ZASPiL Nr. 41 – August 2005

# Articulatory correlates of the voicing contrast in alveolar obstruent production in German

Susanne Fuchs

This work is a slightly modified version of my PhD thesis carried out in collaboration between the Centre for General Linguistics (ZAS) in Berlin and Queen Margaret University College (QMUC) in Edinburgh.

## Acknowledgments

This work was supported by the German Research Council (DFG) and the Engineering and Physical Sciences Research Council (EPSRC).
First, I like to thank Christine Mooshammer, since she was the one who convinced me in a very clever way to take the position at the ZAS and write a thesis. She even did this after I had destroyed the formats of her thesis due to a macro virus. There must have been some belief in me. Thanks Tine!
I like to acknowledge my supervisory team, who supported me wherever I was or they were situated. In addition, I am quite happy that they seem to like each other and although it happened that one said X and the other one Y they were primarily interested in my work. I am also relieved that I had these supervisors since they had to deal with a rather independent, passionate, and sometimes defiant person and I guess it was not always easy.
Thanks to Alan Wrench for giving my the EPSRC grant and staying calm when I got stressed, to Jim Scobbie for his helpful discussions, comments, e-mails and his humour (your humour kept me alive), to Bernd Pompino-Marschall for doing corrections so rapidly, and sorry for calling so long with you that your pancakes got burnt (I will cook for you some day), to Christine Mooshammer for Matlab scripts, for teaching me kinematics from A to Z, e-mails, comments and always keeping interested in this work.
Beside the four official supervisors there were three other important persons, Pascal Perrier, Phil Hoole and Klaus Dahlmeier. Thanks to Pascal Perrier for giving me the CD in memory of Christian Benoit, for the introduction to speech motor control, for an open ear, lots of inspiration, advice and your smile. Thanks to Phil Hoole for his advice concerning the transillumination recording, further post-processing, for quick responses and for letting me copy kilos of articles with his copy card. Thanks to Klaus Dahlmeier for doing the transillumination recording and his straightforwardness.
Thanks to Gordon Ramsay for long e-mail discussions. Your amazing knowledge of literature kept my brain steaming.
Thanks to my subjects CG and JD for participating (you are brave!), to Dirk Fischer and Anke Busler for some help with the acoustic analysis and final editing.
Thanks to Fay Windsor and Laura Koenig for correcting my funny English, to Jörg Dreyer in preparing recordings, to Sabine Koppetsch sharing a room with me and listening in good and bad times, to Mariam Hartinger, Daniel Pape, Jana

Brunner, Christian Geng, Anke Busler, Johanna Creutzburg, Olga Gordeeva, Ben Matthews, Lucy Ellis, and Fay Windsor for being good spirits to me. Thanks to the Multi-Media Centre of the Humboldt-University which digitalised my fiberoptic films.

During the last years I visited some other labs and met people who inspired me in several ways. I like to thank Anders Löfqvist for making my stay at Haskins possible and having a look at my transillumination data, to Mark Tiede, Doug Whalen, Khalil Iskarous, Leigh Lisker (your research idealisms still impresses me) and Arthur Abramson for a nice time there, to Joe Perkell for showing me the speech communication group and their work at MIT, to Osamu Fujimura for inviting me to Columbus/Ohio and Caroline Menezes for her warm welcome. Thanks to my friends who had a hard time with me and probably hope that I'm getting normal again (I am not sure): Hélène Bezencon, Kerstin Köhler, Harald Orth, Karin Igbal Bhatti, and Julia Riebe. Thanks to Kerstin Köhler, Ineke Mennen and Hélène Bezençon for filling my stomach with wonderful food. Thanks to Heidemarie Fuchs, Dieter Fuchs, Irmgard Fuchs, Carl Fuchs, Lena Wünsch, and Heinz Wünsch for being or having been there at several stages of my life, and to my little wonder Jolanda who has arrived.

**“Voicing is something like beauty.”** Jim Scobbie

## Prolog

When somebody would ask me what beauty would mean, I would probably answer: somebody with humour,
somebody who likes love and life,
somebody who smells good,
somebody having a body like the sculpture I’ve got at home.
I would say, well, its difficult to describe that term everybody seems to know.
Sometimes it is “only” the way somebody smiles,
the way somebody walks..........,
in the end it depends on the situation.
Reading journals I get the impression there is some standard about beauty,
but is this real beauty?
Jim, when you told me what voicing is, I had to laugh about it,
but after all, I guess you are right.

# Abstract

This work investigates laryngeal and supralaryngeal correlates of the voicing contrast in alveolar obstruent production in German. It further studies laryngeal-oral co-ordination observed for such productions. Three different positions of the obstruents are taken into account: the stressed, syllable initial position, the post-stressed intervocalic position, and the post-stressed word final position. For the latter the phonological rule of final devoicing applies in German. The different positions are chosen in order to study the following hypotheses:

1. The presence/absence of glottal opening is not a consistent correlate of the voicing contrast in German.
2. Supralaryngeal correlates are also involved in the contrast.
3. Supralaryngeal correlates can compensate for the lack of distinction in laryngeal adjustment.

Including the word final position is motivated by the question whether neutralisation in word final position would be complete or whether some articulatory residue of the contrast can be found.

Two experiments are carried out. The first experiment investigates glottal abduction in co-ordination with tongue-palate contact patterns by means of simultaneous recordings of transillumination, fiberoptic films and Electropalatography (EPG). The second experiment focuses on supralaryngeal correlates of alveolar stops studied by means of Electromagnetic Articulography (EMA) simultaneously with EPG. Three German native speakers participated in both recordings. Results of this study provide evidence that the first hypothesis holds true for alveolar stops when different positions are taken into account. In fricative production it is also confirmed since voiceless and voiced fricatives are most of the time realised with glottal abduction. Additionally, supralaryngeal correlates are involved in the voicing contrast under two perspectives. First, laryngeal and supralaryngeal movements are well synchronised in voiceless obstruent production, particularly in the stressed position. Second, supralaryngeal correlates occur especially in the post-stressed intervocalic position. Results are discussed with respect to the phonetics-phonology interface, to the role of timing and its possible control, to the interarticulatory co-ordination, and to stress as ‘localised hyperarticulation’.

## The structure of the current work

Chapter 1 and 2 provide theoretical background information on the voicing contrast in general and especially in German. They report articulatory investigations which have been carried out. Based on this review three hypotheses are described which have been motivated the current work. Chapters 3 and 4 depict the current investigation and its results. They report 2 experiments, its corresponding methods and results. Chapter 5 and 6 summarise and discuss the results with respect to findings from the literature and finish with concluding remarks and potential future work.

More detailed, **Chapter 1** introduces different phonological feature terms and their phonetic correlates which have been used in order to account for the voicing contrast. In the following section it is proposed that the voicing contrast is a prime example of phonetic research over at least 50 years. Several theoretical concepts in phonetics and their relation to the voicing contrast are discussed. It continues with a report of the peculiarities of the voicing contrast in German with particular attention to its variations considering different syllable, word and prominence positions. Most of the empirical evidence in German is based on acoustic results. Since the acoustic to articulatory relation is not a one to one, an articulatory study of the voicing contrast in German is proposed together with the specific aims of the current work.

**Chapter 2** provides a literature review of articulatory correlates involved in the voicing contrast under three perspectives: laryngeal correlates, laryngeal-oral co-ordination and supralaryngeal correlates. Since articulatory studies of the voicing contrast in German are rather rare, investigations from other languages with a comparable 2-way contrast are included.

**Chapter 3** outlines the underlying methods of the current work. Two different articulatory experiments have been carried out with the same 3 native speakers of German. The choice of subjects as well as the motivation for the speech material is described. In addition, the techniques used in the two experiments, their reliability as well as further postprocessing procedures are reported. It follows an overview of labelling criteria for the recorded acoustic and articulatory data.

**Chapter 4** involves the main results of the current investigation. The structure of chapter 2 has been retained, i.e. results are divided in laryngeal correlates, laryngeal-oral co-ordination and supralaryngeal correlates.

**Chapter 5** starts with a discussion of the main results with respect to the different hypotheses. Further, limits of the current study are taken into account. It continues with a section on understanding the voicing contrast in German and another more general discussion linking the results to several theoretical concepts. Particular attention is given to the influences of stress.

**Chapter 6** summarises the main results, concludes how they support/disagree with the introduced concepts and shows perspectives what future work could investigate.

# Table of Contents

## Chapter 1 Theoretical background 1

# Chapter 1: Theoretical background

> "Modern phonetics integrates measures and numbers with the structural and functional view of speech and language, combining signal and auditory analysis techniques in a unitary speech science. But it is not the aim of this integration to fill existing phonologies with phonetic substance, because this would give no new insights within the established framework of generative phonology and its subsequent developments (e.g. natural, autosegmental, atomic phonologies). What is needed is a new functional approach to speech and language centering on phonetic structure and processes." (Kohler 1984, p. 150)

Chapter 1 is dedicated to the general theoretical background of the voicing contrast. In order to clarify the relevant terminology and the different concepts/findings they are based on, section 1.1. starts with a critical review on definitions of phonological feature terms and their phonetic correlates. Phonological feature terms as for instance [±voice], [tense] versus [lax] or [fortis] versus [lenis] are reported with respect to their phonetic correlates. The definitions do often focus on either laryngeal or supralaryngeal correlates and additionally, they are sometimes rather vague. This calls for the need of a comprehensive investigation, observing the contrast at different levels.

In section 1.2. it is proposed that the voicing contrast is a prime example of phonetic research over at least 50 years. Several theoretical concepts in phonetics and their relation to the voicing contrast are discussed. It is shown that the voicing contrast is a complex phenomenon that has been used to test current concepts on the phonetics-phonology interface, the incorporation of timing into phonology, to study the problem of interarticulatory co-ordination etc.. Additionally it is shown, how the present thesis will address issues pertaining to the voicing contrast in general.

In the following section 1.3. a report of the peculiarities of the voicing contrast in German is presented with particular attention to its variations considering different syllable, word or prominence positions. Most of the empirical evidence in German is based on acoustic results. Since the acoustic to articulatory relation is not a one to one, an articulatory study of the voicing contrast in German is proposed together with the motivation and hypothesis for the current work.

The voicing contrast is probably one of the most investigated issues in phonetics. From a first impression it seems to be a rather simple contrast, which might be described in terms of laryngeal adjustment. The larynx is the source of vocal fold vibrations and thus, it could be simply supposed that voiced obstruents are produced with some amount of vocal fold vibrations and voiceless obstruents with the absence of those (for instance due to glottal opening or due to vocal fold tension). If this holds true for the various phonetic realisations of the contrast, a clear acoustic characteristic (voicing versus voicelessness) could be used as a phonological feature term.

However, when studying the voicing contrast in detail, it becomes apparent that it is in fact a rather complex phenomenon, for instance:

1. Voicing includes an appropriate transglottal pressure difference with a higher subglottal pressure (pressure below the vocal folds) and a lower supraglottal pressure (pressure above the vocal folds).
2. Vocal tract changes affect the transglottal pressure difference. In obstruent production the vocal tract is either constricted (in fricatives) or closed (in stops) which increases the supraglottal pressure and has therefore an effect onto the source, i.e. vocal fold vibrations can easily stop vibrating since the transglottal pressure difference diminishes.
3. The maintenance of voicing during closure is often related to closure duration. A long closure duration coincides with an increasing supraglottal pressure and therefore devoicing. A short closure duration may not allow intraoral pressure to rise so quickly and may guarantee voicing during closure.
4. Temporal differences vary with respect to prosodic requirements such as position in the utterance, word, morpheme, and syllable or with respect to stress. Word initial stressed positions are known to be longer in comparison to word internal post-stressed positions. These temporal variations have an impact on the realisation of the contrast. Consequently, the contrast varies regarding different positions.
5. Voiced and voiceless segments influence the production of their neighbouring sounds. Vowels preceding voiced obstruents are typically longer in duration in comparison to shorter vowels before voiceless obstruents. Thus, specific characteristics occur not only with respect to the contrasting minimal pairs, but also with respect to their environment.

Following just these examples, several theoretical implications are connected to the production and perception of the voicing contrast which will be discussed in 1.2.. But before going into the details, the next section introduces the definition of different phonological feature terms (besides the voice-voiceless distinction) which have been used to describe the contrast. They classify the (main) difference minimal pairs are based on. It is intended to show that already phonological terminology is not as clear as a simple contrast would predict. Terms differ from author to author, the same terms are defined in different ways and different terms are used in the same way. The variation in terminology is due to the findings from various studies, characterising the variety of production mechanisms and perceptual cues involved in the contrast. Additionally, it will be pointed out why and which terminology has been used in this study.

## 1.1. Phonological feature terms of the voicing contrast and their phonetic correlates

First it is reasonable to ask what feature terms should represent: phonetic realisations (acoustic, articulatory, perceptual) or abstract phonological categories? Jakobson and Waugh (1979) pointed out that:

> *"The distinctive features consist of formal oppositions specified and individualized by the phonic prerequisites they are built on. Radical partisans of glossomatics*[1] *have endeavored to extract the system of primitive constituents with no reference to sound substance at all"* (Jakobson and Waugh 1979, p.52).

On the one hand, Jakobson and Waugh's suggestion about distinctive feature terms seems practical, i.e. terms are used in relation to the physical properties or measurements of the relevant phoneme. On the other hand, it implies that a particular 'phonic prerequisite' does always occur. In other words it reflects the presumption of invariance at the phonetic level. As will be shown, differences in terminology exist which already reflect the fact that invariance is difficult to find and that more than one distinction can be made in order to describe the voicing contrast. Some linguists (in Jakobson and Waugh citation the glossomatics) try to avoid terms corresponding to physical values, but generally, terms which can be associated with phonetic output are more frequently used[2].

A third direction is possible, to use feature terms as abstract concepts correspon-ding to a bundle of phonetic characteristics. If feature terms are abstract there is no need that they directly relate to phonetic realisation and they are independent of possible phonetic variations. In the current study the abstract phonological

---

[1] Glossomatics is a linguistic trend of structuralism founded by Hjelmslev.

[2] Jessen (1998) mentioned another important point regarding feature terms and their mixture with respect to production, perception and acoustics.

feature terms voiced and voiceless are favoured, i.e. all phonetic realisations of /p t k/ will be described as phonologically voiceless and all /b d g/ as phonologically voiced. This solution was chosen since many linguists associate voicing contrast to the /p t k/ versus /b d g/ distinction, but may relate the tense-lax contrast or fortis-lenis contrast also to vowel inventories (e.g. Mooshammer 1998, Debrock 1977). Although devoicing of vowels exists too (e.g. in Japanese) it is related to a particular context, and is not contrastive to my knowledge.

### [±voice]

Chomsky and Halle (1968, p.326ff.) subsume under the phonological feature [+voice] vs. [–voice] four different phonetic characteristics: voicing, tenseness, glottal constriction, and subglottal pressure. According to Chomsky and Halle phonologically voiced stops can be described by voicing during oral closure, a low muscular tension in the supralaryngeal articulators, glottal constriction, and a reduced subglottal pressure whereas phonologically voiceless stops show no voicing during oral closure, muscular tension, no glottal constriction and a heightened subglottal pressure. Three years after the publication of “The Sound Pattern of English” Lisker and Abramson (1971) provided a detailed critique on the four characteristics. They noted at first that the description of Chomsky and Halle’s phonology was based on only a few studies: Kim (1965, 1967) and Perkell (1965). Second, experimental evidence was provided for a negligible subglottal pressure difference between the phonologically voiced and voiceless phonemes (Netsell 1969) and thus, the characteristic should not be included in the feature system. Third, a passive pharyngeal enlargement, another characteristic describing the contrast with respect to muscular tension in the oral cavity, would be hampered since it has also been found in /n/ (Perkell 1965). Additionally, Rothenberg (1968) as well as Kent and Moll (1969) proposed that it is an active mechanism. Lisker (1970) has also found no differences in intraoral pressure rise or peak pressure for word-initial /b d g/ and /p t k/ and hence, he assumed that the pharynx is not enlarged and the tensity feature should not be used as a general characteristic. These are only a few examples out of the list of Lisker and Abramson’s critique.

### VOT (Voice Onset Time)

It should be clarified that VOT itself has not been used as a phonological feature term, but as a synonym for the voicing contrast and therefore it is mentioned here. The Voice Onset Time is a temporal characteristic, measured as the duration between the burst and the first occurrence of periodicity in the acoustic signal (known as the onset of the ‘voicebar’). It provides a measure of interarticulatory co-ordination, since the oral release (burst) is realised by supra-laryngeal articulators and the onset of the voicebar is produced by laryngeal activity. In their pioneering work Lisker and Abramson (1964) found in a cross-linguistic study for the word initial position three different categories of VOT.

**[±aspirated]**

The feature aspiration can be associated with the occurrence of an acoustic measurable duration and additionally, with an open glottis at the time of oral release (Kim 1970). Unaspirated stops exhibit aspiration durations approximately below 20ms (Docherty 1992, p. 12 referring to Stevens and Klatt) and no glottal opening at the time of release. The separation between unaspirated and aspirated is based on the fact that aspiration becomes audible above 20ms.

In the generative phonology tradition aspiration has been considered as a feature property rather than an independent feature. It should result from a heightened subglottal pressure and no constriction at the glottis (Chomsky and Halle 1968, p. 326).

**[±spread glottis]**

The term spread glottis implies an articulatory feature at the laryngeal level. In 2002 Jessen changed his former terminology for German (he used tense versus lax) to the [±spread glottis] feature (Jessen and Ringen 2002). From his/their point of view [±spread glottis] does not only describe a typical articulatory adjustment at the glottal level, but also its underlying mechanism (glottal opening due to active muscular contraction versus a glottal opening due to an increase of intraoral pressure). The authors write:

> *“Slighter amounts of glottal opening can also be created passively, i.e. due to biomechanical-aerodynamic factors without muscular activity. Passive glottal opening in non-[spread glottis] stops has been explained on the theoretical level for English by Stevens (1998) and demonstrated empirically for German by Jessen (1998). Consequently, the categorical differences between [spread glottis] and non-[spread glottis] stops in English and German is not so much in terms of presence vs. absence of glottal opening as in terms of active (and large) glottal opening vs. passive (and slight) glottal opening”* (Jessen and Ringen 2002, 189-218).[3]

---

[3] The notion is somehow surprising since Jessen (1998) often refers to Danish work from Hutters, Fischer-Jørgenson and Hirose who found evidence for the activity of the only glottal abductor muscle even with a small amount of glottal opening.

**[tense] vs. [lax]**

The tense versus lax feature terms can be associated with articulatory mechanisms, in particular with muscular tension. By means of a cineradiographic study Perkell (1969) found evidence for a greater tension in the pharyngeal portion of the tongue in 'tense' obstruents. Therefore, he suggests:

> *"The concept of tense seems to be related to greater muscular effort for the consonants as well as for the vowels. For vowels, the greater muscular effort is applied to moving the tongue body, but for consonants it seems to be related to resisting increases in intraoral pressure. The observations of the behavior of the pharyngeal portion of the tongue for the tense-lax consonant pairs /t,d/ and /s,z/ suggests that there is greater tension or tone in the tongue in this region for the tense consonants. This increased tone could be present throughout the entire vocal tract, and it would make the wall more rigid and resistant to deformation as a result of an increased intraoral pressure"* (Perkell 1969, p. 67).

The same terms - tense and lax are used in a different manner by Jessen in order to distinguish German phonologically voiced and voiceless obstruents. They were discussed with respect to temporal characteristics:

> *"On the assumption of Jakobson et al. that the phonetic basis of a feature has to be present across contexts, the evidence argues for the existence of the feature [tense] in German, with a common dominator that is based on duration"* (Jessen 1998, p. 163).

According to Jessen, tense obstruents are often longer in duration in comparison to lax, particularly when considering the acoustic measures aspiration duration and closure duration.

**[fortis] vs. [lenis]**

In a series of studies Malécot (1966a, 1966b, 1968, 1969) used the terms fortis and lenis obstruents. He suggests that the fortis-lenis opposition has nothing to do with articulatory energy (as would be favoured by Perkell 1969, although he called it the tense-lax opposition), but with intraoral pressure variations and its proprioceptive perception:

> *"Force of articulation is a significant attribute of consonants and enters into the lenis-fortis opposition in such pairs as /p/ : /b/ and /s/ : /z/. It is a case of synesthesia, in that it has little or nothing to do with articulatory energy but is rather taken a mistaken proprioceptive impression based primarily on intrabuccal air pressures resulting from the air valving action of the glottis, the occlusion or constriction of the buccal passage, and the velopharyngeal sphincter, and perhaps also involving closure duration. Synesthesia is defined as a subjective sensation of a sense other than the one being stimulated"* (Malécot 1969, p. 1588).

In his 1966 article Malécot goes even further and suggests that the intraoral differences subjects are aware of, could provide possible feedback cues in speech production.

For German obstruents Kohler (1984) described the /b d g/ versus /p t k/ opposition as a contrast of fortis versus lenis. Under fortis versus lenis Kohler subsumes the following criteria:

> *"The feature has been connected with power in the supraglottal movements and in the air stream, and with tension, especially in the larynx. In a somewhat simplified relationship the feature maybe associated with an articulatory timing and with a laryngeal power/tension component. The former relates to the speed of stricture formation and release, and is probably a language universal, the manifestation of the latter (aspiration, voicing, glottalisation) is language specific. The relative weight of the contribution from the two components depends on the position in the syllable/word and on the combination with stop/fricatives"* (Kohler 1984, p. 168).

For some readers it might be easy to follow the different feature terms and their meanings, but others (including myself) can easily get mixed up, ignore the actual meaning and follow their intuition. The different phonetic correlates of the proposed terms already reflect the complexity of the realisation of the contrast. These confusing and sometimes even contradictory proposals call for the need of more data analyses, i.e. investigating the voicing contrast comprehensively on different levels: the laryngeal level, the supralaryngeal level and the laryngeal-supralaryngeal interaction. Several issues and its theoretical implications for phonetic research will be discussed in the next section.

## 1.2. The voicing contrast - A prime example of phonetic research over at least 50 years

After decades of phonological and phonetic research the voicing contrast, its production and perception, its relation to the phonetics-phonology interface are still a matter of debate (e.g. Lisker and Abramson 1964, Malécot 1966, Perkell 1969, Kent and Moll 1969, Hirose, Lisker and Abramson 1972, Bell-Berti 1975, Hutters 1984, Kohler 1984, Lisker 1986, Dixit 1987, Docherty 1992, Brockhaus 1995, Jessen 1998, Cho and Ladefoged 1999, Esposito 2002, Allen, Miller and DeSteno 2003, Fuchs and Perrier 2003, Scobbie in press, Gafos submitted). It is proposed that the voicing contrast is a prime example of phonetic research since its phonetic realisation and its complexity linked/links to several hot topics within our research field, e.g. the phonetics-phonology interface or the discussion on interarticulatory co-ordination. In addition, theoretical concepts in phonetics should be grounded on empirical evidence and the voicing contrast has been used in order to develop and test different theoretical concepts. This section intends to show the link between theoretical concepts and their relation to empirical findings regarding the voicing contrast and to what aspects the present thesis will contribute to it.

### *1.2.1. The voicing contrast and the phonetics-phonology interface*

Since there is an ongoing discussion of an interface between phonetics and phonology, both disciplines have obviously been separated and it seemed to be the case that:

> *"Each group paid little if any serious attention to the problems and findings of the other*" (Lisker and Abramson 1971, p.767).

The strong separation might go back to the influence of structuralism in linguistics (Ferdinand de Saussure). Saussure divided linguistics in 'langue et parole' – 'language and speech' which can be associated with 'substance and form', 'cognition and physical properties' or in Chomsky's words 'competence and performance'. From this traditional point of view (see Chomsky and Halle 1968) speech/performance would be a derivative of language/competence.

Chomsky and Halle (1968) proposed a limited time and spaceless (i.e. they are independent of their realisation in space and time) universal set of phonological features, which should explain the lexically meaningful sounds of the world's languages. Although their work on a formal grammar has been a milestone for scientists working in different fields, it provoked those who had studied and published on variations of time and space related speech production/perception processes which are most of the time language and speaker-specific.

Taking the phonological voicing contrast as an example: it describes a binary categorisation. Phonemes are either voiced or voiceless. Changing from one phoneme to the other changes meaning of a word too. Looking at the phonetic realisation Lisker and Abramson (1964) found an acoustic correlate of the phonological contrast in word-initial position: Voice Onset Time (VOT). By means of three VOT realisations (voicing lead (voiced), short lag (voiceless unaspirated) and long lag (voiceless aspirated)) they were able to differentiate between the phonologically voiced and voiceless stops in several languages. So far evidence from phonetic realisation was found for categorical distinctions, but those distinctions were neither time or spaceless nor universal.

Later work (e.g. Cho and Ladefoged 1999, Docherty (1992), Scobbie (in press)) has shown that categorical boundaries for the VOT measure are less clear cut, since VOT depends on a number of factors such as language, speaker, contextual varieties (position, vowel context), stress, speech rate, place of articulation to name a few. For instance, in German a phonologically voiced stop in word initial position (often realised as a voiceless unaspirated stop) can be produced with a comparable VOT like its phonologically voiceless counterpart in intervocalic position (also realised as voiceless unaspirated) so that VOT values alone do not necessarily refer to the relevant phoneme without any knowledge about the context.

Beckman and Pierrehumbert (2000) have recently noted that

> *"...prosody and other positional information cannot be separated from the specification of phonetic contrast. Children learn the sounds of their native languages in context, and machine systems for synthesis and recognition can be improved by taking position into account"* (Beckman and Pierrehumbert 2000, p.1).

In addition, VOT is not the only phonetic parameter to account for the phonological voicing contrast (e.g. Lisker 1957, 1978, 1986, Jessen 1999, Jessen 2001, Kohler 1984, Luce and Charles-Luce 1985). Lisker (1978, 1986) counted up to 16 acoustic differences for the contrast (and there might be even more), i.e. the voicing contrast includes a set of redundant phonetic correlates. It becomes rather challenging to build a phonetic hierarchy between the correlates which could be appropriate for most of the phonetic realisations (for one new method to build such a hierarchy for the voicing contrast in intervocalic velar stop production in Korean, see Brunner et al. 2003). However, so far it is unclear whether our audiovisual perception takes a hierarchy of phonetic correlates into account or perceives the relevant unit as a whole. Within the later perspective Malécot (1970) pointed out that

> *"A given distinctive feature ideally incorporates a number of redundant cues which constitute a gestalt"* (Malécot 1970, p.1591).

The term 'gestalt' stems from a psychological research direction in the beginning of the 20th century (Berlin School: Wertheimer, Koffka, Köhler, Lewin) investigating mainly the perception of humans. A gestalt in this tradition is more than the sum of its parts.

