examples came from Central Malayo-Polynesian (Dhao), Sino-Tibetan (Thulung), and East Cushitic (Afar, Somali, Rendille). *d is the proto-segment proposed in the existing literature as corresponding to the present-day retroflexes for one language group (Central Malayo-Polynesian). For the other two groups a preglottalized stop is reported, either voiced (East Cushitic) or voiceless (Sino-Tibetan), though the latter has not been given any motivation for the change in voicing (or the drop in glottalization, for that matter).

Furthermore, we conducted an EPG and EMA experiment with four speakers of German to compare the articulation of voiced and voiceless coronal stops. We found that German /d/ shows a more retracted and more variable place of articulation, a smaller percentage of tongue palatal contact patterns, and a lower tongue and jaw position than its voiceless counterpart /t/, especially in

the context of low and back vowels. All these criteria are also used to distinguish retroflex from non-retroflex coronal articulations in languages like Norwegian, Hindi or Tiwi, and support our hypothesis that alveolar and retroflex articulations are similar to each other and can be said to form an articulatory continuum without a sharp boundary. One of our speakers actually produced retroflex [ɖ] as a realization of the voiced alveolar stop phoneme in low and back vowel context.

The articulatory data thus provides us with a phonetic explanation for the typological diachronic findings, an approach that has been forwarded in linguistic theory over the last decades (see, e.g., Ohala 1993 2005; Blevins 2004; 2007). Our study supplements the work by Haudricourt and following on the development of retroflexes from implosives. But whereas the explanations that were provided for the change from implosive to retroflex (such as cavity enlargement, see Ohala 1983) implied a strict direction of sound change (they could not account for the reverse process), the articulatory similarity we propose holds for processes in both directions.

It seems as if the process of retroflexion via implosion would benefit too from an explanation that does not imply a preferred direction, as there is evidence for reverse processes. We gave two potential examples of languages in which a retroflex might have become implosive. In Sabu and Dhao, /ɗ/ is likely to stem from a reconstructed *ɖ in Proto-Malayo-Polynesian (or Proto-Austronesian), and /ɗ/ in Boni might stem from *ɖ in Proto-Sam (as proposed by Heine 1978). An existing example for such a process is the development of Saramaccan, a creole language of Surinam, which has a voiced coronal implosive (Bakker, Smith & Veenstra 1995) which stems from the retroflex voiced stop in the lexical contributor language Fon and other closely related Gbe languages (Smith & Haabo 2007)[8].

All these processes can be acounted for by the articulatory-similarity space proposed in section 2.3, which is based on two gradual continua: One from plain stop to stop with rapidly and strongly lowered larynx (implosive), as proposed by Ladefoged (1964 et seq.), and one from plain tongue shape to retroflex tongue shape (independent of the actual place of articulation) that we proposed in section 2.2. The two continua are orthogonal to each other and create a two-dimensional space, including a gradient continuum from plain to implosive retroflex. Any change from one voiced coronal segment to another within this space is simply due to articulatory similarity (either on one of the two

---

[8] Smith and Haabo (2007) find the circular diachronic development of Proto-Volta-Congo *ɗ to Proto-Gbe/Fon ɖ to Saramaccan ɗ "unexpected" (p.17) and propose that the implosive articulation of this sound has been continous. This contrasts with the general assumption that Proto-Gbe had a retroflex (Capo 1991).

dimensions or on both). The similarity does not imply a preferred direction of change. We have to be aware of the fact that ingressive airflow is not included in this space, though it is a defining criterion of implosives. Future research has to show whether the difference in airflow (from egressive to ingressive) forms a separate dimension and thus makes our similarity space three-dimensional, or whether it correlates with the existing dimension of plain to retroflex implosives.

6   Acknowledgements

7   References

Andersen, Torben (1987). An outline of Lulubo phonology. *Studies in African Linguistics* 18, 39-66.

Anderson, Victoria Balboa & Ian Maddieson (1994). Acoustic characteristics of Tiwi coronal stops. *UCLA Working Papers in Phonetics* 87, 131-162.

Arndt, Paul P. (1933). *Grammatik der Ngad'a-Sprache*. Bandoeng: A.C. Nix & Co.

Baird, Louise (2002). Kéo (Illustrations of the IPA). *Journal of the International Phonetic Association* 32, 93-97.

Bakker, Peter, Norval Smith & Tonjes Veenstra (1995). Saramaccan. In: Jacques Arends, Peter Muysken & Norval Smith (eds.) *Pidgins and creoles: an introduction*. Amsterdam: Benjamins.

Bell-Berti, Fredericka (1975). Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America* 57, 456-461.

Bender, M. Lionel (1992). Central Sudanic segmental and lexical reconstructions. *Afrikanische Arbeitspapiere* 29, 5-61.

Bhat, D. N. S. (1973). Retroflexion: an areal feature. *Working Papers on Language Universals* 13, 27-67.

Bhat, D. N. S. (1974). Retroflexion and retraction. *Journal of Phonetics* 2, 233-237.

Blevins, Juliette (2004). *Evolutionary Phonology*. Cambridge: Cambridge University Press.

Blevins, Juliette (2007). The importance of typology in explaining recurrent sound patterns. *Linguistic Typology* 11, 107-113.

Bliese, Loren F. (1981). *A generative grammar of Afar*. Dallas: Summer Institute of Linguistics.

Capo, Hounkpati B.C. (1991). *A comparative phonology of Gbe*. Berlin: Foris Publications.

Catford, John C. (1939). On the classification of stop consonants. *Le Maître Phonétique* 65, 2-5.

Clements, George N. & Sylvester Osu (2003). Explosives, implosives, and nonexplosives: the linguistic function of air pressure differences in stops. In: Carlos Gussenhoven & Natasha Warner (eds.) *Papers in Laboratory Phonology 7*. Berlin: Mouton de Gruyter, 299-350.

Clements, George N. & Annie Rialland (2005). *Africa as a phonological area*. Manuscript, Paris.

Dart, Sarah N. (1991). Articulatory and acoustic properties of apical and laminal articulations. *UCLA Working Papers in Phonetics* 79, 1-155.

Dart, Sarah N. (1998). Comparing French and English coronal consonant articulation. *Journal of Phonetics* 26, 71-94.

Demolin, Didier & Didier L. Goyvaerts (1986). Some aspects of Madi phonology and morphology. *Antwerp papers in Linguistics* 44, 89-103.

Demolin, Didier (1988). Some problems of phonological reconstruction in Central Sudanic. *Belgian Journal of Linguistics* 3, 53-96.

Dempwolff, Otto (1934). Vergleichende Lautlehre des Austronesische Wortschatzes, Band 1: Induktiver Aufbau einer indonesischen Ursprache. *Beihefte zur Zeitschrift für Eingeborenen-Sprachen* 15.

Dimmendaal, G.J. (1986). Language typology, comparative linguistics and injective consonants in Lendu. *Afrika und Übersee* 69, 161-192.

Dixit, R. Prakash (1990). Linguotectal contact patterns in the dental and retroflex stops of Hindi. *Journal of Phonetics* 18, 189-201.

Dixit, R. Prakash & James Flege (1991). Vowel context, rate and loudness effects of linguopalatal contact patterns in Hindi retroflex /ˇ /. *Journal of Phonetics* 19, 213-229.

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

Driem, George L. van (1987). *A grammar of Limbu*. Berlin: Mouton de Gruyter.

Dyen, Isidore (1971). The Austronesian languages and Proto-Austronesian. In: Thomas A. Sebeok (ed.) *Linguistics in Oceania*. The Hague: Mouton, 5-54.

Ebert, Karen (1994). *The structure of Kiranti languages: comparative grammar & texts*. Zürich: Universität Zürich.

Ebert, Karen (2003). Kiranti languages: An overview. In: Graham Thurgood & Randy L. LaPolla (eds.) *The Sino-Tibetan languages*. London: Routledge, 505-517.

Emeneau, M. B. (1944). *Kota texts. Part 1*. Berkeley & Los Angeles: University of California Press.

Esser, S. J. (1938). Talen. *Altas van Tropisch Nederland*. Amsterdam: Koninklijk Nederlandsch Aardrijkskundig Genootschap.

Ewan, William G. & R. Krones (1974). Measuring larynx movement using the thyroumbrometer. *Journal of Phonetics* 2, 327-335.

Farnetani, Edda (1989). An articulatory study of "voicing" in Italian by means of dynamic palatography. Paper presented at the *Speech Research International Conference*. Budapest.

Farnetani, Edda (1990). V-C-V lingual coarticulation and its spatiotemporal domain. In: William J. Hardcastle & Alain Marchal (eds.) *Speech production and speech modelling*. Dordrecht: Kluwer, 93-130.

Flemming, Edward (2003). The relationship between coronal place and vowel backness. *Phonology* 20, 335-373.

Fox, James J. (2004). Current Developments in Comparative Austronesian Studies. Paper presented at the *Symposium Austronesia*. Bali.

Fuchs, Susanne, Pascal Perrier & Christine Mooshammer (2001). The role of the palate in tongue kinematics: An experimental assessment in VC sequences from EPG and EMMA data. Paper presented at the *Eurospeech*. Aalborg.

Fuchs, Susanne (2005). Articulatory correlates of the voicing contrast in alveolar obstruent production in German. *ZAS Working Papers in Linguistics* 41.

Fuchs, Susanne, Pascal Perrier, Christian Geng & Christine Mooshammer (2006). What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruents. In: Jonathan Harrington & Marija Tabain (eds.) *Towards a better understanding of speech production processes*. New York: Psychology Press, 149-164.

Fujimura, Osamu, I.F. Tatsumi & Ryohei Kagaya (1973). Computational processing of palatographic patterns. *Journal of Phonetics* 1, 47-54.

Fujimura, Osamu & J.E. Miller (1979). Mandible height and syllable-final tenseness. *Phonetica* 36, 263-272.

Gordon, Raymond G. Jr. (2005). *Ethnologue: Languages of the World*. Dallas: SIL International.

Goyvaerts, Didier L. (1986). Glottalized consonants: a new dimension? *Antwerp papers in Linguistics* 44, 105-113.

Gragg, Gene (1976). Oromo of Wellegga. In: M. Lionel Bender (ed.) *The Non-Semitic Languages of Ethiopia*. East Lansing, Michigan: Michigan State University Press, 166-221.

Greenberg, Joseph H. (1970). Some Generalizations concerning Glottalic Consonants, Especially Implosives *International Journal of American Linguistics* 36, 123-145.

Grimes, Charles E. (2006). Hawu and Dhao in eastern Indonesia: revisiting their relationship. Paper presented at the *10th International Conference on Austronesian Linguistics*. Puerto Princessa, Philippines.

Guthrie, Malcolm (1967-1971). *Comparative Bantu*. Farnborough, Hants: Gregg International.

Hamann, Silke (2002). Retroflexion and retraction revised. *ZAS Working Papers in Linguistics* 28, 13-25.

Hamann, Silke (2003). *The Phonetics and Phonology of Retroflexes*. Utrecht: LOT.

Hamann, Silke (2004). Retroflex fricatives in Slavic Languages. *Journal of the International Phonetic Association* 34, 53-67.

Haudricourt, André-Georges (1950). Les consonnes préglottalisées en Indochine. *Bulletin de la Société de Linguistique* 46, 172-182.

Heine, Bernd (1978). The Sam languages: a history of Rendille, Boni and Somali. *Afroasiatic Linguistics* 6.

Jonker, Johan Christoph Gerard (1896). Bimaneesche Spraakkunst. *Verhandelingen van het Bataviaasch Genootschap van Kunsten en Wetenschappen (VBG)* 48.

Jonker, Johan Christoph Gerard (1903). Iets over de taal van Dao. *Album Kern (Opstellen geschreven ter eere van Dr. H. Kern)*. Leiden: Brill, 85-89.

Keating, Patricia A. (1984). Phonetic and phonological representation of stop consonant voicing. *Language* 60, 286-319.

Khatiwada, Rajesh (2007). Nepalese retroflex stops: a static palatography study of inter- and intra-speaker variability. Paper presented at the *Interspeech*. Antwerp.

Klamer, Marian (1998). *A Grammar of Kambera*. Berlin: Mouton de Gruyter.

Krull, Diana, Björn Lindblom, B.-E. Shia & D. Fruchter (1995). Cross-linguistic aspects of coarticulation: an acoustic and electropalatographic study of dental and retroflex consonants. *Proceedings of the XIIIth International Congress of Phonetic Sciences (ICPhS95, Stockholm)* 3, 436-439.

Ladefoged, Peter (1964). *A Phonetic Study of West African Languages: An Auditory-Instrumental Survey*. Cambridge: Cambridge University Press.

Ladefoged, Peter & Peri Bhaskararao (1983). Non-quantal aspects of consonant production: a study of retroflex consonants. *Journal of Phonetics* 11, 291-302.

Ladefoged, Peter & Zongji Wu (1984). Places of articulation: an investigation of Pekingese fricatives and affricates. *Journal of Phonetics* 12, 267-278.

Ladefoged, Peter & Ian Maddieson (1996). *The Sounds of the World's Languages*. Oxford: Blackwell.

Lahaussois, Aimée (2003). Thulung Rai. *Himalayan Linguistics* 1, 1-25.

Lex, Gloria (2006). *Le Dialecte Peul du Fouladou (Casamance - Sénégal): Etude Phonétique et Phonologique*. München: Lincom.

Lindblom, Björn (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-1781.

Lisker, Leigh & Arthur S. Abramson (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word* 20, 384.

Lloret, Maria-Rosa (1995). Implosive consonants: their representation and sond change effects. *Belgian Journal of Linguistics* 9, 59-72.

Löfqvist, Anders & V.L. Gracco (2002). Control of oral closure in lingual stop consonant production. *Journal of the Acoustical Society of America* 111, 2811-2827.

Maddieson, Ian (1984). *Patterns of Sounds*. Cambridge: Cambridge University Press.

Michailovsky, Boyd (1994). Manner vs. place of articulation in the Kiranti initial stops. In: H. Kitamura & et al (eds.) *Current Issues in Sino-Tibetan Linguistics*. Osaka, 766-772.

Moen, Inger & Hanne Gram Simonsen (1997). Effects of voicing on /t, d/ tongue/palate contact in English and Norwegian. Paper presented at the *5th European Conference on Speech Communication and Technology*. Rhodes.

Mooshammer, Christine, Philip Hoole & Anja Geumann (2006). Inter-articulator cohesion within coronal consonant production. *Journal of the Acoustical Society of America* 120, 1028-1039.

Mooshammer, Christine, Philip Hoole & Anja Geumann (2007). Jaw and order. *Language and Speech* 50, 145-176.

Ohala, John 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.

Ohala, John J. (1993). The phonetics of sound change. In: Charles Jones (ed.) *Historical linguistics: problems and perspectives*. London: Longman, 237-278.

Ohala, John J. (2005). Phonetic explanations for sound patterns: Implications for grammars of competence. In: William J. Hardcastle & M. Beck (eds.) *A figure of speech: A festschrift for John Laver*. London: Erlbaum, 23-38.

Perkell, Joseph S. (1969). *Physiology of speech production: Results and implications of a quantitative cineradiographic study*. Cambridge: MIT Press.

Repp, Bruno (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* 22, 173-189.

Ross, Malcolm (1992). The sounds of Proto-Austronesian: An outsider's view of the Formosan evidence. *Oceanic Linguistics* 31, 23-64.

Ross, Malcolm (1995). Some current issues in Austronesian linguistics. In: Darrell T. Tryon (ed.) *Comparative Austronesian dictionary: An introduction to Austronesian studies*. Berlin: Mouton de Gruyter, 45-120.

Sasse, Hans-Jürgen (1979). The consonant phonemes of Proto-East-Cushitic (PEC): A first approximation. *Afroasiatic Linguistics* 7, 1-67.

Schuh, Russell G. (2003). Chadic overview. In: M. Lionel Bender, Gábor Taka\cs & David L. Appleyard (eds.) *Selected comparative-historical Afrasian linguistic studies. In memory of Igor M. Diakonoff*. München: Lincom, 55-60.

Simonsen, Hanne G., Inger Moen & Steve Cowen (2000). Retroflex consonants in Norwegian: are they really? Evidence from EMA and EPG. Paper presented at the *5th Seminar on Speech Production*. Kloster Seeon, Germany.

Smith, Norval & Vinije Haabo (2007). The Saramaccan implosives: tools for linguistic archaeology? . *Journal of Pidgin and Creole Languages* 22.

Starostin, Sergei Anatolyevich (1994). Proto-Kiranti. Paper presented at the *Sino-Tibetan Conference*. Paris.

Steriade, Donca (2001). Directional asymmetries in place assimilation: a perceptual account. In: Elizabeth Hume & Keith Johnson (eds.) *The Role of Speech Perception in Phonology*. San Diego (CA): Academic Press, 219-250.

Stevens, Kenneth, Samuel Keyser & Haruko Kawasaki (1986). Toward a phonetic and phonological theory of redundant features. In: Joseph S. Perkell & D. H. Klatt (eds.) *Invariance and Variability in Speech Processes*. Hillsdale: Erlbaum, 426-449.

Stewart, John M. (1995). Implosive, homorganic nasals and nasalized vowels in Volta-Congo. In: E. 'Nolue Emenanjo & Ozo-Mekuri Ndimele (eds.) *Issues in African languages and linguistics: Essays in honor of Kay Williamson*. Aba: National Institute for Nigerian Languages, 162-169.

Stroomer, Harry (1987). *A Comparative Study of Three Southern Oromo Dialects in Kenya*. Hamburg: Helmut Buske Verlag.

Švarnyˊ, Oldřich & Kamil Zvelebil (1955). Some remarks on the articulation of the 'cerebral' consonants in Indian languages, especially in Tamil. *Archiv Orientalni* 23, 374-407.

Tiede, Mark K., Vincent L. Gracco, Douglas M. Shiller, Carol Espy-Wilson & Suzanne E. Boyce (2005). Perturbed palatal shape and North American English /r/ production. *Journal of the Acoustical Society of America* 117, 2568-2569.

Tosco, Mauro (2001). *The Dhaasanac Language: Grammar, Texts, Vocabulary of a Cushitic Language of Ethiopia*. Köln: Rüdiger Köppe.

Trask, Robert L. (1996). *A Dictionary of Phonetics and Phonology*. London: Routledge.

Tucker, A. N. & M. A. Bryan (1966). *Linguistic analyses, the non-Bantu languages of northeastern Africa*. Oxford: Oxford University Press.

Wakumoto, Masahiko, Shinobu Masaki, Kiyoshi Honda & Toshikazu Ohue (1998). A pressure sensitive palatography: application of new pressure sensitive sheet for measuring tongue-palatal contact pressure. *Proceedings of the 5rd International Conference on Spoken Language Processing*. Sydney, 3151-3154.

Watson, Richard L. (1991). Moru-Ma'di orthographies. In: M. Lionel Bender (ed.) *Proceedings of the Fourth Nilo-Saharan linguistics colloquium (Nilo-Saharan: Linguistic analysis and documentation Vol. 7)*. Hamburg: Helmut Buske Verlag, 273-282.

Westbury, John R. (1983). Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America* 73, 1322-1336.

Wolff, John (1974). Proto-Austronesian *r and *d. *Oceanic Linguistics* 13, 77-121.

Wolff, John (1991). The Proto-Austronesian phoneme *t and the grouping of the Austronesian languages. *Pacific Linguistics (Series C)* 117, 535-549.

Appendix I: Descriptive statistics for horizontal position of tongue tip at the consonantal target

subject	vowel	/d/			/t/		
		N	Mean	Std. dev.	N	Mean	Std. dev.
speaker 1	a	10	2.22	0.12	10	1.86	0.04
	i	10	1.76	0.05	10	1.72	0.02
	u	10	2.17	0.13	10	1.98	0.03
speaker 2	a	10	3.03	0.08	10	2.95	0.11
	i	10	2.67	0.06	4	2.72	0.02
	u	10	3.26	0.14	9	3.13	0.15
speaker 3	a	10	3.23	0.10	10	2.72	0.05
	i	10	2.93	0.11	10	2.73	0.10
	u	10	3.48	0.13	10	2.80	0.06
speaker 4	a	10	1.80	0.03	9	1.70	0.06
	i	10	1.65	0.08	10	1.69	0.10
	u	10	2.12	0.10	10	2.00	0.14

subject	factor(s)	df	F	P
speaker 1	vowel	2	113.04	0.000
	phoneme	1	94.01	0.000
	vowel * phoneme	2	22.11	0.000
speaker 2	vowel	2	75.67	0.000
	phoneme	1	3.04	0.088
	vowel * phoneme	2	2.63	0.083
speaker 3	vowel	2	55.25	0.000
	phoneme	1	370.44	0.000
	vowel * phoneme	2	33.51	0.000
speaker 4	vowel	2	99.86	0.000
	phoneme	1	6.62	0.013
	vowel * phoneme	2	4.24	0.020

Appendix II: Descriptive statistics and univariate ANOVA for overall percent of tongue palatal contact patterns at the consonantal target

subject	vowel	/d/		/t/	
		Mean	**Std. dev.**	**Mean**	**Std. dev.**
speaker 1	a	40.32	4.30	50.81	4.45
	i	52.90	6.49	62.26	4.31
	u	44.84	13.14	59.03	3.82
speaker 2	a	28.23	11.91	45.48	6.62
	i	52.10	11.78	58.87	3.36
	u	38.55	7.03	49.28	7.04
speaker 3	a	40.48	6.24	62.10	3.25
	i	64.84	6.17	75.97	4.13
	u	50.65	4.70	68.23	14.43
speaker 4	a	37.90	1.90	38.71	2.42
	i	52.15	5.23	54.84	6.76
	u	50.97	6.04	56.77	6.49

subject	factor(s)	df	F	P
speaker 1	vowel	2	15.15	0.000
	phoneme	1	40.45	0.000
	vowel * phoneme	2	0.67	n.s.
speaker 2	vowel	2	15.64	0.000
	phoneme	1	19.68	0.000
	vowel * phoneme	2	1.37	n.s.
speaker 3	vowel	2	32.97	0.000
	phoneme	1	75.63	0.000
	vowel * phoneme	2	2.51	n.s.
speaker 4	vowel	2	55.38	0.000
	phoneme	1	5.10	0.028
	vowel * phoneme	2	1.14	n.s.