From the previous perspectives the assumption is drawn that the voicing contrast is one example of phonetic research which can be linked to two major problems in the phonetics-phonology interface:

1. The problem of BINARY phonological features versus CONTINUOUS realisations on the phonetic surface (i.e. language as an abstract cognitive process which is realised by means of physical properties), and
2. second, the problem of distinctive features characterising a minimal pair in ONE domain versus the complex of MULTIPLE combined cues in production (see also the problem with the coordinative structures below) or perception.

For the first problem there have been several attempts to link both areas. Phonetic/phonological research tries to incorporate physical properties as well as time and space into phonology to create an 'embodied linguistics' (Port and Leary 2003, Schwartz et al. 2002). For instance, Ohala (1983) and Maddieson (2003) explain the frequently 'missing /g/' patterns in the sound inventories of the world's languages by means of the difficulty to sustain voicing with a velar oral closure (see also 2.4.1.).

Another research trend describes categorical differences on the one hand and variations on the other hand due to the system properties themselves and not by a translational process from binary features to continuous movements. Two different examples should represent this trend:

1. Pierrehumbert, Beckman and Ladd (2001), Beckman and Pierrehumbert (2000) have considered speech acquisition as a process where the child

acquires the phonemes of its mother tongue due to contextual variations in the production of its caretakers. Additionally, Pierrehumbert, Beckman and Ladd (2001) write:

> *"Variability causes the need for abstraction"* (Pierrehumbert, Beckman and Ladd 2001, p.14).

Under this perspective the variability of performance enforces categorisation. Hence, the distinction between binarity and continuity diminishes and speech performance is no longer a derivative of language competence.

2. Gafos (submitted) modelled speech production, particularly final devoicing in German by means of non-linear dynamics. His work is mainly inspired by Browman and Goldstein's Articulatory Phonology (1986, 1989) in combination with Task dynamics (Munhall and Saltzman 1989). He provides an answer to the question what is a symbol (or phonological feature) with respect to its dynamical formulation.

> *"In the dynamical formulation, the symbol is inseparably linked with its phonetic substance. It is not derivationally antecedent to that substance and therefore it does not need to be translated to that substance"* (Gafos submitted, p.7).

Gafos results provide evidence that non-linear dynamics produces both, continuity and separation between categories.

A similar approach can be found in a definition of perception research (Hawkins 1999):

> *"...the task is to extract meaning from the speaker's acoustic signal. Although this definition is realistic about the nature of the task, it is too broad for a single discipline to cope with at present, for it demands we should examine not only the acoustic cues to speech perception, but also how they interact with higher-order linguistic functions such as grammar and the choice of words, as well as with more intangible influences like speaker's and listener's expectations, which are affected by their shared culture. This comprehensive definition of speech perception, then, acknowledges that speech cannot be separated from language"* (Hawkins 1999, p.198).

The phonetics-phonology interface issue has been one of the central themes in a series of Laboratory Phonology conferences. At present there is no overall accepted theoretical framework for the phonetics-phonology interface. It is rather a bundle of models and theories, partly adapted from other research areas (e.g. statistics/mathematics/computing, biology/psychology, speech technology), modified and improved or rejected over the years, reflecting the interdisciplinary approach of our subject (see Schwartz et al. 2002 for a reflection). The voicing contrast with its variations, dependency on contextual factors and its categorisation is one of the prime examples that can be used for the interface.

### *1.2.2. The voicing contrast and intrinsic timing in phonetics and phonology*

Timing, especially VOT, oral closure duration, and the preceding vowel duration is an important factor for the phonetic realisation of the voicing contrast.

Ohala (1983) pointed out that the longer oral closure duration the greater the likelihood of devoicing and/or voicelessness (for a more detailed description of the phenomenon see Pape et al. 2003). Derived from Ohala's statement and the fact that voicelessness typically coincides with some amount of aspiration whereas voicing does not[4], it can be asked whether the duration of aspiration could be predicted by the duration of oral closure, i.e. aspiration duration would be controlled by oral closure. Such linear relationship may hold true for stops in intervocalic position, but it is not a general phenomenon. For instance, Ridouane (2003) compared single stops with geminates in Berber and found evidence that although geminates have considerably longer closure duration, their aspiration duration is similar to single stops.

Another temporal characteristic well known in the production of the voicing contrast concerns preceding vowel duration. A vowel preceding a voiced stop is often acoustically longer compared to a vowel preceding a voiceless stop. The strength of this effect varies with respect to different languages. Its absence has been found in a few languages like Arabic (de Jong & Zawyadeh 2002), as well as Polish and Czech (Browman and Goldstein 1986). In the Germanic languages such as English and German differences in vowel duration are one of the acoustic correlates of the voicing contrast and might be used as perceptual cues.

Since timing plays a major role for the voicing contrast, Jessen (1998, 2001) has introduced the features tense and lax as broader terms describing mainly temporal differences of the contrast in German. In Jessen's perspective the feature tense corresponds to longer aspiration duration, the primary cue of the

---

[4] This is not true for languages like Hindi where phonologically voiced aspirated stops can be found.

contrast. Secondary cues as for instance closure duration or vowel duration can take over if for some reason aspiration duration is not as pronounced as it should be.

Some investigations, e.g. Docherty (1992) and Allen et al. (2003) consider not only the phonological contrast as causing a temporal distinction, but also speaker-related, regional and social factors which would explain inter-speaker micro-variability. Hence, timing distinctions may not only cue 'what' is said, but also 'who' said it.

Looking from another perspective - from speech motor control - it is still a matter of debate whether timing itself is controlled by the neural nervous system or not (Perrier 2003). In the latter case, timing could be a consequence of the movement from one acoustic/articulatory target region to the next. Regarding the voicing contrast it means that the acoustic/articulatory target regions between phonologically voiced and voiceless stops differ (see Brunner et al. 2004 with experimental data from Korean velar stops and simulations using a 2D biomechanical tongue model), which causes a difference in timing.

In Articulatory Phonology/Task Dynamics (Browman and Goldstein 1989), one of the frameworks of the phonetics-phonology interface, intra-articulatory timing is modelled by means of a stiffness parameter for the appropriate articulator/gesture. With a stiffer articulator a shorter duration can be simulated moving from one articulatory target position to the next.

Hence, timing would be a consequence of different target control or variations in stiffness rather than be controlled itself. However, timing can also be modelled as temporal relations between articulatory gestures (inter-articulatory timing) which will be within the scope of section 1.2.3.

More recent approaches in phonetics take timing differences (wherever they might stem from) into consideration, especially for the description of the voicing contrast. Kohler (1984) has proposed:

> "*The incorporation of the time dimension into phonology is regarded as a necessary prerequisite to the phonological problems in general and to an adequate treatment of the [± voiced] feature in particular, voice onset time being only one temporal aspect* (Kohler 1984, p.150).

However, within the generative phonology tradition (Chomsky and Halle 1968) phonemes correspond to timeless entities as discussed previously with respect to the phonetics-phonology interface. During speech performance phonemes are realised in time, but language competence, so their claim, is independent of time. Fowler (1980) has written one of the key papers proposing the incorporation of intrinsic timing into phonology. She argues that theoretical explanations for phenomena like coarticulation suffer from the notion of a left-to-right array of discrete phonemes, that the array should become translated into time and space during production, and that the plan of an utterance differs from its execution. The distinction between phonological entities and their production call for the necessity of a translation process. In addition, extrinsic timing theories are not able to answer the question how the concept of segments could have been developed in evolution or speech acquisition, if a child cannot learn its mother tongue due to temporal visual and auditory components of speech performance. Therefore (and for many other reasons), Fowler proposes the incorporation of intrinsic timing into phonology.

As it was pointed out timing is an important factor in order to explain the production of the voicing contrast and should be incorporated in phonology (for overview see Port and Leary 2003).

### *1.2.3. The voicing contrast and the problem of interarticulatory co-ordination*

Text book solutions reduce the voicing contrast to a phonological contrast with a feature defined phonetically as being based on laryngeal activity, i.e. vocal folds vibrate or not or the glottis is open and aspiration can be produced or the glottis is closed and no aspiration is possible. However, as it has already been pointed out in the section on timing, the duration of oral closure - a supralaryngeal correlate of the contrast - increases or reduces the likelihood of voicing during closure - a laryngeal correlate of the contrast. Since stops in word internal intervocalic position do generally exhibit a short duration, phonologically voiced stops in this position are most likely to have a longer voicing into closure whereas in word initial position closure duration is rather long, limiting the likelihood of voicing.

Taking aspiration as another example: Aspiration does not only coincide with some amount of glottal opening, it is of further importance that peak glottal opening occurs with respect to oral release. If peak glottal opening occurs during oral closure as in geminates or to some amount in unaspirated stops, there will be no acoustic/auditory consequence except from a silent period during closure. If peak glottal opening is synchronised with oral release, aspiration noise will be heard.

In addition, the velar port should be closed otherwise intraoral pressure cannot rise and a salient burst for phonologically voiceless stops cannot be produced (e.g. in unrepaired cleft palate speech). However, the velar port should also be closed during the production of phonologically voiced stops and hence, it is less likely to differentiate voiced from voiceless stops concerning velar movement.

Looking at the IPA charts one could be led into assuming that the distinction between voiced and voiceless stop is due to differences in laryngeal activity only. However, the contrast is at least an interaction between laryngeal and supralaryngeal articulators within a certain time range (for more details see chapter 2.3.).

An expanded concept of the phonetics-phonology interaction, Articulatory Phonology (Browman and Goldstein 1986, 1989) implements interarticulatory co-ordination, i.e. it takes different articulators[5] into account. The voicing contrast has been modelled as a missing glottal opening gesture plus a velar and an oral gesture for phonologically voiced stops. Phonologically voiceless stops include a glottal abduction gesture synchronised with the supralaryngeal events.

Another intriguing method to describe interarticulatory co-ordination has been offered by the framework of 'coordinative structures' (Kelso, Saltzman and Tuller 1986) where articulators function as a unit within a task-specific manner. One assumption of the 'coordinative structure' framework is that articulators seldom move independently, they are rather grouped together in 'coordinative structures' and their grouping and actions change over time in order to produce the required vocal tract shapes. Thus, the coordinative structures are flexible and task-dependent. Empirical evidence for coordinative structures comes from perturbation studies, where one of the articulators of a coordinative unit gets perturbed and compensatory movement of other articulators within a task specific unit is observed. Compensation is not found for articulators which are not involved in the production of the specific task (see also chapter 2 on perturbation studies). The concept of coordinative structures uses 'task dynamics' (Saltzman and Munhall 1989) to provide evidence from modelling. One of the key examples for coordinative structures of Kelso, Saltzman and Tuller's work is the production of a bilabial stops where jaw, upper lip, lower lip, and velum serve as a functional unit. Based on their data the authors suggest, when changing stress pattern or speech rate the kinematic movement trajectories vary, but interarticulatory time relations are stable.

> *"Thus, the information for "timing" of a remote articulator (e.g. the upper lip) may not be time itself, nor absolute position of another articulator (e.g. the jaw), but rather a relationship defined over the position-velocity state (or, in polar coordinates, the phase angle) of the other articulator"* (Kelso, Saltzman and Tuller 1986, p.42).

---

[5] In Articulatory Phonology jaw has indeed been considered as an articulator, but not as a tract variable in the production of linguistically meaningful units.

Phase angles constitute invariant characteristics of speech production in a task-specific manner. The concept of coordinative structures could be useful in order to explain interarticulatory co-ordination for the voicing contrast (for discussion see Docherty 1992). However, serious doubt on the stability of interarticulatory timing (phase angles) as an invariant property was raised by Alfonso and van Lieshout (1999). They recorded jaw, lower and upper lip movements of seven male subjects in three different rate conditions three times (in approximately 2 weeks distance). Their results provide evidence for a considerable variability for relative timing of jaw and lip movements across subjects and sessions. However, motor equivalence influences (one articulator can compensate for another) were observed so that all gestures were considered ‘well coordinated’ (p.51) although temporal and spatial instabilities occurred.

Further work is necessary to discuss the stability and flexibility of interarticulatory timing patterns, particularly regarding the voicing contrast in different positions in the word or syllable.

### *1.2.4. The voicing contrast and the acoustics/perception to articulation relation*

Articulatory patterns and their resulting acoustics as well as acoustics and their auditory correlates do not behave in a linear relationship. There are several theories/concepts trying to map these relations. For instance, Stevens (1972, 1989) introduced the quantal theory from physics into speech. He defined three different regions: two regions (region I and III) where changes in articulatory patterns do not cause noticeable changes in the acoustics, i.e. acoustic patterns are relatively stable. According to Stevens, stable acoustic regions are the basis for the sounds of the world’s languages. Between the two regions lies region II where small articulatory changes cause an abrupt change or quantal jump in the acoustics. Region II can be considered as a ‘threshold region’, since rapid changes in the acoustics correspond to shifts in auditory responses from one feature to the other. Stable acoustic pattern in a segment can be associated with the distinctive features based on Chomsky and Halle’s framework (1968). Stevens assumes that invariance in the acoustic signal exists and that the invariance can be directly linked to linguistic units. However, the assumption of invariance in the acoustics may be rejected since several researchers found rather variable acoustic patterns regarding phonemes in different positions in a word or syllable, or investigating more natural speech in comparison to classical ‘lab experiments’(see for instance Lavoie (2001) with a comprehensive study on stops produced in spontaneous speech). Since the concept of quantal theory is based on the assumption of relatively invariant acoustic properties, it needs further empirical evaluation, including phonemes and their changes with respect to contextual, position-related and prosodic factors in order to define the precision of stable acoustic pattern. Additionally, not only single but also multiple or combined acoustic characteristics may cause quantal jumps, since phonological minimal pairs are often realised by more than one acoustic difference.

The motor theory of speech perception is another theoretical concept mapping the acoustics-articulatory relations. It was originally developed by Liberman et al. (1967) and later modified by Liberman and Mattingly (1985). In opposition to Stevens, Liberman and colleagues suppose invariant properties at an articulatory level and that speech perception is specific to humans. Listeners would perceive speakers' intended articulatory gestures directly without any translation process, i.e. production and perception share a common link. Studies of the perception of the voicing contrast have contributed to a revision of the motor theory (see Hawkins 1999a, b). Eimas et al. (1971) investigated the perception of VOT differences in babies. Their results provide evidence for babies' categorical perception although babies have still not acquired the articulatory gestures of their mother tongue. In addition, Kuhl and Miller (1975) investigated the perception of /Ca/-syllables in chinchillas. Two different groups of chinchillas were trained to cross a barrier in a cage only by hearing either 0ms VOT (corresponding to phonologically voiced stops in American English) for the one group and 80ms VOT (corresponding to phonologically voiceless stops in American English) for the other group. Animals responded perfectly to these categories. In a next step VOT values were approximated in order to find the boundary condition where animals switched from the voiced to the voiceless perception. Surprisingly, chinchilla's boundary conditions were quite similar to those of humans (33ms VOT for chinchillas and 35ms VOT for human adults). The categorical perception of the voicing contrast in animals provoked the question how far does human perception of speech differ from general abilities of species' perception. Additionally, it questions the original motor theory's assumption that speech perception would be special to humans. However, one advantage of the motor theory is the direct mapping between articulation and perception without any translation process.

Both, the quantal theory and the motor theory of speech perception consider invariant properties, either at an acoustic or at an articulatory level. Both theories could be improved by a more precise definition of consistent versus variable patterns based on empirical findings.

Studying the perception of the voicing contrast led to some other important proposals. Abramson and Lisker (1967) observed variations in the duration between F1 cut relative to F2 and F3. They varied F1 cutback in continuous steps from –150ms to 150ms. Listeners associated voiced stops with a F1 beginning at –150ms to 20ms before F2/F3 onset and voiceless stops with F1 beginning 30ms-150ms after F2/F3 onset. Based on these results the authors proposed categorical perception of continuous physical phenomena. Perception is highly accurate when changes cue the contrast whereas within a category discrimination is rather poor[6]. Categorical discrimination is less accurate for consonants in word medial position and for vowel duration (see Pickett 1999).

As it was already discussed in the paragraph on the phonetics-phonology interface not only one, but multiple acoustic differences can contribute to the voicing contrast. Consequently, several acoustic differences are integrated in units of categorical perception. An integration of cues (also known as phonetic trading relations, see Repp 1982) for the voicing contrast has been shown in Summerfield and Haggard (1975) who combined VOT (longer VOT for voiceless compared to voiced) with F1 onset frequency (higher F1 for voiceless compared to voiced). For low F1 values (cueing voiced stops) longer VOT values were required to be perceived as voiceless. Discrimination can be poor when two conflicting cues interact. Diehl, Kluender and Walsh's (1990) 'auditory enhancement theory' emphasizes the role of the combination of such cues and their perception within a unit. Particular attention has been dedicated to the voicing contrast since several redundant acoustic cues are known for it.

A last example is dedicated to a principle called 'motor equivalence'. Motor equivalence describes a phenomenon where different articulatory strategies can be used in order to produce the same acoustic target. It has been intensively studied in the production of /u/ where lip and tongue movements interact and can compensate for each other (e.g. Perkell et al. 1993, Savariaux, Perrier and Orliaguet 1995). The principle of motor equivalence may be seen within the concept of articulatory movements working towards acoustic goals (Perkell et al. 1997), i.e. speakers learn acoustic targets rather than articulatory gestures. It should be noted that although this framework is influenced by Stevens' quantal theory, it does not assume invariants in the acoustics and it is rather based on empirical data showing speaker dependent patterns with respect to articulatory movements. For the voicing contrast it is supposed that motor equivalence contributes to its production. Particularly, voicing during oral closure can be produced with different articulatory strategies (see chapter 2 on cavity enlargement strategies).

---

[6] Kuhl (1991) showed for the internal structure within a category that discrimination is poorer when sounds are closer to a 'prototype' in comparison to sounds further apart from the prototype. She called this phenomenon the 'perceptual magnet effect'. Most of her work observes vowels. Further work could continue to investigate the magnet effects in consonant perception as well as vowels in different prosodic conditions.

In summary, the voicing contrast has been studied for many decades. It was/is particularly interesting to investigate the voicing contrast with respect to the phonetics-phonology interface, the incorporation of timing into phonology, and to discuss interarticulatory co-ordination as well as the acoustics/perceptive-articulatory relations. The current thesis aims to contribute to the following general topics:

1. Phonetics-phonology interface: Similarly to the work on several acoustic characteristics of the voicing contrast, the current work investigates articulatory correlates of the contrast at the laryngeal and supralaryngeal level. First, it will discuss whether there are stable and invariant articulatory correlates which determine the voicing distinction. Second, by studying different contextual variations, this work contributes to the question which characteristics of the contrast occur more frequently than others and could be used in order to build an abstraction of the contrast (following Pierrehumbert, Beckman and Ladd (2001). Third, the main articulatory findings will be seen in the light of several phonological feature terms and which of them characterise the contrast properly.
2. The role of timing and its stability: Considering the voicing contrast in different positions provides the right frame in order to observe the stability of timing patterns such as aspiration, closure, and vowel duration. Additionally, this study will offer not only insights into the temporal stability at an acoustic level but also at an articulatory level, since both levels do not behave in a one-to-one relationship. Based on the articulatory dataset the current work is discussed with respect to speech motor control and the raised question whether timing is controlled or whether it is a consequence of articulatory movements from one target to the next.
3. Interarticulatory co-ordination: Since distinct laryngeal and supralaryngeal production mechanisms of the voicing contrast are taken into account here, this work offers new insights into the stability of laryngeal-oral co-ordination in general and specifically into the stability of interarticulatory co-ordination with respect to different positions in the word.

In the next section the voicing contrast in German is described, since this language is particularly interesting. German exhibits a variety of positional effects regarding the contrast. It will be the language taken into account in the current study.

## 1.3. The voicing contrast in German

### *1.3.1. The phoneme inventory of German consonants*

In any further explanation words in *Italics* are used to describe graphemes, the parentheses // are used to describe phoneme representations and brackets [] are used to describe particular phonetic realisations.

The Standard Modern German consonant inventory consists of the following IPA phonemes (IPA = International Phonetic Alphabet), adapted from Kohler (1990). Three stop pairs show a voicing contrast: /b/ vs. /p/, /d/ vs. /t/ and /g/ vs. /k/. The glottal stop is written in brackets since it is only a potential sound feature before the initial vowels of words and stem morphemes (Kohler 1994). In the German fricative inventory four voice-voiceless pairs can be seen: /f/ vs. /v/, /s/ vs. /z/, /ʃ/ vs. /ʒ/ and /χ/ vs. /ʁ/. For the latter, no minimal pairs are known since they occur in different positions in the word. /χ/ occurs most frequently as a variation after back vowels in morpheme or word final position, similar to /x/. In opposition, the uvular voiced fricative /ʁ/ is commonly produced in word, morpheme or syllable initial position, e.g. in *Rat* (advice) realised as [ˈʁaːt]. In word final position (e.g. in the suffix *er*) [ʁ] becomes vocalised as [ɐ] (e.g. *Vater* (father) [ˈfaːtɐ]) or is even deleted (e.g. Bar (bar) [ˈb̥aː]).

**Table 1.1:** German consonant system, after Kohler 1990.

| | **Plosiv** | **Nasal** | **Fricative** | **Approxi-mant** | **Lateral** |
|---|---|---|---|---|---|
| **Bilabial** | p b | m | | | |
| **Labio-dental** | | | f v | | |
| **Alveolar** | t d | n | s z | | l |
| **Post-alveolar** | | | ʃ ʒ | | |
| **Palatal** | | | (ç) | j | |
| **Velar** | k g | ŋ | (x) | | |
| **Uvular** | | | χ ʁ | | |
| **Glottal** | (ʔ) | | h | | |

The palatal and velar fricatives /ç/ and /x/ are allophones and written in brackets in Table 1.1. Their production is ruled by the surrounding vowel environment, i.e. /ç/ is realised after front vowels and consonants as well as morpheme initial, and /x/ as well as /χ/ after back vowels. However, both of them have no voiced counterpart. In a newer version of the IPA (1999) /x/ and /χ/ are allophones of

/ç/. The voiced /ʒ/ exists only in non-native vocabulary such as the French [ˈgaʁaːʒə].

In summary, in the German vocabulary /p t k f s/ and /b d g v z/ are distinctive obstruents, distinguished by the voicing contrast.

### *1.3.2. The voicing contrast in different positions*

Extensive acoustic studies of stops have been carried out investigating how voiced/voiceless obstruent pairs are realised, in particular with respect to their position in the word, morpheme or syllable. Less work of this nature has been carried out for fricatives.

Since the current work is dedicated to articulatory correlates of the voicing contrast it will primarily discuss articulatory production mechanisms (chapter 2) and reference at this point is made to Jessen (1998) who provides an excellent summary of acoustic studies, particularly on aspiration duration in German obstruents. In Jessen (1998) single obstruents mostly surrounded by vowels are taken into account (this will also be the case for the present study). Jessen considers three different positions in his literature review:

a) The intervocalic position (V_V), where the obstruent is surrounded by vowels without an intervening morpheme boundary (e.g. *ibe).* The first vowel belongs to the stressed syllable and the second is often a schwa, which is typical in native German words.
b) The utterance-initial position (##_V), where the obstruent is preceded by a pause or by silence and occurs in the beginning of a stressed syllable, i.e. it refers to a word spoken in isolation (e.g. *Biene*).
c) The post-voiceless position (#_V), where the obstruent occurs in word initial stressed position and is preceded by a word ending with /ʃ/ (e.g. *rasch Bier).*

Jessen summarised the previous empirical work, discussing the voicing contrast in German **stops**, as follows in Table 1.2. Aspiration duration and voicing were taken into account as possible acoustic correlates to distinguish between /b d g/ and /p t k/ in German. The term ‘dominant’ corresponds to the question whether a difference in the majority of the data could be found or not (not all of the references Jessen mentions report statistical significance, but a general trend). The term ‘variations’ was used with respect to the amount of token-to-token variation or interspeaker variations. Three degrees of variation have been reported: minimal, small and large.

**Table 1.2:** Summary of acoustic differences between /b d g/ and /p t k/, adapted from Jessen (1998), p. 67, see text for description

| | | **Intervocalic** | **Utterance-initial** | **post-voiceless** |
|---|---|---|---|---|
| **Aspiration** | dominant:<br>variations: | difference<br>large | difference<br>minimal | difference<br>minimal |
| **Voicing** | dominant:<br>variations: | difference<br>small | no difference<br>large | no difference<br>minimal |

Results in Table 1.2 provide evidence that aspiration duration is the primary acoustic characteristic to distinguish between /b d g/ and /p t k/ when different positions are taken into account.

It can be generalised that the phonological voicing contrast is realised in different ways and varies according to the position of the obstruent:

The stops /b d g/ and /p t k/ can be distinguished primarily by aspiration duration rather than voicing in **utterance initial stressed position** (##_V) and also in the **post-voiceless position** (#_V). The phonologically voiced stops /b d g/ are often devoiced in these positions. Devoiced is a phonetic term which describes that a token is realised without sufficient voicing during oral closure. Different degrees of devoicing can occur. Phonological minimal pairs regarding the word initial position are for instance:

| | |
|---|---|
| *Daten* (data) vs. *Taten* (doings) | [ˈd̥aːtən] vs. [ˈtʰaːtən] |
| *Gipfel* (peak) vs. *Kipfel* (croissant) | [ˈg̊ɪpfəl] vs. [ˈkʰɪpfəl] |

Similar examples also occur in the syllable and morpheme initial positions, e.g.:

| | |
|---|---|
| *gebackt* (baked) vs. *gepackt* (packed) | [g̊əˈb̥akt] vs. [̥g̊əˈpʰakt] |

The stops /b d g/ and /p t k/ can be distinguished by voicing and by aspiration in **intervocalic V_V position**. Typically, the first vowel belongs to the stressed syllable and the second to the unstressed syllable in German. When the utterance consists of a CVCV-sequence and the first vowel is tense and stressed, then the following consonant will be syllable initial. Syllable initial /b d g/ can be come devoiced too. Their phonological voiceless counterparts /p t k/ might be realised with aspiration.

Minimal pairs regarding the post-stressed intervocalic position are e.g.:

| | |
|---|---|
| *Lieder* (songs) vs. *Liter* (litre): | [ˈliːdɐ], [ˈliːd̥ɐ] vs. [ˈliːtɐ], [ˈliːtʰɐ] |
| *Made* (maggot) vs. *Mate* (tea): | [ˈmaːdə], [ˈmaːd̥ə] vs. [ˈmaːtə], [ˈmaːtʰə] |

Coming from the German stop inventory to the fricatives, it turns out that the voicing contrast in fricatives is a particular challenge, since minimal pairs for the voicing contrast occur with different frequencies with respect to place of articulation.

In **utterance initial position** (##_V) minimal pairs for the labiodental fricatives can be found rather frequently, e.g.
*fahl* (sallow) vs. *Wahl* (election) [ˈfaːl] vs. [ˈvaːl].