Appendix III: Descriptive statistics and univariate ANOVA for vertical position (in cm) of tongue dorsum (td) and jaw at the consonantal target

| tongue dorsum | | /d/ | | | /t/ | | |
subject	vowel	N	Mean	Std. dev.	N	Mean	Std. dev.
speaker 1	a	10	0.56	0.10	10	0.92	0.11
	i	10	1.67	0.04	10	1.56	0.08
	u	10	1.30	0.07	10	1.32	0.09
speaker 2	a	10	0.11	0.09	10	0.30	0.15
	i	10	0.66	0.08	4	0.68	0.15
	u	10	0.43	0.08	9	0.47	0.08
speaker 3	a	10	0.93	0.23	10	0.92	0.06
	i	10	1.47	0.11	10	1.50	0.07
	u	10	1.35	0.06	10	1.43	0.06
speaker 4	a	10	0.12	0.07	9	0.23	0.06
	i	10	1.02	0.11	10	0.98	0.04
	u	10	0.81	0.06	10	0.72	0.06

| jaw | | /d/ | | /t/ | |
subject	vowel	Mean	Std. dev.	Mean	Std. dev.
speaker 1	a	-1.67	0.09	-1.15	0.06
	i	-1.35	0.06	-1.22	0.06
	u	-1.36	0.09	-1.11	0.07
speaker 2	a	-1.36	0.03	-1.32	0.06
	i	-1.27	0.03	-1.26	0.02
	u	-1.23	0.03	-1.22	0.02
speaker 3	a	-0.94	0.03	-0.93	0.01
	i	-0.97	0.04	-0.93	0.02
	u	-0.93	0.03	-0.91	0.02
speaker 4	a	-1.10	0.05	-1.01	0.05
	i	-1.09	0.03	-1.05	0.03
	u	-0.98	0.02	-0.94	0.02

subject	factor(s)	df	F (td)	P (td)	F (jaw)	P (jaw)
speaker 1	vowel	2	555.05	0.000	31.04	0.000
	phoneme	1	18.18	0.000	256.12	0.000
	vowel * phoneme	2	41.00	0.000	37.95	0.000
speaker 2	vowel	2	78.71	0.000	54.25	0.000
	phoneme	1	7.64	0.008	3.04	n.s.
	vowel * phoneme	2	3.60	0.035	0.80	n.s.
speaker 3	vowel	2	134.16	0.000	5.70	0.006
	phoneme	1	1.44	n.s.	9.35	0.003
	vowel * phoneme	2	0.73	n.s.	1.19	n.s.
speaker 4	vowel	2	755.37	0.000	53.16	0.000
	phoneme	1	0.15	n.s.	41.43	0.000
	vowel * phoneme	2	11.23	0.000	3.01	n.s.

The representation of turbulent sounds in German: a government approach

Anna Bloch-Rozmej
John Paul II Catholic University of Lublin, Poland

1   Introduction

The unfolding discussion will focus on the internal representation of turbulent sounds in the phonology of German as well as pinpoint the special status of the prime defining the quality of turbulence. It will also be argued that this prime is capable of entering into special types of licensing relations, which results in specific phonetic manifestations of forms. We shall compare the effects of two processes attested in German: consonant degemination and spirantisation with a view to revealing the role of the turbulence-defining element in the two operations. Furthermore, our attention will be focused on the workings of the *Obligatory Contour Principle* which, as will be shown below, exerts decisive impact on prime interplay and consequently the phonetic realization of sounds and words. We shall see that segmental identity is contingent on the language-specific interpretation of inter-element bonds.

Aware of the importance of prime autonomy in determining the manifestation of sounds, let us start with a brief outline of the fundamental segment structure principles offered by the theory of *Phonological Government*.

2   Elements and segments in Government Phonology

Within the framework of *Government Phonology* (Kaye, Lowenstamm and Vergnaud (1985; 1990), Harris (1994), or Gussmann (2002)), the representation of segments is strictly dependent on the number, type and configuration of phonological primes. Put differently, the melodic plane constitutes a hierarchical multi-tiered construct, where each autonomous element, residing on its own tier, is directly associated with a skeletal position. The establishment of an association line between a prime and a slot results in the phonetic manifestation of the element. We shall illustrate this situation with a structure of vocalic expressions as proposed in Harris (1994, 101):

(1) *The melodic plane*
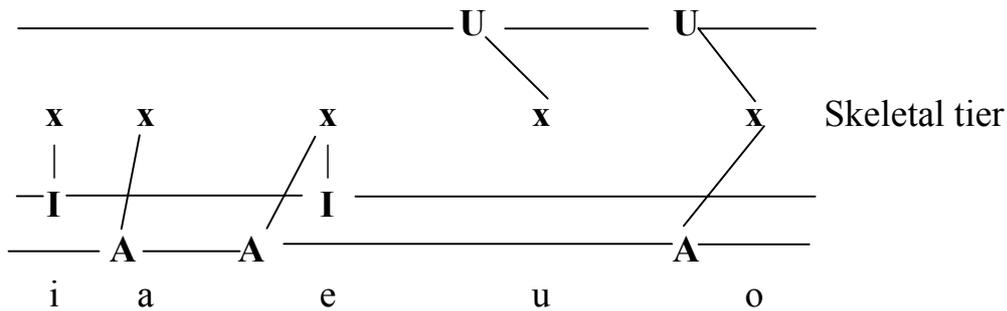
As depicted in (1), a melodic expression can consist of one or more elements. *Government Phonology* proposes that primes constituting melodic entities should correspond to identifiable phonetic properties, each having its own unique interpretation at any level of derivation. In the case of multi - element combinations, the phonetic outcome is contingent on the properties contributed by their constituent members. Below in (2), a list of phonological primes is supplied together with their salient properties.

(2) <u>Element</u> <u>property</u>

A	low (in vowels), coronality (in consonants)
I	high
U	round
?	occlusion (abrupt and sustained decrease in overall amplitude)
h	noise (aperiodic energy)
L	slack vocal cords / low tone
H	stiff vocal cords / high tone
R	coronality[1]
N	nasality[2]
@	velarity[3]

It needs to be observed that whenever a given segment is defined by means of more than one prime, one of the elements can perform the function of the expression head, whereas the others are operators / dependents. The head prime

---

[1] Recently, the necessity to recognise the element **R** has been undermined in various studies. However, it is used both in KLV (1985) and Harris (1994). In the present work, we follow authors who represent coronality with **A**.

[2] In recent phonological studies attempts are being made to replace the nasal element with the low tone element **L**, or rather fuse the two elements, for example Ritter (1997).

[3] In fact, most *GP*-based analyses today have abandoned the neutral element and represent velarity as an empty head.

specifies the salient property of the melody. In terms of its formal capabilities, the head licenses all the remaining primes constituting a segment. The combinatorial capabilities of elements are constrained by the so-called licensing constraints which are language-specific.

Furthermore, individual elements are organised under class nodes: the Root, Place/Resonance and Laryngeal. Consequently, a phonological process can involve either a single prime or target the whole class node, thus affecting all the elements gathered under it. The structure in (3) shows the geometric configuration of primes that potentially build phonological expressions (Harris (1994, 129)):

(3)

```
                        x
                        | ⁄ ?
            Root  ──── • ──── h
     Laryngeal  •           N
                          •        Place
               H  L
                         A  I  U   (R)
```
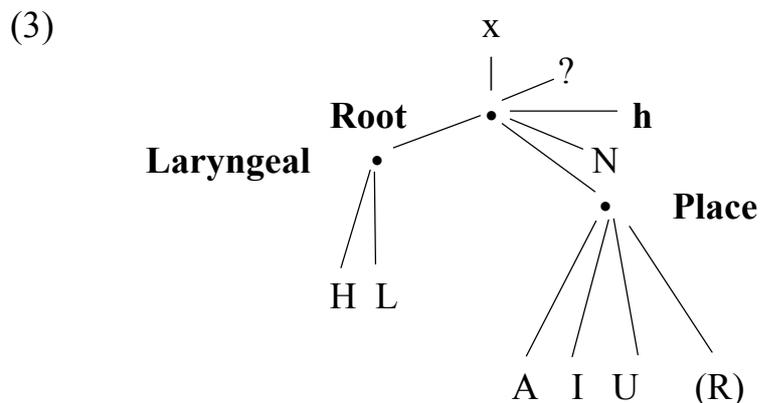
As mentioned above, each single prime corresponds to a unique, identifiable phonetic reality and when uttered in isolation, possesses its own unique manifestation. Below, a brief summary of the *GP* proposals concerning the nature of elements and segments is presented:

❖   Elements are autonomous
❖   Segments are composed of elements
❖   Elements are fully interpretable at any stage of derivation
❖   Elements have unique phonetic manifestations
❖   Each element, when licensed, is directly linked to a skeletal position
❖   Elements can enjoy either head or operator status
❖   Elements are phonetically manifested once attached to a skeletal slot.

2.1   The representation of turbulence

Turning now to turbulent sounds, the basic prime that will concern us in their representation will be the **h** whose elemental pattern is defined as **noise**. Harris (1994, 123) explains that '*it maps onto the speech signal as aperiodic energy. The articulatory targeting of this effect involves a narrowed stricture which produces turbulent airflow.*' The **h** prime is responsible for adding a noise

component to obstruents, whereas sonorants, characterised by periodicity, will lack it. In fricatives and affricates, noise is commonly recognised as phonologically significant. However, the evidence concerning consonant lenition proves that **h** is phonologically active also in stops, characterising the release stage of genuine plosives. When manifested phonetically in isolation, the noise yields a glottal fricative. When paired with a place-defining prime, the latter will point to the location of the noise-producing gesture. Consequently, [s], for instance, will arise due to the combination of **h** with the resonance element **A** defining coronality, while [f] will be a fusion of **h** with **U**.

Relying on the above insights, in (4) below the representations of German voiceless fricatives are depicted (partly based on Brockhaus (1995, 110)):

(4)[4]   [f]   [s]   [ʃ]   [χ]
          x   x   x   x
          | | | |
          U   R/A   I   @
          | | | |
          h   **h**   **h**   **h**
          | | | |
          H   H   H   H

The truly voiced series will contain the 'slack vocal cords' element **L** instead of **H**. The process of *Final Obstruent Devoicing* attested in the phonology of German will be represented as **L**-delinking, thus producing a neutral series without any laryngeal element.[5]

As regards affricate sounds, Harris (1994) treats them as contour structures, where a single melodic expression possesses one resonance node (element) and two root nodes, dominating the elements of occlusion and noise. This representation is provided in (5) below and can be applied to the structure of German affricates.

---

[4]   Recall that in the present study, **R** has been replaced with **A**, whereas **@** abandoned altogether.

[5]   For an extensive discussion of the *FOD* phenomenon in the phonology of German, see Brockhaus (1995).

(5)

[ts]
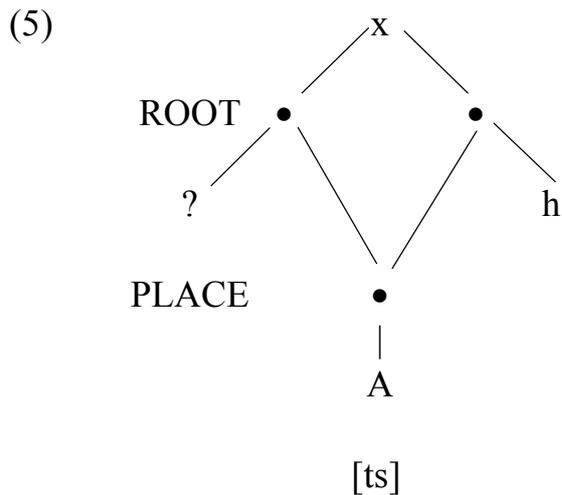
Thus, the availability of two independent manner elements: occlusion and noise, each attached to a separate root node, will characterise the structure of all German affricates.

Keeping the above theoretical considerations in mind, let us turn to the processes of consonant degemination and the spirantisation of [g] in Modern Standard German which, on the one hand, point out to the necessity of recognizing the presence of the noise element in the representation of obstruents, and on the other, bring out the difference between fricatives and affricates in the system of German.

3   Consonantal identity and the OCP effects

As far as consonant degemination is concerned, the process involves the identity of the adjacent segments but not necessarily strict identity. Consequently, sequences of two [t]s or two [d]s will undergo reduction as readily as those of [s]+[z] or [d]+[t]. To illustrate the process of degemination, a number of examples are provided in (6):[6]

(6)	(er) rät	[RE:t]	'he advises'
	(er) lädt	[lE:t]	'he loads'
	(du) liest	[li:st]	'you read'
	Nam+en+n >	Nam+en	'name/dat.pl.'
	Orden+n >	Orden	'order/dat.pl.'
	Esel+lein >	Eselein	'donkey/dim.'

---

[6]   The evidence has been taken from Wiese (1996, 230-231). Following Wurzel (1970) and Lieber (1987), Wiese (1996, 230) assumes that 'all nouns in their plural form (except for those plural-marked by /-s/) receive a suffix /-n/ as a marker for the dative case'. Thus, in *Namen*, for instance, the first [n] signifies plurality, whereas the second one the dative case.

wahl+los   'indiscriminately'
un+nahbar   'unapproachable'

The handful of illustrative examples supplied above are those where degemination is either an obligatory or preferred outcome. Below in (7), we present a list summarising degemination-displaying contexts (based on Wiese (1996)):

(7) The contexts of degemination
  a. in verbal inflection, in a sequence of identical coronals
  b. in noun inflection, in a sequence created by the dative marker /n/ and [n]-ending nouns
  c. in the suffixation of the diminutive /-lein/ to a syllable ending in [l] and whose nucleus dominates schwa
  d. in compounds and prefixation in a sequence of identical consonants

What needs to be added at this point is that the suffixation of a verbal ending opening with a segment identical with that of the root does not always exhibit the degemination effect. For a change, in some verbal forms, schwa epenthesis is attested. Consider forms such as:

(8) (er) wartet [vartWt] 'he waits'
  (er) endet   [EndWt] 'he finishes'
  (sie) berichtet [bW}rç'tWt] 'she informs'

Interestingly, in the case of affricates, consonant reduction will affect only those affricate-final verbs which are followed by a suffix beginning with [s] and [z] but never with [t] or [d]. In the present analysis we shall argue that the peculiar behaviour of affricates indicates a specific role of the manner-defining elements in the process of degemination. It will be maintained that the primes in question contract special binding relations which make them susceptible to the operation of the *Obligatory Contour Principle*. Before we unfold the details of the analysis, however, let us look at the evidence illustrating the behaviour of affricates with respect to degemination (Wiese (1996: 41)):

(9) sitzt /zits/ + /st/ > [zitst] 'sit/2 sg. pres.'
  tanzt /tants/ + /st/ > [tantst] 'dance/2 sg. pres.'
  tanzt /tants/ + /t/ > [tantst] 'dance/3 sg. pres.'

Summarising the facts about consonant degemination in German, we need to observe that the cluster simplification occurs when consonants are identical and

strictly adjacent on the melodic plane. Interestingly, however, complete identity is not expected for reduction to take place. Above all, degemination seems insensitive to the voicing specifications of the segments involved. What is important, however, sequences of affricates and following fricatives are regarded as identical. On the other hand, affricates followed by stops behave as distinct units by resisting the simplifying effects. Another significant requirement for degemination to occur is that the melodies affected by the process must be separated by an empty nucleus.[7] Should we be having a schwa intervening between them, no degemination will ever be found. Thirdly, we need to observe a correlation between the presence of consonant degemination and the effect of vowel umlaut, which, because of its multi-aspectual nature, will not be studied in detail here.

A *GP*-based account of consonant degemination is tightly connected with the operation of the *Obligatory Contour Principle* (Goldsmith (1976; 1990), McCarthy (1986)) which as a strong cross-linguistic tendency enforces the reduction of identical segment sequences whose members are adjacent on the melodic level. Language-specific *OCP* effects reveal that identity can mean different things in different systems. In some languages, the *OCP* will target segments whose elemental make-up is identical, whereas in others, it is sufficient for adjacent melodies to share only a subset of primes, e.g. head elements (as in Japanese (Yoshida (1991)),[8] to become liable to the *OCP* reduction. In German, cluster simplification is naturally induced by the complete identity of the segments that are immediately adjacent in the melodic plane. In consequence, sequences of [t]s, [d]s, [s]s or [z]s will be reduced to single occurrences of these melodies. A more interesting situation arises when the *OCP* becomes triggered in the case of affricate – fricative strings yet blocked whenever affricate - stop sequences are encountered. The autosegmental mode of representation and the geometry of prime configuration advocated by *Government Phonology* make it possible for us to account for that discrepancy in the behaviour of different obstruent strings. The difference in question can be illustrated with two structures of *tantzt*, one being the realisation of the 2 sg. pres. and the other manifesting the 3 sg. pres. of *tantzen* 'dance'. Notice that in the former, the fricative [s] deletes as a result of the *OCP* requirements, whereas

---

[7]   *Government Phonology* recognises the possibility of the presence of empty positions in the phonological representation. An empty position is a timing slot unattached to any melodic material. In order to remain phonetically unrealised, this slot has to be either licensed by parameter (e.g. word-finally) or properly governed (*the Empty Category Principle*).

[8]   In Connemara Irish segments which share manner-defining primes are considered as identical (Bloch – Rozmej (1998)).

no such phenomenon is observed in the latter. Below we depict the effect of degemination:

(10)

a.   O   N   O   b.   O   O

$[ts] + [s]$ $[ts] + [t]$
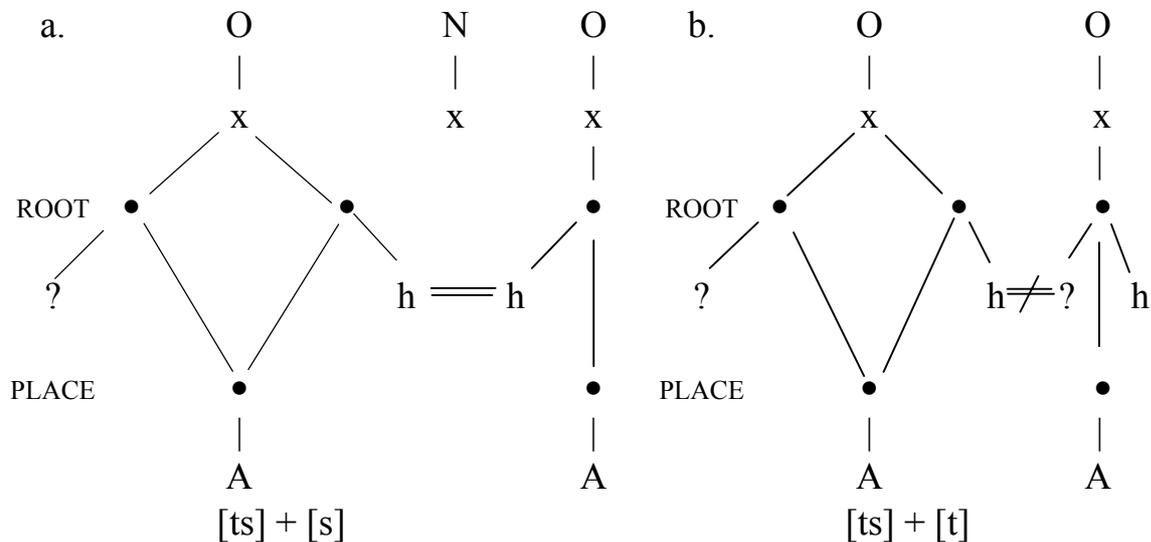
Looking at the above structures, one cannot fail to observe that in (10a) in the sequence of an affricate followed by a fricative, the noise elements (**h**) are strictly adjacent unlike in (10b), where noise borders on occlusion. Since the *OCP* reduction effect targets the former, we have to conclude that German interprets sequences of homorganic affricates and fricatives as identical. Consider (11) below:

(11)   O   N   O
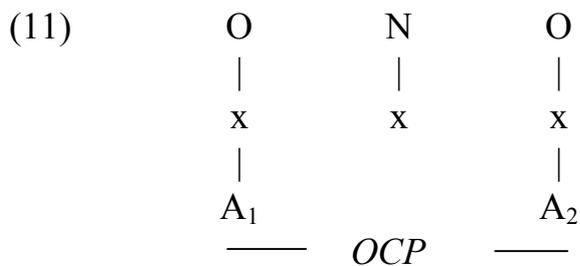
A_1 and A_2 have the same Root elements.