Alveolar fricatives are phonologically always voiced (assuming a following vowel), but there are a few minimal pairs when including foreign vocabulary, e.g.
*sechs* (six) vs. *Sex* (sex) [ˈzɛks] vs. [ˈsɛks].

In the **intervocalic position** V_V both labiodental and alveolar voiced and voiceless fricatives occur. Since grapheme representations are of further importance for the recorded speech material (see 3.3.) they are also reported here. For the alveolar voiceless fricatives the graphemes *ß/ss*[7] were introduced into orthography to mark the voicelessness of /s/. In comparison to English where the grapheme *z* holds for the phonologically voiced fricative and the *s* for phonologically voiceless, a similar graphematic opposition cannot be used in German, since *z* is pronounced [ts] (e.g. people write the name *Suzanne* to indicate the voiced pronunciation, but *Suzanne* in German would be pronounced [zuˈtsanə]).
Mangold (1978) noted for the intervocalic /s/ vs. /z/ opposition that there are 56 minimal pairs in the German lexicon including words with flection, but only 11 real word pairs, whereas for /f/ vs. /v/ there are 8 to 9 times more minimal pairs. One of the 11 real word pairs for /s/ vs. /z/ is for instance:
*Muse* (muse) vs. *Muße* (leisure) [ˈmuːzə] vs. [ˈmuːsə].

Even though Mangold found only a few minimal pairs in the lexicon, his review gives evidence about the /s/ versus /z/ distinction in intervocalic position. Following Mangold, the voicing distinction in intervocalic position is related to the preceding vowel. If the vowel is long (tense[8]) or a diphthong then /z/ occurs. If the vowel is short (lax) then /s/ occurs:

---

[7] In new German orthography ß is written after tense vowels and diphthongs and ss after lax vowels.

[8] In Mooshammer it is referred to Meier (1907) that the distinction between vowels in German is not only quantitative, but also qualitative (tense-lax), with some exceptions regarding /a/. Mooshammer (1998) gives an extensive overview on the theoretical background and the tense-lax distinction in German.

*lasen* (read, Past) vs. *lassen* (let) [ˈlaːzn̩] vs. [ˈlasn̩].

However, there are other authors who described /s/ and /z/ as allophones of /s/ dependent on the position in the word and varying according to regional dialect[9] (for the first argument see Mangold 1978 citing to Bithell).

Generally, there seems to be a lack of exhaustive acoustic investigations on the voicing contrast in German fricatives. Most studies are relatively old. Differences which have been reported are voicing and duration of the fricatives, with longer duration for the voiceless (Jessen 1998).

The voicing contrast in word final position was not discussed so far, although German is a key case study for the phonological rule called ‘Auslautverhärtung’. Since it is of particular importance the following section is dedicated to this issue.

### *1.3.3. Issues around final devoicing*

‘Final obstruent devoicing’ or ‘Auslautverhärtung’ or ‘final devoicing’ has been described as the “most popular of German phonological rules”(Giegerich, 1989, p.51). The phonological rule itself refers to the loss of voicing for phonologically voiced obstruents in word, morpheme or syllable final position, for example:

| | |
|---|---|
| *Rad* (wheel) vs. *Rat* (advice) | [ˈʁaːt], [ˈʁaːtʰ] vs. [ˈʁaːt], [ˈʁaːtʰ] |
| *lies* (read!) vs. *ließ* (let) | [ˈliːs] vs. [ˈliːs] |

Historically, it seems to have emerged during the transition from Old High German (approximately 750-1050) to Middle High German (1050-1350). Literary manuscripts from the Middle High German period, the only source to study sound changes, show changes in spelling where word final *b, d, g* became *p, t, k* e.g. *tac* [ˈtak] ‘day’; gen. sg. *tages* [ˈtagəs] (Brockhaus, 1995, referring to König 1978, p.73).

Phonologically, the contrast attracted so much attention since several theoretical implications were linked to the rule (for an extensive phonological overview see Brockhaus 1995). Given the binary feature framework of generative phonology (Chomsky and Halle 1968) final obstruent devoicing would be described as

[- sonorant] → [- voiced] / __#

i.e. a non-sonorant (= obstruent) becomes non-voiced (= voiceless) under the condition that it occurs in a word final position. The underlying form [- voiced] would be an underspecified feature [0 voice] which becomes specified in word

---

[9] “*Die Unterscheidung des stimmhaften von dem stimmlosen Laute ist in Mittel- und Süddeutschland unbekannt*” (Mangold 1978, p.35 citing Viëtor 1923).

final position and thus, the process of final devoicing has been described as a feature specifying rule. Brockhaus (1995) refers to it as the ‘text book solution’. Alternative phonological explanations have been given for instance by Mascáro (1987) in autosegmental phonology who described devoicing as a loss of a phonological property, i.e. the autosegment [+ voice] will be deleted from the representation of an obstruent or Lombardi (1991) who refers to the privative feature[10] [voice] which would be licensed from syllable onset.

The final devoicing rule is still a major issue in the phonetics-phonology interface. The focus changed from questions of a feature specifying rule, feature deletion, licensing etc. to the question whether the devoicing in word final position is complete (= full neutralisation) or not (= partial neutralisation). The concept of neutralisation itself goes back to Trubetzkoy[11].

Experimental phoneticians do often favour the solution of partial neutralisation, which might be based on the fact that there is generally a very small likelihood that one phoneme behaves as another, given the complex mechanisms underlying the speech apparatus and their empirical knowledge about the (non-existing) uniformity of acoustic or articulatory data in general. However, evidence was provided in both directions – for full (e.g. Fourakis and Iverson 1984) and for partial neutralisation (e.g. Port and Crawford 1989, Port and O’Dell 1985) so that the debate became more and more a debate about methodological problems. Results were dependent on statistical methods, speech corpora (read versus natural speech, isolated words versus words in a frame sentence, nonsense words versus real words, frequently occurring words versus not frequently occurring words) and experimental set-ups (the subjects were aware/not aware of the research question, hyperarticulated versus more natural speech conditions etc.). In addition, a recent acoustic study of Piroth and Janker (2004) finds differences in final devoicing with respect to regional variations (South German subjects preserved the voicing contrast whereas speakers from the area around Cologne and from the area Berlin-Brandenburg did not).

Taken together, based on acoustic results both, significant variations or no difference can occur between the phonologically voiced and the phonologically voiceless obstruents in a final position. It seems to be of particular challenge to explain neutralisation and the different empirical findings within a phonological theory. One phonological concept which can deal with these variations is the one of archiphonemes (Jakobson 1929). Archiphonemes are defined as an

---

[10] Privative features are not binary, they can be present or not, and do not carry the +/- value since the absence of a ‘+’ does not necessarily implies a ‘–‘.

[11] Brockhaus (1995) refers to 2 articles from 1933, I found a later article “Die Aufhebung der phonologischen Gegensätze” Trubetzkoy (1936).

> *“èlèment commun de deux ou plusieurs phonèmes corrélatifs, qu’on peut concevoir abstraction faite des propriétés de corrélation”* (Martinet 1936, p.46).

Within the archiphoneme concept neutralisation is not a question of one phoneme becoming like its phonological counterpart (Trubetzkoy 1939). There is a much finer differentiation, since an archiphoneme does not necessarily have to be realised identical to one of the phonemes of a contrast. It can be either one of the phonemes of a contrast or a third segment.

Some other important points which should be made here are:

1. Even though total neutralisation can be found in the acoustic signal, it is unlikely that this holds for articulatory movements (Fuchs and Perrier 2003).
2. Even though there might be partial neutralisation in the acoustic signals, it is not clear whether listeners perceive the remaining differences. Assuming the word final position would be a ‘weak’ position, Trubetzkoy would doubt the necessity of the perception of contrasting phonemes:

> *“Die psychische Deutung der besprochenen Aufhebungsarten ist nicht schwer. Das Wesen der Aufhebung besteht darin, dass gewisse Eigenschaften, die bei den Oppositionsmitgliedern in der Relevanzstellung [position where a phonological contrast is of particular importance, S.F.] deutlich wahrgenommen werden müssen, weil sie dort einen phonologischen Wert besitzen, in der Aufhebungsstellung [position where a phonological contrast diminishes/is eliminated S.F.] nicht wahrgenommen zu werden brauchen, weil sie phonologisch irrelevant sind. Die Aufhebung bewirkt also ein Nachlassen der Aufmerksamkeit, ein Sinken der Beachtungsschwelle“* (Trubetzkoy 1936, p.44).

Hence, the issue of final obstruent devoicing in German is not only a major topic in the phonetics – phonology interface, but also in the speech production – perception domain.

## 1.4. Specific aims of the current study

The current work will focus on German, since this language is known for a voicing contrast varying with position in the word or syllable. The distinction between German voiced and voiceless obstruents has recently been described as being primarily based on aspiration duration. This distinction has been experimentally supported by means of mainly acoustic data from Jessen (1998) who found besides the primary aspiration distinction, other differences which often relate to the supralaryngeal level. Supralaryngeal correlates were also found in different studies of other languages, increasing the complexity of the description of the contrast.

Thus, the current study aims at providing a comprehensive description of the voicing contrast in German by:

1. Collecting not only acoustic, but especially ARTICULATORY DATA at both levels: the LARYNGEAL, and the SUPRALARYNGEAL level,
2. Studying THE LARYNGEAL-SUPRALARYNGEAL CO-ORDINATION,
3. Increasing the number of subjects in comparison to Jessen's preliminary work on laryngeal abduction,
4. Considering three different positions of obstruents in a word, including the word final position which is of particular interest in German.

Given the findings published in the literature, this thesis attends to the correlates involved in the voicing contrast at the laryngeal and supralaryngeal levels. Therefore, a combination of articulatory investigations must be carried out in combination with acoustic analysis. In this thesis, a range of complementary experimental techniques will be used, each of which can reveal different aspects of the phenomenon simultaneously. The techniques selected are:

TRANSILLUMINATION, where an endoscope is inserted in the pharynx. The tip of the endoscope is placed above the glottis and provides the relevant light source for illuminating the glottis. Photosensors are placed outside at the neck in order to detect the light if the glottis is open. The detected light level correlates with glottal opening, giving evidence of the existence of glottal opening, a general idea about the magnitude of glottal opening, and a precise information about the timing of glottal abduction gestures.

FIBEROPTIC FILMS, where a video camera is attached to the endoscope and records the illuminated glottis. Fiberoptic films are used as an additional information and support the findings from the transillumination data qualitatively.

ELECTROPALATOGRAPHY, which records the contact of the tongue against the alveolar ridge and the hard palate. Such data provide evidence for: the actual

on- and offsets of tongue tip closure, to what extent the tongue touches the hard palate at which places, and how tongue-palate contact patterns change in time.
ELECTROMAGNETIC ARTICULOGRAPHY, where coils are attached to the tongue, jaw and some reference points in the midsagittal plane. Coils are situated in a magnetic field and the articulatory movements induce an electric current which can be transformed into a 2-dimensional space showing horizontal and vertical movements of the coils. By means of this technique a kinematic analysis of tongue and jaw movements and their corresponding movement amplitudes, velocities, and timing characteristics can be observed in detail.

Such a multichannel analysis presents many difficult methodological problems to solve, but has the power to reveal supralaryngeal and laryngeal articulation in a co-ordinated way. In particular, such an experimental set-up will let me examine the following hypotheses, which get to the heart of the issue of the multiple phonetic correlates for what may be regarded as a simple one-dimensional phonological contrast of the larynx:

1. The presence/absence of glottal opening is NOT a consistent correlate of the voicing contrast in German.
2. SUPRALARYNGEAL correlates are also INVOLVED in the contrast.
3. SUPRALARYNGEAL correlates can COMPENSATE for the lack of distinction in laryngeal adjustments.

Extending the characterisation of the voicing contrast from the glottal to the supralaryngeal domain offers a new insight into the final devoicing phenomenon. Phonologically, it is expected that neutralised voiced obstruents should have laryngeal and supralaryngeal correlates similar to their voiceless counterparts.

# Chapter 2: Articulatory correlates of the voicing contrast: state of the art

> "Claims about the innate capabilities of man are uninteresting
> if they are not made in terms of measurable physical or physiological events."
> (Ladefoged, 1973, p.73)

## 2.1. Introduction

Chapter 2 goes more into articulatory details and reports previous investigations on the voicing contrast under three perspectives: laryngeal correlates, laryngeal-oral co-ordination and oral correlates. The report focuses on investigations of single obstruents in German or comparable languages with a phonological 2-way voicing distinction. It is suggested that even though most effort has been devoted to observations of laryngeal correlates, supralaryngeal correlates play an important role too and cannot be neglected.

1. The first section "laryngeal correlates" starts with a brief anatomical introduction in order to explain intrinsic and extrinsic laryngeal muscles and their involvement in the voicing contrast. It continues with a summary of rare and invasive experiments, which study laryngeal muscular activation during the production of voiced and voiceless obstruents. Afterwards, a short overview including other, more straightforward techniques examining laryngeal adjustments is provided. More emphasis is given to the following subsection: studies investigating laryngeal opening gestures. It focuses on the occurrence and the amount of glottal opening in single obstruent production, first in German and later in comparable languages.
2. In the second section 'laryngeal-oral co-ordination' the co-ordination between the larynx and the upper articulators is discussed considering different timings. The emphasis is laid on differences between aspirated versus unaspirated stops. There follows a short 'walk on the experimental wild side' regarding perturbations and laryngeal-oral co-ordination in stop production. The section finishes with a description of laryngeal-oral co-ordination in fricative production which differs from the one found in stop production.
3. The third section 'supralaryngeal correlates' is divided into three parts. The first part is dedicated to strategies possibly involved in enlarging the oral cavity in order to maintain voicing during oral closure. The second

part summarises experiments discussing tongue and jaw movements and their potential involvement in the voicing contrast. However, it is unclear whether tongue and jaw movements are concomitants for cavity enlargement strategies or independent mechanisms. The third part summarises investigations regarding tongue-palate interactions and their possible role in the production of the voicing contrast.

## 2.2. Laryngeal correlates

### *2.2.1. Functional anatomy of the larynx*

Leaving communicative factors aside, the general biological basis of speech production per se consists of muscle activities, biomechanics and aerodynamics. Regarding the voicing contrast it is necessary to give a brief introduction to the morphology of the larynx, since it is the source for vocal fold vibrations (phonetically associated with voicing) as well as the lack of vibrations (phonetically associated with voicelessness or devoicing). This section relates anatomical structures of the larynx to their function with respect to the voicing contrast.

In anatomy, muscles are named after the structures according to where they start and where they end. For instance the cricothyroid muscle starts at the cricoid cartilage and ends at the thyroid cartilage. The origin of the muscle has to be a fixed point. The decision was made to start with a summary of the related structures, i.e. with the laryngeal cartilages and to continue with the relevant muscles.

Figure 2.1 exhibits laryngeal cartilages in different perspectives: a) for the sagittal view, b) for the frontal view c) for the back view, and d) for the top view. The epiglottis is a cartilage protecting trachea and lungs from any foreign body, particularly during ingestion. Backward movement of the tongue causes downward movement of the epiglottis (which can become a problem in transillumination, see 2.2.3.). On top of the larynx the hyoid bone is situated with its horseshoe-shaped form. The hyoid is further connected via different muscles with the mandible, the cranium, the sternum and the clavicle. Cartilages which are primarily important for speech production are the thyroid cartilage, the arytenoid cartilages, and the cricoid cartilage. The thyroid cartilage looks like two shields which are connected to each other. The angle of these shields differs between males and females.

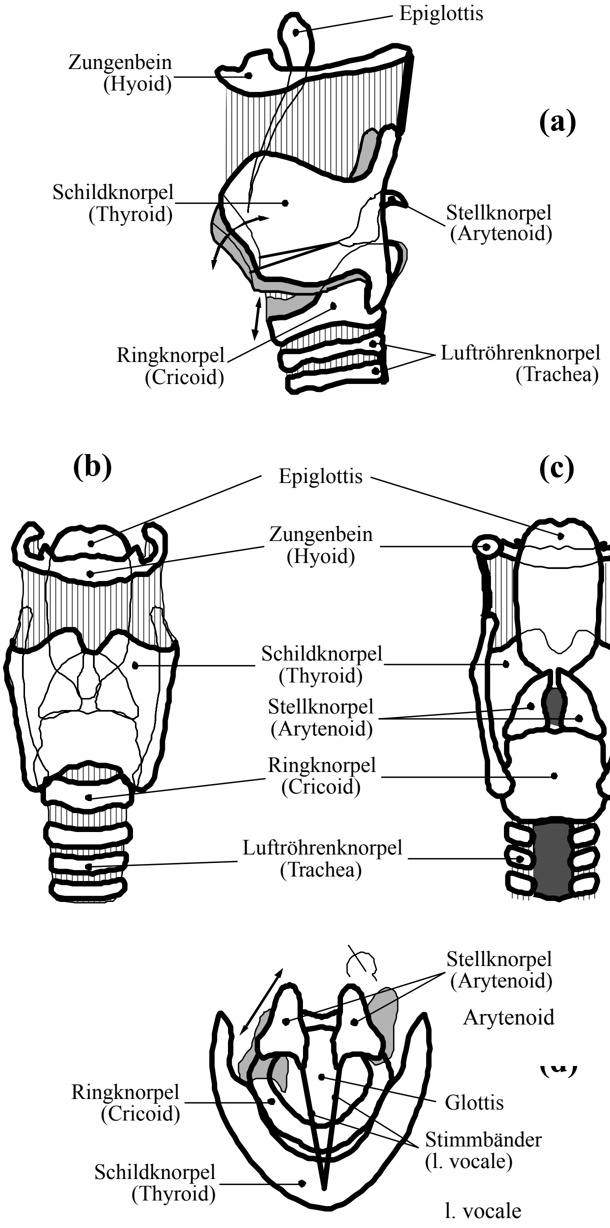

**Figure 2.1:** Main cartilages of the larynx and its surrounding structures; a) sagittal view, b) frontal view, c) back view, d) top view; membranes hatched vertically, ligaments in dark grey, hidden structures are displayed due to fine contours, possible movements of the cartilages are marked by arrays, from Pompino-Marschall (1995), p.33, fig.10 (copyright with permission from Walter de Gruyter, Berlin)

For males it is an acute angle which can be seen in the front of the neck and is popular under the name ‘adams apple’. The cricoid cartilage is the cartilage of the trachea in cranial direction. The cricoid and the thyroid are linked with each other due to the cricothyroid joint. Two different movements are generally possible (Hardcastle 1976): an upward movement of the anterior part of the cricoid or a forward and downward movement of the thyroid (see grey areas in Figure 2.1. (a)). Those movements can be used to tense the vocal folds passively. The pair of arytenoid cartilages is located at the cricoid via the cricoarytenoid joints. These joints allow the rocking and gliding motion of the arytenoid cartilages. Since the vocal cords are directly linked to them, any movement of the arytenoids will have an effect onto the vocal folds.

The muscles connecting the cartilages (thyroid, cricoid, arytenoid) are listed in the left column of Table 2.1 with their corresponding function on the right. They are intrinsic laryngeal muscles (situated in the inside of the larynx) and divided into two functional groups, the abductors and the adductors. An abductor is defined as a muscle involved in the movement where the vocal folds move apart (glottal opening, glottal abduction, glottal aperture) and an adductor is a muscle involved in the movement where vocal folds approach each other (glottal closing). Abduction and adduction are considered with respect to the glottis. The glottis is the space between the vocal folds (m. vocalis) and is not itself an anatomical structure.

**Table 2.1**: Main laryngeal muscles and their functions

| Muscles (Abbreviation) | Function |
|---|---|
| m. posterior cricoarytenoideus (PCA) | abductor |
| m interarytenoideus (INT)<br>(also m. arytenoid or transversus) | adductor |
| m. lateral cricoarytenoideus (LCA) | adductor, medial compression |
| m. thyroarytenoid with the m. vocalis (VOC) | adductor, longitudinal tension |
| m. cricothyroideus (CT)[12] | longitudinal tension |

One of the main characteristics of glottal opening and closing is the fact that there are several muscles which can adduct/close the glottis[13], but only one muscle which abducts/opens it. The greater number of adductors could have been developed during evolution, since the primary function of the larynx as a valve sitting on top of the trachea is to protect the lungs from any external bodies or fluids[14]. For protection it was necessary to adduct the glottis, i.e. to

---

[12] In some classifications the CT belongs to the extrinsic laryngeal muscles.

[13] Note, all described laryngeal muscles are muscle pairs and do also often show different branches.

[14] During evolution the larynx has been developed in parallel with the migration from aquatic to terrestric environment. First larynges were found in the family of the African ray finned fish (polypteridae) and the African lung fish (protopterus) which had gills and lungs (Hirose 1975).

close the valve. Glottal adduction is a characteristic of phonation too, the secondary function of the larynx.

Glottal abduction can be seen during breathing or during the production of voiceless sounds. During quiet breathing the amount of glottal abduction is considerably larger compared to the production of voiceless speech sounds (see Figure 2.2).

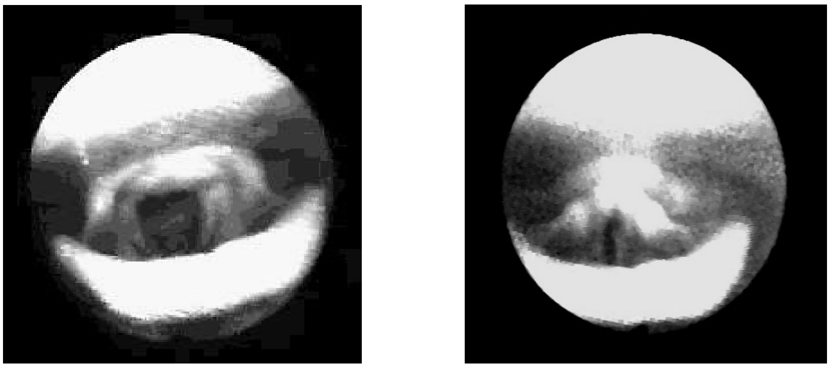

**Figure 2.2**: Fiberoptic film data of subject CG (current study), left: quiet breathing; right: glottal opening during the production of stressed /t/ followed by /u/

Hence, there are differences in glottal opening between the speech mode and breathing.

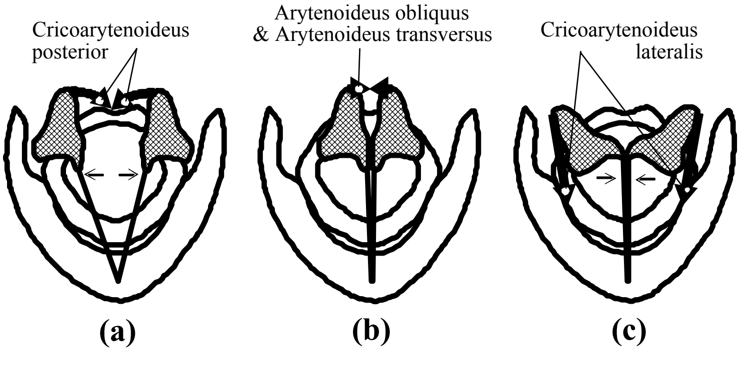

**Figure 2.3**: Schematic view of different laryngeal adjustments and the involvement of the appropriate muscles: (a) during breathing (abducted), (b) during phonation (adducted), (c) during whispering (closed glottis with open triangle at the posterior part of the glottis), from Pompino-Marschall (1995), p. 35, fig.13 (copyright with permission from Walter de Gruyter, Berlin)

Sawashima (1997) described glottal opening in the speech mode as an intermediate position of the vocal folds between maximum glottal abduction and glottal adduction.

Figure 2.3 shows basic manoeuvres due to intrinsic laryngeal muscle activity. Activity of the posterior cricoarytenoideus muscle (PCA) causes an opening of the glottis (a), activity of different parts of the arytenoideus muscles close the glottis (b), and activity of the cricoarytenoideus lateralis causes a wedged shape orifice by rotating the arytenoid cartilages. As a result vocal folds are adducted, but there is still an open part between the rotated arytenoid cartilages.

The larynx is connected to the jaw or cranium via the hyoid in cranial direction and connected to the sternum and clavicle in caudal direction. Muscles which build these connection are the extrinsic laryngeal muscles (Figure 2.4).

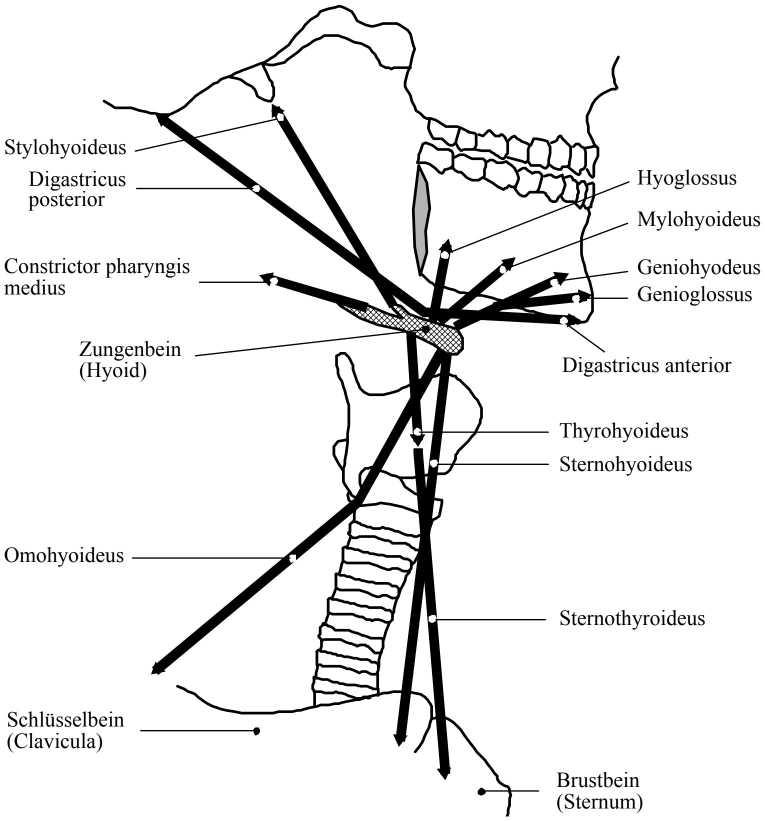

**Figure 2.4:** Schematic view on extrinsic laryngeal muscles, arrays correspond to the direction of muscle contraction with respect to the hyoid bone; from Pompino-Marschall (1995), p.42, fig. 16 (copyright with permission from Walter de Gruyter)

All the muscles (Figure 2.4) above the hyoid bone are potential laryngeal elevators (also called suprahyoid musculature) since their contraction causes a shortening of muscle length in cranial direction (with all other muscles being deactivated). The muscles below the hyoid bone (so called infrahyoid musculature) can potentially lower the larynx position (with all other muscles being deactivated). Laryngeal elevation and depression are mechanisms used in vowel production (e.g. laryngeal lowering in /u/) or laryngeal up and downward movements in swallowing, but it is unclear to what extent they could be involved in the production of the voicing contrast (see also 2.4.1.). Hardcastle (1976) pointed out that extrinsic laryngeal muscles would primarily be responsible for gross movements of the larynx. However, some suggestions of a potential involvement were also made by the author, e.g.: contraction of the digastricus raises the larynx and may be involved in voiceless stop production to guarantee a high intraoral pressure, and increasing activities of the geniohyoideus, digastricus, genioglossus, and mylohyoideus could be involved in raising fundamental frequency (which is a shared correlate of the voicing contrast).