The schematic structure in (11) reveals that in German a hypothesis can be issued to the effect that two immediately adjacent segments whose manner-defining elements (noise) gathered under the Root nodes are the same are regarded as identical. Furthermore, the analysis offered in Bloch–Rozmej (2002) formulates another important condition that has to be satisfied for the *OCP* to be triggered. More specifically, the noise or occlusion primes residing in neighbouring melodies have to contract a specific bridging relation, which represents a special type of licensing, that has the capacity to expose the two elements to the operation of the *OCP*. Support for the existence of interelement

bridging in German comes from the observed correlation between degemination and Umlaut.[9] In brief, the manner elements will be capable of entering into a bridging relation once a nuclear position separating them remains empty. Otherwise, should schwa appear between them, which is interrelated with the occurrence of the stem vowel umlaut, no bridging is found and hence no degemination is attested. Compare *(er) rät* [RE:t] 'he advises' and *(er) wartet* [vartWt] 'he waits'. In the former, the stem vowel [a] is umlauted to [E], which correlates with an absence of schwa in the verbal suffix and degemination of the [tt] cluster. In [vartWt], on the other hand, the root vowel remains unchanged and hence schwa is realised in the suffix, thus preventing the degemination of the consonant sequence.

In what follows, we shall look into the problem of consonant spirantisation attested in German words which sheds some light on the exceptional character of manner elements in this language.

4 The process of spirantisation

Within the model of *Government Phonology*, spirantisation represents one of the weakening processes. Harris (1994) mentions the following opening trajectories resulting in segment decomposition:

(12) Opening trajectories (Harris, 1994: 120)

❖ *Spirantisation*   >   Aspiration   >   Deletion
   plosive > fricative   >   **h**   >   **Ø**
❖ *Glottalling*   >   Deletion
   plosive > **?**   >   **Ø**
❖ *Vocalisation*   >   Deletion
   non-continuant   >   resonant   >   **Ø**

Turning now to the process of [g]–spirantisation, it has to be noted that it reveals new facts concerning a special status of the turbulence-defining prime as well as that of occlusion in the phonology of German. More specifically, when promoted to the status of the segment head, the primes are interpreted as contributing to their phonetic manifestation their salient phonetic properties of both noise and **voicing**. Additionally, there emerges a significant difference between the dialects of German with reference to consonant weakening, instantiated by spirantisation. It is significant that Northern Standard German

---

[9]   This problem has been analysed in detail in Bloch–Rozmej (2002). Because of its great complexity, the question of Umlaut in German will not be dealt with in the present study.

(NSG) differs from Hochlautung with respect to the spirantisation of devoiced velar stops. While Hochlautung seems to have no truly general spirantisation process, NSG speakers 'systematically spirantise all those underlyingly voiced velar stops which occur in *FOD* environments' (Brockhaus (1995: 252)).[10] Once again, we shall adhere to the mechanism of element bridging to account both for the process of spirantisation and its correlation with final devoicing. Evidence illustrating the process of spirantisation is provided in (13) (based on Wiese (1996)):

(13) Tei[Å]   Tei[g]e   'dough/pl.'
     Zwei[Å]   Zwei[g]e   'branch/pl.'
     but
     Tei[Å]   Teiche   'pond/pl.'

Thus, [g] is subject to spirantisation and becomes turned into a palatal fricative if preceded by [i] and word-finally. The change is not reserved for the suffix /-ig/ exclusively. In the formal register of Modern Standard German, the non-syllabic [i`] (as in *Teig,* or *Zweig*) will not trigger the spirantisation of the following velar. Nonetheless, in northern colloquial speech, the narrow context for spirantisation includes also the non-syllabic [i`], other vowels than [i] and even consonantal contexts.

(14) We[Å]   We[g]e   'way/pl.'
     Zu[X]   Zü[g]e   'train/pl.'
     Ta[X]   Ta[g]e   'day/pl.'
     Sar[Å]   Sär[g]e   'coffin/pl.'

The alternations depicted in (13) and (14) above point out that the lexical realisation of the palatal fricative must be [g]. In terms of segment-internal changes as perceived by *GP*, spirantisation will consist in the delinking of the occlusion prime. It is also noteworthy that lexical voiceless velar stops do not subdue themselves to spirantisation (Brockhaus (1995: 253)), e.g:

(15) buk [bu:k]   '(he/she/it) baked'
     Pik [pi:k]   'spades' (cards)

This can indicate that their phonological constitution in terms of elemental material should be different. One hypothesis aiming at resolving this problem is that proposed by Brockhaus (1995) introducing a lexical distinction between

---

[10]   Recall that *FOD* stands for 'final obstruent devoicing'.

truly voiced stops and truly voiceless ones by assigning the **L** element to the former and **H** to the latter. In such a case, spirantisation will be analysed as the delinking of the **L** element and occlusion in the context characteristic of *FOD*, i.e. before a word-final empty nucleus, and the following [i]. This situation is depicted in (16) below:

(16)
```
     N   O←  N              N   O←  N
     |   |   |              |   |   |
     x   x   x    ⟶        x   x   x
     |   |                  |   |
     I   _                  I >  >>
         |                       |
         h                       h
         ┼
         ?
         ┼
         L
```
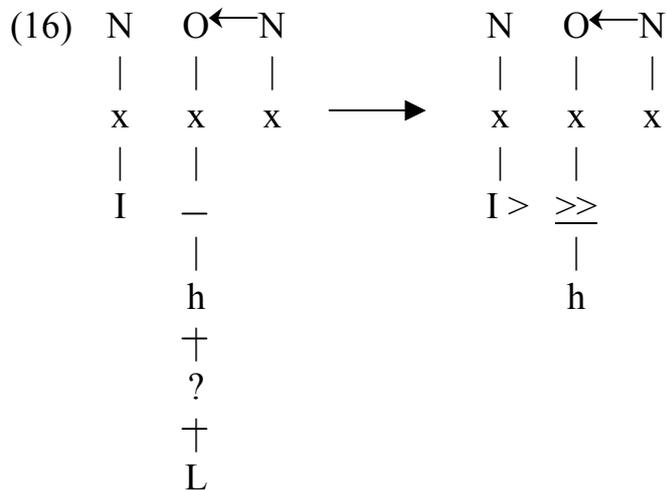
Government Phonology recognises the position licensed by an empty word-final nucleus, which because of its lack of melodic material is the weakest licenser, as one of the primary lenition sites. Hence, elements that have not been autosegmentally licensed, i.e., **L** and **?** undergo detachment from their skeletal position, and in consequence, remain unrealised phonetically. As a result of that operation, the noise element becomes the sole resident of the consonantal structure, which is realised as a voiceless spirant. Simultaneously, in the dialect where a high front vowel precedes the lenition site, the palatal prime **I** spontaneously spreads into the head position of the spirant, which is realised as [Å] (a voiceless palatal fricative).

When analysing the change of the voiced velar plosive into a voiceless palatal fricative, three separate steps can be distinguished: devoicing, spirantisation and palatalisation. However, as observed before, the loss of the occlusion element necessarily accompanies the delinking of **L**. In other words, spirantisation and final devoicing apply in a joint manner, both being instances of segmental weakening. We have identified the trigger of the lenition as a final empty nucleus, an inherently weak prosodic licenser. *GP* proposes that a phonological process targets one prime at a time, which in the light of the facts concerning spirantisation, whereby two primes become delinked from the slot, suggests that both elements can be intimately related. In fact, we shall argue that the loss of both primes represents the same mechanism. Moreover, we should keep in mind that in the autosegmental mode of representation primes are autonomous in the sense of the absence of mutual conditioning of elements'

behaviour. This requirement, however, seems to be violated in the case of spirantisation since the detachment of occlusion is always bound up with the loss of the laryngeal element.

To resolve this dilemma, the present analysis proposes that in German the voicing dimension of segments should be represented as headedness. More precisely, the occurrence of the occlusion element in the head position of a given melody is interpreted simultaneously as both stopness and voice. Therefore, in NSG voiceless consonants are either underlyingly marked for the **H** element or when **H**-less, possess occlusion in other than the head position. Voicing, on the other hand, will be the absence of **H** on the one hand, and the headedness of the manner element on the other. These proposals will be illustrated with the contrast among [k]/[g] and [X].

(17) a.

O	O	O	b.	O
|	|	|		|
x	x	x		x
|	|	|		|
H	**?**	—		**?**
|	|	|		|
?	h	h		h
|				
h				
[k]	[g]	[X]	devoiced	[*k*]

In Northern Standard German voiced segments will lack the **H** element and one of the manner elements will occupy the head position. Such melodies will be attested in environments other than those typical of *FOD*. In these sites spirantised variants occur. By representing the voiced velar plosive as a segment in which the headedness of **?** manifests itself as voicing, we are able to account for both devoicing and spirantisation by means of one operation. Specifically, the licensing of this segment by an empty nuclear position confers too little autosegmental potential to occlusion, thus causing its delinking.[11] This, in turn,

---

[11] The autosegmental licensing potential is conferred by a licenser onto a licensee. Consequently, depending on the strength of the licenser, the number of the elements capable of being sanctioned by the licensed position, will be directly contingent on the amount of the potential that this position is endowed with. Melodically-filled positions will be considered as normally stronger licensers than empty ones.

produces a spirant and simultaneously devoices the melody as the manner element ceases to be the head. Since in NSG spirantisation is optional, in cases where only devoicing occurs, the manner element will become demoted to the dependent status without being detached from the skeletal position. This will produce a headless and hence voiceless plosive.

Finally, let us observe that German has two dorsal spirants [Å] and [X], either of them being a potential outcome of spirantisation, depending on the dialect. In terms of contextual realisation, [Å] is invariably found after a palatal glide or a front vowel, while [X] is attested following a back vowel or glide. Distributionally, the manifestation of the velar fricative appears more severely restricted. Its dependence on the occurrence of the preceding back vowel reflects a phonological backness sharing generalisation operative in German. Precisely, this condition is put forward in Gussmann (2002: 61) as *German Backness Sharing.*

(18)   x   x
          | |
        vowel   spirant
                     |
                   dorsal
              back
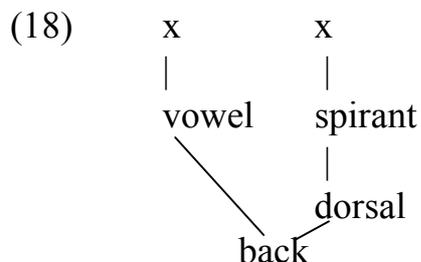
The palatal fricative, on the other hand, occurs in the remaining contexts Therefore, it can be maintained, as in Gussmann (2002), that [Å] is merely a phonetic effect.

5 Conclusion

Government Phonology represents the quality of turbulence with an independent phonological prime of noise which is defined as aperiodic energy. Apart from contributing its unique phonetic characteristics to the interpretation of fricative and affricate sounds, noise, similar to occlusion, can enjoy the status of the segment head that is capable of performing   language-specific roles.

The analyses of the two processes of degemination and spirantisation emphasise the importance of the manner-defining elements in the representation of fricatives and affricates, their specific function in the melodic plane, i.e. a possibility to become segment heads with the accompanying effect of voicing, and possibilities of contracting special types of licensing/bridging relations that bring about significant changes in the phonetic manifestation of the relevant segments, e.g. leading to the *OCP*-enforced reduction effects. Noise and occlusion sharing (bridging) that obtains between melodically adjacent

consonants is interpreted as the identity of the segments involved, thus triggering the *OCP*.

6   References

Bloch–Rozmej, Anna (1998). *Element Interactions in Phonology. A Study in Connemara Irish*. Lublin: Redakcja Wydawnictw KUL.

Bloch–Rozmej, Anna (2002). Degemination in German. In: Elżbieta Mańczak-Wohlfeld (ed.). *Proceedings of the 10th Annual Conference of the Polish Association for the Study of English.* Kraków: Jagiellonian University Press, 11 – 21.

Brockhaus, Wiebke (1995). *Final Devoicing in the phonology of German*. Tübingen: Max Niemeyer Verlag.

Goldsmith, John A. (1976). An Overview of Autosegmental Phonology. *Linguistic Analysis* 2, 23 – 68.

Goldsmith, John A. (1990). *Autosegmental and Metrical Phonology*. Oxford: Blackwell.

Gussmann, Edmund (2002). *Phonology. Analysis and Theory*. Cambridge: Cambridge University Press.

Harris, John (1994). *English Sound Structure*. Oxford: Blackwell Publishers.

Kaye, Jonathan, Lowenstamm, Jean and Vergnaud, Jean-Roger (1985). The Internal Structure of Phonological Elements: a Theory of Charm and Government. *Phonology Yearbook* 2, 305 – 328.

Kaye, Jonathan, Lowenstamm, Jean and Vergnaud, Jean-Roger. (1990). Constituent Structure and Government in Phonology. *Phonology* 7, 193 – 231.

Lieber, Rochelle, (1987). *An Integrated Theory of Autosegmental Processes*. Albany, NY: State University of New York Press.

McCarthy, John (1986). OCP Effects: Gemination and Antigemination. *Linguistic Inquiry* 17, 207 – 63.

Ritter, Nancy (1997). Headedness as a means of encoding stricture. In: G. Booij and J. van de Weijer (eds.). *Phonology in progress – progress in phonology*. The Hague: Holland Academic Graphics. 333 – 365.

Wiese, Richard (1996). *The Phonology of German*. Oxford: Clarendon Press.

Wurzel, Wolfgang, U. (1970). *Studien zur deutschen Lautstruktur*. Berlin: Akademie-Verlag.

Yoshida, Shohei (1991). *Some Aspects of Governing Relations in Japanese Phonology*. Doctoral dissertation, London: London School of Oriental and African Studies.

Laryngeal articulations of /x/ in Southern Polish

Grzegorz Nawrocki
Higher Vocational School, Tarnów, Poland

The present paper offers a summary of the results of two earlier experiments (Nawrocki and Gonet 2004; Nawrocki 2004), in which acoustic properties of the voiceless velar fricative phoneme /x/ in Southern Polish were investigated.

As is found in both studies (Nawrocki and Gonet 2004; Nawrocki 2004), speakers of both genders favour glottal articulation, with partial or full voicing. Word final contexts are decisively in favour of [x]. The word initial, prevocalic positions seem to allow quite a number of allophonic variants of /x/ . These are: [x], [ɦ], [ç] and, additionally, the voiceless glottal, the pharyngeal or the epiglottal [h]/[ħ]/[ʜ]. Another factor taken into account is the coarticulation effect of the vocalic context on the choice of articulation. Based on the results of the experiments, a reformulated allophonic composition is proposed for Polish /x/. It makes room for previously unconsidered pharyngeal and glottal allophones.

In order to inspect the acoustic properties of the allophones of Polish /x/ further, their static and dynamic spectral features are compared to those of phonetically similar sounds in other languages where they have the status of independent phonemes. Special attention is paid to the distribution of spectral peaks and their intensity. The fact that in Polish there are no 'back' fricative phonemes that would contrast with /x/ creates a wide range of acceptable allophonic articulations that cannot be challenged from either articulatory or perceptual points of view.

1   Introduction

The studies by Nawrocki and Gonet (2004), and Nawrocki (2004) investigated the Polish phoneme /x/ only in the positions close to vowel sounds. Such contexts are known to bring about lenition in many languages.

In his presentation of one-strength hierarchy, Lass (1984:178) shows the direction of the change in manner and, ultimately, place of articulation from plosives to affricates, then fricatives, initially oral, where place of articulation is maintained, to more retracted, glottal. The final stage is the deletion of a segment. Green (2003), in his discussion of phonological lenition in Manx (Green 2003:52-53), includes an example of the process involving Manx /x/ in

word-internal V_V contexts : x > ɣ ~ h ~ ø (e.g. le:xən ~ le:ɣən ~ le:hən ~ le:ən 'days'). The variants of Polish /x/ that prevail in word-internal intervocalic positions seem to follow the above principles. Lenited allophones of /x/ are more sonorous because of a wider articulatory channel (Lavoie 2002:44) in contrast to word-initial allophones which are likely to emerge as more salient sounds. The findings of Nawrocki and Gonet (2004) present a challenge to the descriptions of Polish /x/ that view it as a homogeneous element.

1.1 /x/ in Standard Polish

Polish phonetic and phonological literature provides descriptions of Polish /x/ as the voiceless velar fricative and from the prescriptivist point of view, such articulation is most desirable in native speakers and recommendable for foreign learners of Polish. The only major allophone of /x/ that occurs before /i/ and /j/ is the voiceless palatal fricative [ç] (Wiśniewski 1998:52), which Wróbel (1995:142) regards as an independent phonological entity. Both authors mention a less common voiced velar allophone [ɣ], in which voicing occurs due to regressive voicing assimilation, both at word boundaries and morpheme boundaries. Figure 1 presents the classic range of positional variants of Polish /x/.

/x/

[x]	[ç]	[ɣ]
chleb /xlɛp/ 'bread' chata /xɑtɑ/ 'hut' mucha /muxɑ/'a fly' ruch /ɾux/ 'movement'	Chiny /çinɨ/ 'China' hiena /çjenɑ/ 'hyena'	dach domu /dɑɣˈdɔmu/ 'a roof of a house' tych ludzi/ tɨɣˈludʑi/ 'of those people'
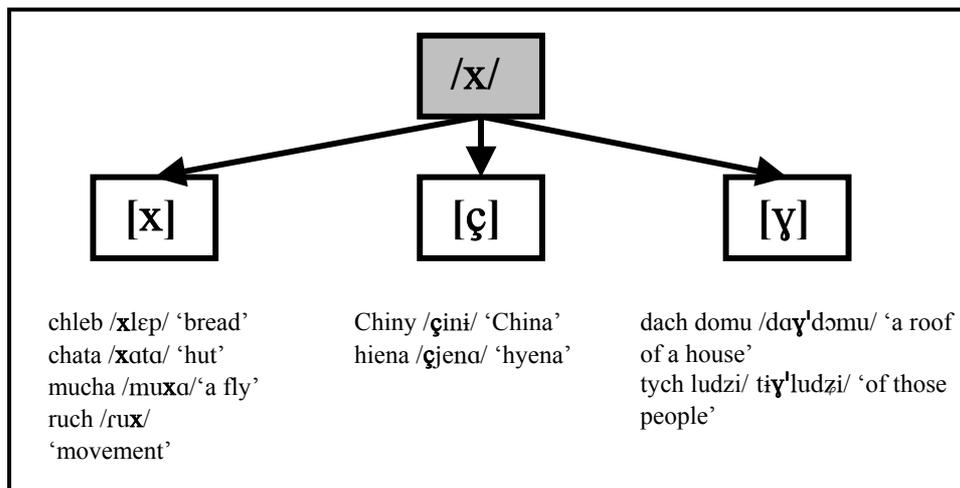
Figure 1: Composition of Polish /x/.

The use of the voiced velar [ɣ] is more widely commented on in earlier Polish dialectological literature (Nitsch 1994) together with minor, but more debatable pronunciations, chiefly in relation to {h}-spelling and pre-vocalic positions of /x/. Nitsch (1994:174), however, attributes the differentiation in the production of written {ch} and {h} only to the speech of inhabitants of eastern borderlands of Poland, mainly people of Ukrainian or Belo-Russian origin.

On the whole, {h} is considered by Nitsch as the spelling form confined mainly to loan words (e.g. from Ruthenian 'hałas' 'noise', from German 'hebel' 'plane').) or onomatopoeic expressions (e.g. 'hm', 'hop',) (1994:175). In terms of articulation, Ruthenian 'h' is classified as the voiced glottal fricative /ɦ/, whereas the German fricative consonant as the voiceless glottal /h/. Nitsch treats the glottal variants as rather unlikely in Polish, and argues against the instances of the glottal variants, reportedly attested in the region of the Tatra Mountains, defining the articulations as 'weak velar' ones.

The proposed 'weakness' of the Polish /x/ receives thorough attention in Stieber's account of Proto-Slavic /x/ (1974:104,105). Since, historically, /x/ had no contrastive correspondent phoneme in the velar region, spirantization of /g/ took place in order to fill the gap (see Figure 2).

$$
\begin{array}{c|c}
\mathrm{k} \!-\!\!-\! \mathrm{g} & \mathrm{k} \\
| & | \\
| & | \\
\mathrm{x} & \mathrm{x} \!-\! \mathrm{ɣ(ɦ)}
\end{array}
$$
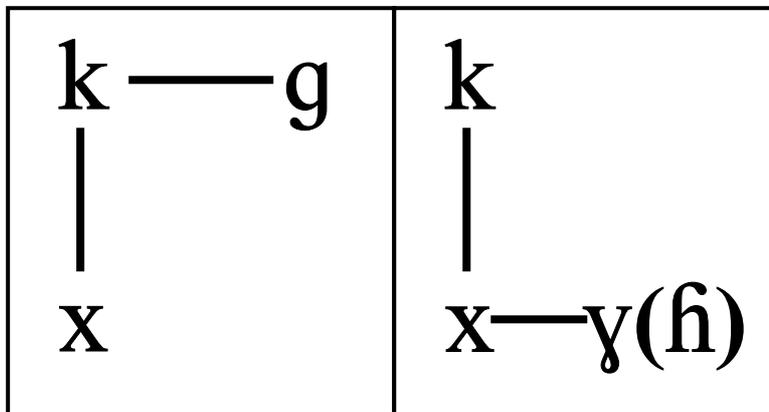
Figure 2: Development of Proto- Slavic /x/ and the change of /g/ to /ɣ/(/ɦ/). The left box shows the situation in proto-Slavic and Modern Polish. The right box shows the change in Byelo-Russian, Ukrainian, Southern Russian, Chech, Slovak, Upper Sorbian dialects (Stieber: 1974:104).

In Polish, however, the change from /g/ to /ɣ/(/ɦ/) did not occur and /x/ remained uncontrasted, which led to its variable articulations. Stieber defines Polish /x/ as a weakly pronounced lenis velar consonant (1974:105). This 'weakness' of /x/, according to Stieber, applies to the whole of the Polish territory. Stieber comments on the fact, that Czechs and Slovaks automatically identify Polish /x/ 'rather with their own voiced laryngeal lenis *h* than with the Chech velar unvoiced and fortis *x*'(1974:105). Such an identification seems to support the findings of Gonet and Nawrocki (2004) and Nawrocki (2004).

Stieber goes on to mention the loss of word-initial pre-consonantal, word-medial intervocalic, and word-final /x/ in the most southern part of Poland (1974:105). An alternative possibility is mentioned, namely the substitution of /x/ with /f/, but only in word-initial and final position. /x/ may also be replaced with /k/ in word final position (Stieber; 1974:105). No examples of a

replacement of intervocalic word-internal /x/ with /f/ have been reported, however. I remember hearing pronunciations like ' w Tychach [f ˈtɪfɑf] ('in Tychy'), while visiting the area of Polish Spisz, or 'pojechał do nich' [ˌpʷojɛ.ɔw ˈdɔɲik] ('he went to them') which I have heard in broad local dialects outside Tarnów.

1.2 Acoustic features of the voiceless velar fricative [x]

The voiceless velar fricative sound [x] is defined as articulated with the back part of the tongue forming a narrow channel with the velum. The air pushed vigorously through the narrow dorso-velar channel produces sharp friction.

Cross linguistic studies (Gordon, Barthmeier, and Sands 2001) reveal close similarity of the spectral features of [x]. The data for Western Apache, Scots Gaelic, Western Aleut, Hupa show that the maximum energy is concentrated at or slightly below 2 kHz. Jassem (1972) points at the second formant of [x] as the dominant among all fricatives as regards its intensity. The second visible, but a less intense frequency peak is located at 4kHz for Western Apache, Scottish Gaelic, Western Aleut and at 5kHz for the Hupa, whereas the last observable energy maximum falls between 6 and 7 kHz (Gordon et al. 2001:8).

Very similar properties are mentioned by Lavoie (2002:48) to be typical of the Spanish [x], where the fricative noise is said to be concentrated at 2 and 4 kHz. Jassem additionally refers to the first formant of [x] at 0.5kHz about which the others remain silent, probably because its intensity is the weakest (1973:223) and therefore it may not be visible in the spectrogram. According to Jassem, formant level values for F_1, F_2, F_3, F_4 of [x] are -12 , 0, -10, and -17dB respectively (1973:225). Jassem does not, however, mention formants at values higher than 4 kHz and allocates the frequency level for F_3 at 2.5 kHz. Such a peak is omitted in the descriptions presented by Gordon et al. (2001).

What seems to transpire from the accounts presented above is a peculiar spectral image of [x]. Depending on the intensity of articulation, individual properties of a speaker's articulators and on the quality of the recording one should see 1, 2, or at best 3 isolated bands in the spectrum of the voiceless velar fricative. They should be located at approximately 2, 4-5 and 6-7 kHz. Jassem's study also includes noise range values for the spectrum of [x]. At -20 dB the span of visible frequencies' extends from 0.3 – 7 kHz, and noise 30 dB weaker than the maximum peak is detectable in the range of frequencies from 0 – 12 kHz (1973:225).

As the two studies (Nawrocki and Gonet, 2004; Nawrocki, 2004) show, what speakers in Southern Poland articulate where [x] is expected is often radically different from the ideal pattern.

2   Results of the two studies of Southern Polish /x/

2.1   Informants, method and equipment

The study by Nawrocki and Gonet (2004) based its findings on audio material gathered from 16 informants, 8 female and 8 male speakers, aged between 20-40. All of them were students of English at the Higher Vocational School in Tarnów, Southern Poland. Most of the informants were inhabitants of the town and its vicinity at the distance of up to 20 km. Among the women, 3 were born and lived in towns quite considerably removed from Tarnów, namely Dębica, Rzeszów, and Przemyśl. The areas represented by these locations stretch to the east and, ultimately, the south-east of Tarnów in the direction of the Ukrainian border.

For the first study, the informants recorded 27 Polish words, all containing the phoneme /x/, spelt with{ch} in word-medial position between vowels. The items, placed in random order, were read by the informants and recorded in a small classroom setting using a dictaphone, at a sampling frequency of 20 kHz. The samples were later digitalized at 44 kHz in preparation for acoustic analysis, which was preliminarily carried out using SFS/Wasp 1.0 software. A more detailed spectrographic analysis was subsequently performed with the use of Speech Analyzer 1.5.

Due to the fact that the group of the female informants for the first experiment was not quite uniform as regards their whereabouts, the study by Nawrocki (2004) was focused only on female speakers. Furthermore, the impact of the vocalic environment was investigated. The group of women was this time made fairly homogeneous, representing the areas close to Tarnów, not more than 20km away from the place. The number of informants was increased to 11 in order to allow for more reliable statistical analyses. The reading list contained 69 words in which /x/ was found in all possible vowel contexts and desirable word positions (i.e. #_V, V_V, V_#). There was a fairly representative list of words with word-initial {h} to check the influence of spelling on the choice of articulation. In order to enhance the quality of the tested material, a W-10 Olympus Digital Voice Recorder was used at the sampling frequency of 15.5 kHz. Preliminary auditory and visual identification of the fricative articulations was carried out using Speech Analyzer 1.5. The articulations for which less typical spectrograms were obtained had to be further looked into using Praat 4.2.05 software. Spectral slices were obtained for the central part of each fricative in order to measure the frequencies of spectral peaks.

2.2 Predominant pronunciations among male and female speakers

Both studies (Nawrocki and Gonet 2004; Nawrocki 2004) revealed an overwhelming preference for non-velar pronunciations in most positions in a word. The prevailing variant among these allophones appeared to be the glottal fricative [ɦ], either partially or fully voiced. Its spectra were relatively easy to identify because they are meant to mirror the formant structure of the neighbouring vowel sounds (Gimson 1989:180).

A two-way ANOVA test ('Sex ' by 'Pronunciation type') showed highly significant variation for pronunciation types: F=37.63 F(2:42)=8.17 at α = 0.001. The interaction between sex and pronunciation type was also significant: F=9.97 F(2:42)= 8.17 at α = 0.001. Figure 3 illustrates means for each pronunciation type in both groups of speakers in the first study. Articulations that were not classified as either glottal or velar were labeled 'other'.
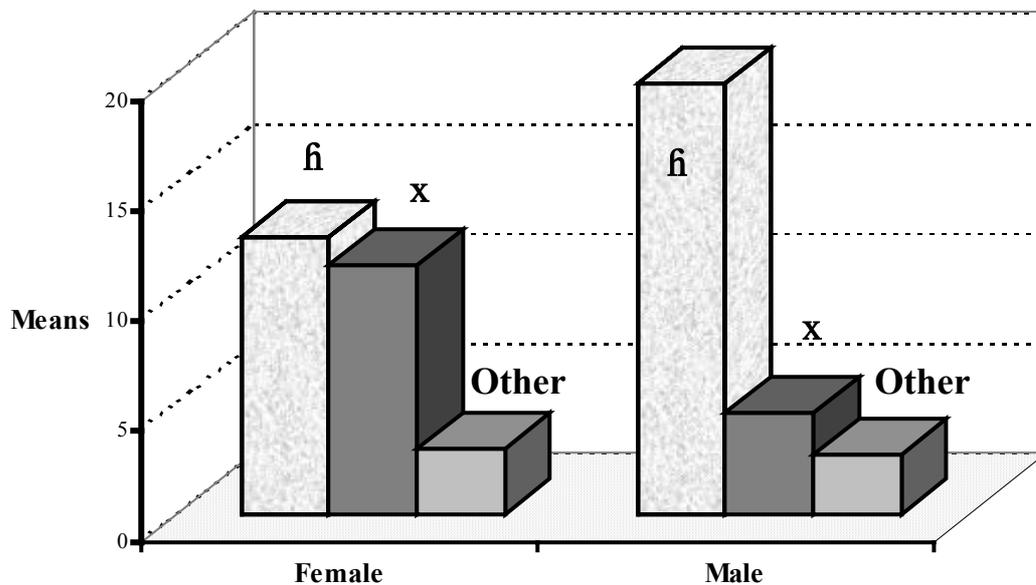
Figure 3: Means of the number of tokens for each fricative articulation in intervocalic contexts in the Interaction 'Sex' by 'Pronunciation Type'(Nawrocki & Gonet; 2004).

As can be seen in Figure 3 the glottal pronunciation appear to be characteristic of men, whose speaking habits have been observed before in various studies by Cruttenden (2001), Gonet and Gawrońska (2003), Gonet and Różańska (2003) as more innovative and less conservative than women's.

The results of the study by Nawrocki (2004) radically undermine the low occurrence of the glottal /ɦ/ in intervocalic position in female speakers obtained in the study by Nawrocki and Gonet (2004). Means for articulation types

(Figure 4) indicate clear dominance of the glottal allophone [ɦ] in female speakers.
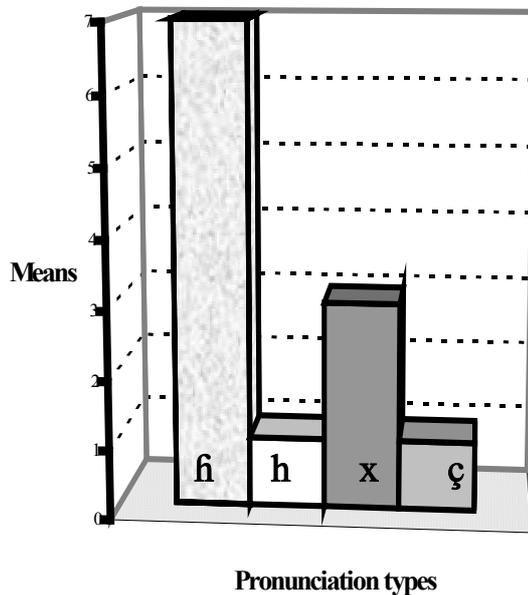
Figure 4: Means of the number of tokens for each fricative articulation in all contexts in female informants (Nawrocki; 2004).
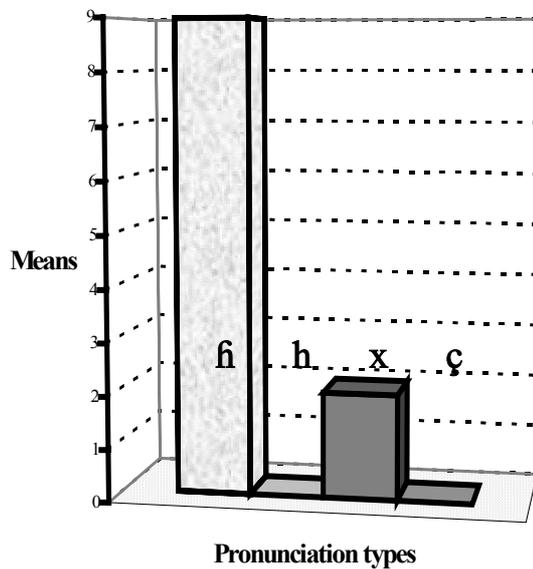
Figure 5: Means of the number of tokens for each fricative articulation in for V_V contexts in female informants (Nawrocki; 2004).

In word-internal intervocalic contexts, the glottal variant towers over the remaining variants (Figure 5) to the degree that resembles the data for males from Nawrocki and Gonet (2004). It seems, after all, that both genders prefer the glottal articulation in V_V contexts. There is a total lack of either fortis

voiceless glottal /h/, or the palatal /ç/. A one factor ANOVA test has shown highly significant difference for pronunciation types: F=55.19 F(3:204)=5.62 at α =0.001.

Other articulations that were identified in both studies, especially the voiceless velar [x], the voiceless palatal [ç], and the voiceless fortis [h] were found more frequently in #_V and V_# contexts, the latter definitely favouring [x].

2.3 Minor articulation types and the influence of word and vocalic context

The identification of the velar fricative [x] from its spectrograms was fairly straightforward. It was based on the existence of an intense peak at around 2kHz, and occasionally an extra one at approximately 4kHz. The spacing of the peaks of [x] was not influenced by the vocalic context.

Although [x] occurred relatively rarely in V_V contexts (cf. Figure 5), it was the most common variant in word-final position. Non-word-internal locations of /x/ were the subject of investigation in Nawrocki (2004). As may be seen in Figure 6, the mean for the velar variant is 9 as compared to 2 for the glottal one, with other articulations scoring zero. A one-factor ANOVA test shows highly significant differences for pronunciation types in word-final position: F=315.35 F(3:24)=7.55 at α =0.001.
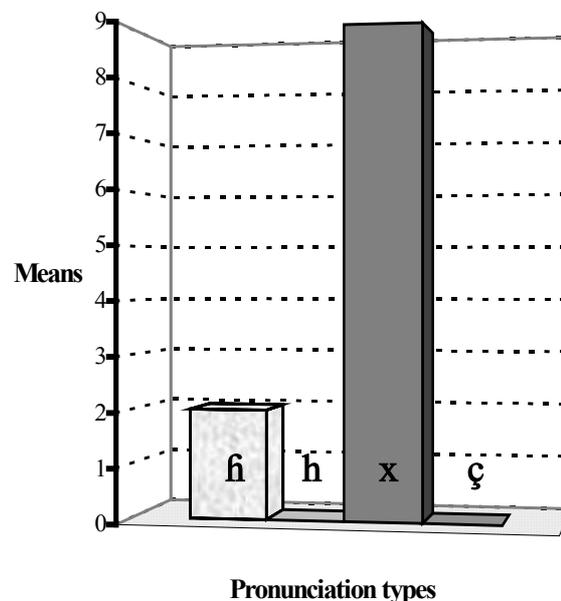
Figure 6: Means of the number of tokens for each fricative articulation in word-final position in female speakers (Nawrocki; 2004).

The variation within individual speakers is not significant. As a matter of fact, 7 out of 11 speakers pronounced the final /x/ as invariably velar, and for 2 speakers there were single instances of the glottal fricative. Two remaining speakers pronounced mostly [ɦ] with more or less voicing.

The context that was the richest in allophones of /x/ was the word-initial pre-vocalic position. There were four major articulation types, namely [x], [ɦ], the voiceless palatal [ç] and also voiceless fortis glottal [h]. Figure 7 presents means for each articulation type.
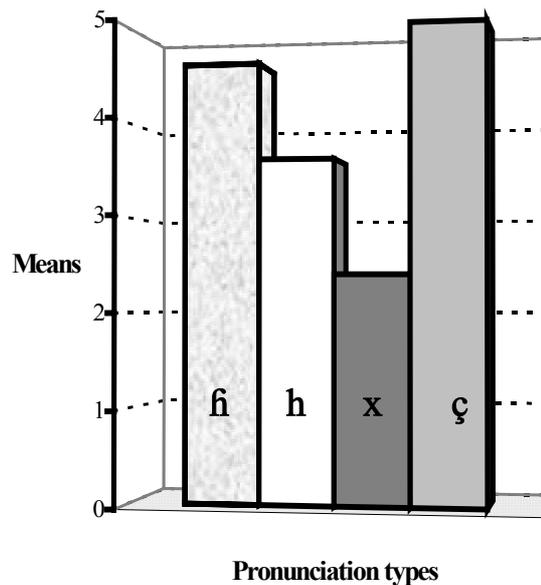
Figure 7: Means of the number of tokens for each fricative articulation in word-initial position in female speakers (Nawrocki; 2004).

A two-way ANOVA ('Speaker' by 'Word-initial articulation type') has proved significant variation for the initial {h}-spelling; F=4.91 F(3:30)=2.92 at α =0.05. The difference between the informants and within the {ch}-spelling was not significant. An additional one-factor ANOVA test proved that the overall impact of the two spelling forms on the choice of pronunciation was not significant (for {ch} F=0,21 F(3:15)=2.48 at α =0.1and for {h} F=0.78 F(3:24)=2.32 at α =0.1).

A relatively high occurrence of [ç] in the initial position may be owing to quite a number of words with /j/ following initial {h}; 4 vocabulary items per 9 contained the palatal glide. This factor might have caused the significance in the differences in pronunciations with initial {h}, but not with {ch}, after which the sequences /ja/, /je/, /jo/ simply do not occur and hence the number of palatalizing contexts is limited. In the vocabulary list there were 4 items where /ç/ occurred invariably in all 11 speakers. These were: 'Chiny'('China'),

'hiacynt'('hyacinthus'), 'hiena'('hyena') and 'Hiob' ('Job'). Before /i/ and /j/, [ç] appears regularly, and alternative pronunciations are hard to come across.

Yet another articulation type, common in all word positions, the voiceless fortis glottal fricative /h/, has so far been left unaccounted for. Judging by the difference in the spectrograms of the two 'glottal' allophones there would be no reason to treat them as identical pronunciation type. As can be noticed in Figure 8 the label 'h' seems to stand for two separate sounds.
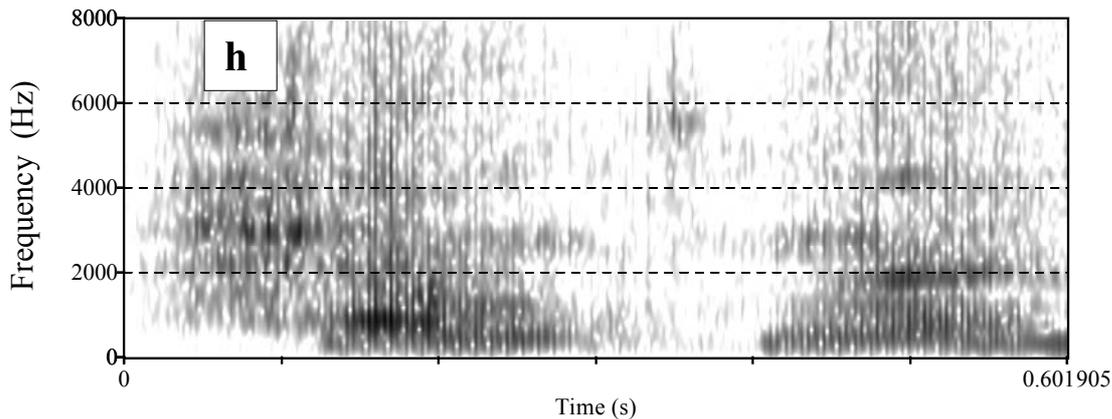
Figure 8: Spectrogram of 'chętny'('willing') pronounced by a female speaker.

Both subvariants, still labeled as 'h' in the second study are characterized by strong noise frequencies and a complete lack of voicing, which completely distinguishes them from [ɦ] . They are clearly voiceless and fortis, but the fortition or intensity of friction is markedly stronger in the variant presented in bottom left part of Figure 8. What this might suggest is a slightly different place of articulation of the second 'h'. While the first one has the noise bands that imitate the formant structure of the following vowel, the other one has much more densely concentrated noise peaks with a noticeably robust peak at around 3kHz. Compared with the acoustic properties of the voiceless pharyngeal fricative [ħ] given by Jassem (1973:225), the sound seems to be a fricative articulated with a considerable constriction of the pharynx, and hence it is classified as [ħ] in later part of the present paper.

3   Further analysis – Polish versus foreign 'back' fricatives

3.1   Method

In order to have a closer insight into the spectral properties of the individual allophones of /x/, audio samples were selected from the material from Nawrocki (2004). For each allophone, three vowel contexts were chosen: mid-front /ɛ/, low /ɑ/, and mid-low back /ɔ/. Each sample had an average duration

of 100 ms and represented the most stable part of the spectrum. The fricatives were isolated from words using Praat 4.2.05. Spectrograms were obtained using Wavesurfer 1.6.4 in Hamming window of around 60 points size. Grid frequency spacing was set at 1000 Hz for easier identification of noise peaks. Then FFT spectra were obtained from individual samples with the use of Praat for clearer recognition of spectral peaks' frequency values. Lower frequency peaks could easily be traced in logarithmic type of FFT spectra. Next, Center of Gravity values were calculated from FFT spectra. Additionally LPC smoothing was performed on individual FFT spectra with the number of peaks set at 5 in order to compare the spectral images of all samples in individual group.The samples representing fricative consonants from other languages were drawn from audio files that accompany 'Handbook of the IPA' (2002).

3.2   Spectrograms and LPC spectra

3.2.1 Samples representing the velar variant [x]

Samples representing the velar variant [x] were isolated from words where they appeared in the final position after /ɨ/, /ɛ/ and /ɑ/. The number of peaks that were clearly visible in spectrograms ranged from 1-2, at times 3. This complies with the number of frequency bands visible in the spectrograms representing phoneme /x/ in five randomly chosen languages (see Figure 9).
The FFT and LPC spectra of Polish [x] reveal the highest peaks located at 1.8-2 kHz and around 5 kHz. The lower noise peak has an average amplitude from -15 to -30 dB. The spectra of the foreign [x] differ only slightly from Polish. The first visible peak is found even at 1.5 kHz and there is also an additional peak at 3.5 kHz (German, Scottish Gaelic) or 4.5 kHz (Irish and Slovak). Logarithmic FFT spectra show yet another markedly noticeable spectral peak at 0.5 kHz (F_1), mentioned by Jassem (1973:225). Its close proximity to F_2 makes it hard to distinguish in normal spectra and spectrograms of [x].