Kent and Moll (1969) have also suggested that hyoid bone depression and thus, laryngeal lowering could be one mechanism in the production of voiced stop. A more detailed discussion about the role of laryngeal lowering in the voicing contrast will be given later.

Surprisingly in surgery it is quite common (in personal communication with Dr. Dahlmeier) that a removal of the body of the hyoid bone (due to thyroglossol duct cystes) and hence the cut of all the muscles which are connected to the hyoid bone body does not cause changes with respect to speech production or swallowing.

The current literature review will further concentrate on experimental investigations of intrinsic laryngeal muscles and their role in the production of the voicing contrast. It will not be speculated about potential mechanisms regarding extrinsic laryngeal muscles and their interaction. The only motor control model for this multi-muscle complex of human jaw and hyoid movements (I am aware of) was described by Laboissière et al. (1996), but it is not related to the voicing contrast and therefore not further discussed here.

### *2.2.2. Laryngeal activity in the voicing contrast*

In the following section various laryngeal manoeuvres of the intrinsic laryngeal muscles will be described which participate in the production of the phonological voicing contrast.

**Principles to produce voicing or phonation**: It is assumed that the adductors (INT, LCA, VOC, CT) are involved in order to close the glottis and sufficiently tense the vocal folds. Since there are at least three intrinsic laryngeal muscles for vocal fold adduction (INT, LCA, VOC) and two for tension (VOC, CT), single muscles might be activated or a combination of all five could be involved. Phonation can *NOT* only be explained by muscular activity induced by neural impulses (see 1.3). Aerodynamic factors have to be involved too. As a pendant to the neuro cronaxic theory van den Berg (1958) postulated the myoelastic theory, a combination of neural activity and aerodynamic factors.

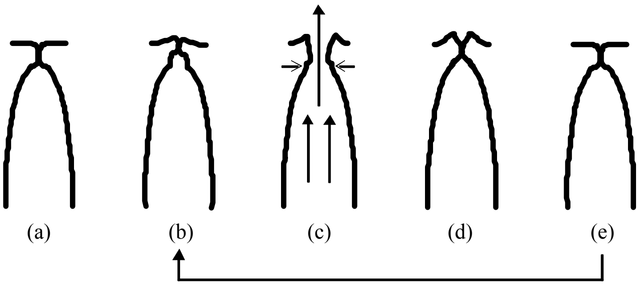

**Figure 2.5**: Schematic frontal view of vocal fold motions during phonation: (a) closed glottis, (b) influenced by subglottal pressure (c) released vocal folds due to subglottal pressure (filled arrays correspond to airflow and open arrays to Bernoulli forces), (d) elastic closing glottis, (e) onset of a new phonatory cycle; from Pompino-Marschall (1995), p.34, fig.12 (copyright with permission from Walter de Gruyter)

Figure 2.5 exhibits a phonatory cycle from a frontal view. Glottal adduction and appropriate tension via muscular activity, plus a transglottal air pressure difference are the pre-conditions for phonation[15]. If vocal folds are adducted the subglottal pressure can raise to such an amount that it overcomes vocal fold resistance and forces the folds to move apart (c). Simultaneously with the release of the vocal folds, subglottal pressure decreases again and in addition, elastic tissue forces of the vocal folds suck the folds together again. This phonatory cycle can be repeated. One problem arises when the vocal tract is closed or constricted as it is the case in obstruents since intraoral pressure rises and the transglottal difference diminishes. If voicing should be maintained, strategies for enlarging the oral cavity have to be used (for more details see 2.4.1.).

**Principles to produce voiceless aspirated or devoiced sounds**: It is assumed that abductor activity (PCA) is involved in order to open the glottis. This mechanism is essential to produce aspirated sounds. However, it is well known that voicelessness does not necessarily involve an open glottis. Vocal fold tension to a degree which does not allow phonation (CT, VOC) as well as a sufficient increase in intraoral pressure would also explain the absence of vocal fold vibrations. Voicelessness can occur in both, phonologically voiced (which are realised as voiceless unaspirated) as well as in voiceless obstruents.

Empirically, the involvement of various laryngeal muscles in the production of the voicing contrast has been measured in terms of the bioelectrical activity by

---

[15] However, phonation can also take place with an abducted glottis, as for example in /h/, but generally phonation occurs more frequently with vocal fold adduction.

means of electromyography (hereafter EMG). For many muscles in the human body, bioelectrical activity can be recorded by electrode placement on the surface of the skin, above the relevant muscle (Zemlin 1988, p.140). Surface electrodes cannot be used to measure laryngeal activity reliably, since the signal would be influenced by adjacent muscles of the neck, and intrinsic muscles are protected by the thyroid (except the CT). Therefore, hooked-wire electrodes are used in EMG recordings of laryngeal muscles. These electrodes are inserted into the laryngeal muscles either perorally, i.e. through the oral cavity or percutanably, i.e. from the outside through the neck (e.g. Hirano & Ohala 1969, Hirose 1979). Special practise and expertise for correct electrode placement is necessary, since one of the risks, especially of percutan insertion, consists in damaging the recurrent nerve, which innervates the vocal folds. Damaging the recurrent nerve can cause vocal fold paralysis. Hence, only a few studies have been undertaken, most of them at the University of Tokyo and at Haskins Laboratories, in close co-operation with researchers who have been highly specialised in this field like Hirose, Hirano, Sawashima, Yoshioka and more recently with Honda and Murano. The majority of EMG investigations were published between the early 60s and the late 80s. It appears that there is only one publication, (Hoole et al. 2004), on the voicing contrast using EMG recordings for German subjects. However, other languages have been studied more intensively which show similar contrasts as German. In particular, the literature on American English and Danish provides a useful comparison based on what is known already about German acoustics.

**PCA and INT activity**: In 1972 Hirose and Gay investigated laryngeal muscle activity by recording two speakers of American English. Their speech material consisted of voiced and voiceless obstruents in stressed[16] and post-stressed (the consonant following a stressed syllable) syllable position. For voiceless stops they observed increasing PCA activity[17] compared to the voiced items. The INT was suppressed at the time when the PCA reached its maximum patterns for voiceless stops, whereas in voiced stops it showed nearly continuous activity. Some years later Hirose et al. (1978) confirmed the reciprocal relationship between the PCA and INT during the production of voiceless obstruents in a cross-language study for American English (from Hirose & Gay 1972), Danish (from Fischer-Jørgensen & Hirose 1974), Japanese (from Hirose & Ushijima

---

[16] Since stress belongs to the whole syllable, not only to the vowel – this will be called the stressed position in my latter study.

[17] Muscle activity is associated with increasing frequency of firing motor units (for further explanation see Hardcastle, 1999).

1978) and French (from Benguerel et al. 1978). Similar data were also reported for Danish (Hutters 1984), and Dutch (Collier et al. 1979).

These investigations provide evidence that the PCA is activated and INT is suppressed in order to guarantee the relevant glottal abduction for phonologically voiceless obstruents, at least in word initial stressed position.

The averaged peak activity of the PCA was correlated with fiberoptic films showing the amount of glottal opening (Hirose 1975). A positive correlation was found with r = 0.86 at a 0.01 confidence level. Hence, a strong relationship between PCA activity and the amount of glottal opening was suggested. However, the influence of INT suppression was not taken into account.

Concerning the neural innervation of the PCA and the INT, Sanders et al. (1994) found that the nerval branch innervating the horizontal compartment of the PCA are connected with the branch going to the INT.[18] One assumption could be that when the PCA is activated, the INT is inhibited (suppressed) due to this branch.

**Positionally and mode induced variation of PCA and INT activity**: Sawashima et al. (1975) compared voiced and voiceless labial stops in absolute utterance initial position for two speakers of Japanese. They reported a temporal delay of vocal fold adduction. Results from the first subject showed a high PCA activity and a delay of the onset of INT activity. For the second subject a delay of INT activity was found too, but without any particular PCA activity. Sawashima et al. (1975) suggested that the speaker dependent differences were related to utterance initial position, where the first speaker was already in ‘speech mode’ which supports the results of high PCA activation and INT suppression for voiceless stops. Results from the second subject were explained in terms of the ‘respiratory mode’, which is known as a state where adductors and abductors are minimally activated and the glottis is considerably open.

If PCA activity correlates with the amount of glottal opening, an interesting question is whether the two subjects also differ in terms of their glottal width. The results of Hirose (1975) would predict such a difference.

Hirose and Gay (1972) compared voiceless stops in different positions. Obstruents in stressed syllable positions showed higher PCA activity than those in the post-stressed position. Similar results were also found by Hirose and Ushijima (1978) comparing voiceless word-initial and word-medial stops in Japanese. They reported less activity for the word medial position. Differences in activity regarding different positions were less extreme for voiceless fricatives.

Lisker and Baer (1984) compared among other consonants word final /p/ with the following word initial /p/, /i/ and /h/. In the utterance *kee**p** **ea**rrings* the PCA

---

[18] Both branches are parts of the superior recurrent laryngeal nerve (see Sanders et al. 1994).

showed some action and INT ‘slackening’, the glottis was closed and the vowel glottalised. In the utterance *kee**p** **p**eering* the typical patterns of PCA activation and INT suppression were shown, but for a longer duration than would be the case in single word initial stops. In the utterance *kee**p** **h**earing* a delay of PCA activity was found, but the INT showed values similar to the p#p utterance.

Even though these experiments are very rare it can be concluded that the activity of the laryngeal abductor muscle (PCA) and the inhibition of the adductor muscles, in particular the INT, can be influenced by word borders and the following initial phonemes. In addition, the level of activity seems to be related to the position of the obstruent in the syllable or word.

**PCA and INT activity for stops versus fricatives**: Hirose & Gay (1972) found similar patterns with respect to PCA activity comparing voiced and voiceless stops with fricatives. A considerable amount of activity was shown for both voiceless stops and fricatives. However, INT activity differed with respect to manner of articulation, i.e. a higher INT activity was found for /z/ compared to /s/, but this difference was less marked than during stop production. These results concerned the stressed position. For the post-stressed position, PCA activity differed between voiced stops and fricatives. Some activity was found for the fricative. This finding was related to a former transillumination study where voiced fricatives showed some opening.

Collier et al. (1979) investigated the activity of the laryngeal adductor muscles of a single Danish speaker. They reported a difference between INT activity in stops and fricatives. INT was most relaxed (suppressed) in voiceless fricatives, less suppressed for voiced fricatives, and also for voiceless stops. It did not show a reduction of activity/ suppression for the voiced stops.

**CT activity**: Examinations of EMG patterns for the VOC and the LCA for the voicing contrast produced inconsistent results (e.g. Hirose and Ushijima 1978, Hirose et al. 1978, Benguerel et al. 1978). VOC and LCA were more often related to prosodic factors such as pitch.

Further interest was directed to the activity of the CT. This muscle was mostly associated with changes of longitudinal vocal fold tension in rotating the cricoid and thyroid cartilages relatively to each other (e.g. Haberman 1986, Löfqvist et al. 1989). An active mechanism of the CT could explain the higher fundamental frequency (F0) which is often found after voiceless obstruents in comparison to the lower F0 after voiced obstruents.

By means of EMG, Kagaya and Hirose (1975) investigated CT activity in one speaker of Hindi, a language with a four-way distinctive voicing contrast. Their results provided evidence that the CT contributes to the voicing contrast by tensing the vocal folds. They proposed that this mechanism would help to facilitate voicelessness in unvoiced stops, i.e. CT activity is higher in voiceless obstruents compared to voiced.

Dixit and MacNeilage (1980) confirmed the findings of Kagaya and Hirose (1975) for another speaker of Hindi.

For Dutch, a language with a two-way contrast, Collier et al. (1979) described no differences in CT activity. Hirose and Ushijima (1978) found a temporary decrease in CT activity for voiced as well as voiceless stops in Japanese.

To further clarify the role of the CT for the voicing contrast, Löfqvist et al. (1989) examined two speakers of American English and one of Dutch. They found consistently higher CT activity in voiceless obstruents than in their voiced counterparts and a higher F0 at the beginning of the following vowel. Löfqvist et al (1989) suggested that the terms stiff folds versus slack folds for the voiceless-voiced contrast as proposed by Halle and Stevens (1971) would be in disagreement with these physiological data.

The most recent study on CT activity for 3 German speakers (Hoole et al. 2004) provides evidence for higher activity levels during voiceless consonant production compared to voiced. However, the timing of the higher activity relative to the consonant was quite variable and hence it was not always obvious that CT activity was contributing to a suppression of voicing or may have been planned directly to generate higher F0 in vowels following voiceless consonants (F0 differences usually extended throughout the vowel).

**Summary**: Describing the anatomical structure and the associated laryngeal muscles in the speech mode, two main manoeuvres concerning the voicing contrast have been assumed:

1. the abduction of the glottis, caused by activity of the posterior cricoarytenoid muscles (PCA) and
2. the adduction of the glottis due to the activity of the interarytenoid muscles (INT), the lateral cricoarytenoid muscles (LCA), the thyroarytenoid muscles, and the vocalis muscles (VOC).

In order to prove empirically the particular involvement of several muscles during the production of the voicing contrast, a number of EMG investigations were discussed (Hirose and Gay 1972, Sawashima et al. 1975, Kagaya and Hirose 1975, Hirose and Ushijima 1978, Hirose et al. 1978, Benguerel 1978, Collier et al. 1979, Dixit and MacNeilage 1980, Lisker and Baer 1984, Löfqvist et al. 1989, Hoole et al. 2004).

For different languages, EMG results provide evidence that the PCA participates actively in the production of voiceless obstruents and the INT is simultaneously suppressed. The voiced counterparts do not exhibit these patterns of high PCA activity and suppression of the INT (for review on PCA activity see also Hirose 1975). Concerning the influence of context, PCA activity patterns is higher in word initial or stressed position than in the word medial or post-stressed position.

In addition, higher CT activity has been found for voiceless obstruents which could explain the higher fundamental frequency often occurring in the following vowel.

### *2.2.3. Techniques to investigate laryngeal production mechanisms*

All techniques have their advantages and disadvantages depending on the aim of the observation. The previously described technique of electromyography is difficult in terms of the necessary practise for inserting electrodes, and of a possible migration effect of electrodes during a recording session (which can change the EMG output). It is also difficult to interpret the EMG output with respect to the special function of the relevant muscle (Hardcastle 1999, p.273). The technique is uncomfortable for the subject as well as invasive. Nevertheless obtaining reliable EMG data is important for the study of neuromuscular control mechanisms and many EMG studies have improved our knowledge of the physiology of speech production.

There are other techniques which are more frequently used to study laryngeal kinematics. Kinematics refers to the exploration of spatiotemporal movement sequences without looking at the forces causing the kinematic output (Hauger et al. 1999). Observing the underlying forces corresponds to dynamics[19]. The following techniques allow the investigation of devoicing gestures (specifically glottal abduction) in order to obtain a more detailed picture of articulatory mechanisms involved in the production of voiceless obstruents. The techniques are also used to study the voicing contrast in two directions: first, the occurrence of glottal openings and second, looking at differences in the amount of glottal opening.

The most common techniques used to investigate laryngeal adjustment are a combination of fiberoptic filming and transillumination or Photoelectroglottography (Hoole 1999). Since glottal abduction cannot be detected in the acoustic signal or by the laryngographic wave form, another technique had to be found to observe glottal opening. For example, in the production of a voiceless stop, glottal opening might start with the beginning of the oral closure, but there is no unique information in a spectrogram about glottal opening during oral closure. It could be either maximally abducted, partially open or not open.The

---

[19] However, there are differences in the definition of dynamics. Some other researchers define dynamics as the change of movements in time without going into the underlying forces.

same is true for the data obtained by the Laryngograph[20]. The laryngograph has two electrodes which are placed on both sides of the thyroid cartilages. If the vocal folds are in contact, an electrical current applied to the electrodes flows via the vocal fold contact. If the vocal folds do not touch each other, no electrical current traverses the glottal space. The advantages of this method are: it is straightforward, non-invasive, and easy to combine with other techniques. However, it does not give any information about the degree of glottal opening or the relevant start of glottal opening when the vocal folds are not in contact (see Baken 1987, p. 224-225). The signal can also be influenced by vertical movement of the larynx and by tissue or fatty layer under the skin, since fat is a poor conductor. Because the phonological voicing contrast in German is often produced with aspiration (Jessen 1998), laryngography is not an appropriate technique for such an investigation. Instead fiberoptic films, transillumination and possibly pulse-echo ultrasound are more reliable techniques to study the distinction since they permit the measurement of glottal abduction.

Fiberoptic filming, first presented by Sawashima and Hirose (1968), is a technique whereby a flexible fiberscope is inserted through the nose and placed in the pharyngeal cavity with its tip close to the glottis. The fiberscope consists of glassfibers with one bundle acting as a light guide to illuminate the glottis and another acting as an image guide. Motion pictures with a rate of 50-60 frames per second (Sawashima and Ushijima 1971) were achieved in older studies. The fiberscope can be connected to a video camera or even a high speed filming camera. The latter has developed rapidly and can, for instance, record 2500 frames per second (Kiritani and Niimi 1997, studying vocal fold vibration) or even more.

To derive estimates of glottal abduction from fiberoptic images it is necessary to study the images frame by frame throughout the relevant interval and measure glottal width as the maximal distance between the vocal folds. Glottal width values are arbitrary values, i.e. they depend on the distance between the tip of the fiberscope and the glottis. To solve this problem Kiritani (1971) proposed to monitor the fiberscope position by means of computer controlled x-ray. Another proposal was made by Fujimura, Baer and Niimi (1979) and Sawashima and Miyazaki (1974). They developed a stereo-fiberscope, consisting of two fiberscopes placed in a special stereo-appliance. The fiberscopes are inserted each one into a nostril, hereafter connected and calibrated by pulling it out from the mouth and then inserted again into the hypopharynx. The problem with this stereofiberscope is that it is difficult to use, it would be rather expensive compared to a usual fiberscope, and fixing the end of the two cables in the back of the mouth requires a lot of skill and patience (Osamu Fujimura in personal

---

[20] The Laryngograph was developed in 1957 by Fabre.

communication). Up to now, this problem has not been solved and all values describing glottal width have to be handled with care. However, as a device to investigate the timing of glottal abduction and adduction, and in combination with Photoelectroglottography, it is a reliable technique and used quite routinely. Photoelectroglottography (PGG) works with a similar principle: the glottis is illuminated, but a phototransducer picks up the light. There have been several developments, one with the light source inserted via the mouth or nostril and the phototransducer placed externally below the thyroid. Another possibility consists of placing the light source below the glottis and inserting a phototransducer through the nostril into the pharyngeal cavity to pick up the light above the glottis (for an early review of the different types of glottographs see Frøkjær-Jensen 1967). When PGG together with fiberoptic films became common in experimental phonetics, it seemed reasonable to choose the first version: illuminating the glottis from the pharyngeal cavity, measuring the amount of light through the glottis by externally placed phototransducers, and additionally obtaining images of laryngeal behaviour by means of fiberoptic filming[21] (Hoole 1999).

The combination of transillumination and fiberoptic films increases reliability in the sense that fiberoptic films enable control of possible influences on the transillumination signal. Such influences could be: epiglottal movement due to tongue retraction which causes a shadow on the glottis, saliva production or incorrect placement of the fiberscope. The problems of the PGG are extensively discussed in Hutters (1976). She focused mainly on the relationship between glottal width and the amplitude of glottal opening by means of synchronous transillumination and fiberoptic film recordings during the production of different voiceless consonants. The amplitude of glottal opening varied with the position of the endoscope and the position of the phototransducers. Hutters' (1976) results often exhibited a nonlinearity between glottal width and amplitude. Therefore the amplitude of glottographic curve has to be interpreted with caution.

However, Löfqvist and Yoshioka (1980) found a high correlation between results from both techniques. This was achieved by placing the phototransistor below the cricoid cartilage. Placement of the sensor on the cricothyroid membrane would not result in a good correlation, since patterns from this sensor showed baseline shifts related to intonation and abnormally large glottal

---

[21] The combination of high speed fiberoptic filming and transillumination during the production of steady phonation was also tested (Baer, Löfqvist and McGarr 1983). Both techniques give essentially the same information about maximum of glottal opening and glottal closure in the phonatory cycle.

openings for velars. Hence, choosing to place the transistor below the cricoid gave a reliable estimate of glottal opening in comparison to fiberoptic films.

Another technique to study the devoicing gesture is pulse-echo-ultrasound (PEU) as presented in Munhall (1984). PEU involves two ultrasound transducers, each placed on one side of the thyroid lamina. The technique is based on ultra high frequency sound waves produced with an electric crystal. The crystal emits sound waves and receives returning echoes (Stone 1997). The extent to which sound is reflected depends on the tissue, its density and the compressibility of the medium. Differences between media and/or tissue have an effect on the reflection of the sound, i.e. the larger the difference between two media at the tissue-air boundary of the vocal folds, the larger the amount of reflected energy (Munhall 1984). The advantage of the technique is that it is applied externally, but compared to transillumination and fiberoptic filming its reliability is less proved (Hoole 1999).

Some of the older investigations used x-ray or cineradiographic filming. Because of the danger of the radiation it is not used as frequently as during the time when the technique was first developed.

Following the description of different techniques which allow investigations of the laryngeal devoicing gesture, a combination of fiberoptic filming and transillumination seems to represent a good compromise between invasive versus non-invasive methods, and user-friendly techniques versus those which require special expertise. It is also reliable and accurate concerning the timing of glottal abduction, although relative values of the amplitude of glottal opening have to be handled with care.

### *2.2.4. Observing glottal opening: a review*

The following section gives an overview of the main transillumination and fiberoptic film studies which have already been carried out. To provide an overview of these studies a table (Table 2.3) was created which was organised according to publication year. It includes the authors, the language studied, the techniques used, the number of subjects, and the speech material and its position in the syllable, word or sentence. Some general remarks onto the history of investigations of laryngeal adjustment can be derived. Table 2.2 summarises the abbreviations used in Table 2.3 and hereafter.

**Table 2.2:** Abbreviations used in Table 2.3 and throughout the present study

| **Abbreviations** | |
|---|---|
| Airfl | Airflow |
| Ac | Acoustic |
| Cineradio | Cineradiographic filming |
| EGG | Laryngography/ Electroglottography |
| Electr Transconduc | Electrical Transconductance Technique (electrodes of a modified EGG) |
| EMG | Electromyography<br>INT: Interarytenoid muscle (peroral)<br>CT: Cricothyroid muscle (percutan)<br>LCA: Lateral Cricoarytenoid muscle (percutan)<br>PCA: Posterior cricoarytenoid muscle (peroral)<br>SH: Sternohyoid muscle<br>VOC: Musculus vocalis (percutan) |
| EMA/EMMA | Electromagnetic Articulography |
| EPG | Electropalatography |
| Fiberop | Fiberoptic filming |
| LED | Infrared light-emitting diodes |
| LGG | Laryngography |
| Pala | Palatography |
| PEU | Pulse-echo-ultrasound |
| Press Trans | Pressure Transducer |
| Rothenberg | Rothenberg Mask |
| Trans | Transillumination/ Photoelectroglottography |
| Velo | Velograph |

Table 2.3 was also created as a reference whenever results from another experiment are reported. The reader can look into the Table in order to get an overview about the relevant experiment which is referred to. Table 2.3 might also help in further experimental work to find out quite quickly what would be interesting for the planned project.

The marked column “Position/s” in Table 2.3 provides information about the studied position of the consonants in the syllable, morpheme, word or utterance. The original terminology used by the authors is maintained, except the pre-stressed position is called stressed position. This column also included particular effects that were the focus of the study such as the influence of word boundary on glottal gestures, the influence of speech rate or loudness, or the influence of perturbations. Table 2.3 does not include articles which summarised the studies named earlier like e.g. Hirose (1975), Hirose and Sawashima (1981, 1983) and Löfqvist (1995).