Generally speaking, Polish [x] seems to share spectral cues with similar fricatives that exist in other languages. Its spectral parameters generally coincide with the cross-linguistic data presented in 1.2 above.
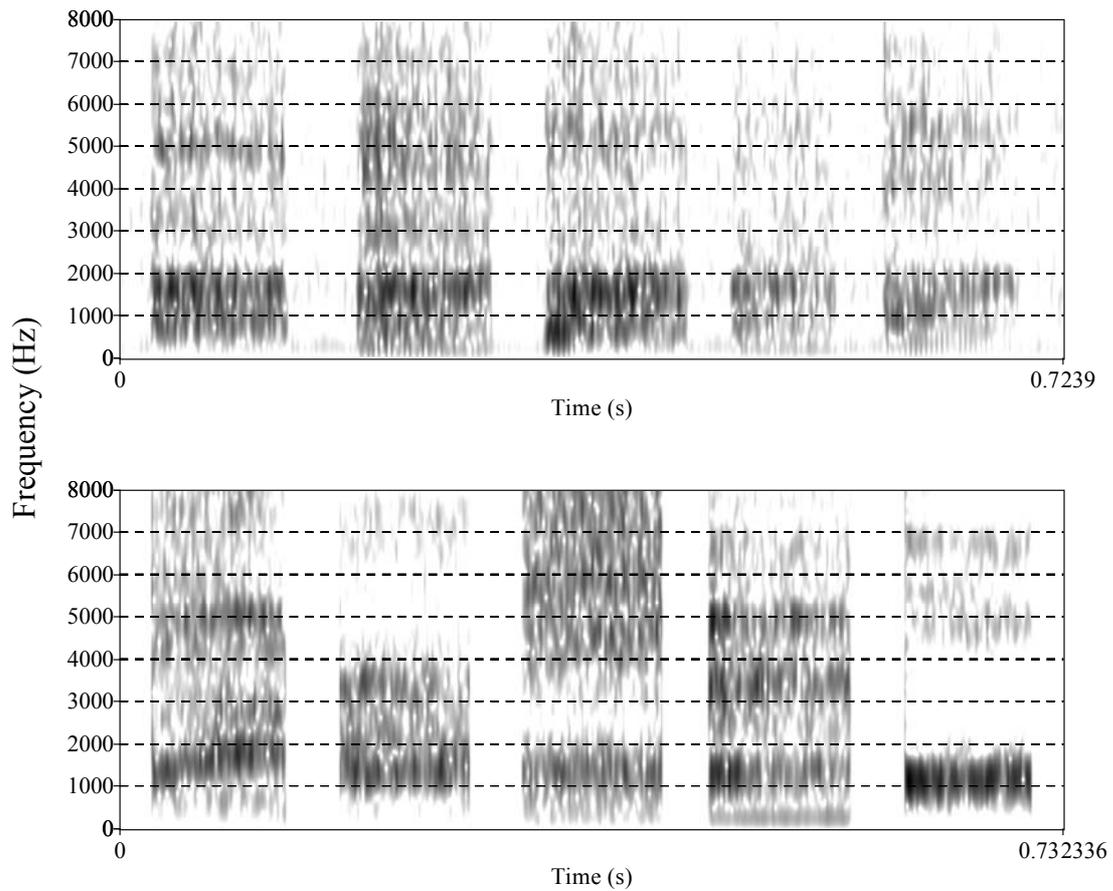
Figure 9: Spectrograms of Polish and foreign [x]. Top part: Polish [x] in the position after /ɑ/ for five speakers. Bottom part: foreign [x] (Cz. 'moucha''fly', Ger.'Macht''power', Ir. 'chaol' 'thin(fem.)', Sc.G.'biodach' 'tiny' Sl.'dvoch' 'two(dat.)'.
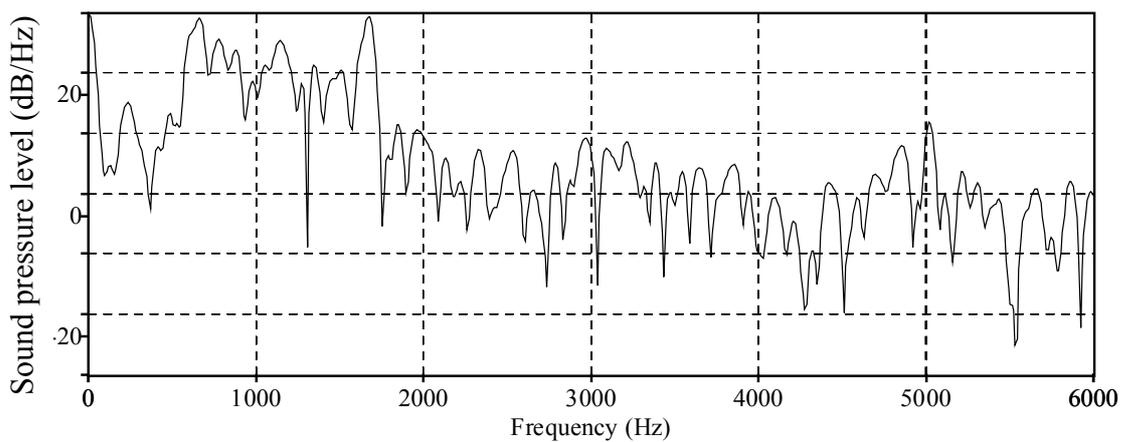
Figure 10: FFT spectrum of Polish [x] after /ɑ/.

3.2.2 Samples representing the voiceless palatal fricative [ç]

The spectrograms of [ç] are radically different from those representing the velar variant (see Figure 11). There is barely any noise below 3 kHz and the friction range stretches from 3-6 kHz. The intensity of friction varies depending on the force of articulation of individuals. The faintest bands are at 20 dB, while the strongest ones reach even 80 dB. F_1 does not exist, or is not detectable (cf. Jassem 1973:225). The number of visible peaks varies from 2-3.
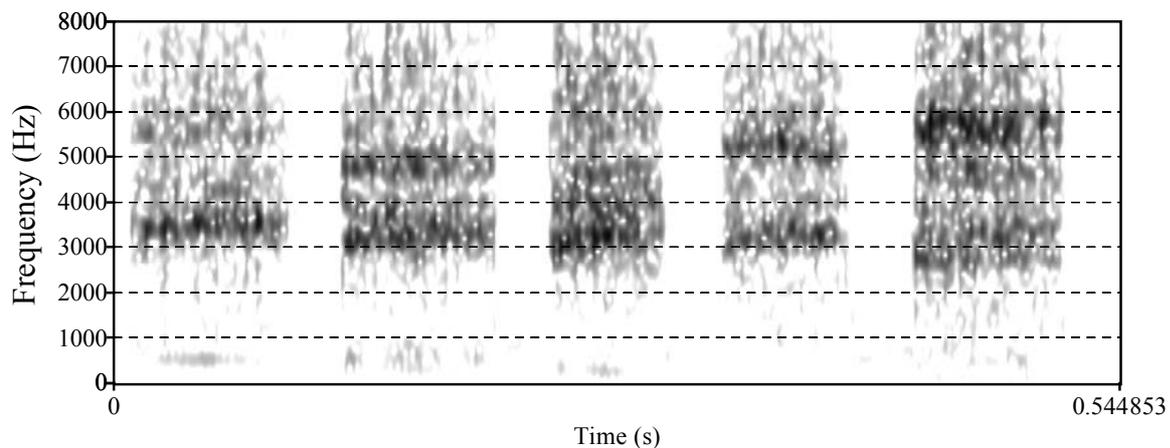
Figure 11: Spectrograms of Polish [ç] in the position before /jo/.

Smoothed LPC spectra of Polish [ç] display mostly two peaks located at 3-4 and 5-6 kHz (Figure 12). There is not too much difference in intensity between the frequency maxima and the distance between them is rather small.
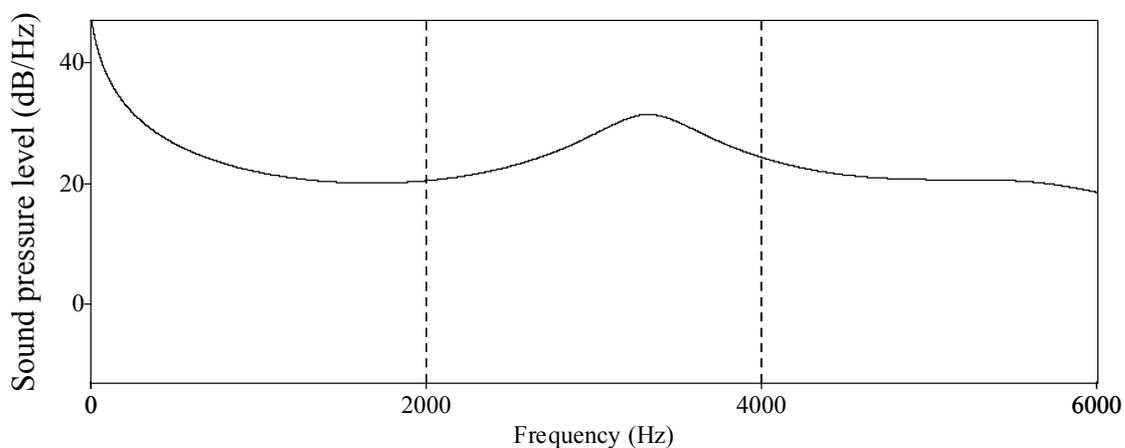
Figure 12: Smoothed LPC spectrum of Polish [ç] in the position before /ja/.

3.2.3 Samples representing the voiceless glottal fricative [ɦ]

Polish voiceless glottal fricative [h] was identified only in word-initial, stressed position. The sound was analyzed only in the position before vowels /ɛ, ɑ, ɔ/. As can be noticed in Figure 13, the pattern appears to be scattered and variable even before vowel of very similar quality. Such an irregular arrangement of spectral peaks and their relative robustness make this variant markedly different from the palatal and velar allophones of /x/ that are more stable in this respect. Strong noise and syllable-onset position of voiceless glottal /h/ may be dictated by high articulatory effort of the speakers who produce this variant.
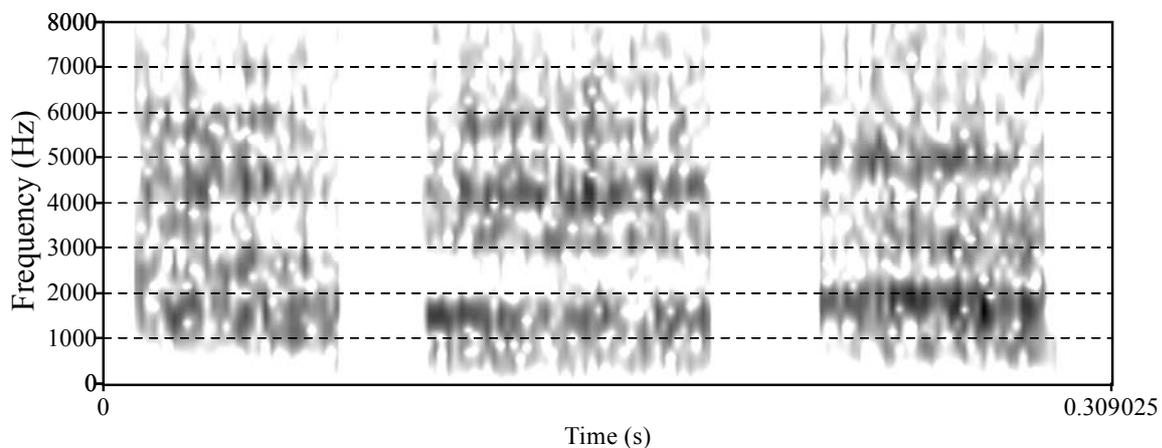
Figure 13: Spectrograms of Polish [h] in word-initial position before /ɑ/.

The smoothed LPC 10-peak spectra of Polish [h] indicate a peculiar spectral nature of the sound (see Figure 14). The sound appears in word-initial stressed position before a vowel. Therefore, a voiceless breathy onset of the following vowel sound seems to form strong articulatory grounds for the glottal voiceless [h] to figure as an allophone of /x/ in this context. Spectral features of [h] are quite outstanding in comparison to the variants that have been discussed so far.
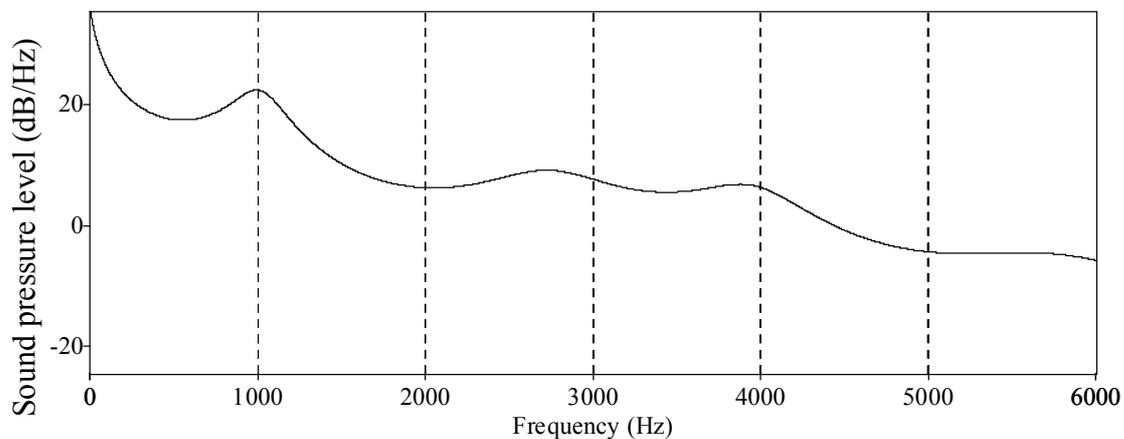
Figure 14: Smoothed LPC spectrum of Polish [h] in the position before /ɔ/.

3.2.4 Samples representing the voiced glottal fricative [ɦ]

The samples for spectrograms presented in Figure 15 were isolated from the intervocalic position in the investigated material. The flanking vowels that were chosen were /ɛ/, /ɑ/, and /ɔ/. As can be observed in both parts of Figure 15, friction in [ɦ] presents two different patterns. In the left-hand part the fricative stands in the ɛ_ ɛ context and may be interpreted as a breathy voiced [ɛ̤] . The right-hand part of Figure 15 shows the formant structure of the fricative resembling the mid-low back vowel /ɔ/, which precedes and follows [ɦ]. Hence the fricative may be regarded as breathy voiced [ɔ̤].
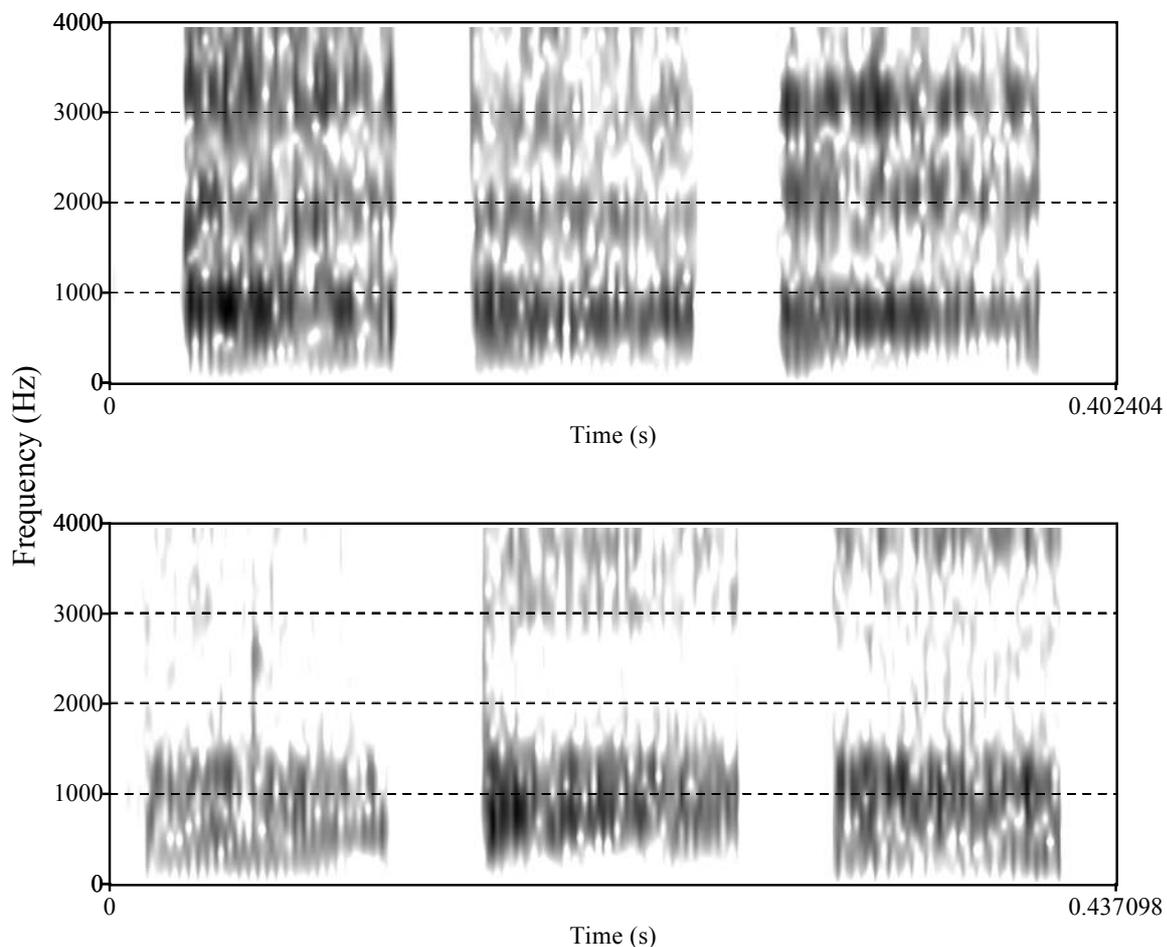
Figure 15: Spectrograms of Polish [ɦ] in word-medial position in two different vocalic contexts. Top part : [ɛ_ ɛ], bottom part: [ɔ_ɔ].

Unlike in the spectra of [h], the voicing in [ɦ] is present in the form of an additional peak at around 200 Hz. The top part of Figure 16 presents the logarithmic FFT spectrum of an example of Polish voiced glottal fricative in the ɑ_ɑ environment. The bottom part shows the spectrum of a similar sound taken from a recording of the Dutch word 'hoed' ('hat'). The peak representing a glottal pulse in the Dutch fricative is situated at 100 Hz. The intensity of F_0 is

very high, 50 dB, and dominates the remaining spectral peaks. The spectrum of the female voice pronouncing the Polish fricative has F_0 located at 200 Hz, but of lower intensity. It manifests a higher fundamental frequency.
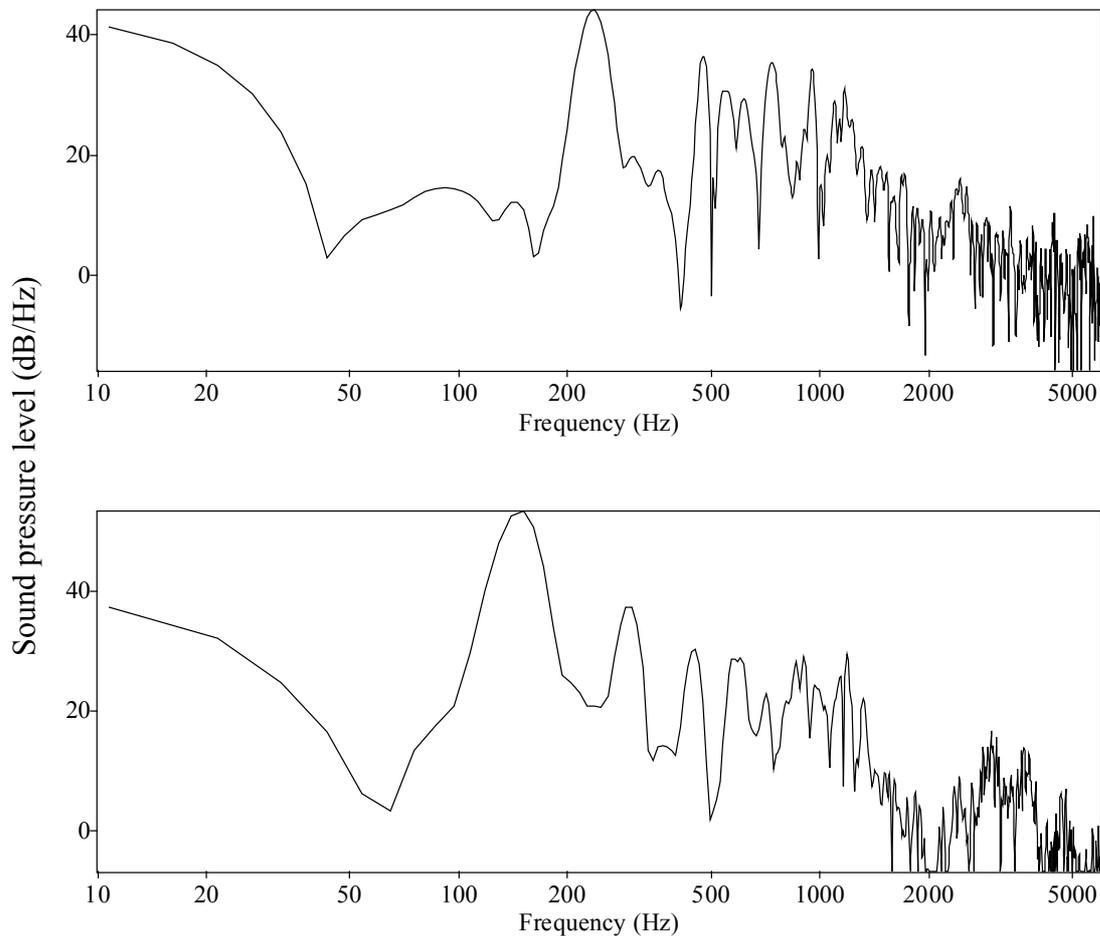
Figure 16: Logarithmic frequency FFT spectra of Polish (top) and Dutch (bottom) [ɦ].