**Table 2.3:** Summary of the main investigations

| Publ. year | Author/s | Language | Technique/s | Subject/s | Position/s | Speech material |
|---|---|---|---|---|---|---|
| 1960 | Sonesson | | Trans (Development of PGG) | | | |
| 1965 | Malécot & Peebles | | Trans (Testing PGG) | | | /p b m / |
| 1967 | Slis & Damsté | Dutch | Trans | 1 | Word initial & intervocalic position | /p t k b d f s x v z w h m n l r h/ |
| 1967 | Frøkjær-Jensen | | Trans (Development of another PGG) | | | |
| 1968 | Sawa-shima & Hirose | | Trans (Testing PGG) | | | |
| 1968 | Sawa-shima | Japanese | Ac, Trans | 1 | CVCV sequences | /p b t d s z h kk/ |
| 1969 | Lisker et al. | American English | Ac, Trans | 1 | Stressed and unstressed position | /p t k v ð z ž s š/ |
| 1970 | Kim | Korean | Ac, Cineradio | 1 | Word initial position | /p p’ $p^h$ k k’ $k^h$ t t’ $t^h$/ |
| 1970 | Lisker et al. | American English | Ac, Cineradio | 3 | Preceding unstressed vowels | /b p g k/ |
| 1970 | Sawa-shima et al. | American English | Ac, Fiberop | 3 (1 discus-sed here) | Word initial & final position | /p t k s v z h ǰ/ |
| 1970 | Sawa-shima | American English | Ac, Fiberop | 3 | Word initial & final position | /p k ð v z h ǰ/ |
| 1971 | Frøkjær Jensen et al. | Danish | Ac, Trans, Airfl | 3 | Word initial, stressed position | /b p h f/ |
| 1971 | Fujimura & Sawa-shima | American English | Ac, Fiberop | 1 | Word final with follo-wing word initial posi-tion (boun-dary effects) | /t d/ |
| 1971 | Kagaya | Korean | Ac; Fiberop, Trans | 1 | Word initial and medial | /p p’ $p^h$ k k’ $k^h$ t t’ $t^h$/ |
| 1972 | Hirose et al. | American English | Ac, Fiberop, EMG (INT, PCA,LCA, CT,SH) | 1 | Intervocalic, stressed position | /b ɓ $b^h$ $p^h$ p/ |
| 1972 | Lindqvist | Swedish | Ac, Fiberop, Trans | 1 | Utterance initial & final position | /b d g p t k v j f s ç ɧ m n ŋ r l/ |

| **Publ. year** | **Author/s** | **Language** | **Technique/s** | **Subject/s** | **Position/s** | **Speech material** |
|---|---|---|---|---|---|---|
| 1973 | Sawashima & Miyazaki | Japanese | Ac, Fiberop | 1 | Word initial & medial position | /s k t/, geminates |
| 1974 | Sawashima & Niimi | Japanese (Tokyo dialect) | Ac, Fiberop | 3 | Word initial & medial position | /p t k s ts/, geminates |
| 1974 | Kagaya | Korean (Seoul dialect) | Ac, Fiberop | 2 | Word initial and medial position | /p p’ p$^h$ k k’ k$^h$ t t’ t$^h$ č č’ č$^h$ s’ s$^h$/ |
| 1974 | Hirose et al. | Korean | Ac, EMG (VOC, LCA, CT), Fiberop | 1 | Word initial before unstressed vowel | /p p’ p$^h$ k k’ k$^h$ t t’ t$^h$/ |
| 1974 | Fischer-Jørgensen & Hirose | Danish | Ac, EMG (PCA, INT, VOC, LCA, CT+lab.musc.) | 6 | Word initial position | /p b t d k g s f l m h/ |
| 1975 | Kagaya & Hirose | Hindi | Ac, Fiberop, EMG (INT, LCA,CT,VOC) | 1 | CVCVˈCV | /b p b$^h$ p$^h$ d t d$^h$ t$^h$/ |
| 1975 | Sawashima et al. | Japanese (Tokyo dialect) | Ac, Fiberop, EMG (VOC, INT, PCA) | 2 | Word initial | /p b s z h m/ |
| 1976 | Iwata & Hirose | Mandarin | Ac, Fiberop | 2 | Word initial position | /t t$^h$ tɕ tɕh/ |
| 1977 | Pétursson | Icelandic | Ac, Trans | 1 | Initial clusters & word final, boundary effects | /st/ |
| 1977 | Kjellin | Tibetan | Ac, Fiberop, EMG (CT, SH, VOC) | 1 | Word initial | /p p$^h$ b s z ʔ ɦ h/ |
| 1977 | Butcher | German | Ac, Trans, LGG | 1 | VCV-sequences | /t t$^h$ s p p$^h$ f/ |
| 1978 | Sawashima et al. | Japanese (Tokyo dialect) | Ac, Fiberop EMG (PCA, INT) | 2 | Word initial & medial position | /t s z/, geminates |
| 1978 | Hirose & Ushijima | Japanese (Tokyo dialect) | Ac, EMG (PCA, INT, VOC, LCA, CT), Fiberop | 1 | CVn, CVCV-sequences | /p b t d k g s z h/ |
| 1978 | Benguerel et al. | French | Ac, Fiberop, EMG (INT, PCA, LCA,CT, SH) | 2<br>1 | Utterance initial, medial & final | /p b t d k g f v s z ç/ |
| 1978 | Löfqvist & Péturrson | Swedish & Icelandic | Ac, Trans, Press Trans | 1<br>1 | CVCV-sequences | /p b t d k g/<br>/p p$^h$ t t$^h$ c c$^h$ k k$^h$/ |
| 1979 | Iwata et al. | Fukinese | Ac, Fiberop | 3 | Morpheme initial & final position | /b p p$^h$ l t t$^h$ g k k$^h$ ʔ tʃ dʒ/ |
| 1979 | Sawa- | Korean | Ac, Fiberop | 2 | Syllable final | /k k’ k$^h$/ |

| Publ. year | Author/s | Language | Technique/s | Subject/s | Position/s | Speech material |
|---|---|---|---|---|---|---|
| | shima & Park | | | | | |
| 1979 | Collier et al. | Dutch | Ac, EMG (PCA, INT, VOC, ICA, CT) | 1 | Intervocalic position | /p b t d k f v s z x ɣ H/ |
| 1980 | Benguerel & Bhatia | Hindi | Ac, Fiberop | 2 | Word inital, medial & final position | /b p b$^h$ p$^h$ d t d$^h$ t$^h$ g k g$^h$ k$^h$ dʒ tʃ dʒ$^h$ tʃ$^h$/ |
| 1980 | Dixit & MacNeilage | Hindi | Ac, EMG (CT) | 1 | Stressed initial medial; post-stressed final position | /p t tš k$^h$ p t$^h$ tš$^h$ k$^h$ b d ḍ dž g b$^h$ d̊$^h$ dž$^h$ gb/ |
| 1980 | Löfqvist & Yoshioka | Swedish | Ac, Trans, Fiberop, EMG (PCA, INT) | 1 | Initial & final position, boundary effects | /s k st ks sts sp p$^h$/ |
| 1980 | Löfqvist & Yoshioka | Icelandic | Ac, Trans, Fiberop | 1 | Word initial, medial & final position | /s b p h d t/ /st ks sts s k sp/ |
| 1980 | Löfqvist | Swedish | Ac, Trans, Airflow, Press Trans | 2 | Stressed, unstressed & initial, medial position | /t/ |
| 1981 | Löfqvist et al. | - | Ac, Trans, Fiberop, Press Trans | 2 | Testing the control of glottal opening with & without visual feedback | |
| 1981 | Iwata et al. | Cantonese | Ac, Fiberop | 1 | Syllable initial, medial & final, sentence final, boundary effects | /p t k/<br>/t k/ + /y l tʃ p$^h$ tʃ$^h$ h ʃ f/,<br>/k/ + /a ts ts$^h$/,<br>/p/ + /s/ |
| 1981 | Yoshioka et al. | American English | Ac, Trans, Fiberop, EMG (PCA) | 1 | CVCV, CVCVC, VCV, boundary effects | /s k sk ks kk ksk ss ssk sks skk sksk kss kssk skss skssk/ |
| 1981 | Anders | Danish | Ac, Trans, Fiberop,<br>EMG (INT, PCA) | 2<br>1 | Word initial position, variations of loudness and rate | /p/ |
| 1983 | Fukui & | Danish | Ac, Fiberop, | 2 | Word initial, | /sp sb/ |

| Publ. year | Author/s | Language | Technique/s | Subject/ s | Position/s | Speech material |
|---|---|---|---|---|---|---|
| | Hirose | | EGG | | medial & final position, boundary effects | |
| 1983 | Hoole et al. | German | Ac, Trans, Press Trans | 3 | Stressed & post-stressed position | /p t st/ |
| 1984 | Löfqvist & Yoshioka | American English | Ac, Trans, Fiberop, Pala | 2 | Initial stressed & medial un-stressed position; effect of rate | /t s/ |
| 1984 | Munhall | Canadian English | Ac, PEU | 2<br>2 | Intervocalic, rate & stress effects | /s t/<br>/s t st/ |
| 1984 | Yoshioka | Japanese | Ac, Trans, Fiberop, Press Trans | 3 | Utterance medial | /g k s/ |
| 1984 | Lisker & Baer | American English | Ac, EMG (PCA, INT), Trans, Press Trans | 1 | Word initial and final position; boundary effects | /b p/<br>/p/+/ e p h/<br>/s p t/ + /p/<br>/b p/ + /h/ |
| 1984/ 1985 | Hutters | Danish | Ac, Fiberop, Trans, EMG (PCA; INT; VOC, CT) | 5<br>7 | word initial | /p t k b d g f s h/ |
| 1987 | Löfqvist & McGarr | American English | Ac, Trans, Fiberop | 2 | Initial stressed & unstressed position, rate effects | /t s/ |
| 1987 | Hoole | Icelandic | Ac, Trans, Velo | 1 | Word medial post-stressed position | /t nt nd p mp mb/ |
| 1987 | Ni Chasaide | Icelandic<br>Irish | Ac, Trans, Airflow | 1<br>1 | CVC-sequences; stress effects | /$^{h}$p p/<br>/t$^{h}$/ |
| 1989 | Löfqvist et al. | American English<br>Dutch | Ac, EMG (CT) | 2<br>1 | Word initial position | /p b t d k g f v θ ð s z ʃ ʒ tʃ dʒ/ |
| 1989 | Dixit | Hindi | Ac, Trans | 1 | CV, VCV, VC-sequences | /p p$^{h}$ b b$^{h}$/ |
| 1991 | Cooper, (A.M.) | English | Ac, Trans | 2 | Word initial & medial, stressed & | /p t k/ |

| Publ. year | Author/s | Language | Technique/s | Subject/s | Position/s | Speech material |
|---|---|---|---|---|---|---|
| | | | | | poststressed | |
| 1992 | Löfqvist & McGowan | American English<br>Swedish | Ac, Rothenberg, Trans, Fiberop | 1<br>1 | Syllable initial | /b m h v s p sp/ |
| 1992 | Munhall & Löfqvist | English | Ac, Trans, Fiberop | 2 | Word final position, boundary effects & rate | /s/+/t/ |
| 1994 | Munhall et. al. | English | Ac, Trans, LED, Lip paddle, Press Trans | 3 | Unstressed, word medial position, effects of jaw perturbation | /p/ |
| 1995 | Jessen | German | Ac, Trans, Fiberop | 1 | Word initial & medial position | /p t k b d g f v s z/ |
| 1998 | Jessen | German | Ac, Trans, Fiberop | 1 | Word initial and medial position | /p t k b d g f v s z/ |
| 1998 | Jessen | German | Ac, Trans, Fiberop | 1 | Word final, boundary effects | /ʃ/+/p t k f s/ |
| 1998 | Saltzman et al. | American English | Ac, Trans, Fiberop, LED, Lip paddle | 2 | Discrete & repetitive, effects of perturbation | /p/ |
| 1999 | Jessen | German | Ac, Trans | 1 | Word initial & medial position | /p t k/ |
| 1999 | Romero | Castilian Spanish | Ac, EMA, Trans, Fiberop | 1 | Word medial, boundary effects | /s/+/p t k b d g/ |
| 2000 | Fujino et al. | Japanese | Ac, Trans, Fiberop, Press Trans, EMA | 1 | VCVCVCV & VQCV utterances | /p t s/ |
| 2001 | Ridouane | Berber | Ac, Fiberop | 1 | | Voiceless words (C - clusters) |
| 2003 | Ridouane et al. | Berber | Ac, Trans, Fiberop | 1 | Influence of position and word boundaries | Clusters with 2-5 consonants /s/ and /k/ |
| 2003 | Hoole et al. | German | Ac, Trans, Fiberop, EPG | 3 | Word initial position | /p t f ʃ pf ps ts tʃ ʃt ʃt pl fl ʃl ʃpl pfl/ |
| 2004 | Hoole et al. | German | Ac, EMG (CT) | 3 | Word initial and medial position | /f p b/ |

Table 2.3 shows that extensive work has been carried out on laryngeal production mechanisms. Earlier work was mainly dedicated to develop and test different techniques as well as to investigate laryngeal adjustment during the production of the voicing contrast (i.e. the occurrence and amount of glottal opening in the voiced versus voiceless obstruents). To date, American English and Japanese are the most frequently studied languages. In most of the studies reported in Table 2.3 one to three subjects were recorded. An exception is the extensive work of Fischer-Jørgensen and Hirose (1974) and Hutters (1984, 1985) on Danish.

With respect to different speech material, Table 2.3 also emphasises the fact that during the 1980s and 90s the focus of research changed from the voicing contrast (i.e. questions concerning the phonetics-phonology interface) to work in the area of speech motor control. Studies included voiced obstruents less frequently. The focus was more related to the influence of speech rate, word boundaries, consonant clusters, perturbations of laryngeal adjustment and its timing. Such change was likely connected to the question of variance versus invariance in speech production (e.g. Gracco and Abbs 1986, Perkell and Klatt 1986). However, since the current work is primarily dedicated to the voicing contrast in German, the following section will discuss this issue in terms of studies which already exist for German and afterwards for other languages with a similar contrast.

### *2.2.5. Studies investigating single obstruents in German*

**Butcher (1977)**: The first transillumination experiment concerning the voicing contrast in German was described by Butcher (1977). One subject was recorded. Butcher compared the amount of glottal opening in voiceless aspirated and unaspirated stops with voiceless fricatives. It was found that the size of glottal abduction was greatest for voiceless aspirated stops, least for voiceless unaspirated stops, and in between for voiceless fricatives. With respect to laryngeal-oral co-ordination, differences were found with respect to manner of articulation. Butcher's results provide evidence that glottal opening for fricatives started before supraglottal constriction was produced whereas for stops the onset of glottal opening occurred afterwards.

**Hoole et al. (1983)**: Although Hoole et al. (1983) studied only voiceless /p t s/ in 3 subjects and not their voiced cognates some of their findings are of relevance here:

- Peak glottal opening occurred before oral release for stressed /p t/ with smaller durational differences between peak glottal opening and oral release for /t/ compared to /p/.

- Vowel context did not have any particular effect on the patterns (it varied between /a e i/).
- Results from the post-stressed position showed larger inter-subject differences. For one subject only a very weak glottal opening was observable and no reliable measurements could be made. The authors proposed an active slackening process of the vocal folds with only very minor glottal abduction. For the two other subjects a reduced glottal aperture was found, with more or less comparable laryngeal-oral timing values.
- Differences in laryngeal-oral timing between stops and fricatives confirmed Butcher's (1977) findings.

**Jessen (1995, 1998)**: An investigation exclusively dedicated to the voicing contrast in German, including articulatory data, was reported by Jessen (1995, 1998). Jessen's primary interest was the degree of maximum glottal opening (see Jessen 1998, p.197). All the voiced and voiceless obstruents (or lax and tense obstruents in Jessen's terms) were recorded either in stressed, absolute word initial position or in post-stressed word medial position. The following stressed vowel (for obstruents in word initial position) or preceding stressed vowel (for obstruents in word medial position) was either /i/ or /a/. In most cases it was the high front vowel in order to get reliable transillumination results. Regarding the occurrence of glottal opening, Jessen reported only a few tokens with glottal opening for the word initial lax stops and fricatives, while in most cases a clear opening could not be demonstrated, since the transillumination signal was weak. Jessen associated these tokens with 'a grey area', an area where the signal is between presence and absence of glottal abduction. Additionally, lax obstruents often did not show the typical bell-shaped (=ballistic) glottal opening and sometimes multiple velocity peaks were found which made the segmentation procedure rather difficult (segmentation was based on the velocity signal). Word initial lax fricatives were excluded from further analysis since glottal opening was missing in most tokens.

Surprisingly, Jessen found glottal abduction for /b d g/ in the intervocalic post-stressed position in more than 50% of the cases. This position is typically known for the greatest likelihood to be produced with vocal fold vibrations and voicing through closure. The question arises whether this effect might be related to the bisyllabic word material, e.g. [ˈiːb̥ə] where the phonologically voiced stop is syllable initial. Depending on the cohesion of the first with the second syllable, the voiced stops could be produced rather independently of the first syllable. Assuming such a low cohesion, the /b/ could behave like a /b/ in word initial position. In a word with more than two syllables, voicing and no glottal opening could be suspected. However, this explanation is only speculative.

Concerning the amount of glottal opening, significant differences between tense and lax stops were described for all the positions. For tense stops the amplitude of glottal opening and its duration was consistently larger/longer.

As expected, laryngeal-oral timing for tense stops shows a tight co-ordination between peak glottal opening and oral release. Peak glottal opening occurred shortly after oral release. Onset of glottal abduction and onset of oral closure were tightly coupled too. The glottis started to open approximately 5-15 ms after oral closure. With respect to laryngeal-oral timing for lax stops peak glottal opening preceded oral release, particularly in word initial position. This pattern can be associated with unaspirated stops (see 2.3.1.). Glottal opening commenced later than onset of oral closure.

Jessen ranked different characteristics of the tense-lax contrast, to estimate their statistical stability. His ranking (in order of increasing importance) was: the acoustic aspiration duration; the overall glottal opening duration; the amount of glottal opening; and the duration between closure onset and peak glottal opening.

For the intervocalic fricatives the amount of glottal opening was consistently larger for tense compared to lax, but it was not accompanied by a longer overall glottal opening duration. Ranking the different parameters again, the amount of glottal opening was the most consistent cue to the contrast, followed by overall fricative duration, and the duration between peak glottal opening and vowel onset. Jessen concluded:

> *"Evaluating both stops and fricatives together, we find that the single parameter that is maximally reliable across stops and fricatives is Gmx [SF: the amount of glottal opening]. Thus, tense obstruents are reliably produced with a larger maximum of glottal opening than lax obstruents in German, according to the present results"* (Jessen 1998, p.227).

He noted that the primacy of glottal size over laryngeal-oral timing might be influenced by the problem defining on- and offset of glottal gestures with a small peak.

He also reported that his results were not in opposition to those of Löfqvist (1995) who claimed that interarticulatory timing would be a more reliable correlate of the voicing contrast than the amount of glottal opening.

Summarising Jessen's work on laryngeal adjustment and the tense versus lax contrast of German:

- Glottal abduction occurred in both tense and lax obstruents, but it was optional for the lax.

- The most consistent differences regarding the contrast were related to the amount of glottal opening. Laryngeal-oral timing differences were found too, but they occurred less consistently.

Jessen's work was based on one subject, and hence the current study will extend on Jessen's findings with additional comparisons with other languages (American English, Danish, Mandarin Chinese, Swedish, Japanese) having a two-way distinction. The next section is divided into two parts, one referring to studies investigating voicing contrast and the amplitude of glottal openings and the other referring to work on laryngeal-oral timing since both seem to play a major role in the production of the contrast.

### *2.2.6. The occurrence and amplitude of glottal opening*

**Results from American English**: Lisker et al. (1969) reported glottal abduction for /p t k/ in stressed position which was accompanied with the absence of glottal pulsing. In unstressed position glottal abduction was also found, but less frequent. For /b d g/ no glottal opening occurred. Fricatives showed a slightly different picture. In most of the cases glottal abduction was found (large glottal opening amplitude for voiceless and small for voiced), but with uninterrupted voicing for the phonologically voiced fricatives.
In another study by Sawashima et al. (1970) it was tentatively pointed out that:

> *"(1) Variations in glottal opening occur in running speech, and these are effected by controlling the arytenoid cartilages. (2) Speech sounds with a predominant non-transient noise are produced with opening of the arytenoids [SF: i.e. with glottal abduction]. (3) Speech sounds having predominant voicing are produced without opening of the arytenoids [SF: i.e. with glottal adduction]. (4) Other speech sounds are produced with a variety of glottal openings, ranging from those for which there is clearly separation of the arytenoids to those which are not distinguishable from phonatory position. Such sounds are the voiceless unaspirated stops, certain of voiced fricatives, and those varieties of English /b, d, g/ in which voicing is interrupted"* (Sawashima et al. 1970, pp.198-199).

In the same year Sawashima (1970) confirmed these preliminary results analysing three subjects.

Lisker et al. (1970) also published their results relating to the previously described variable cases of unaspirated stops. Their aim was to answer the question whether or not there is a separation between the arytenoids (opening gesture) for the unaspirated stops (which were phonologically either voiced or voiceless). But again, no general conclusion could be derived for all subjects, the authors found speaker-dependent differences.

Irrespective of the general occurrence of glottal opening with respect to different phonemes, another fact should be discussed here. Glottal opening, if it occurs, is for example affected by the position of the relevant obstruent in the word, by stress, and by speech rate. One of the experiments (see Lisker and Baer 1984) described in section 2.1.2. demonstrated how the amplitude of glottal opening in word final /p/ was influenced by the following word initial phoneme. When the following word started with another voiceless stop, the glottal amplitude behaved similarly to a voiceless aspirated stop (i.e. a large opening was found), but when word final /p/ was followed by a word initial vowel, the glottal gesture disappeared. Comparable findings were observed by Fujimura and Sawashima (1971) when word final /t/ was followed by word initial /d/ or /t/. When the second word started with a /d/, the glottis was found to be ‘nearly closed’. The authors also reported that the false vocal cords above the real vocal folds participated during the production of word final /d/ and also for /t/.

Cooper (1991) showed that voiceless stops in word initial stressed position showed larger glottal opening than voiceless stops in word initial unstressed position. This findings also applied to voiceless stops in word medial position. Munhall (1984) also found evidence for temporal and spatial influence of laryngeal gestures due to stress. A similar study varying stress and rate was carried out by Löfqvist and Yoshioka (1980). They suggested that a smaller glottal opening for the fast rate could explain the shorter overall glottal opening duration, but the authors did not present the relevant values. However, using the same speech material Löfqvist and McGarr (1987) described larger glottal abduction for voiceless stops in stressed position compared to the ones in unstressed position. Additionally, glottal opening amplitude was reduced at higher speech rates compared to the slower rates, but only for one subject.

Comparing peak glottal opening amplitude between stops and fricatives, Löfqvist and McGarr (1987) found a larger glottal opening for the fricative /s/ than for the voiceless stop /t/ in American English.

**Results from Danish:** Frøkjær-Jensen et al. (1971), Fukui and Hirose (1983) and Hutters (1984, 1985) reported a large amount of glottal abduction for the voiceless aspirated stops and a small amount for voiceless unaspirated. The small amount of glottal abduction occurred relatively consistently in word initial

position for /b d g/ (produced as voiceless unaspirated). Frøkjær-Jensen et al. (1971) suggested that it could be explained by aerodynamics, i.e. by an increase of intraoral pressure whereas Hutters (1984, 1985) explained it with neural activity which was also found in Fischer-Jørgensen and Hirose (1974).
Comparing fricatives and stops, Hutters (1984) reported a trend for a slightly smaller amount of glottal opening in fricatives compared to the voiceless aspirated stops, but she also noted that the amount of glottal opening in both cases was very similar.

**Results from Swedish**: For Swedish, Lindqvist (1972) described a closed glottis and voicelessness for the voiced stops in utterance initial position. He also found a large glottal opening for the voiceless aspirated stops. Comparable results were shown in Löfqvist and Pétursson (1978).
Regarding influences on the amplitude of glottal opening, Löfqvist (1980) investigated laryngeal adjustment and different levels of stress. The glottal opening amplitude was speaker-dependent: for one subject the size of glottal opening covaried with stress degree and for the other subject the amplitude of glottal opening was related to oral closure duration (positive correlation). It should also be noted that for one subject during the production of the second /t/ in *tetteteten* and for the other subject during the third /t/, glottal opening was reduced to such an extent that no analysis could be made.

**Results from Japanese**: In Sawashima (1968), glottal abduction for voiceless /p s/ occurred with cessation of vocal fold vibrations whereas /h/ showed glottal opening too, but vibrations continued. During the production of /b/ and /z/ vocal folds were in the adducted position and vibrated during oral closure. Only a slightly open glottis was reported for initial /z/. Generally, these results are in agreement with Sawashima et al. (1975).
Sawashima and Miyazaki (1973) compared voiceless /k t s/ in word initial and word medial position. A reduction of glottal opening in medial position was reported for /k/, and for /t/ no glottal opening at all was observed in medial position. In Sawashima (1968) and Sawashima et al. (1978) a considerable reduction in glottal opening for /t/ in word medial position was found too. The voiceless fricatives showed a small reduction of glottal amplitude, but far less compared to the stops (Sawashima 1968, Sawashima & Miyazaki 1973, Sawashima and Niimi 1974).

**Results from Mandarin Chinese**: Iwata and Hirose (1976) compared word initial aspirated voiceless stops with their unaspirated counterparts. The authors found a large glottal aperture for the aspirated stop, and a ‘spindle shaped gap’ between the vocal folds for the unaspirated stops resulting in a negligible glottal opening amplitude. Both stops were produced without vocal fold vibrations during oral closure.

**Glottal opening amplitude and its duration**: The question arises, whether the relationship between glottal opening amplitude and glottal opening duration could be a linear function in single consonant production. Sawashima and Miyazaki (1973) investigated this issue. They found no linear relationship, but a rather complexly regulated mechanism. Hutters (1984) reported results which support Sawashima's and Miyazaki's findings. The amplitude of glottal opening for voiceless aspirated stops was reduced to a quarter in voiceless unaspirated stops whereas the duration of glottal opening was only reduced by half (Hutters 1984). It might be possible that these findings were influenced by the techniques, since the distance between the tip of the fiberscope and the glottis in transillumination and fiberoptic filming is not controlled for. However, similar results have often been found, even when not explicitly discussed (e.g. Kim 1970). Thus, they seem to be quite robust.

Another interesting experiment which referred to the amount of glottal opening was organised by Löfqvist, Baer and Yoshioka (1981). They investigated to what extent the amount of glottal aperture can be controlled by the speaker under static and dynamic speech and nonspeech conditions with and without visual feedback. A previous pilot experiment had shown that voluntary control of glottal opening in isolation was very difficult or impossible to produce. Therefore the authors changed the task towards a more manageable speech-like procedure, i.e. the subjects produced CV syllables. Visual feedback consisted of an oscilloscope put on a screen displaying four different equidistant levels of glottal opening. Neither of the two subjects could accurately produce the different levels. The authors reported a tendency towards an overshoot for the smaller target levels (i.e. the produced position was above the intended target) and an undershoot for the larger ones (the produced position was below the aim). It was also difficult to differentiate the amount of glottal opening for the subjects without visual feedback control. Löfqvist et al. suggested that the voluntary control of laryngeal opening is rather poor, i.e. not as accurate as lip or tongue movements, but similar to movements of the velum. They concluded that glottal opening movements in speech production are commonly related to supralaryngeal events, so that in addition to the degree of glottal opening the timing between laryngeal and supralaryngeal events is important for the production of the relevant phonological distinction.

So far it has been pointed out that:

1. A large amount of glottal opening is a necessary articulatory characteristic for the production of voiceless aspirated stops. This seems to be independent of the language, i.e. it applies to languages with a two-way (e.g. Hoole et al. 1983, Hutters 1984, 1985), three-way (Kim 1970, Kagaya 1971, Iwata et al. 1979) or four-way contrast (Dixit 1989).
2. For the voiceless unaspirated stops (phonologically either voiced or voiceless) either small laryngeal abduction or a closed glottis can be seen. In Danish, voiceless unaspirated stops were consistently produced with some glottal abduction. The same result was also found by Jessen (1998) regarding stops in word initial and medial position for German.
3. Tokens which were produced with vocal fold vibrations did not show glottal abduction, except /h/ and some voiced fricatives. For the voiced fricatives a high airflow rate could be responsible for a small amount of glottal opening.

It can be assumed that the amount of glottal opening plays an important role in the voicing contrast, but on the other hand, the timing of the laryngeal gestures with respect to supralaryngeal events is another important issue. The importance of laryngeal-oral co-ordination can be seen from the following points:

- A significantly larger glottal abduction for voiceless aspirated stops (or voiceless fricatives) has been found for geminates (Ridouane, 2003), but the larger glottal abduction did not affect aspiration. Hence, a large glottal opening on its own does not unambiguously explain long aspiration, it needs to be synchronised with supralaryngeal articulators.
- The amount of glottal opening seems to be most strongly affected by stress, and position, (also loudness, see Anders 1981) compared to the duration of glottal opening. Hence, glottal abduction changes with its environment and it is a variable rather than a consistent articulatory characteristic.

Consequently, aspects of laryngeal-oral co-ordination are considered in the following section.

## 2.3. Laryngeal-oral co-ordination

This section is divided in three different parts: first, results for laryngeal-oral co-ordination in aspirated and unaspirated stops, second, perturbation studies and third, laryngeal-oral co-ordination in fricative production.

### *2.3.1. Aspirated and unaspirated stops*

During the production of obstruents oral events are co-ordinated with laryngeal events. For an aspirated stop, a sufficient amount of glottal opening has to be produced approximately at the time of oral release. This co-ordination makes it possible to produce the relevant aspiration noise and to perceive it.

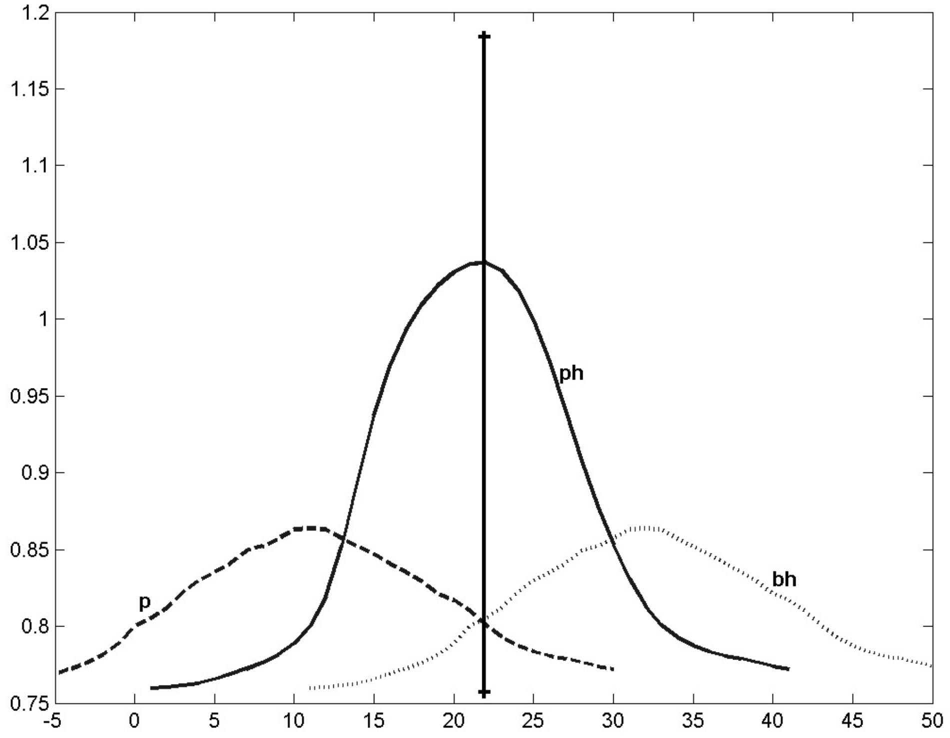

**Figure 2.6:** Schematic laryngeal-oral co-ordination for voiceless aspirated, unaspirated and voiced aspirated stops lined up with oral release = vertical line, solid line = glottal opening for /$p^h$/, dashed line = glottal opening for /p/ and dotted line = glottal opening for /$b^h$/, y = amount of glottal opening (arbitrary scale), x= time in samples (based on Hirose, Lisker and Abramson 1972, p.190)

Glottal opening onset and oral closure onset are also tightly coupled. However it is still unclear which time landmarks are controlled exactly, and how such control might work (see Löfqvist 1980, or Hoole et al. 2003 and section 2.3.2).

For unaspirated stops glottal abduction is nearly closed at oral release and hence no aspiration noise can either be produced or perceived. Both types of co-ordination are shown in Figure 2.6 together with an example for the voiced aspirated stops (as it occurs in Hindi).

These differences in laryngeal-oral co-ordination have been observed in various studies (e.g. Iwata et al. 1979, Fukui & Hirose 1983, Dixit 1989). Laryngeal-oral co-ordination can vary gradually between unaspirated and aspirated. (Perceptually it might be more of an either/or process, but from a production point of view it is not binary)[22]. For instance in Löfqvist's (1980) data on /t/ in different stressed and unstressed positions the general trend can be seen that the longer the aspiration duration, the smaller the duration from peak glottal opening to oral release. The shorter the aspiration duration (/t/ becomes unaspirated), the longer the duration from peak glottal opening to oral release (peak glottal opening shifts away from oral release as in Figure 2.6 to the left). This description provides evidence about a general trend and does explicitly consider that peak glottal opening is produced slightly after the burst. The latter shows some differences regarding voiceless stops at different places of articulation (see Jessen 1999).

Speaker dependent differences for laryngeal-oral co-ordination were found in Iwata and Hirose (1976) analysing two subjects of Mandarin Chinese. Subject A produced the unaspirated stop with an earlier timing of the small glottal amplitude with respect to oral release. A small glottal abduction was only seen in the membranous portion of the glottis, but in most cases the glottis was closed. Subject B showed comparable patterns in laryngeal-oral co-ordination for the aspirated and unaspirated type. The glottis was more abducted than for subject A and the differences between the aspirated and the unaspirated stop were related more to the size of glottal opening and to its duration than to timing.

Thus, the authors proposed two models for possible control strategies for the unaspirated type, one in terms of a reduction in glottal opening amplitude and the other with respect to laryngeal-oral co-ordination (see Figure 2.7 with a reduction in amplitude in the upper track and with a different timing in the lower track).

---

[22] Although for German and English it can be assumed that aspiration duration is a primary cue to distinguish between /b d g/ and /p t k/ when both consonant types are realised as voiceless, it could also be due to relational differences between a longer aspiration versus a shorter average aspiration.

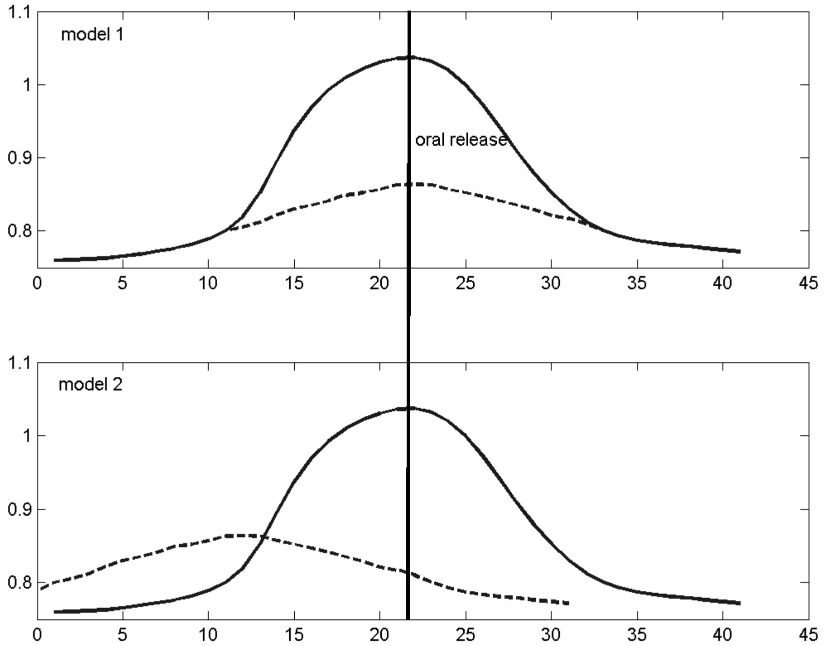

**Figure 2.7:** Laryngeal-oral co-ordination for aspirated /t/ (solid lines) and unaspirated /t/ (dashed lines) lined up with oral release (vertical line), y-axis = amount of glottal opening (arbitrary scale) and y-axis = time in samples, for the description of the 2 models see text (based on Iwata and Hirose 1976, p. 49)

For German, Jessen (1998) reported a similar laryngeal-oral co-ordination for the lax stops in /ibə/ which would correspond to Iwata and Hirose's unaspirated type from their subject B (model 1)[23]. However, he also pointed out that it was rather difficult to describe reliable values for the co-ordination between glottal opening onset and oral closure onset, because the onset of the very small glottal amplitudes were difficult to label (Jessen 1998).

### *2.3.2. Laryngeal-oral co-ordination and perturbations*

When considering interarticulatory co-ordination it is interesting to report perturbation studies. The aim of perturbation studies is to modify specific movements and examine possible consequences of such modification in order to search for the underlying control strategies used in the relevant tasks[24]. Munhall et al. (1994) pointed out that:

---

[23] Again, these findings are quite surprising since the lax stops were produced in word medial position. Since Jessen reported only one subject, it might be a speaker dependent characteristic.

[24] It is worth to note, that perturbed kinematic output could be either a consequence of compensatory behavior or modifications of the primary intended behavior. Both are difficult

> *"Understanding this coordination has proven to be a difficult task, since we have no direct window into the planning or control processes. One valuable experimental paradigm for examining such coordination is the introduction of unexpected perturbations to ongoing motor acts….The rationale for this research is that the nature and time course of responses to the load [i.e. to the perturbations S.F.] are thought to reveal the motor organization and reflex structure of the ongoing act"* (Munhall et al. 1994, p.3605).

Perturbation studies are not limited to speech production, they are quite common in motor control research in general, e.g. in studies of posture control. The assumptions underlying perturbation studies in speech production are:

1. different articulators are co-ordinated/coupled with each other and
2. the co-ordination/coupling of articulators depends on the relevant task, i.e. interarticulatory co-ordination is functional, task specific and the output carries lexical information. For example during the production of an alveolar voiceless stop, onset of glottal opening has to be temporally co-ordinated with onset of tongue tip closure and velum elevation.

Perturbations are divided into two types (Gracco 2001): first, static perturbations, i.e. modifications of size, shape or mechanical environment e.g. bite blocks (Fowler and Turvey 1980), extension of the upper incisors (Jones and Munhall 2003), modification of the palate (Hamlet and Stone 1978, McFarland et al. 1996, Baum and McFarland 1997), thus requiring an adaptation of learned motor programs, and second, dynamic perturbations (e.g. Munhall et al. 1994, Honda and Kaburagi 2000, Honda et al. 2002), i.e. unanticipated mechanical loads which are applied to an articulator during a very short period. These modifications require an immediate reaction of the motor system.

With regards to the role of laryngeal-oral co-ordination in the production of voiceless stops a few articulatory studies were carried out in the 1990s using dynamical perturbations. Munhall et al. (1994) investigated lip-larynx co-ordination during the production of voiceless bilabial stops. Their work was based on earlier findings from Folkins and Abbs (1975) and Shaiman and Abbs (1987) who found a delay of voicing as a response to lower lip perturbation. Regarding laryngeal movements, earlier work was recorded by means of

---

to separate from each other (R. Kent in his talk given at the Conference on speech motor control in normal and disordered speech, Nijmegen 2001).

electroglottography, which cannot provide reliable information about glottal abduction. Therefore, Munhall et al. investigated glottal abduction by means of transillumination simultaneously with pharyngeal air pressure by means of a pressure transducer inserted in the nose, and lip and jaw movements by means of LED's (light emitting diodes). Unexpected mechanical perturbations were applied to the lower lip via a paddle resting on the subjects lip. Three subjects repeated /i ˈpip/ about 400 times, where the stressed /p/ was taken into account. Lower lip perturbation was applied at three different points in time: just before the preceding vowel offset, early during oral closure, and just before bilabial release.

In general, acoustic results showed a shorter closure and a longer VOT duration for all subjects in the perturbed condition. Additionally, a longer vowel duration was found for two subjects. Intraoral pressure values did not differ significantly between perturbed and control trials. Results from two subjects showed some upper lip responses onto lower lip perturbation, i.e. the upper lip compensated immediately by means of larger lowering movements.

Glottal opening onset showed the expected delay in the perturbed condition. Such delay was discussed in terms of preserving laryngeal oral timing at oral closure onset:

> *"movement of individual articulators within a coordinative structure are adjusted in response to perturbations so that the goal of the coordinative structure is achieved"* (Munhall et al. 1994, p. 3615).

A similar adjustment was not found for the time of oral release. Peak glottal opening occurred later with respect to oral release in the perturbed conditions, which caused a longer VOT.

The authors suggested that:

> *"the failure of the larynx to initiate adduction movements earlier in response to the early oral release and the increased duration of the laryngeal adduction are less easily understood within this framework"* (Munhall et al. 1994, p. 3615).

One possible explanation for these results (i.e. the tight timing for onset of oral closure and glottal opening and the rather inconsistent timing between oral release and peak glottal opening after unexpected perturbations) could be: A single glottal opening seems to be a 'ballistic' movement, i.e. once it is initiated (here due to firing of neurons innervating the PCA and suppressing INT), it moves through different stages of opening and closing relatively automatically. Thus, glottal abduction would not change after its initiation took place and does not adapt to a shorter oral closure, i.e. glottal abduction onset would be only triggered by closure onset. Hence, the tight coupling of peak glottal opening and oral release would be a result of the initial interarticulatory co-ordination.

Perturbation studies are useful in order to explain control strategies of a co-ordinative task specific system which would be difficult to find in the kinematics without perturbing the system. However, most of the time specific expertise or technical equipment is required to do such experiments (e.g. an inflatable palate as in Honda et al. 2002), but as a result they provided some thought provoking results.

### *2.3.3. Laryngeal-oral co-ordination in fricatives*

Fricatives differ considerably from the production of voiceless stops. They are realised by a sufficient airflow through an oral constriction in order to produce the appropriate frication noise. Additionally, fricatives do not show complete oral closure with an abrupt rise in intraoral pressure as stops do.

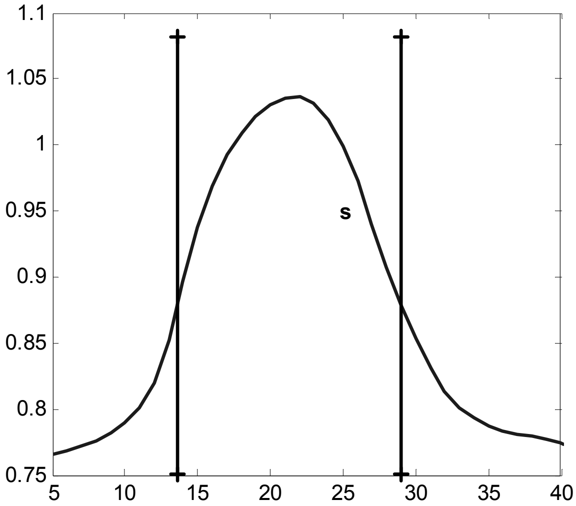

**Figure 2.8:** Schematic laryngeal-oral co-ordination for /s/, first vertical line = frication onset, second vertical line = frication offset; y-axis = amount of glottal abduction (arbitrary scale), x-axis = time in samples

The voiceless fricatives are produced with a different laryngeal-oral timing than stops (see Figure 2.8 for /s/). The beginning of glottal opening is realised before onset of oral constriction, defined as the onset of frication noise in the acoustic signal. Hereafter peak glottal opening is produced during the frication period,

and the glottis is closed after constriction offset. The described pattern were reported quite consistently in Butcher (1977), Jessen (1998), Hoole et al. (1983), Hutters (1984). Figure 2.8 exhibits this laryngeal-oral relationship.
For the voiced fricatives some amount of glottal abduction was most frequently found in the word medial position in Jessen’s study (1998). Since weak glottal opening was difficult to label, no precise laryngeal-oral co-ordination pattern were presented by Jessen. Figures 7.25a and b in his work (Jessen 1998, p. 244) show a relatively similar co-ordination comparing tense and lax fricatives.

It has been shown that glottal abduction is always co-ordinated with respect to supralaryngeal articulators. In the following paragraph, studies are considered which discuss several supralaryngeal articulatory mechanisms that may be involved in the voicing contrast.

## 2.4. Supralaryngeal correlates

> "I found myself wondering just why the larynx seems to attract so much attention than any of the other parts of the speech-producing apparatus. Not that we have no questions about the functioning of the velum, tongue, jaw, lips and respiratory musculature, but the larynx is especially provocative of question and debate"
> Lisker (1977, p.304).

### *2.4.1. Strategies for cavity enlargement*

The discussion about supralaryngeal correlates that may be involved in the voicing contrast is intimately connected with the process called 'cavity enlargement' or 'cavity expansion'. It is based on the following principle: Phonation requires a sufficient tension of the vocal folds and a sufficient transglottal pressure drop (>2000 dyn/cm$^2$, for review see Westbury 1983), i.e. the pressure below the glottis (subglottal pressure) has to be higher than the intraoral pressure. In the production of an obstruent, transglottal pressure differences are affected by the production of the appropriate oral closure or constriction which causes a rise in intraoral pressure. Therefore, transglottal pressure differences become equalised and vocal folds are likely to stop oscillating (for a study on devoicing in German see Pape et al. 2003). This is true when subglottal pressure and vocal fold tension are relatively stable for the voicing contrast. The assumption for subglottal pressure was confirmed e.g. by Netsell (1969), and Löfqvist (1975). Ohala and Riordan (1980) reported a duration of 5 to 10 ms for conservative estimates in which voiced stops would become devoiced after closure onset.

For the production of voiced stops it seems to be quite common that vocal folds still oscillate up to 100 ms after oral closure onset (see e.g. Lisker 1977, Westbury 1983). Such a phenomenon is only possible when either oral closure is incomplete, the velar port is still open or transglottal pressure differences are maintained due to strategies expanding the oral cavity and therefore prohibiting the increase of intraoral air pressure. The latter are known as strategies for cavity enlargement.

> *"The cumulative effect of articulatory movements on volume of the cavity above the glottis is more relevant to the problem of voicing maintenance during consonantal closure than are the direction and extent of movements of any single articulator"* (Westbury 1983, p.1331).

In Figure 2.9 a schematic midsagittal view through a vocal tract is depicted during the production of an alveolar stop. The arrows in the oral cavity hold for possible directions of cavity expansion. Theoretically, laryngeal lowering (downward arrow), velar elevation (upward arrow), advanced tongue root movement (arrow to the left) or movement of the posterior part of the pharyngeal walls (arrow to the right) could be possible. Tissue compliance, e.g. of the cheeks during the production of bilabials would be another possibility (not included in Figure 2.9, depicting only midsagittal aspects).

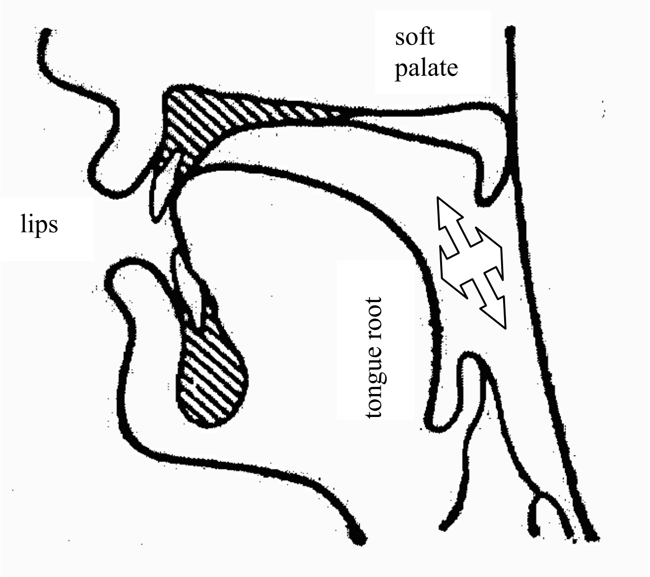

**Figure 2.9:** Midsagittal view through the vocal tract; left = lips, jaw, teeth; right = beginning of the spine); hatched area = bony structure (based on Fiukowski 1992, p. 161)

From an experimental phonetics point of view Westbury's statement is a warrantable issue, but from a technical point of view it is recently not feasible to control for the whole complex and additionally guarantee the subject's well-being. Magnetic Resonance Imaging (MRI) might be an appropriate technique to analyse cavity enlargement, but currently new MRI techniques are still under construction which might produce images from very fast movements with a good time resolution.

Methodologically it still seems reasonable to search for possible cavity enlargement strategies included in the whole complex. Such strategies were supposed and observed, e.g. laryngeal lowering together with hyoid bone depression (Kent and Moll 1969), velar elevation (Bell-Berti 1975), vocal tract wall compliance (Perkell 1969, Ohala and Riordan 1979), tongue compliance (Svirsky et al. 1997). Jaw lowering could be another strategy, but to my knowledge empirical evidence is missing. In the following section several strategies are discussed which have been related to cavity enlargement in the literature.

**Laryngeal height**: The larynx is the lower boundary of the vocal tract and it is relatively flexible in its ability to move upwards (towards the jaw) or downwards (towards the sternum). Laryngeal downward movement increases the volume of the vocal tract all other things being equal. It was discussed as one possible strategy for cavity enlargement in order to maintain a sufficient transglottal pressure difference and thus voicing during oral closure (Kent and Moll 1969, Bell-Berti 1975, Westbury 1983). Laryngeal height for voiced and voiceless stops was investigated by a few researchers using various techniques, e.g.:

By means of an thyroumbrometer, a photoelectric device with 16 overlapping rectangular photocells, Ewan and Krones (1974) recorded the vertical laryngeal movements (of the incisura thyroidea superior) during the production of VCV-sequences. The intervocalic consonants consisted of voiced or voiceless stops from different languages. Six English, one French, one Thai and one Hindi speaker served as subjects. They were recorded in supine position in order to guarantee a relatively stable and comfortable position for the subjects. The accuracy of the technique was quite low with respect to position (approximately 2mm), but on the other hand a good time resolution could be reached (2ms intervals). Results from Ewan and Krones provide evidence that voiceless stops have a higher larynx position than their corresponding voiced stops, in particular at the end of oral closure. These findings support the idea of an involvement of laryngeal lowering during the production of the voicing contrast.

Gandour and Maddieson (1976) also found a low larynx position for voiced stops compared to voiceless stops, but only at vowel onset following oral release. They used a mechanical device, a cricothyrometer to record one speaker of Standard Thai. However, their primary goal was to test effects of pitch, consonant, phonation type, vowel quality, tonal categories and position in the utterance onto laryngeal height rather than to study cavity enlargement strategies. It turned out that position in the utterance affected laryngeal height most clearly (with a low laryngeal height utterance finally). Other effects were more weak or not significant.

Petersen (1983) did not only study laryngeal height in aspirated versus unaspirated obstruents (the phonological contrast is based on aspiration in Danish), but also in nasals. Petersen's primary aim was to relate laryngeal height with fundamental frequency, especially to explain the lower F0 after unaspirated compared to higher F0 after aspirated stops. By means of video images and acoustic data he analysed three Danish subjects. The position of the larynx was measured with an accuracy of 0.5 mm and with 50 frames per second. Petersen observed a trend for a higher F0 and an elevated laryngeal position for aspirated obstruents, but only 17 percent of the F0 variation could be explained by variations of laryngeal height. Hence he concluded that it is questionable to explain F0 differences of varying consonant types by the vertical position of the larynx. Additionally, the nasals showed the lowest laryngeal position, they were even lower than the unaspirated obstruents. Petersen assumed the lower larynx position for nasals could be hardly the purpose to preserve a sufficient pressure drop to guarantee voicing for the nasal consonants.

The previously discussed investigations differ considerably in their methodology, and with respect to the languages taken into account. An effect which could explain the different findings is: the larynx moved together with the jaw via the hyoid bone. To my knowledge laryngeal position was taken as an absolute measure with no fixed references. It might enhance the discussion if laryngeal position would be recorded together with jaw movement and a reference point (e.g. the sternum or the middle of the ear).

A fixed reference point was taken in Westbury's investigation (1983). By means of cinefluorographic films he measured among other things laryngeal position for one subject as the distance between the end of the upper incisors and the anterior edge of the laryngeal ventricle. Even though Westbury reported laryngeal downward movement for utterance initial /b d g/ and medial /b d/ as well as small movements during closure of medial /g/, he also noticed laryngeal upward movement for utterance-final /g/. He concluded:

> *"Thus, voice-sustaining movements of the larynx were observed during some instances of voiced stops, but such movements were neither characteristic of nor unique to the closures of those segments in all phonetic environments"* (Westbury 1983, p.1327).

In other words, laryngeal height is not necessarily a mechanism involved in the voicing contrast.

Serious doubt on the hypothesis of the involvement of laryngeal height in cavity enlargement strategies was raised in Riordan (1980). Using the thyroumbrometer in combination with a nasal catheter (for intra-oral pressure measurements) he recorded two subjects producing VCV-sequences. The medial consonant varied between /p b m/. Even though Riordan found a small difference in laryngeal height between /b/ and /p/[25], which

> *"only minimally increases the capacity of the vocal tract to absorb glottal air flow, contributing little to an explanation of voiced closure duration's commonly observed in speech. Finally, the data for the voiced stop reveal no distinctively characteristic relationship between larynx movement and intra-oral air pressure"* (Riordan 1980, p.359).

The weak correlation between laryngeal height and intra-oral air pressure should provide evidence that laryngeal lowering can not count on its own for cavity enlargement, and other mechanisms might be involved too. Riordan's results for nasals showed a relatively low laryngeal position too, which were comparable to the voiced stops.

**Velar height and the pharyngeal cavity**: Perkell (1969), Kent and Moll (1969), Bell-Berti and Hirose (1972) found a higher velum for voiced stops in comparison to a lower velum for voiceless stops. Slis (1967, 1970) even went so far as to propose a theoretical articulatory model for the voice-voiceless distinction only based on the activation of pharyngeal muscles. He assumed that a contraction of the pharyngeal walls would pull the larynx upward and decrease the volume of the cavity for voiceless stops. The opposite effect would occur in voiced stops, i.e. the larynx would be pulled down which increases the cavity together with its volume.

Since Perkell (1969) concluded that the enlarged cavity for voiced stops would be due to a reduced tension of the pharyngeal wall musculature, and Kent and Moll (1969) explained the same effect with an activity of muscles causing hyoid bone depression, the question of active versus passive cavity enlargement strategies was raised. In a series of studies Bell-Berti and Hirose (1971, 1972) and Bell-Berti (1975) investigated this issue. Three American English speakers were recorded by means of EMG (Bell-Berti 1975). Greater activities of the levator palatini and sternohyoid muscles were associated with active expansion, and inhibition of the constrictor muscles with a passive expansion mode. Results differed within and between speakers, i.e. both strategies were used.

---

[25] It seems surprising to me that the main results of laryngeal lowering (between 1mm and 1.5 mm) can be discussed when the thyroumbrometer had an accuracy of „1mm or more". Results were most of the time close to the accuracy limit.

Thus

> *"an adequate description of pharyngeal cavity expansion for voiced articulation is neither exclusively active nor an exclusively passive one. Each speaker uses both modes of enlargement, while apparently favoring one mode over the other……It is clear, then, that the feature [±tense] is inadequate for describing the pharyngeal volume change concomitant with voicing distinctions, as that feature at best explains the larger portion of some speakers pharyngeal adjustments and never explains the full measure of enlargement"* (Bell-Berti 1975, p.460).