One might assume that, because of the fact that F_0 is not the dominant peak in the example of the Polish [ɦ], voicing is not the distinctive feature in the pronunciation of the Polish sound

Similar maximum F_0 was observed in the spectra of the remaining foreign voiced glottal fricatives, taken from Slovak, Czech, and Ukrainian. All have the status of a phoneme in those languages. Unlike in English, where [ɦ] is merely an intervocalic allophone of /h/, in the three languages (i.e. Slovak, Czech, and Ukrainian) / ɦ/ acts as an independent phoneme. The voiceless /h/ is, in fact, absent in phoneme inventories of the three languages. /ɦ/ originates here from proto-Slavic /g/, and is phonetically enhanced by contrast with /x/ (Stieber: 1974:104). That is why Czech, Slovak, or Ukrainian /ɦ/ may strike the ear as 'more voiced' than the Polish intervocalic allophone [ɦ].

Recapitulating, as regards Polish and foreign spectral images of [ɦ], the Polish variant differs from the latter only in voicing, fortition (or articulatory strength) being rather of secondary importance. In Polish, however, /ɦ/ was the most frequent variant in the inter-vocalic position, to which it owes its voiced character.

3.2.5 Samples representing the voiceless pharyngeal fricative [ħ]

The voiceless pharyngeal fricative /ħ/ exists as a phoneme in only a few languages. Three such languages have been chosen for spectrographic illustration, namely Arabic , Hebrew and Agul (see Figure 17). The location of fricative noise agrees only roughly with the data from Jassem (1973:225), who locates the most intense peaks at 3.1 and 4.2 kHz with a weak one at 0.9 kHz. There seem to be other intense peaks in the spectrum of the Arabic consonant.
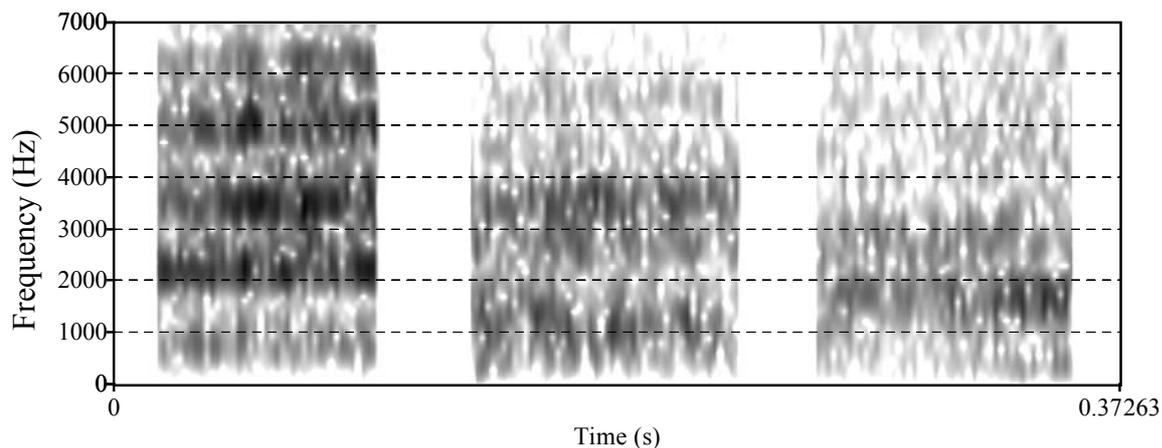
Figure 17: Spectrograms of foreign [ħ] (Ar. 'ħuruwb' 'wars', Heb. 'maħar' 'tomorrow', Ag. 'muħar' 'barns').

The study by Nawrocki (2004) attempted to label quite a large number of variants of [h] as 'epiglottal'. Yet, in view of the fact that hardly any language, apart from Agul, contrasts epiglottal and pharyngeal fricatives, the label has been discarded. Nevertheless, in word-initial pre-vocalic stressed positions, spectra of a fricative consonant of a more salient type than that of [h] were discovered. That is why, though not a very typical articulation in Polish speakers, [ħ] must not be ignored in the present study.

The sound was examined in the position before /ɛ, ɑ, ɔ/. As seen in Figure 18 the fricative energy is thickly distributed all over the spectrum and most of the peaks are of high intensity.
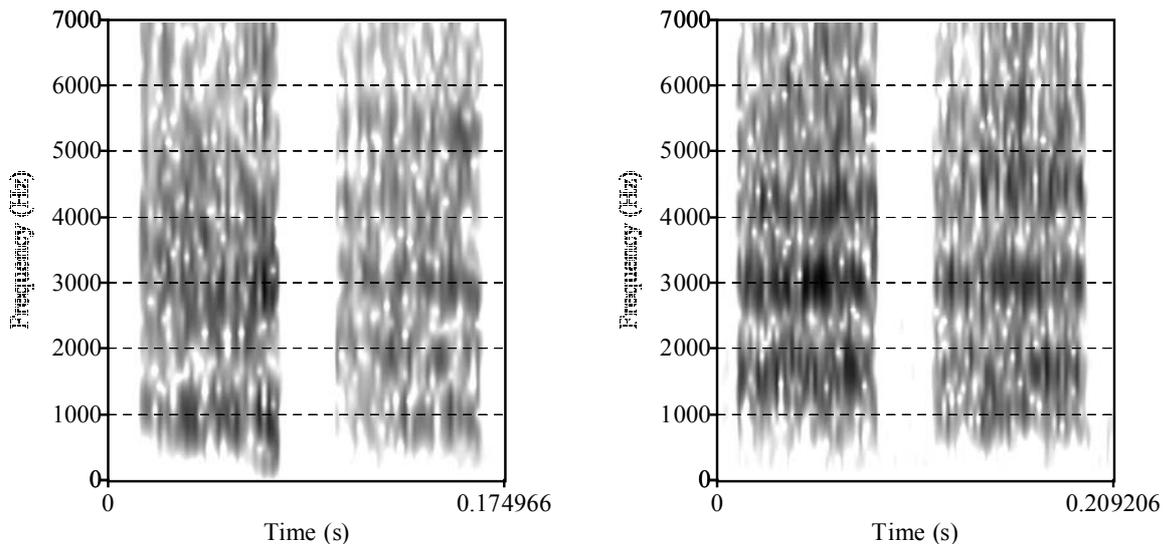
Figure 18: Spectrograms of Polish [ħ] in the position before /ɛ/ (left) and /ɑ/(right).

Maximum energy seems to be located at around 3kHz, which coincides with frequency values of F_3 and F_4 in Jassem's description of [ħ] (1973:225). Compared with the glottal [h] discussed earlier, [ħ] has a spectrum independent of the vowel context and the intensity of the noise is much higher, 40-50 dB. For [h], noise values were lower, 15-45 dB.

4 Conclusions

4.1 The proposed labeling of allophones of Polish /x/

Among the five allophones suggested for Polish /x/ in the present study, the voiceless palatal [ç] is invariably determined by the presence of the following high vocoid. As can be seen in Figure 19, the difference between the spectrograms of [ç] and [x] is quite radical. The movement of the tongue towards the palate in preparation for /i/ or /j/ appears to be a type of assimilation found in all speakers.

The relative difficulty of pronunciation of [x] makes it an unlikely candidate for all possible contexts. In fact, in both studies for male and female speakers (Nawrocki and Gonet 2004, Nawrocki 2004), [x] was the dominant type of articulation chiefly in the post-vocalic word-final position. In V_V context, [x] was preferably replaced by the voiced glottal [ɦ].
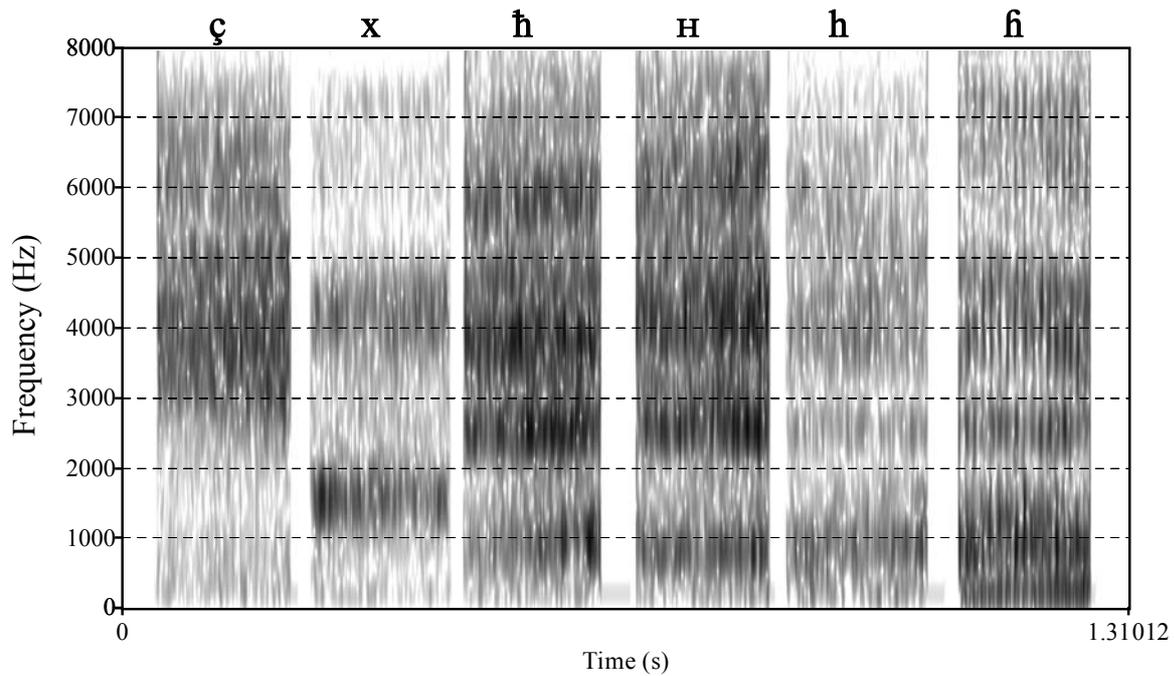
Figure 19: Spectrograms of six model fricatives pronounced by the author in ɑ_ɑ context.

Word-initial positions require more salient allophones, and the strongest friction noise distributed over a relatively wide frequency range appears in the pharyngeal fricatives. Even the voiceless glottal [h] wins over [x] in this respect. Polish speakers opted for the glottal, at times pharyngeal, variant. One might assume that, in many cases this was a glottal friction, reinforced by a variable, and perhaps partial, pharyngeal constriction. The strength of the friction achieved in this manner, made the word-initial variant easily perceptible. Such an allophony might be expected from the reports of authors like Nitsch and Stieber, although they are not clear about the possibility of intervocalic laryngeal pronunciations of /x/. Yet, they report the loss of /x/ in such context, which might not possibly occur without an intermediate phase of laryngeal articulation (Lass 1984:178).

4.2 Suggested composition of Southern Polish phoneme /x/ in vocalic contexts

If we take into consideration only those contexts where /x/ appears close to vowel sounds, namely #_V, V_V, and V_#, the distribution of allophones for Southern Polish /x/ might look as in Figure 20.

The frequency of occurrence is reflected in the sequence of allophonic symbols. Thus, the most likely variants stand first. The symbols in brackets represent possible, but relatively uncommon articulations. The distribution in #_V has been slightly altered in relation to the data in 2. It gives priority to

163

fortis glottal and pharyngeal allophones and places the voiced glottal variant as the final alternative.
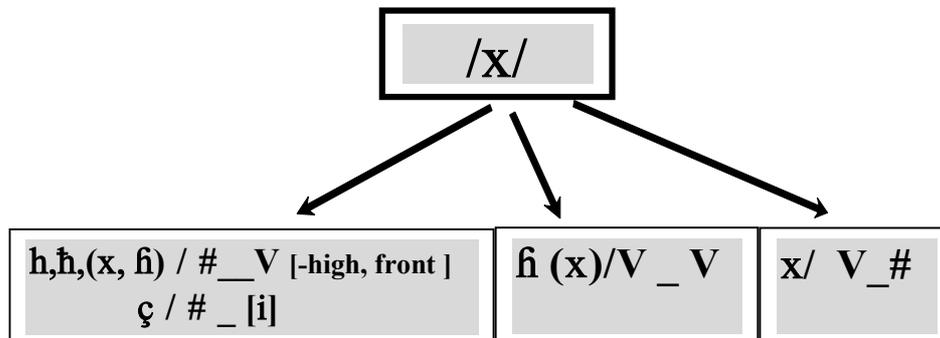
Figure 20: Phoneme /x/ in Southern Polish in the positions adjacent to vowel sounds.

It might be assumed, that Southern Polish /x/ should be underlyingly represented as /h/, in order to reflect the dominance of the 'back', or laryngeal sounds. In this way, the velar fricative [x] would become its allophone in word-final and pre-consonantal contexts. Of course, such reasoning would have to put aside all historical and most phonological evidence. If /x/ is a historically viable representation of the phoneme, the variants described above may constitute phonetic means of some 'linguistic therapy' to save the seriously weakened and unspecified phoneme (cf. Stieber 1974:107 for the discussion of the development of Upper Sorbian aspirated /k^{x}/ in early 18th century). However, more research is necessary to find out to what extent [h] might represent /x/ in vocalic contexts in other parts of Poland.

References

Cruttenden, A. (2001): *Gimson's Introduction to the Pronunciation of English.* 7th ed. London: Arnold Publishers.

Green, A.D. (2003): The independence of phonology and morphology. *ZAS Papers in Linguistics* 32, 47-86.

Gimson, A. C. (1989): *An Introduction to the Pronunciation of English*. London: Arnold Publishers.

Gonet, W. & Gawrońska, E. (2003): Vowel quality in the speech of NBC presenters. *Speech and Language Technology. Papers, Reports and Technical Notes* 7, 21-34.

Gonet, W. & Różańska, K. (2003): Vowel quality in the idiolects of four BBC world news presenters. *Speech and Language Technology. Papers, Reports and Technical Notes* 7, 53-70.

Gordon, M., Barthmaier, P., Sands, K. (2002): A cross-linguistic study of voiceless fricatives. *Journal of the International Phonetic Association* 32, 141-174.

Gordon, M., Potter B., Dawson J., de Reuse, W. & P. Ladefoged (2001): Phonetic structures of Western Apache. *International Journal of American Linguistics* 67, 415-448.

International Phonetic Association (1999): *The Handbook of the IPA.* Cambridge University Press.

Jassem, W. (1972): *Podstawy fonetyki akustycznej.* Warszawa: PWN.

Lavoie, L.M. (2002): Subphonemic and suballophonic consonant variation: The role of the phoneme inventory. *ZAS Papers in Linguistics 28, 39 – 54.*

Lass, R. (1984): *Phonology.* Cambridge: CUP.

Nawrocki, G. & Gonet, W. (2004): Realizations of /x/ in intervocalic contexts in Southern Polish. *Speech language and Technology. Papers, Reports and Technical Notes* 8, 87-106.

Nawrocki, G. (2004): Southern Polish /x/-subphonemic variation. A study in female speakers. *Speech language and Technology. Papers, Reports and Technical Notes* 8, 107-136.

Nitsch, K. (1994): *Świat mowy polskiej.* Warszawa: PWN.

Stieber, Zdzisław (1974): *Swiat językowy Słowian.* Warszawa: PWN .

Wiśniewski, M. (1998): *Zarys fonetyki i fonologii współczesnego języka polskiego.* Toruń: Wydawnictwo Uniwersytetu Mikołaja Kopernika.

Wróbel, H. (ed.) (1995): *Gramatyka współczesnego języka polskiego*. Vol. 3. *Fonetyka i fonologia*. Kraków: Wydawnictwo Instytutu Języka Polskiego PAN.

Focus and Phrasing in Shingazidja

Cédric Patin
Zentrum für Allgemeine Sprachwissenschaft, Berlin &
Laboratoire Phonétique et Phonologie, Paris

1   Introduction[1]

In his seminal PhD thesis (Kanerva 1990), Jonni Kanerva has demonstrated that focus influences the prosodic phrasing in Chicheŵa, a Bantu language of Malawi. The core data are exemplified in (1). The symbol ')' signals the end of a phonological phrase[2].

(1)   Chicheŵa (Kanerva 1990: 98)
        a. (anaményá   nyu^{m}bá ^{n}dí   mwáála)   (VP focus)
            he-hit   House   with   rock
        he hit the house with a rock

        b. (anaményá   nyuú^{m}bá)   (^{n}dí   mwáála)   (Object focus)
            he hit THE HOUSE *with a rock*

        c. (anaméenyá)   (nyuú^{m}bá)   (^{n}dí   mwáála)   (Verb focus)
            he hit the house with a rock

In (1-a), all the elements of the VP (i.e. the verb *anaményá* 'he hit', the object *nyu^{m}bá* 'house' and the oblique PP *^{n}dí mwáála* 'with a rock') form a single phonological phrase. In (1-b), however, the object is focused in-situ and is then followed by a prosodic break. Finally, in (1-c), the verb is focused and both the verb and the object are followed by a prosodic boundary. Similar results were independently observed in another dialect of Chicheŵa (Downing et al. 2005).

---

[2]   Reffered to as *focus phrase* in Kanerva (1990) – see Downing (2002) for a discussion.

It has been established since Kanerva's work that focus conditions phrasing – directly or indirectly – in several other Bantu languages, e.g. Chimwiini (Kisseberth 2007, Downing 2002, Kisseberth & Abasheikh 2004), Xhosa (Jokweni 1995, Zerbian 2004), Chitumbuka (Downing 2006, 2007), Zulu (Cheng & Downing 2006, Downing 2007), Bemba (Kula 2007), etc.

In this paper, I will argue that focus also conditions phrasing in Shingazidja, a Bantu language[3] spoken on Grande Comore (or Ngazidja, the largest island of the Comoros).

Many works have been dedicated to the tonology of Shingazidja. The bases of the system were firstly identified by Tucker & Bryan (1970) and reanalysed by Philippson (1988). Later, Cassimjee & Kisseberth (1989, 1992, 1993, 1998) provide a very convincing analysis of the whole system of the language, and my own research (Patin 2007a) shows a great correspondence with their results. However, little attention has been paid by these authors or others (Jouannet 1989, Rey 1990, Philippson 2005) to the phonology-pragmatics interface, especially on the relation between focus and phrasing. This paper thus proposes to explore this question. It will be claimed that focus, beside syntax, has an influence on phrasing in Shingazidja.

In his seminal thesis (Truckenbrodt 1995) and in a following article (Truckenbrodt 1999), Truckenbrodt has provided a very convincing analysis, sketched within Optimality Theory (now OT; Prince & Smolensky 1993, McCarthy & Prince 1993a, 1993b, and an abundant literature), of the Chicheŵa facts identified by Kanerva. I will adopt Truckenbrodt's model in this paper, showing that a minimal reranking of the hierarchy proposed by Truckenbrodt (1995, 1999) to account for Chicheŵa derives the correct phrasing of Shingazidja.

The paper is organised as follows: in section 2, I present the phonological processes that are relevant to identify prosodic boundaries in Shingazidja, and I identify some of the contexts in which the phrasing is syntactically determined. In section 3, I show how the focus influences the phonological phrasing in Shingazidja. In section 4, an OT analysis of the data presented in sections 2 and 3 is proposed.

---

[3] Shingazidja is coded G44a in Guthrie (1967-71)'s referential classification. The language belongs to the Sabaki group, which also notably contains Kiswahili. Data for this paper, which were partially recorded, were mostly obtained from the native speaker Ibrahim Barwane in Paris (France), between June 2006 and January 2007.

2 Phrasing in Shingazidja

Evidence for phrasing in Shingazidja comes from the shift of tone. In this language, a high tone will shift to the syllable preceding another tone bearing syllable (see section *2.1*), the last syllable of a phonological phrase (see section *2.2*) or the penultimate syllable of an intonational phrase (see section *2.3*). Indeed, Shingazidja is typologically interesting since it clearly presents two prosodic levels above the level of the prosodic word, while such an evidence does not exist in many other Bantu languages (see Zerbian 2006: 130 for a discussion). Most of the results of this section were previously identified by Cassimjee & Kisseberth (1993). Due to space restrictions, I will not discuss in this paper all the contexts in which phrasing is conditioned by morphosyntactic parameters – see Patin (2007a: chapter 5) for a detailed list of the relevant contexts.