On the other hand, Westbury 1983 reported an EMG-study by Minifie et al. (1974), who recorded five American speakers and found less pharyngeal muscle activity for voiced stops compared to voiceless stops.

However, no consistent differences were found with respect to velar height in Westbury (1983) for one speaker of American English and in Ushijima and Sawashima (1972) for two speakers of Japanese. Results from Westbury were based on high speed cinefluorographic films and Ushijima's and Sawashima's results on fiberoptic observations. The latter also included voiced and voiceless fricatives in their study.

Investigations in velar height showed predominantly a higher velum position for voiced stops and the reverse for the voiceless. This process was explained by active as well as passive mechanisms. Similar to laryngeal height, findings differed with respect to the subjects, languages, and techniques used. None of the described experiments recorded velar height simultaneously with intra-oral pressure. The investigators assumed the relevant conditions, i.e. higher intraoral pressure for voiceless obstruents and lower intraoral pressure for the voiced obstruents. However, interspeaker variations could be due to variations in intraoral pressure differences as well.

**Tissue compliance**: Ohala and Riordan (1980) tried to answer the question of how long voicing could persist if cavity enlargement is a passive process due to vocal tract compliance. In order to do so, they applied the following experiment: One subject produced an abnormally long voiced stop embedded in different vowel contexts. By means of a nasal catheter intraoral air pressure was vented to the atmosphere. The catheter, inserted in the upper pharynx, could be closed externally at unpredictable times, during the long oral closure production. The analysis concerned the duration how long voicing could be maintained after closing the catheter tube at unpredictable times. In general, results provided evidence that the duration of voicing was longer the more forward the place of articulation of the stop and the higher the vowel the stop was coarticulated with (with one exception). The authors tentatively concluded that without active cavity enlargement but with tissue compliance voicing was maintained for approximately 70 ms. There were no data available on how different parts of the vocal tract differ with respect to compliance. Cheeks should have a relatively high compliance, whereas teeth or the hard palate should have a very low one.

Tongue displacement in relation to intraoral pressure estimates were observed by Svirsky et al. (1997). Both measurements were used to assess the validity of a tongue compliance model. The mode of tongue displacement for the consonants in /aba/, /apa/, and /ama/ in relation to intraoral pressure changes could explain a rather active or passive expansion strategy. Since oral closure is produced by the lips, the tongue would be more free to keep its position or to move independently, e.g. in order to prevent an intraoral pressure rise.

The magnitudes of peak tongue dorsum displacement estimated by an EMMA-system were significantly larger during the production of voiced bilabials compared to smaller magnitudes in voiceless bilabials. It seemed surprising that such tongue dorsum differences occurred during a bilabial when surrounded by the same vowel context. The displacement was close to zero during the nasal. Svirsky et al. reported:

> *"It is interesting to observe that the relatively sharp, fast downward tongue dorsum displacement during /apa/ or /aba/ were generally close to the rise in intraoral pressure"* (Svirsky et al. 1997, p.565).

Using a lumped parameter circuit model Svirsky et al. estimated tongue compliance and found much higher values for the voiced stops than for the voiceless. They concluded the tongue would be actively stiffened for voiceless stops. However, relaxation of the tongue for voiced stops did not explain all the results. Hence Svirsky et al. proposed a combination of intentional relaxation of tongue muscles with an active displacement for the voiced.

Lindblom et al. (2002) investigated another phenomenon, the so called "trough effect" using similar speech material as Svirsky et al. (1997), i.e. VCV-sequences with C being /b, p/ and V= /i/. The trough effect was associated with a momentary deactivation of tongue movement during the bilabial stop, which is thought to be unspecified for vowel production. Lindblom et al. conclude that the trough would provide evidence for a segment-by-segment activation of the neural nervous system and could not be attributed to aerodynamics. Lindblom et al.'s suggestion is surprising since they didn't include nasals to control for aerodynamic differences and additionally, they mention a language-specific effect (For more on troughs in German, see Fuchs et al. 2004a).

Similar tongue patterns were investigated under different perspectives. They were identified as either tongue compliance due to cavity enlargement strategies or the trough effect as evidence for segment-by-segment activation. Up to now it is unclear whether there are two underlying strategies or whether the same phenomenon was explained by different research directions.

**Tongue root movements**: Westbury (1983) stated that changes in the pharyngeal width could be particularly attributed to anterior-posterior movements of the tongue root. Tongue root movement was found to be more advanced/forward for voiced compared to voiceless which increased the back oral cavity. Again, this characteristic did not uniquely occur with respect to voiced stops, but it was generally attributed to cavity enlargement strategies.

### *2.4.2. Tongue and jaw movements*

It is still a matter of debate whether tongue and jaw movements are involved in the voicing contrast. Most consistent finding for an involvement was presented by Fujimura and Miller (1979) for the jaw. Generally, jaw movements have been examined with regards to vowel height, accent, as well as sonority hierarchy etc. but concerning the voicing contrast it was of less interest. However, by means of an x-ray microbeam system Fujimura and Miller (1979) recorded 3 American speakers producing /d/ and /t/ in syllable and word final position. Their results were most consistent for the mandible and gave evidence that /d/ was produced with a lower jaw position and a lower velocity compared to /t/. For /t/ the jaw moved more vigorously. These results could explain the production of a salient burst in /t/ due to a high jaw position (Mooshammer et al. 2003).

Kent and Moll (1969) compared homorganic supralaryngeal articulations /p b m/, /t d n/ and /k g/ and reported similar closure and release gestures, but they also noted a tendency for voiceless stops to have slower articulatory movements than their voiced counterparts. Westbury (1983) presented similar results and did not find evidence for tongue tip or tongue dorsum differences between the voiced and voiceless cognates.

Regarding velar differences in the elliptical trajectories, so called “looping patterns” have been observed with respect to the voicing contrast (Mooshammer et al. 1995, Löfqvist and Gracco 1994). In Mooshammer et al. (1995) it was rather the movement amplitude which was larger for voiceless and in Löfqvist and Gracco (1994) tongue body raising was longer for voiced since the tongue also started at a lower position. Both results can not be related to cavity enlargement. It might be subject dependent as well as dependent on the laryngeal-oral co-ordination whether tongue and jaw are involved in the contrast. Jaw positioning could play an important role, but less is known about it and its potential influence.

### *2.4.3. Tongue palate contacts*

By means of EPG, differences of tongue palate contact can be considered, in particular for alveolar obstruents. Regarding supralaryngeal mechanisms and the voicing distinction differences in palatal contacts could be discussed under several perspectives:

1. A greater amount of anterior contact could hold for a relatively high tongue pressure against the palate for the voiceless stops.
2. Less posterior contact could provide evidence for a strategy involving cavity expansion for the voiced stops.
3. Incomplete alveolar closures during alveolar stop production could be another strategy to maintain transglottal pressure differences and thus, voicing during oral closure.

Many EPG studies investigated place of articulation, effects of coarticulation or the influence of prosodic constituents on tongue palate contact, but less is known about differences with respect to the voicing contrast.

Substantial work has been carried out by Dagenais et al. (1994), who recorded a variety of single voiced and voiceless obstruents and consonant clusters in CV-sequences from 10 American speaking adults. Regarding single alveolar stops the midline lengths of contact (= length of alveolar closure within the relevant EPG pattern) were consistently greater for voiced stops compared to voiceless averaged over all speakers. The authors explain these findings with intraoral pressure differences. For the voiced stops the tongue would be more relaxed or spread out, because low intraoral pressure was assumed whereas the voiceless stops would have a greater tension in order to resist intraoral air pressure. Dagenais et al. also found a consistent, but weak significant difference in groove width for alveolar fricatives. The voiced sibilants were produced with a narrower groove width compared to the voiceless.

Moen and Simonsen (1997) and Moen et al. (2001) analysed tongue palate contact patterns for /d/ versus /t/ in Norwegian (1997, 2001) and English (1997). For both languages they reported a tendency for a greater amount of contact for /t/ than for /d/. It was unclear whether these results were statistically significant or not. They explained their results due to differences in intraoral pressure:

> *"In order to prevent air from escaping between the tongue and the palate, which would cause a fricative instead of a plosive articulation, a firmer contact is needed for voiceless than for voiced stops"* (Moen and Simonsen 1997, p.2401).

In 2001 Moen et al. additionally observed intraoral air pressure. They found the typical patterns, i.e. higher intraoral pressure for voiceless and lower pressure for voiced stops.

As can be seen results from tongue palate contact patterns differ considerably, yet surprisingly, they have been given similar interpretations with respect to intraoral pressure differences.

Some other EPG studies can be included here, but most of them put their main focus on something different:

- Fletcher (1989), who recorded American English speaking children and found no significant differences between voiced and voiceless alveolar stops
- Tabain (2002), who investigated co-articulatory effects on various consonants recording Australian English speaking females. She found almost no differences between voiceless and voiced cognates.
- Dixit (1990), who observed voiced and voiceless dental stops and retroflexes in Hindi and found generally that voiceless stops showed a significantly greater overall contact compared to voiced.
- Shockey (1991) and Shockey & Gibbon (1993) reported a high amount of incomplete closure during the production of /t/ in conversational speech, but /d/ was not observed.

New results considering tongue palate contact and voicing distinction were brought in with the development of techniques for measuring tongue palatal contact pressure. The devices are still prototypes and not commercially available (except for Honda and Wakumoto's wireless pressure sensor system to control an electric wheelchair with tongue tip movements). However, some preliminary results should be described since they are valuable regarding the voicing contrast.

Matsumura et al. (1994) developed an artificial palate including 5 strain gauge transducers along the midsagittal line. Results from the most anterior sensor give evidence that /d/ and /t/ showed similar pressure values between 5 to 6 kPa (see figure 6 in Matsumura et al.'s paper), whereas they were lower during the production of /n/. The differences between /d/ and /t/ seem not to consist in maximal tongue pressure but rather in the timing of peak tongue pressure with respect to the burst. In /t/ production maximal tongue palate pressure preceded the oral release (acoustic bursts) more than 100 ms. In /d/ the two landmarks occurred closer to each other. Additionally the palatolingual contacts were larger (more contacts) during /d/. Results could be influenced by the position of the consonants. They were produced in absolute syllable initial position (/ta, /da/ and /na/).

A more robust pressure sensitive palate was developed by Wakumoto et al. (1998). Results for 10 Japanese speakers saying /ada/ and /ata/ provide evidence for a higher tongue pressure during voiceless stop production compared to voiced. Similar values have been found in Tiede et al. (2003) for one speaker of American English using a capacitive device for the pseudo palate. The voiced alveolar stop was produced with significantly smaller pressure and the same was true for the voiced alveolar fricative compared to its voiceless counterpart.

**Summary**

1. Generally, laryngeal correlates are involved in the voicing contrast. In particular glottal abduction is usually produced in phonologically voiceless obstruents and glottal adduction in phonologically voiced obstruents. For the first, the glottal opening amplitude is often reduced according to context, e.g. in post-stressed word medial position. For the second, vocal fold vibrations do often disappear after oral closure onset, since intraoral pressure increases and the transglottal pressure diminishes below the level necessary to sustain vocal fold vibrations. Consequently, differences concerning laryngeal correlates between phonologically voiced and voiceless cognates become smaller and may even disappear.
2. Laryngeal abduction is co-ordinated with oral closure/constriction, assuming glottal abduction is found. The more stable laryngeal-oral co-ordination can be seen for aspirated stops and voiceless fricatives in comparison with unaspirated stops or voiced fricatives. The latter coincide with a weak amount of glottal abduction. For aspirated stops the onset of oral closure is well synchronised with the onset of glottal abduction. In addition, peak glottal opening is tightly coupled with oral release. Results from perturbation studies provide evidence that onset of laryngeal abduction can be immediately adapted when a supralaryngeal articulator involved in the specific stop production is perturbed. I.e. laryngeal movements are well synchronised with supralaryngeal articulators at the beginning of oral closure (in particular for aspirated stops). Unaspirated stops do not show a similar stability of laryngeal-oral co-ordination. Two different control strategies have been proposed. One strategy involves a reduction in glottal opening amplitude and the other strategy involves a different laryngeal-oral co-ordination. For the latter

   the small peak glottal opening occurs well before the point of oral release. In both cases the glottis is mostly adducted at the moment of oral release. Fricatives show a different laryngeal-oral co-ordination compared to stops. The onset of glottal abduction precedes the onset of oral constriction. Peak glottal opening is produced approximately at the middle of oral constriction, but there is a tendency for peak glottal opening to be more closely related to the onset of glottal abduction. The offset of glottal abduction delays offset of oral constriction. If voiced fricatives show some amount of glottal opening, then there seems to be a similar laryngeal-oral co-ordination. However, there is a lack of studies investigating this issue.

3. Supralaryngeal correlates of the voicing contrast have been attributed to strategies for cavity enlargement. Several strategies have been proposed, e.g. laryngeal lowering, velar raising, tongue root advancing, tissue compliance. Speakers can use these strategies differently, i.e. they can prefer one particular strategy over another. The involvement of tongue and jaw movement with respect to the voicing contrast is still a matter of debate. There seems to be a trend that the jaw shows more consistent involvement than the tongue. A hypothesis has been proposed that a high jaw position is necessary to produce a salient burst for voiceless /t/. Some evidence that voiceless alveolar stops are produced with a higher tongue pressure against the palate and with more percent of contact in the anterior region has also been presented. A lower tongue palate pressure and a smaller percentage of contact in the anterior region was found for the voiced cognates. It is supposed that the supralaryngeal articulators do participate in the voicing contrast, i.e. they regulate the intraoral pressure with respect to the subglottal pressure in order to produce the relevant output.

# Chapter 3: Methods

## 3.1. Introduction

The aim of this work is to observe laryngeal and supralaryngeal correlates which could be involved in the voicing contrast in German alveolar obstruent production. Two separate experiments were carried out with the same three subjects.
**Experiment 1** is dedicated to laryngeal adjustment as well as laryngeal-oral co-ordination. It consisted of simultaneous recordings of glottal abduction by means of combined transillumination (hereafter PGG) and fiberoptic filming (FF) as well as tongue-palate contact patterns by means of Electropalatography (EPG, Reading system, EPG3).
**Experiment 2** is addressed to supralaryngeal production mechanisms, i.e. several temporal and spatial parameters of tongue and jaw movements as well as tongue-palate contact patterns by means of simultaneously recorded Electromagnetic Articulography (EMA, Carstens Medizinelektronik AG100) and EPG (Reading system, EPG3) data.
Additionally, acoustic data were collected in the two experiments. Both experiments were conducted in the phonetics laboratory at the Centre for General Linguistics (ZAS) in Berlin. Special expertise for the first experiment was provided by Dr. Phil Hoole from the Institute of Phonetics and Speech Communication at the Ludwig-Maximilians-Universität in Munich and the Otorhinolaryngologist Dr. Klaus Dahlmeier from Berlin.

## 3.2. Subjects

A limitation of most articulatory studies is that it is commonly not feasible to record and analyse many subjects in a reasonable time. To my knowledge data from about 3-5 subjects are often analysed in articulatory EMA studies. The number of subjects recorded by means of transillumination was already summarised in chapter 2 (see Table 2.3). Hence, the three subjects recorded here are approximately average for such a procedure. Any conclusions derived from such an amount of subjects can not generally be transferred to a whole speech community. However, results from this study can be enhanced by comparing it with other investigations or other languages with a comparable voicing contrast (e.g. Danish).

It is intended to record the same three subjects for both experiments in order to describe each individual's laryngeal production mechanisms in relation to her/his individual supralaryngeal characteristics.
Subjects were chosen in accordance with the following selection criteria:

- they are native speaker of German,
- they were willing to volunteer,
- they have no known history of speech or hearing disorders,
- they are not especially sensitive in reacting with a gag reflex,
- during a initial endoscopic inspections Dr. Dahlmeier did not find any particular anatomical constrictions or special characteristics which could affect the well-being of the subject and/or the reliability of the recordings
- these subjects already had a custom-made artificial EPG-palate.

Subjects are:
**JD** (male) = 38 years old at the time of the recording, grown up in the north of the former GDR (Rostock),
**CG** (male) = 31, born and grown up in the south of the former FRG (Lake Constance),
**SF** (female) = 30, born in the north of the former GDR (Greifswald), grown up in the south (former Karl-Marx-Stadt).

All three subjects lived for at least 10 years in Berlin and are colleagues in the phonetic lab. JD served several times as a reference speaker for acoustic, EMA, EPG and MRI recordings. SF (the author) has received training in Standard German due to her courses in speech sciences at the university. CG has some slight southern German dialectal accent which is usually not perceivable under conscious control in the experimental conditions. SF and CG are familiar with EPG and EMA recordings too. All subjects are native Germans with different levels of proficiency in second languages: JD in Russian and English, CG in English and French, SF in Russian and English.

### 3.3. Linguistic material

When searching for relevant linguistic material appropriate to the two different articulatory investigations, a number of factors have to be taken into account. The German obstruent inventory contains the following stops and fricatives:

**Table 3.1:** German obstruent inventory divided by place of articulation in columns and voicing contrast in rows (Glück 1993, p.326 based on Kohler 1990); upper table: stops; lower table: fricatives; consonants within the grey background are used in this study

| stops | labial | alveolar | velar | glottal |
|---|---|---|---|---|
| voiced | b | d | g | |
| voiceless | p | t | k | (ʔ) |

| fricatives | labio-dental | alveolar | post-alveolar | palatal | velar | uvular | glottal |
|---|---|---|---|---|---|---|---|
| voiced | v | z | ʒ | | | ʁ | |
| voiceless | f | s | ʃ | (ç) | (x) | χ | h |

The German stop system consists of bilabials, alveolars, velars and the glottal stop and the fricative system of labiodentals, alveolars, postalveolars, palatals, velars, in most dialects of the uvular /ʁ/ (in Bavarian it would be an apical trill), /χ/ which is an allophonic variation of /x/, and the glottal /h/ (see Table 3.1).
In order to study voicing contrast correlates, all phonemes with a missing counterpart are excluded, i.e. /ʔ, ç, x, h/. Additionally, the voiced /ʒ/ was excluded since it exists only in loan words, e.g. [gɑˈʀɑːʒə] (garage) a loan word from French[26]. The uvular fricatives /ʁ/ and /χ/ are also excluded since they are allophones. The voiceless uvular fricative is an allophonic variation of /x/ and it occurs after low back vowels (see also chapter 1).

Since in this study tongue palate contact pattern (EPG) are simultaneously recorded with transillumination in experiment 1 and with EMA in experiment 2, the following phonemes are excluded too /b p g k v f/. EPG would not provide reliable tongue-palate contact patterns for these phonemes. The bilabials and labiodentals are excluded, because there is obviously no contact between the tongue and the palate. The velars are not taken into account because some posterior contacts might not be registered and they can be problematic for transillumination too. Hence, this study will focus on the alveolar stops /d t/ and the alveolar fricatives /s z/ (see grey background Table 3.1). They are also examined frequently in other languages.

---

[26] On the other hand, word initially [s] is not allowed except in loan words like sex [sɛks] in contrast to sechs [zɛks] (six).

### *3.3.1. Vowel environment*

The choice was rather difficult concerning the vowel context in which these consonants should be placed in. Since the transillumination signal can be influenced by tongue retraction which causes epiglottal movement and hence a shadow onto the glottis, many studies consider obstruents only in the high front vowel /i/-context.

However, the /i/ context does not suit EMA. Since movements from a high front vowel towards an alveolar stop are quite small they are difficult to label consistently. Previous results in our lab showed that small fluctuations or noise have a bigger influence in small movements or even two gestures merge into one. Labelling becomes nearly impossible when jaw kinematics in small movements are taken into account. Thus, it seems reasonable to record linguistic material for large vertical movements, e.g. /t/ in /a/-context.

It is most common in articulatory studies to use the tense corner vowels /a/, /i/ and /u/ as the appropriate environment, because they are extreme values and for this reason should give the clearest insights into the ways in which obstruents can be influenced by surrounding vowels. Therefore the tense vowels /a/, /i/ and /u/ are used here, but caution is taken in terms of discussing results from transillumination recordings in /a/-context as well as EMA data in /i/-context. In fricative production also lax vowel context /a, ɪ, ʊ/ is taken into account (see 3.3.3.).

### *3.3.2. Nonsense words*

Using nonsense words in speech production experiments is motivated by the desire to use “speech-like” sounds with a precise structure. Methodologically it is helpful in two ways: in terms of searching for variance versus invariance to understand basic principles of articulation; and in controlling coarticulation and avoiding reduction effects which are common in real speech. Some examples which can be simply addressed using nonsense words, are:

1. to keep the context constant and vary the voicing contrast, e.g. /ata/ vs. /ada/,
2. to vary the vowel context and keep the rest constant, e.g. /ate/ vs. /ite/ vs. /ute/,
3. to keep the material constant, but vary the symmetry, e.g. /ati/ vs. /ita/,
4. to keep the material constant and vary suprasegmentalia, e.g. to vary syllable position or stress.

The voicing contrast in German varies with the position in an utterance and possibly with stress and vowel context. To control for this the nonsense words *geCVCe* and *geCVC* were created, for example *getiete* [gəˈtʰiːtə]. The structure of these nonsense words is in agreement with German phonotactical rules (Eisenberg 1991). Bi- and trisyllabic words are most frequent in German with the word accent on the first syllable, except when preceded by a prefix like *ge* as in this study. Target words are constructed with a prefix to guarantee a gesture towards the first consonant (C), otherwise the beginning of the consonant, in particular the beginning of the oral closure might be influenced by the preceding word boundary, pauses or lengthening effects. The first syllable after the prefix *ge* has always the main word stress, because the prefix *ge* and final /ə/ are always unstressed, and additionally they are common in the German word structure.

The nonsense words are embedded in the carrier sentence: "*Ich habe geCVCe nicht geCVC erwähnt.*" (I said geCVCe not geCVC). Using carrier sentences is a common strategy to provide similar lexical surrounding and suprasegmental conditions for the target words. Note, that for CG the carrier sentence differs slightly: "*Ich habe geCVCe wie geCVC erwähnt.*" (I said geCVCe like geCVC), because the /n/ in *nicht* influenced the transillumination signal for this subject. Such influences are often subject dependent, and since CG was recorded as the last subject, the carrier phrase could not be changed for previously recorded speakers.

### *3.3.3. Position of the consonants*

From previous descriptions of the target words it can already be seen that the alveolars are placed into various syllable or word positions:

1. The first consonant (hereafter C1) in /geˈC1VCe/ marks the intervocalic position without an intervening word boundary. C1 occurs at the beginning of the stressed syllable. Its position is syllable and morpheme initial too.
2. The second consonant (hereafter C2) in /geˈCVC2e/ marks the intervocalic position without an intervening word boundary. It occurs after the stressed syllable, i.e. in post-stressed position, which is also the onset of the third syllable when the preceding vowel is tense. It is ambisyllabic when the preceding vowel is lax. Ambisyllabic means that it is still unclear whether the consonant belongs to the previous or the following syllable and it is assumed that it belongs to both.
3. The last consonant (hereafter C3) in /geˈCVC3/ marked the intervocalic position with an intervening word boundary. It occurs after the stressed

syllable, i.e. in post-stressed position, which is also syllable and word final.

All /d t/ as well as /s z/ are single consonants and surrounded by vowels. Table 3.2 shows the spelling of the target words as they were presented to the subjects. The spelling of tense vowels in German is associated with either a doubling of the vowel as in *gedaade*, with the vowel and an *h* as in *geduhde* or for /i/ with a following *e* or *eh* as in *gediede* or *gesiehse*. In a test session subjects were asked to choose the spelling which would fit most with their imagination of the relevant tense vowel.

The voiceless alveolar fricatives are placed in lax vowel context. Lax vowels are always produced when the following consonant is represented by the same two graphemes as in *ss*.

Combinations as *getaade* or *gedaate* with varying /d/ or /t/ within a target word are not included for this study.

**Table 3.2**: Spelling of nonsense words used in the experiment; C1, C2 and C3 characterise the different positions of the observed consonants, relevant phonemes are written in bold

| | **Spelling of relevant target words** | | |
|---|---|---|---|
| | **C1** | **C2** | **C3** |
| **/t/** | | | |
| /a/ – context | ge**t**aate | getaa**t**e | getaa**t** |
| /i/ – context | ge**t**iehte | getieh**t**e | getieh**t** |
| /u/ – context | ge**t**uhte | getuh**t**e | getuh**t** |
| **/d/** | | | |
| /a/ – context | ge**d**aade | gedaa**d**e | gedaa**d** |
| /i/ – context | ge**d**iede | gedie**d**e | gedie**d** |
| /u/ – context | ge**d**uhde | geduh**d**e | geduh**d** |
| **/z/** | | | |
| /a/ – context | ge**s**ahse | gesah**s**e | gesah**s** |
| /i/ – context | ge**s**iehse | gesieh**s**e | gesieh**s** |
| /u/ – context | ge**s**uhse | gesuh**s**e | gesuh**s** |
| **/s/** | | | |
| /a/ - context | | gesa**ss**e | |
| /ɪ/ - context | | gesi**ss**e | |
| /ʊ/ - context | | gesu**ss**e | |

## 3.4. Experiment 1: Transillumination, fiberoptic films, and Electropalatography

### *3.4.1. Experimental set-up*

The experimental set-up is schematically depicted in Figure 3.1. A standard endoscope of the type Olympus ENF (type P3) was attached to a camera and connected to a video recorder (Hitachi CCT) with a monitor. The video images enabled the Otorhinolaryngologist Dr. Dahlmeier to control the position of the tip of the endoscope and to identify whether saliva production influenced the signal. In the latter case the subject was asked to swallow and repeat the relevant sentence again. The video signals were taped to enable qualitative interpretation of the transillumination data. The frame rate is 25 per second. The video tape of CG was digitised and some qualitative analysis was undertaken.

To provide the relevant amount of cold light for the tip of the endoscope, an external light source was attached to the endoscope. The external light source had to be plugged into the main power supply, because of a missing special adapter to connect it to a battery. The connection to the power caused a 50 Hz noise (frequency of the power) onto the transillumination data (filtering will be described in the postprocessing section).

Two phototransistors, PGG1 and PGG2, were glued externally onto the subjects neck (see Figure 3.1) and connected to the Photoelectroglottograph (type LG 900 No. 13). PGG1 was placed between c. thyroid and c. cricoid and PGG2 below c. cricoid. PGG1 is more sensitive to the anterior part of the glottis. Its advantage is that it shows higher amplitudes, but the signal is also more sensitive to vertical laryngeal movements, e.g. during lip rounding. PGG2 is more sensitive to the posterior part of the glottis, which is related to devoicing and hence, it is often preferred in investigations of the voicing contrast. Löfqvist and Yoshioka state:

> *"The results of the present study show a high correlation between measures of glottal area variations obtained by fiberoptic filming and by transillumination. For this to hold true, it was necessary to position the phototransistor just below the cricoid cartilage"* (Löfqvist and Yoshioka 1980, p. 798).