2.1 Tone phenomena

Shingazidja, similarly to other genetically related languages like Mijikenda (E72-73, a set of nine dialects mostly spoken in eastern Kenya; see Philippson 1991, Cassimjee & Kisseberth 1998), exhibits an 'unbounded' tone shift phenomenon. In (2-a), the noun *waleví* 'drunkards' has a high tone on its final syllable in isolation[4]. In (2-b-i), however, the tone originally associated to the last syllable of this same noun *waleví* shifts onto the penultimate syllable of the adjective *wailí* 'two'. The noun, which underlyingly bears a high tone on its last syllable, is then fully low. In (2-c-i), the noun receives the shifted tone of the verb *hawóno* 'he saw'.

(2) a. wa-leví
         2[5]-drunkard
         drunkards

     b. i. wa-levi   wa-íli
             2-drunkard 2-two
             two drunkards

         ii. *waleví wailí*

---

[4] I mark the trigger syllable which underlyingly bears the tone by underlining it.
[5] Numbers refer to noun classes.

c. i. ha-w<u>o</u>no wa-lév<u>i</u>

  3sg(past)-see 2-drunkard

  he saw drunkards

  ii. *haw<u>ó</u>no walev<u>í</u>

The shift of the first tone leads to the deletion of the second one. In (2-c-i), the tone of the last syllable of *walev<u>í</u>* 'drunkards' is deleted because it is adjacent to the tone of the verb, which surfaces on the penultimate syllable of the noun. This deletion is triggered by the Obligatory Contour Principle (now OCP), which prevents the adjacency of two identical elements – here, two high tones. In a sequence of high tones, then, the first one shifts to the syllable preceding the following one, and every other tone is deleted – *haw<u>o</u>no walév<u>í</u>*.

Compare for instance (2-b-i) with (3). In (2-b-i), the shift of the tone of the noun *walev<u>í</u>* 'drunkards' causes the deletion of the tone of the adjective *wail<u>í</u>* 'two'. In (3), however, the tone of the adjective is free to appear in surface since the tone of the noun is deleted.

(3) ha-w<u>o</u>no walév<u>i</u> wail<u>í</u>

  3sg(past)-see 2-drunkard 2-two

  he saw two drunkards

If the noun does not underlyingly bear a high tone – e.g. *ndóvu*[6] 'elephant' in (4) –, the tone of the verb shifts to the adjective. As with the noun, the lexical tone of the adjective *mɓil<u>í</u>* 'two' is then deleted.

(4) ha-w<u>o</u>no n-dovu m-ɓíl<u>i</u>

  3sg(past)-see 10-elephants 10-two

  he saw two elephants

2.2 Phonological phrases

In Shingazidja, the maximal syntactic phrase and the phonological phrase are coextensive. For instance, in (5), the whole NP[7] forms a single phonological

---

[6] In Shingazidja, a phonological phrase has to present at least one tone. The tone that appears on this noun when it is isolated is a default one.

[7] In Shingazidja, an isolated NP is frequently preceded by the so-called 'stabilizer' (here *ndé*), which also introduces clefts. What is important here is that no obligatory prosodic boundary has to be inserted within the NP.

phrase, despite its heaviness[8]. The symbol ')$_\Phi$' signals the end of a phonological phrase.

(5) (nɗe zɛ́ m-ɓuɗa m-ɓilí n-jɛu halisi z-á hahe)$_\Phi$

      STAB AT$_{10}$ 10-stick 10-two 10-white very 10-of his

these are his two very white sticks

Similarly, a VP constitutes a single phonological phrase (6-a), even if it contains two object NPs – (6-b).

(6) a. (tsi-wọno n-dovu y-a wá-mẹzi)$_\Phi$

        1sg(past)-see 9-elephant 9-of 2-beggar

I saw an elephant of the beggars

    b. (ha-nịka ɲama m̩-ɖu-ḿ̩-ʃe)$_\Phi$

        3sg(past)-give 9/meat 1-person-1-woman

he gave (some) meat to a woman

In (6-a), the tone of the verb *tsiwóno* shifts onto the first syllable of the genitive noun *wamẹ́zi*. This means that no prosodic boundary separates the verb from the NP.

    The same goes for (6-b), where the tone of the verb *hanị́ka* shifts onto the penultimate syllable of the second object *m̩ɖu-m̩ʃĕ*.

    However, a subject NP phrases separately from the VP. In (7-a), the tone of the subject noun *mlimạ́dji* 'farmer' does not shift on the first syllable of the verb.

(7) a. (m̩-limạdjí)$_\Phi$ (ha-rẹmé pạha)$_\Phi$

        1-farmer 3sg(past)-beat (5-)cat

a farmer beat a cat

    b. *(m̩limạdji hárẹme pạ́ha)$_\Phi$

---

[8]   Some boundaries may be inserted depending on rhythm, speech rate, word order and branchingness parameters that I have not identified precisely enough at this point of my research. Still, the presence of these boundaries is not obligatory.

The prosodic system also presents several complications since some function words are preceded by a phonological phrase boundary. In (8-b) and (8-c), one of the objects is preceded by a so-called augment, a deictic that is related to definiteness/referentiality in Shingazidja. In (8-b), the tone of the verb cannot shift onto the augment or the following noun. This indicates that the augment is preceded by a phonological phrase boundary.

(8) a. (ha-ni̱ka ɲúmɓa̱ m-leví)Φ
3sg(past)-give (9-)house 1-drunkard

he gave a house to a drunkard

b. (ha-ni̱ká)Φ (e ɲumɓa̱ m-lévi̱)Φ
3sg(past)-give AT₉ (9-)house 1-drunkard

he gave the house to a drunkard

c. (ha-ni̱ka ɲúmɓa̱)Φ (e m-leví)Φ
3sg(past)-give (9-)house AT₁ 1-drunkard

he gave a house to the drunkard

Those facts are beyond the scope of this paper and will not be discussed here in detail (see Patin 2007a for this issue).

2.3 Intonational phrases

In (9-a), the tone of the verb – which is prosodically separated from the subject – shifts on the *penultimate* syllable of the object *ʈúnɗa* 'orange', not on its *last* syllable (9-b). On the contrary, the tone of the penultimate syllable of the subject noun *mwá̱na* 'child' shifts to the following syllable, which is the last of the phonological phrase.

(9) a. (e mw-a̱ná)Φ (ali̱ ʈúnɗa)Φ
AT₁ 1-farmer 3sg(past)-eat (5-)orange

the child ate an orange

b. *(e mwa̱ná)Φ (ali̱ ʈunɗá)Φ

The last syllable of a clause, in Shingazidja as in many other Bantu languages, has a special prosodic status. In Shingazidja, the last syllable of a clause cannot be a target for a shifted tone[9]. This restriction, refered here as *extraprosodicity*, indicates that, in Shingazidja, two different prosodic levels exist above the prosodic word: the *phonological phrase* and the *intonational phrase*. The former is the domain of tone shift. The later is characterised by (i.) the extraprosodicity of its last syllable (ii.) by the lowering of the last tone of the group.

In (9-a), the fact that the tone of the subject can shift on its last syllable shows that the subject NP and the VP are not separated by an intonational phrase boundary. Besides, the extraprosodicity does not characterise the end of the VP. If the subject is postposed to the VP[10] (10), the last syllable of the former – being then the last syllable of the clause – is thus extraprosodic, while the last one of the latter – i.e. the object – receives the tone of the verb.

(10) (al<u>i</u>   ṭundá)$_\Phi$ (e   mw-ána)$_\Phi$

   3sg(past)-eat (5-)orange   AT$_1$ 1-child

   the child ate an orange

The domain of the intonational phrase is then wider than the syntactic phrase (now XP) in broad-focus sentences. Most of the time, the intonational phrase corresponds to the whole utterance but its length varies considerably, depending on focus-related parameters (cf. section 3.2).

A more precise representation of the examples (9-a) and (10) is then proposed in (11). ')$_I$' signals the end of an intonational phrase.

(11) a. ((e   mw<u>a</u>ná)$_\Phi$ (al<u>i</u> ṭúnḓa)$_\Phi$)$_I$

   the child ate an orange – cf. (9-a)

   b. ((al<u>i</u> ṭundá)$_\Phi$ (e mw<u>á</u>na)$_\Phi$)$_I$

   the child ate an orange – cf. (10)

To sum up, it has been said in this section that two different prosodic levels exist above the prosodic word level in Shingazidja. The first one is the phonological phrase level, generally corresponding to the XP, and it is determined by tone

---

[9]   Note however that a lexical tone associated to the last syllable of a clause, if not deleted following OCP phenomena, is maintained – cf. (3) or (8-a).

[10] SVO is the unmarked word order. VOS, even though rarely found, is also possible. Other orders are only possible if the verb presents an object marker – i.e. a resumptive pronoun.

shift. The second one is the intonational phrase level and is most of the time coextensive with the whole utterance. However, morphosyntactic parameters are not enough to account for all the phrasing situations observed in Shingazidja, as it will be demonstrated in the following section.

3   Focus and Phrasing

In the preceding section, I showed that phrasing is conditioned by syntactic considerations in Shingazidja. This section explores the influence of focus on phrasing. Following Jackendoff (1972), Watters (1979) and many others, I will consider in this paper that focus *denotes the information that is assumed by the speaker not to be (entirely) shared by him and the hearer*.

First, I will show that focus is expressed by phonological phrasing in Shingazidja. Then, I will briefly explain how intonational phrases are driven by the contrastive focus.

3.1   Narrow focus

This section discuss the influence of focus on phrasing in Shingazidja. At this point of my research, I did not found evidence that focus may be expressed by a specific word or morpheme, or by a specific word order[11].

3.1.1   *Narrow focus and phonological phrases*

In the previous section, it has been claimed that a maximal syntactic phrase basically corresponds to one phonological phrase in Shingazidja. However, any lexical item[12] may be focused in situ by being followed by a phonological phrase boundary.

For instance, while the VP constitutes a single phonological phrase in (12-a), the in situ focused noun is prosodically separated from its genitive modifier in (12-b).

(12) a. (tsi-w<u>o</u>no   n-dovu   y-a wá-m<u>e</u>zi)_Φ

       1sg(past)-see 9-elephant 9-of 2-beggar

    I saw an elephant of the beggars

---

[11]   Variations in word order seem to be linked to topicalisation or stylistic parameters.

[12]   To focus a grammatical element (e.g. an augment) or a morpheme, other strategies – the insertion of a tone (briefly discussed in Patin 2007b), or an intensity peak – seem to be selected. Further research will be necessary to identify precisely the nature and the range of these alternative strategies.

b. (tsiw<u>o</u>no ndovú)_Φ (ya wam<u>é</u>zi)_Φ

I saw an ELEPHANT *of the beggars*

(answering the question: *what did you see of the beggars?*)

In (12-b), the tone of the verb *tsiwóno* shifts to the last syllable of the head noun *ndóvu* while it shifts up to the genitive noun in (12-a). This means that there is a phonological phrase boundary following the head noun in (12-b). The presence of this boundary seems to be obligatory in this context.

3.1.2   *Object focus*

To focus the first object of a VP, it is necessary to place a phonological phrase boundary after it. Compare for instance (13-a) with (13-b).

(13)   a. (ha-n<u>i</u>ka   n-dovu   mlév<u>i</u>)_Φ   (VP focus)

        3sg(past)-give   9-elephant 1-drunkard

        he gave an elephant to a drunkard

b. (han<u>i</u>ka   ndovú)_Φ   (mlev<u>i</u>[13])_Φ   (object focus)

        he gave an ELEPHANT *to a drunkard*

(answering the question: *what did he gave to a drunkard?*)

In (13-a), the tone of the verb *haníka* shifts to the penultimate syllable of the second object of the VP. In (13-b), however, the shift stops on the last syllable of the first one, indicating that the first object is followed by a prosodic break.

This result is comparable to what was observed in Chicheŵa – see (1-b) above – or in Bemba. In this language, a low tone marks the end of the phonological phrase[14]. In (14-b), a low tone appears on the last syllable of the focused object, indicating that it is followed by a prosodic break.

---

[13]   When a single word follows a focused element, it tends to loose its lexical tone – but see (15-a-ii).

[14]   Referred to as a *major phrase* in Kula (2007). In (14), SM is used for subject marker and TNS for tense.

(14) Bemba   (Kula 2007: 203)

  a. (n-alí-mónóómwááná   mucímuti)$_{MaP}$ (VP focus)

     1SM-TNS-see.1child   18tree

    I saw a child in the tree

  b. (n-alí-mónóómwáána)$_{MaP}$ (mucímuti)$_{MaP}$ (object focus)

     1SM-TNS-see.1child   18tree

    I saw a child in the tree

3.1.3 Verb focus

As expected from the data above, a focused verb is also separated from its complements with a prosodic break. Compare for instance (15-a-i) and (15-b-i) with, respectively, (15-a-ii) and (15-b-ii).

(15) a. i. (ha-pi̱ha   djánze̱)$_Φ$

       3sg(past)-cook (5-)crab

      he cooked a crab

    ii. (hapi̱há)$_Φ$ (djanzé̱)$_Φ$

      he COOKED *a crab*

  b. i. (tsi-wo̱no   má-βa̱ha ma-ili̱ y-á ha̱he)$_Φ$

      1sg(past)-see 6-cat   6-two   6-of his

     I saw his two cats

    ii. (tsiwo̱nó)$_Φ$ (maβa̱ha maíli̱   ya   há̱he)$_Φ$

     I SAW *his two cats*

In both (15-a-ii) and (15-b-ii) the tone of the focused verb is not allowed to shift on the object noun. Again, this is consistent with what has been said in the previous sections, or with what was observed in other Bantu languages.

    However, Shingazidja differs from Chicheŵa on a particular point. Remember that in this language – see (16), extracted from (1) – a prosodic break is inserted between two objects when the verb is focused.

(16) Chicheŵa

(anaméenyá) (nyuúⁿbá) (ⁿdí mwáála)

<small>HE HIT</small> *the house with a rock* (Kanerva 1990: 98)

In Shingazidja, however, the first object of a VP does not phrase separately from the second one if the verb is focused – (17-b).

(17) a. (ha-ni̱ka ɲúm6a̱ m-leví̱)_Φ (VP focus)

3sg(past)-give (9-)house 1-drunkard

he gave a house to a drunkard

b. (hani̱ká)_Φ (ɲum6a̱ mlévi̱)_Φ (verb focus)

<small>HE GAVE</small> *a house to a drunkard*

In (17-b), the tone of the verb does not shift on the first object – as in (17-a) – since the verb is followed by a prosodic break. However, the tone of the first object *ɲum6á* shifts on the penultimate syllable of the second object *mleví*.

On this particular point Shingazidja differs from Chicheŵa and behaves like Chitumbuka. In this language too, both objects of a VP will phrase together (18).

(18) Chitumbuka (Downing 2006: 72)

([β]a-mbwéengu) ([β]a-ku-lyáa-*so*) (ma-gwáfya m-ma-kúuni)

2-monkey 2-tam-eat-again 6-guavas in-6-trees

The monkeys are eating again the guavas in the trees

The major correlate of phrasing in Chitumbuka is the lengthening of the penultimate syllable of a phonological phrase. In (18), the subject has a long penultimate, since it does not phrase with the VP as in Shingazidja. The last word of the clause also presents a long penultimate, so does the verb, because it ends with the focus particle *–so*. However, the absence of length on the penultimate syllable of the object *magwáfya* indicates that this word is not followed by a prosodic break.

3.2 Contrastive focus

In the previous section, it has been said that a focused lexical word is followed by a phonological phrase boundary (19-b), which stops the shift of the tone.

However, if a word bears a *contrastive* focus, a shifted tone does not stop on its last syllable, but on its penultimate syllable (19-c). In Shingazidja, an *intonational phrase boundary* is added at the right of a word that bears a contrastive focus.

(19) a. ((tsi-w<u>o</u>no   n-dovu   y-a wá-m<u>e</u>zi)$_\Phi$)$_I$

        1sg(past)-see 9-elephant   9-of 2-beggar

    the child ate an orange – cf. (9-a)

   b. ((tsi-w<u>o</u>no ndovú)$_\Phi$ (ya wa-m<u>é</u>zi)$_\Phi$)$_I$

    I saw an ELEPHANT *of the beggars*

    (answering the question: *what did you see of the beggars?*)

   c. ((tsi-w<u>o</u>no ndóvu)$_\Phi$)$_I$ ((ya wa-m<u>é</u>zi)$_\Phi$)$_I$

    No !) I saw an <u>ELEPHANT</u> *of the beggars*

    (answering the question: *did you see the horse of the beggars?*)

In the section 2.3, I claimed that the intonational phrase level is marked in Shingazidja by the extraprosodicity of its last syllable. In (19-c), the tone of the verb *tsi-w<u>ó</u>no* stops on the penultimate syllable of *ndóvu*, and does not shift on the last syllable of the noun. Since *ndóvu* has no underlying tone on its last syllable, it means that the last syllable is 'extraprosodic'. Thus, the word *ndóvu* is the last word of an *intonational phrase*. Evidence for such an analysis is the optional presence of a pause after the focused item in sentences like (19-c).

4   An OT analysis of phrasing in Shingazidja

In this section, I will provide an OT analysis of the data presented in the previous sections. I will only discuss here broad-focus and narrow-focus situations.

   First (sections 4.1 and 4.2), I will discuss broad-focus sentences, showing that Shingazidja may be accounted for using alignment constraints (Selkirk 1995) and the constraint WRAP-XP proposed by Truckenbrodt (1995, 1999). Then (sections 4.3 and 4.4), situations that involve a narrow focus will be analysed, using another constraint proposed by Truckenbrodt: ALIGN FOC. I will insist on the verb-focus sentences, where Shingazidja differs from Chicheŵa.

4.1 Broad- focus sentences: basic phrasing

Based on previous works by Clements (1978) and Chen (1985), Selkirk (1986) and Selkirk & Shen (1990) proposed a model called *end-based theory*. This model aligns the right or left[15] edge of phonological phrases with the right or left edge of syntactic XPs. Later, Selkirk (1995) has reanalysed her model in the format of Optimality Theory, and has reformulated the parameters as alignment constraints. These constraints are presented in (20).

(20) Alignment constraints (Truckenbrodt 1999: 223, following Selkirk 1995: 469, 477)
    a. ALIGN-XP, R: ALIGN (XP, R; P, R)

        For each XP there is a P such that the right edge of XP coincides with the right edge of P

    b. ALIGN-XP, L: ALIGN (XP, L; P, L)

        For each XP there is a P such that the left edge of XP coincides with the left edge of P

In Shingazidja, the subject NP phrases separately from the VP (21).

(21) (m-leví)_Φ   (ha-djá)_Φ
     1-drunkard   3sg(past)-come
   a drunkard came

To account for the prosodic pattern of sentences like (21), I will consider that the constraint ALIGN-XP,R (cf. 20-a) is dominant in Shingazidja. Such an analysis was previously proposed to account for several other Bantu languages, such as Chicheŵa (Truckenbrodt 1999), Chitumbuka (Downing 2006), Chimwiini (Kisseberth 2000) or Xhosa (Zerbian 2004). As a first step, I will simply say that ALIGN-XP,R is ranked higher than the structural constraint *P-PHRASE, defined in (22).

(22) *P-PHRASE (Truckenbrodt 1999: 228)
    Avoid p-phrases altogether

---

[15] Depending on the language. For instance, the left edge is the relevant parameter In Ewe (Clements 1978), while the right edge is the relevant parameter in Chimwiini (Selkirk 1986).

Cédric Patin

In the tableau (23), the candidate *b* is rejected because it violates the constraint ALIGN-XP,R. The candidate *a*, which presents a phonological phrase boundary after the subject NP, emerges as the optimal candidate.

(23) Shingazidja

[NP VP]_FP	ALIGN-XP,R	*P-PHRASE
a. ☞ ( )_Φ ( )_Φ		**
b. ( )_Φ	NP !	*

There is no prosodic break between the verb and an indefinite object – see (15-a-i), repeated in (24).

(24) (ha-pi̱ha   djánze̱)_Φ

    3sg(past)-cook (5-)crab

  he cooked a crab

In (25), the proposed ranking selects the appropriate candidate *b* over the candidate *a*, which splits the VP.

(25) Shingazidja

[V NP]_VP	ALIGN-XP,R	*P-PHRASE
a. ( )_Φ ( )_Φ		** !
b. ☞ ( )_Φ		*

However, this ranking alone is not sufficient to account for all the attested phrasing patterns of Shingazidja. When a VP contains two objects – see (6-b), repeated in (26), the proposed ranking will incorrectly[16] predict that a phonological phrase boundary is inserted after the first object *ɲáma* (27).

(26) (ha-ni̱ka   ɲama   m̩-d̪u-m̍-ʃe̱)_Φ

    3sg(past)-give 9/meat   1-person-1-woman

  he gave (some) meat to a woman

(27) *Shingazidja

[V NP NP]_VP	ALIGN-XP,R	*P-PHRASE
a. 💣 ( )_Φ ( )_Φ		**
b. ( )_Φ	* !	*

---

16   The selection of the 'wrong' candidate is marked by a bomb.

To account for double-object constructions, another constraint is thus necessary: WRAP-XP, which will be introduced in the next section.

4.2 Broad-focus sentences: double object constructions

To account for the fact that the two objects of a VP phrase together in Chicheŵa and Papago (Uto-Aztecan; Hale & Selkirk 1987), Truckenbrodt (1999: 228) introduced the constraint WRAP-XP, defined in (28).

(28) WRAP-XP (Truckenbrodt 1999: 228)

Each XP is contained in a phonological phrase

When WRAP-XP is ranked higher than alignment constraints, it prevents the insertion of a prosodic boundary inside any maximal XP. Then, the whole VP, or the whole NP, will phrase together. When WRAP-XP is ranked lower than alignment constraints, prosodic boundaries will be inserted inside the maximal XP, e.g. after each NP that it may contain.

However, a kind of prosodic structures does not violate either WRAP-XP nor alignment constraints: *recursive structures*. The recursive structure proposed in (29) does not violate WRAP-XP, since the maximal XP (XP$_1$) is contained in a single phonological phrase. Furthermore, it does not violate ALIGN-XP,R since a phonological phrase boundary follows the first embedded XP (XP$_2$). It has been demonstrated that recursive structures exist in the Bantu languages Kimatuumbi (Truckenbrodt 1999) and Chimwiini (Kisseberth 2000).

(29)   [X$_1$ XP$_2$ XP$_3$]XP$_1$
       ((   )$_\Phi$)$_\Phi$

In Chicheŵa and Shingazidja, however, such a structure must be excluded, since no prosodic break separates the first object from the second object. To avoid recursive structures in Chicheŵa, Truckenbrodt (1999) proposed that the constraint NONRECURSIVITY, defined in (30), is high-ranked in this language.

(30)   NONRECURSIVITY (Truckenbrodt 1999: 240)

Any two p-phrases that are not disjoint in extension are identical in extension

This constraint "punishes recursive structure to the extent that the two elements of the recursive structure differ in extension" (Truckenbrodt 1999: 240-241). For instance, in (29), XP_3 constitutes a violation of the constraint NONRECURSIVITY. In (31), the tableau proposed by Truckenbrodt (1999) to account for the phrasing of double-object constructions in Chicheŵa is presented. Candidates that present a phonological phrase boundary after the first object of the VP are ruled out by the constraints WRAP-XP and NONRECURISVITY.

(31) Chicheŵa (Truckenbrodt 1999: 246), Shingazidja

	[V NP YP]$_{VP}$	WRAP-XP	NONREC	ALIGN-XP,R	*P-PHRASE
a.	( )$_p$ ( )$_p$	VP !			**
b.	(( )$_p$)$_p$		*!		**
c. ☞	( )$_p$			NP	*

The candidate *a* violates WRAP-XP since the VP is split in two different prosodic phrases. It is then excluded. The candidate *b* does not violate WRAP-XP since the whole VP is contained in a single phonological phrase. However, it violates the constraint NONRECURSIVITY, and it is then excluded. NONRECURSIVITY has to be high-ranked in Chicheŵa to prevent the insertion of a prosodic boundary after the first object of the VP.

Since Chicheŵa and Shingazidja do not differ vis-à-vis the phrasing pattern of double-object constructions, I will retain temporarily Truckenbrodt's analysis.

4.3 Narrow focus

In Shingazidja, as in Chicheŵa, a focused word is followed by a phonological phrase boundary – see the example (15-a-ii), repeated in (32).

(32) (hap<u>i</u>há)$_Φ$ (djanz<u>é</u>)$_Φ$

    3sg(past)-cook (5-)crab

    he COOKED *a crab*

To capture this fact, I will consider – following previous analyses of similar facts in several other Bantu languages[17] – that the constraint ALIGN-FOC, proposed by Truckenbrodt (1999) and defined in (33), is ranked higher than *basic alignment* constraints in Shingazidja.

---

[17] E.g. Truckenbrodt (1999) on Chicheŵa, Downing (2006, 2007) on Chitumbuka, Zerbian (2004) on Xhosa, etc.

(33)   ALIGN-FOC = ALIGN(FOC, R; P, R) (Truckenbrodt 1999: 248)
         Each focused constituent is right-aligned with a p-boundary

The analysis of VPs presenting a focused element is presented in the tableaux (34) and (36). At this point of the discussion, for reasons that will be clear later (cf. the section 4.4), I will not represent here the constraint NONRECURSIVITY.

   In (34), the candidate *b*, which respects basic alignment constraints, is ruled out since it does not present a phonological phrase boundary after the focused element, i.e. the verb. The candidate *a*, even if it violates the constraint WRAP-XP, does not violate the high-ranked constraint ALIGN-FOC. Then, it is selected.

(34) Shingazidja

[V_FOC NP]_VP	ALIGN-FOC	WRAP-XP	ALIGN-XP,R	*P-PHRASE
a. ☞ ( )_Φ ( )_Φ		*		**
b. ( )_Φ	* !			*

To allow the insertion of a prosodic boundary after a focused word, ALIGN-FOC has to outrank WRAP-XP.

   The proposed ranking also selects the 'good' candidate when the first object is focused – see the example (13-b), repeated in (35). In the tableau (36), the high-ranked constraint ALIGN-FOC selects as optimal the candidate *a*.

(35) b. (han<u>i</u>ka   ndovú)_Φ   (mlev<u>i</u>)_Φ   (object focus)
          3sg(past)-give   9-elephant   1-drunkard

      he gave an ELEPHANT *to a drunkard*

(36) Shingazidja

[V NP_FOC NP]_VP	ALIGN-FOC	WRAP-XP	ALIGN-XP,R	*P-PHRASE
a. ☞ ( )_Φ ( )_Φ		*		**
b. ( )_Φ	* !		*	*

However, this ranking alone is not sufficient to account for all the situations that have been discussed in this paper. When the verb of a double-object VP is focused – see (17-b), repeated in (37) –, the proposed ranking will incorrectly predict that a prosodic boundary has to be inserted after the first object of the VP – cf. (38).

(37)   (haniká)_Φ   (ɲumɓa   mlévi)_Φ   (verb focus)

         3sg(past)-give   (9-)house 1-drunkard

         HE GAVE *a house to a drunkard*

(38) *Shingazidja

[V_FOC NP NP]_VP	ALIGN-FOC	WRAP-XP	ALIGN-XP,R	*P-PHRASE
a. ( )_Φ ( )_Φ	* !	*		**
b. ( )_Φ	* !		*	*
c. ☌ ( )_Φ ( )_Φ ( )_Φ		*		***
d. ( )_Φ ( )_Φ		*	* !	**

WRAP-XP is a *categorical* constraint (Truckenbrodt 1995, 1999; Samek-Lodovici 2005): it can only be violated once, since its effect is neutralised as soon as a maximal XP is split in two different prosodic phrases. In (38), the candidate *c* is selected because it does not violate the constraint ALIGN-XP,R, even if the whole VP is split in three different phrases. The 'good' candidate *d* is ruled out because it violates this same constraint ALIGN-XP,R, even if the VP is minimally split.

       Evidence for the categorical aspect of WRAP-XP comes from Chicheŵa data. In Chicheŵa (cf. 3.1.3), both the focused verb and the first object of a double-object construction are followed by a prosodic boundary, while all the words of the VP phrase together in broad-focus sentences. The model proposed by Truckenbrodt correctly predicts that the boundary that follows the first object will be allowed to emerge as soon as WRAP-XP is neutralised.

To account for the specificity of Shingazidja, I will then propose that focused elements are embedded in recursive structures.

4.4 Verb focus in double object constructions: the hidden recursivity hypothesis

The analysis of Chicheŵa proposed by Truckenbrodt is presented in (39).

(39) Chicheŵa (Truckenbrodt 1999: 249)

[V_FOC NP PP]_VP	ALIGN-FOC	NON REC	WRAP-XP	ALIGN-XP,R	*P-PHRASE
a. ( )_p	*!			NP	*
b. ( )_p ( )_p			VP	NP!	**

c. ☞ ($)_p$ ($)_p$ ($)_p$		VP		***
d. (($)_p$ $)_p$	NP ! PP		NP	**
e. ((($)_p$ $)_p$ $)_p$	NP ! PP PP			***

The tableau (39) is similar to the one proposed in (38) except that the constraint NONRECURSIVITY is added. The critical point, here, is the *ranking of NONRECURSIVITY*. Truckenbrodt (1999: 249-250) precise:

> "To secure this result, a crucial ranking needs to be added that was not previously established for Chicheŵa: NONRECURSIVITY here has to outrank WRAP-XP. This can be seen by comparing candidates (c) and (d). The recursive structure in (d) allows ALIGN-FOC and WRAP-XP to be met simultaneously. However, candidate (d) must be ruled out since it does not include a right p-boundary after the first object, contrary to fact. Any constraint ruling out candidate (d) has to do so against the demands of WRAP-XP. Thus, (d) cannot be ruled out by ALIGN-XP,R, which is ranked below WRAP-XP for independent reasons […]. NONRECURSIVITY, then, is the only constraint at hand that can rule out the phrasing in (d). For it to do so, NONRECURSIVITY has to be ranked above WRAP-XP […]. This ranking then also eliminates other recursive candidates such as (e)."

To account for Shingazidja facts, I will then adopt the ranking rejected by Truckenbrodt in his analysis of Chicheŵa. I propose that Shingazidja presents recursive phonological phrases as soon as focus is involved.

Formally, I will suppose that WRAP-XP is ranked higher than NONRECURSIVITY in Shingazidja. This ranking selects as optimal the *recursive* candidate that does not include a prosodic break after the first object of the VP – cf. (40).

(40) Shingazidja

[V$_{FOC}$ NP$_1$ NP$_2$]$_{VP}$	ALIGN-FOC	WRAP-XP	NONREC	ALIGN-XP,R	*P-PHRASE
a. ( )$_\Phi$	*!			NP$_1$	*
b. ( )$_\Phi$ ( )$_\Phi$		VP !		NP$_1$	**
c. ( )$_\Phi$ ( )$_\Phi$ ( )$_\Phi$		VP !			***
d ☞ (( )$_\Phi$)$_\Phi$			NP$_1$ NP$_2$	NP$_1$	**
e. ((( )$_\Phi$)$_\Phi$)$_\Phi$			NP$_1$ NP$_2$ NP$_2$!		***

In (40), the candidates *b* and *c* are excluded because they violate the high-ranked constraint WRAP-XP. To respect both ALIGN-FOC and WRAP-XP, a candidate thus has to present a recursive structure. Then, the candidate *d* is selected because it minimally violates the constraint NONRECURSIVITY.

When the first object bear a focus (cf. 35 above), the ranking also selects as optimal a recursive candidate. In (41), the candidate *b*, which was selected as optimal in the tableau (36), is now rejected because it violates the constraint WRAP-XP.

(41) Shingazidja

[V NP$_{FOC}$ NP$_2$]$_{VP}$	ALIGN-FOC	WRAP-XP	NONREC	ALIGN-XP,R	*P-PHRASE
a. ( )$_\Phi$	*!			NP$_{FOC}$	*
b. ( )$_\Phi$ ( )$_\Phi$		VP !			**
c. ☞ (( )$_\Phi$)$_\Phi$			NP$_2$		**
d. (( )$_\Phi$)$_\Phi$			V NP$_2$!		**

Two candidates – *c* and *d* – respect both the constraint ALIGN-FOC and the constraint WRAP-XP. The candidate *d* violates twice the gradient constraint NONRECURSIVITY: both the verb and the second object are excluded from the embedded phonological phrase. The candidate *c* is then selected.

To sum up, I assume that the difference in phrasing between Chicheŵa and Shingazidja is related to the respective ranking of WRAP-XP and NONRECURSIVITY. While the former is dominated by the later in Chicheŵa (42-a), the opposite ranking is proposed to account for Shingazidja data (42-b).

(42) Chicheŵa and Shingazidja rankings

   a. Chicheŵa (Truckenbrodt 1995, 1999):

ALIGN-FOC >> NONRECURSIVITY >> WRAP-XP >> ALIGN-XP,R
>> *P-PHRASE
b. Shingazidja:
ALIGN-FOC, WRAP-XP >> NONRECURSIVITY >> ALIGN-XP,R >>
*P-PHRASE

6   Summary

In this paper, I have presented the main correlates of phrasing in Shingazidja. I have shown that focus conditions the phonological phrasing in Shingazidja. A phonological phrase boundary is inserted after a focused lexical item, resulting in restrictions on tone shift. This result is consistent with similar results obtained in several other Bantu languages.

The data was analysed in an OT-based framework, following Truckenbrodt's famous analysis of Chicheŵa. However, Shingazidja differs from Chicheŵa by phrasing together the two objects of a verb-focus sentence. I showed that the model can account for this specificity by reranking the constraint WRAP-XP over the constraint NONRECURSIVITY.

References

Cassimjee, Farida & Charles Kisseberth (1998). Optimal Domains Theory and Bantu Tonology: a Case Study from Isixhosa and Shingazidja. In: Larry Hyman & Charles Kisseberth (eds.), *Theoretical Aspects of Bantu Tone*. Stanford: CSLI, 33-132.

Cassimjee, Farida & Charles Kisseberth (1993). The phrasal tonology of Shingazidja. 24th Annual Conference on African Linguistics. Ohio State University, July 23-25 1993.

Cassimjee, Farida & Charles Kisseberth (1992). Metrical Structure in Shingazidja. *CLS* 28, 72-93.

Cassimjee, Farida & Charles Kisseberth (1989). Shingazidja Nominal Accent. *Studies in the Linguistic Sciences* 19.1., 33-61.

Clements, Georges N. (1978). Tone and syntax in Ewe. In Donna Napoli (ed.), *Elements of Tone, Stress, and Intonation*. Washington, DC: Georgetown University Press, 21-99.

Chen, M. (1985) The syntax of phonology: Xiamen tone sandhi. MS, University of California, San Diego.

Cheng, Lisa & Laura J. Downing (2007). The prosody and syntax of Zulu relative clauses. *SOAS working papers in linguistics* 15, 51-63.

Downing, Laura (2007). Focus prosody divorced from stress and intonation in Chichewa, Chitumbuka and Durban Zulu. ICPhS 2007 Satellite Meeting, Intonational Phonology: Understudied or Fieldwork Languages, Saarbrücken, August 5 2007.

Downing, Laura (2006). The prosody and syntax of focus in Chitumbuka. *ZAS Papers in Linguistics* 43, 55-79.

Downing, Laura (2002). Fitting focus into the prosodic hierarchy. *SOAS Working Papers in Linguistics*, 111-133.

Downing, Laura, Al Mtenje & Bernd Pompino-Marschall (2005). Prosody and Information Structure in Chichewa. *ZAS Papers in Linguistics* 37, 167-186.

Guthrie, Malcolm (1967–71) *Comparative Bantu: an introduction to the comparative linguistics and prehistory of the Bantu languages*. 4 vols. Farnborough: Gregg Press.

Hale, Ken & Elisabeth Selkirk (1990). Government and Tonal Phrasing in Papago. *Phonology Yearbook* 4, 151-183.

Jackendoff, Ray (1972). *Semantic Interpretation in Generative Grammar*. Cambridge: MIT press.

Jokweni, Mbulelo (1995). *Aspects of Isixhosa Phrasal Phonology*. PhD dissertation, Urbana-Champaign, University of Illinois.

Jouannet, Francis (1989). *Des tons à l'accent. Essai sur l'accentuation du comorien*. Université de Provence Aix-Marseille.

Kanerva, Jonni (1990). *Focus and Phrasing in Chichewa Phonology*. New York: Garland.

Kisseberth, Charles (2007). Optimality Theory and the Theory of Phonological Phrasing: The Chimwiini Evidence. *The Sound Patterns of Syntax*, Research workshop of the Israel Science Foundation on the Syntax-Phonology Interface, Ben-Gurion University, June 11-13, 2007.

Kisseberth, Charles & Mohammad Abasheikh (2004). *The Chimwiini Lexicon Exemplified*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa (ILCAA).

Kula, Nancy (2007). Effects of phonological phrasing on syntactic structures. *The Linguistic Review* 24.2-3: *Special Issue : Prosodic Phrasing and Tunes*, 201-231.

McCarthy, John & Alan Prince (1993a). Generalized Alignment. *Yearbook of Morphology* 1993, 79-153.

McCarthy, John & Alan Prince (1993b). *Prosodic Morphology I: Constraint interaction and satisfaction*. Ms, University of Massachusetts, Amherst & Rutgers University.

Patin, Cédric (2007a). *La tonologie du shingazidja, langue bantu (G44a) de la Grande Comore : nature, formalisation, interfaces*. Thèse de doctorat, Université Paris 3.

Patin, Cédric (2007b). Shingazidja focus hierarchy. *Nouveaux Cahiers de linguistique francaise* 28 *Interfaces discours-prosodie*, actes du 2ème Symposium international. Geneva, Switzerland, September 12-14 2007, 147-154.

Philippson, Gérard (2005). Pitch accent in Comorian and Proto-Sabaki tones. In Koen Bostoen & Jacky Maniacky (eds.), *Studies in African Comparative Linguistics with special focus on Bantu and Mande: Essays in Honour of Yvonne Bastin & Claire Grégoire*. Tervuren, Musée Royal de l'Afrique Centrale, 199-220.

Philippson, Gérard (1991). *Tons et accent dans les langues bantu d'Afrique orientale : étude typologique et diachronique*. Thèse d'Etat, Université Paris V René Descartes, Paris.

Philippson, Gérard (1988). L'accentuation du comorien : essai d'analyse métrique. *Etudes Océan Indien* (Paris) 9, 35-79.

Prince, Alan & Paul Smolensky (1993) *Optimality Theory: Constraint Interaction in Generative Grammar*. Rutgers University Center for Cognitive Science Technical Report 2.

Rey, Véronique (1990). *Approche phonologique et expérimentale des faits d'accent d'une langue africaine, le shingazidja (parler de la Grande Comore)*. Thèse de doctorat, Université Aix-Marseille 1.

Samek-Lodovici, Vieri (2005). Prosody Syntax Interaction in the Expression of Focus. *Natural Language and Linguistic Theory* 23, 687-755.

Selkirk, Elisabeth (1995). The prosodic structure of function words. In Jill Beckman, Suzanne Urbanczyk & Laura Walsh (eds.), *Optimality Theory Occasional Papers: UMOP* 18. Amherst: UMAS, 439-470.

Selkirk, Elisabeth (1986). On derived domains in sentence phonology. *Phonology* 3, 371-405.

Selkirk, Elisabeth & Tong Shen (1990). Prosodic domains in Shanghai Chinese. In Sharon Inkelas & Draga Zec (eds.), *The Phonology-Syntax Connection*. Chicago: Univerity of Chicago Press, 313-338.

Truckenbrodt, Hubert (1999). On the Relation between Syntactic Phrase and Phonological Phrases. *Linguistic Inquiry* 30, 219-255.

Truckenbrodt, Hubert (1995). *Phonological Phrases: Their Relation to Syntax, Focus, and Prominence*. Ph.D. dissertation, MIT.

Tucker, Archibald & Margaret Bryan. (1970). Tonal classification of nouns in Ngazija. *African Language Studies* 11, 351-383.

Watters, John (1979). Focus in Aghem: a study of its formal correlates and typology. In Larry Hyman (ed.), *Aghem grammatical structure. Southern California Occasional Papers in Linguistics* 7. Los Angeles: University of Southern California, 137-97.

Zerbian, Sabine (2006). *Expression of Information Structure in the Bantu Language Northern Sotho*. PhD dissertation, Humboldt University, Berln.

Zerbian, Sabine (2004). Phonological Phrasing in Xhosa. In Susanne Fuchs & Silke Hamann (eds.), *ZAS Papers in Linguistics* 37, 71-100.

ZAS Papers in Linguistics were originally published by the Forschungsschwerpunkt Allgemeine Sprachwissenschaft, Typologie und Universalienforschung (FAS, Research Center for General Linguistics, Typology and Universals). The Center is now known as *Zentrum für Allgemeine Sprachwissenschaft, Typologie und Universalienforschung* (ZAS) under the auspices of the Deutsche Forschungsgemeinschaft (The German Research Foundation) and the State of Berlin. The Center currently has research projects in syntax, semantics, morphology, phonology, phonetics as well as language contact and language acquisition. ZAS provides a forum for the exchange of ideas in the academic community of the Berlin area through lectures, seminars, workshops and conferences. The Center cooperates with other universities in Germany, and sponsors visits by scholars from Europe and America.

Director: Manfred Krifka

For further information about ZAS, please consult our website:

http://www.zas.gwz-berlin.de

or write to:

Manfred Krifka, Director
Zentrum für Allgemeine Sprachwissenschaft
Schützenstr. 18
D-10117 Berlin
Germany

E-mail: krifka@zas.gwz-berlin.de

ZAS Papers in Linguistics reflect the ongoing work at ZAS. They comprise contributions of ZAS researchers as well as visiting scholars. Issues are available on an exchange basis or on request. For further information, please write to:

Sekretariat
Zentrum für Allgemeine Sprachwissenschaft
Schützenstr. 18
D-10117 Berlin
Germany

E-mail: sprach@zas.gwz-berlin.de
Phone: +49 30 20 19 24 04
Fax: +49 30 20 19 24 02

Later issues can also in part be downloaded at the ZAS website:

http://www.zas.gwz-berlin.de/index.html?publications_zaspil

Cover design: Mathias Krüger, Mechthild Bernhard and the CMS, HU Berlin.