They report baseline shifts for PGG1 and spuriously high glottal opening amplitudes for /k/. Hence PGG2 is the most stable sensor, but it is also weaker in amplitude (Hoole 1999).

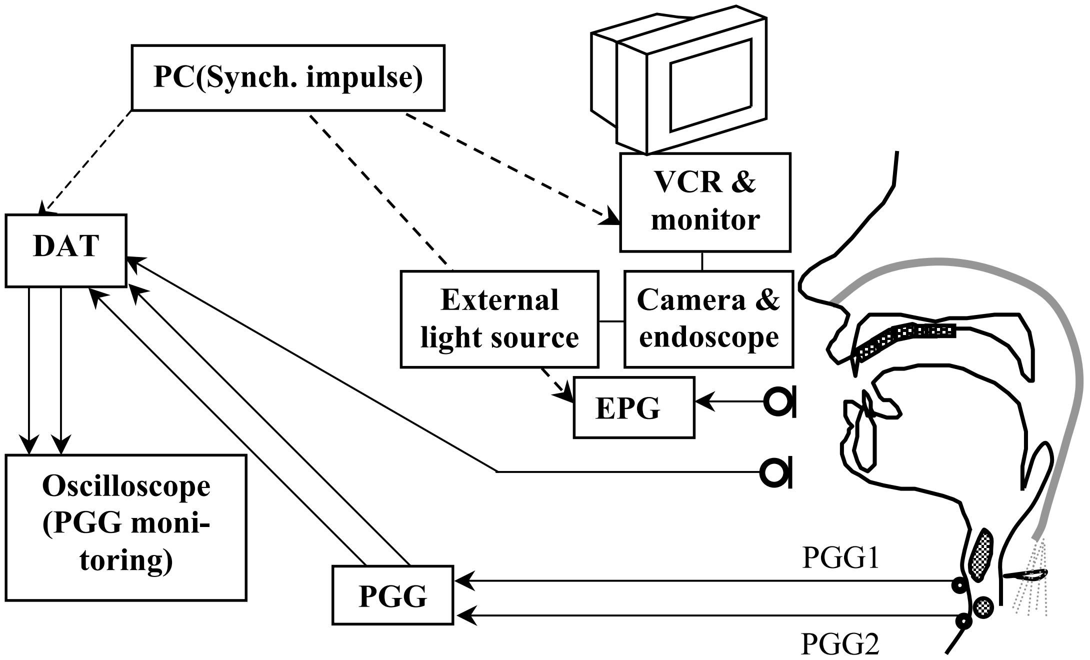

**Figure 3.1:** Schematic overview about the experimental set-up in experiment 1

Most lights in the room were switched off during the recording sessions, curtains were put in front of the windows and a towel was wrapped around the subject's neck to guarantee that photo transducers would only register light from the endoscope.

Photoelectroglottographic data and acoustic data (with a Sennheiser microphone MKH 20) were recorded on a multichannel DAT-recorder (Sony PC 208 Ax) and the PGG signals were checked by monitoring the data through an oscilloscope. Each subject wore her/his own artificial palate and EPG data were stored on the first channel of the EPG3-system simultaneously with the audio speech signal on the second channel. The audio signal for the EPG-PC was recorded by another high quality microphone.

A rectangular synchronisation impulse was sent from another PC (see Figure 3.1) to several channels: one to the audio channel of the video recorder, one to the audio channel of the EPG-PC and one to the multichannel DAT-recorder. The synchronisation impulse allowed the data to be chunked into sections with a duration of 3 seconds. The time information of the synchronisation impulse was used to align the recorded data from different channels and techniques (see Hoole 1996, p. 92-94).

Six people participated during the recordings:

1. one person at the EPG-PC,
2. one person at the PC with the synchronisation impulse,

3. one person holding cards with the speech material (one card per sentence) at a comfortable height for the subject,
4. one person to attach phototransistors and to monitor the PGG signal,
5. the subject,
6. and the ENT-specialist to do the insertion and control the position of the endoscope.

Since so many people participated, the recording could not take place in a small sound proof room and even then, the involved PC's would have made some background noise. Steps were however taken to minimise the background noise.
Before inserting the endoscope into the subjects pharynx a nose spray (Gelonasal) was inserted in one of the subjects nostrils to reduce possible sensitivity and swelling. Another spray with local anaesthetic (Xylocain) was inserted into the pharyngeal cavity especially to inhibit sensitivity to the gag reflex in order to guarantee the subjects well-being. It neither affects the movements of the vocal folds nor of the tongue. These small amounts of anaesthetic are used in endoscopy: in the clinical practice as well as in phonetic experiments.

### *3.4.2. Postprocessing*

Transillumination data with a sampling frequency of 24000 Hz were further postprocessed using the Matlab scripts implemented by Phil Hoole and Christine Mooshammer. In a first step a program searched for the synchronisation impulse in order to generate „cut" files. It became necessary to chunk the whole recording into trials of 3 seconds, since such a procedure allows a precise post-synchronisation with EPG data. In a second step all trials got labels (called code in the cut-file), e.g. A+T1 would correspond to the first repetition (1) of the sentence with /t/ in /a/-context. Code names are easy to handle, when editing files.
In a third step a program extracted channels (AUDIO, PGG1, PGG2) and sweep numbers from cut files and multiplexed Sony input files, digitally transferred from the multichannel DAT-recorder. For reasons of data reduction and further velocity calculations, data were downsampled in a first step to 3000 Hz, and additionally, they were low pass filtered using a transition band from 30 to 90 Hz. In a fourth step data were filtered again from 15 to 60 Hz, and they were downsampled to 200 Hz. This was necessary for the following reason: transillumination data were still relatively noisy due to the connection of the Photoelectro-Glottograph to the power system. Since the calculation of the velocity signals of PGG signals was useful in order to label the glottal events by

means of extreme values at the velocity signal, PGG data had to be smoothed. The velocity is the first derivative of the movement signal, consequently a noisy movement signal would result in an even more noisy velocity signal which would be worse for labelling using extreme values.
To visualise the acoustic, transillumination and EPG data simultaneously the Matlab software package "Artmat" written by Christine Mooshammer is used.
To summarise, acoustic data had a sampling frequency of 24 kHz, EPG data 100 Hz and PGG data 200 Hz respectively.

### *3.4.3. Reliability*

No interference between EPG and transillumination recordings are found during the recordings. Both systems are sensitive to different kinds of energy (to light for PGG or to electric current for EPG) and they were not in contact with each other during the recording, hence, it is not likely that they interfere.
The advantages and disadvantages of transillumination recordings are already pointed out in section 2.2.3.. At present it is not possible to calibrate the system properly, hence the distance between endoscope and glottis can vary. Nevertheless, a comparison of different amplitudes can be made by normalising transillumination data (Hutters 1984, Löfqvist and Gracco 1994).
To control for positioning of the endoscope in the pharynx and possible epiglottal movement or saliva production, fiberoptic films were taped on video and inspected later on. Two PGG phototransistors instead of one were used, but in this study it is focused on the sensed light from PGG2.
The reliability of the EPG data depends on the tuning of the system. To tune the system, the subject had to put the whole tongue on the EPG palate and the system was adjusted so that all contacts were on. In a next step the subject had to produce an /a/ with no contacts with the palate. The system was tuned that no contact patterns were visible on the computer screen. Different speech samples were produced and the relevant contact patterns were checked.
It could be suspected that pseudo-palates might interfere with the sensory feedback of the palate and therefore influence normal articulation. However, these arguments can be rejected since it is unlikely that the sensory feedback of the palate itself would play a major role in speech production. It is more likely that the sensory feedback of the tongue is of primary importance, since the tongue can adapt to different palate shapes (e.g. Hamlet and Stone 1976, 1978, Baum and McFarland 1997) or even to unexpected palate perturbation (Honda and Kaburagi 2000, Honda, Fujino and Kaburagi 2002). In several experiments it has been shown that wearing a pseudo-palate does not perturbate tongue palate contact patterns significantly or change intelligibility significantly (for review

see Byrd et al. 1995). Note, most investigators ask their subjects to wear the pseudo-palate before the recording takes place in order to take some time for speakers adaptation.

The reliability of EPG patterns can be impaired by excessive saliva production, i.e. more contacts would be seen than normal[27].

Another factor which should be taken into account concerning the reliability of the temporal data in general is the speech rate of the subjects. Subjects had 3 seconds to produce the relevant sentence, a window which has been determined previously by repeating different sentences at a comfortable normal speech rate. The beginning of the recording was initiated by an audible beep tone (the synchronisation impulse). The subject was told to repeat the sentence again or to increase rate when the sentence was produced too slowly.

To test for rate effects the Kolgomorow-Smirnoff test was applied. The time interval from the beginning of C1 in the first target word to the end of C3 in the second target word was chosen, i.e. the interval from: *Ich habe ge**{CVCe nicht geCVC}** erwähnt,* shown in brackets. The test was calculated for each subject separately and split by sentences including C = /s z/ and sentences including C = /d t/. No effects of speech rate were found for the two different experiments.

### *3.4.4. Fiberoptic films*

The fiberoptic video films serve as an additional control to check whether there are possible influences on the transillumination signal, e.g. saliva or epiglottal movements. In a broader sense they are also a good "reminder" of what was done during the whole recording.

The quality of our fiberoptic films changed considerably between the second and third recording. For the first and second recording (JD and SF) images were bright enough to check for epiglottal movement or saliva at the tip of the endoscope, but they were too dark for further detailed analysis concerning the amount of glottal opening. Before the third recording took place (the one for CG), the automatic gain control of the camera was switched on which gave a brighter image.

Then, the fiberoptic filming material from subject CG was further processed at the Multi Media Centre of the Humboldt-University Berlin. Images were digitised and afterwards processed using an Apple Macintosh and the software Final Cut Pro, Version 1.2.5. The two audio channels (first channel = synchroni-

---

[27] This was the case for the second experiment (EMA/EPG) for speaker JD and hence we repeated the recording again at another date.

sation impulse, second channel = spoken material) served as references to cut the speech material into a file for each sentence.

The film consisted of 720 x 576 pixels per image. The number of pixels was further reduced to 250 x 250 pixel per image (parts of the left and right side which did not belong to the glottal picture were cut off so that the image became a square). The resulting image sequences were saved as quick time movies onto CD-ROM. In addition, images from certain time points were collected (see 3.6.3.). To enhance the quality of the images, they were post-processed in brightness and contrast with the software Paint Shop Pro 6.0.

## 3.5. Experiment 2: Electromagnetic Articulography and Electropalatography

### *3.5.1. Experimental set-up*

Electromagnetic Articulography (EMA) tracks midsagittal fleshpoint movements by measuring induced current from receiver sensors moving in a magnetic field.

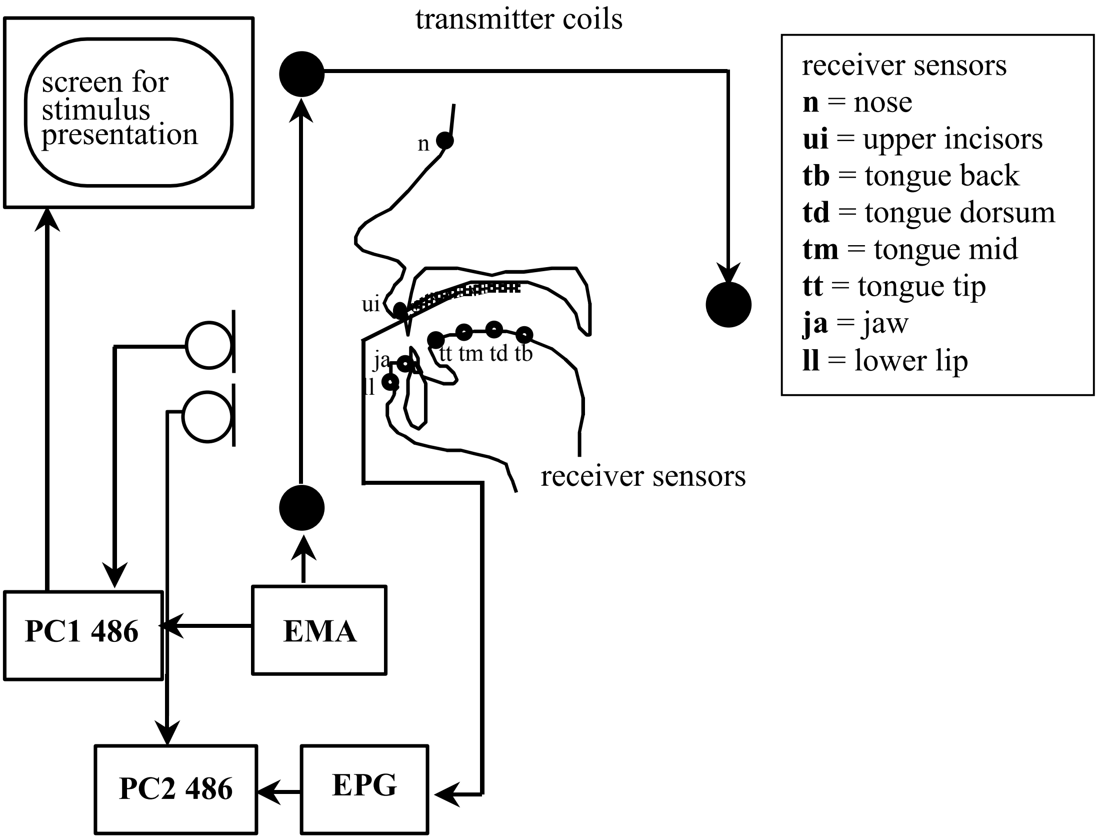

**Figure 3.2:** Simultaneous EMA and EPG experimental set-up

The magnetic field is generated by transmitter coils. The transmitter coils are fixed midsagittally on a plastic helmet, which subjects were wearing during the recordings.

Transmitter coils are configured on an equidistant triangle. Receiver sensors are attached to different fixed points (reference points) or articulators (see Figure 3.2). Two receiver sensors, one at the bridge of nose (n) and one at the upper incisors (ui) served as reference coils to enable compensation for head movements in post-processing.

To place the sensor coils at the right position on the tongue (see Figure 3.2) and normalise for interspeaker differences in tongue size, the following procedure was applied: A purple non-toxic coloured disinfectant was painted on the posterior end of the subjects artificial palate. Then, the palate was fixed in the subjects mouth and the subject was asked to press the whole tongue against the palate. The purple line corresponding to the end of the hard palate could be seen on the subjects tongue.

The tongue dorsum (td) sensor was placed at the purple line and the tongue tip coil (tt) was placed approximately 1 cm posterior to the tip of the tongue. Tongue mid (tm) was attached half-way between tt and td. The tongue back (tb) sensor was placed posterior to td using the same distance as from tt to tm and tm to td.

Before attaching sensors onto the articulators, they were first glued on small pieces of silk in order to increase the area of contact and therefore the stability of the sensor on the tongue. This procedures were adopted from Phil Hoole and helped to keep the sensor as long as possible onto the tongue. Extreme salivation can be responsible for coils coming off during the recording and a repetition of the whole experiment is reasonable[28].

Weismer and Bunton (1999) have acoustically and perceptually evaluated the potential influence of lingual coils or pellets used in x-ray and EMA investigations. Results from 21 subjects showed no consistent effects of pellets across speakers, but certain effects were consistent within a speaker. In a perception test listeners were not able to identify reliably between the stimuli with pellets on and off. The authors suggest a small screening test, where subjects would produce some speech material including high vowels and fricatives with pellets on and off. But criteria for rejecting a potential subject should depend on the aim of the study.

All subjects included here do not show any particular sensibility to the EMA coils or the pseudo-palate. They have often participated in EMA/EPG sessions and their results do not show any obvious experimental artefacts.

---

[28] This was the reason to repeat subject SF's recording.

Before gluing the sensors onto the tongue, the positions were marked using the purple colour again, in order to guarantee accuracy of midsagittal placement. It has to be as exact as possible, since displacement of about 5 to 10 mm from the mid-line will cause an error of 1.5 to 4 mm (Hoole & Nguyen 1999 referring to Honda & Kaburagi 1993). The jaw sensor was placed below the lower incisors and the lower lip sensor at the vermilion border. The recording took place in a sound proof room. One sentence corresponding to one stimulus of 3 seconds was displayed on a screen for the subject. The subject got a “beep” signal as a prompt to start speaking.

The post-processing procedure and synchronisation with acoustic and EPG signals was described in detail in Hoole (1993, 1996). The procedure was implemented at the phonetics lab in Berlin in co-operation with Christine Mooshammer.

Original sampling frequencies were 16 kHz for acoustic data, 100 Hz for EPG data and 400 Hz for EMA data. EMA data were further downsampled to 200 Hz. For x and y co-ordinates of tt, tm, td, tb, ll sensors a low pass filter was applied with a bandwidth of 18 Hz with a damping of 50 dB at 52 Hz. Velocities and accelerations of the same sensors were low pass filtered with a band width of 15 Hz and a damping of 50 dB at 25 Hz.

### *3.5.2. Reliability*

The measurement of the sensor position in a two-dimensional co-ordinate system depends on the conversion from induced voltage of the sensor into distance from the corresponding transmitter coil (Hoole 1996). The voltage-distance relation is theoretically estimated by the equation V = k*1/D**3 with k as an amplification factor. However, in practice the exponent of D is not exactly 3 and it has to be estimated empirically by calibration. To calibrate the system the manufacturer developed a device in which each sensor can be placed systematically in a variety of possible exactly predefined positions in the measurement field. Hereafter the exponent can be determined by regression. The accuracy of the data decreases away from the centre of the triangle of the transmitter coils. Hoole (1996) reports an error of 0.67 mm +/-0.42 for positions more than 6 cm away from the centre (in the midsagittal plane) and 0.2 mm +/-0.13 for positions up to 6 cm. In practice, tongue sensor coils are usually placed close to the centre of the helmet with small error values.

Another important influence concerning the reliability of EMA data are rotational misalignments (twist and tilts of the receiver coils). The rotational misalignments are called “tilt” by the Carstens software and they are automatically detected. The closer the average value of the tilt to 100 percent and the smaller the standard deviation, the higher the reliability of the EMA data (Mooshammer 1998, p.104). Tilt values for each sensor coil were recorded per sweep. Since the tongue back coil from subject JD fell off, no tilt values could be reported. Only sweeps which were further analysed are included in calculating the mean tilt values for each coil and their standard deviations.

**Table 3.3**: Averaged tilt values with their variability for each coil and all subjects

| Receiver sensor | n | ui | tb | td | tm | tt | ja | ll |
|---|---|---|---|---|---|---|---|---|
| CG tilt (n=162)[29] | 97.0 | 97.2 | 98.1 | 100.6 | 93.6 | 91.8 | 99.1 | 94.6 |
| CG std. | 0.01 | 0.36 | 0.7 | 0.69 | 1.28 | 1.46 | 1.08 | 0.62 |
| JD tilt (n=162) | 97.0 | 96.3 | # | 100.6 | 100.3 | 93.8 | 97.9 | 94.4 |
| JD std. | 0 | 0.37 | # | 0.5 | 0.97 | 1.12 | 0.21 | 0.42 |
| SF tilt (n=161) | 97.9 | 97.1 | 96.6 | 96.8 | 95.0 | 95.5 | 102.6 | 97.0 |
| SF std. | 0.29 | 0.07 | 0.58 | 0.38 | 0.75 | 0.73 | 0.64 | 0.02 |

As can be seen in Table 3.3 all tilt mean values are above 90 percent. The reference coils (n, ui) exhibit high values between 96 and 97 percent and small standard deviations. For all subjects the lowest tilt values, i.e. highest rotational misalignment were found for the tongue tip coil. Since it is consistent for all the subjects and varies more than most of the tilt values from other coils, it could be based on the fact that the tongue tip is most flexible and changes alignment of the midsagittal plane more frequently than other articulators.

## 3.6. Labelling criteria

### *3.6.1. Acoustical time landmarks for stops*

For acoustic analysis the software package Praat (version 4.0.11.) was used. An oscillogram and a derived spectrogram were plotted for each sentence. Since the spectral information was used for temporal segmentation a broad band sonogram was used. Its is based on an FFT analysis using a 5ms Gaussian window with a 6 dB/Oct pre-emphasis.

For visualisation a 45 dB dynamic range was taken into account over the 0 to 8 kHz frequency range. The following acoustic time landmarks were labelled for stops (for examples see Figure 3.3):

---

[29] The number of trials included here consists of all tokens recorded in the experiments, but not all tokens will be taken into account, only the speech material reported here.

- **t_f2off** was defined as the second formant offset of the vowel preceding the consonant. For many tokens of /d/ in word medial position the second formant does not end, but a clear reduction in intensity is observed. For these cases the end of the high intensity of the second formant was chosen as an equivalent to t_f2off. If there was no remarkable time point for a change in intensity, no label was placed.
- **t_burst** was defined as the beginning of the burst spectrum. For cases with multiple bursts, the one with the highest intensity was chosen.
- **t_noisoff** was defined as noise offset, i.e. high frequency noise offset in the spectrogram. The offset of high frequency noise is easier to detect and more reliably defined than onset of phonation for the following vowel, because acoustic data of the transillumination recording were very noisy.
- **t_f2on** was defined as the onset of the second formant of the following vowel.
- For position C3 the onset of the second formant of the preceding vowel **t_f2on_prec** was additionally labelled for the calculation of the vowel duration.

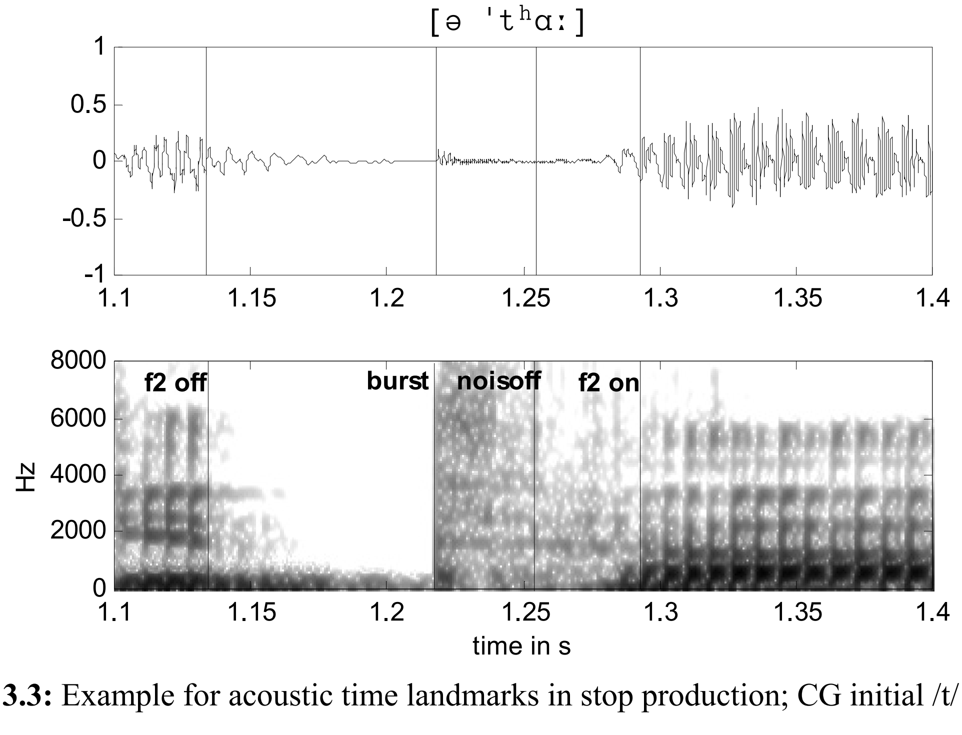

**Figure 3.3:** Example for acoustic time landmarks in stop production; CG initial /t/ in getahte

After examining different files, the decision was made that no label could be placed at the voicing offset during oral closure, since the quality of the acoustic signals, particularly for the first experiment, was not good enough for this to be a reliable measurement criterion. Comments were also noted for special characteristics, e.g. glottalisation of the following vowel.

Additionally, in Appendix 1 an overview is given whether the relevant trials are perceived as phonetically voiced or voiceless. Since the author of this study has some theoretical background knowledge and her perception might be influenced by it, a more naive listener, a student from the lab has been chosen. The student got the task to label a token with ‘0’ when she perceived an obstruent without voicing and to label it with ‘1’ when she perceived it with voicing.
Time landmarks are used to calculate the following acoustic durations related to stop production (see Table 3.4).

**Table 3.4:** Calculated acoustic durations for stops

| Calculated durations | in | formula |
|---|---|---|
| vowel duration | ms | t_f2off (C2)–t_f2on (C1) for the vowel between C1 and C2<br>t_f2off (C3)–t_f2on_prec for the vowel before C3 |
| closure duration | ms | t_burst–t_f2off |
| noise duration | ms | t_noisoff–t_burst |

The relevance of these durations for the voicing contrast will be discussed in the result chapter. The term noise duration might be somewhat unusual, since the term aspiration duration is more often used. There were two reasons why noise duration is chosen: first, one of the time landmarks of noise duration is the offset of high frequency noise, not the beginning of phonation or periodicity (which would correspond to aspiration duration or VOT). Second, the term aspiration is commonly associated with noise produced by an open glottis. Other noise characteristics found in the spectrogram could be produced due to the release of the oral closure, vocal tract walls going back to their rest position after oral release, any vocal tract constrictions or incomplete closures (see Stevens, 1998, p. 347-348). On the basis of acoustic data alone it is difficult to decide where noise comes from (problem of source-filter separation). However, if a considerable amount of noise occurs for more than 20-30ms, then it is probably due to aspiration, i.e. due to an open glottis (e.g. Stevens and Klatt 1974, Kim 1970).
In the beginning of the analysis Voice Onset Time (VOT) based on Klatt 1975 was computed too (Fuchs et al. 2002, poster presentation). The measurement is defined as the difference between burst and the beginning of the second formant of the following vowel. It turned out that results for voice onset time in word final position showed a considerable variability which was related to the measurement criterion. Since the vowel following the consonant is always glottalised, the onset of the second formant occurred quite late and hence, VOT was relatively long and variable. Therefore noise duration served as a more reliable measurement in this position, it was less affected by glottalisation.

However, for word initial position only, other labelling criteria might be used as recently discussed in Francis, Ciocca and Yu (2003), but since different positions are taken into account the offset of noise duration was chosen.

### *3.6.2. Acoustical time landmarks for fricatives*

Fricatives are labelled according to second formant offset, frication onset, and frication offset (see Figure 3.4). The defined labels are:

- **t_f2off** = offset of the second formant of the preceding vowel,
- **t_fricon =** frication onset (onset of high frequency noise in the spectrogram and noise in the oscillogram),
- **t_fricoff =** frication offset (offset of high frequency noise in the spectrogram), and
- **t_f2onset** = onset of the second formant of the following vowel.

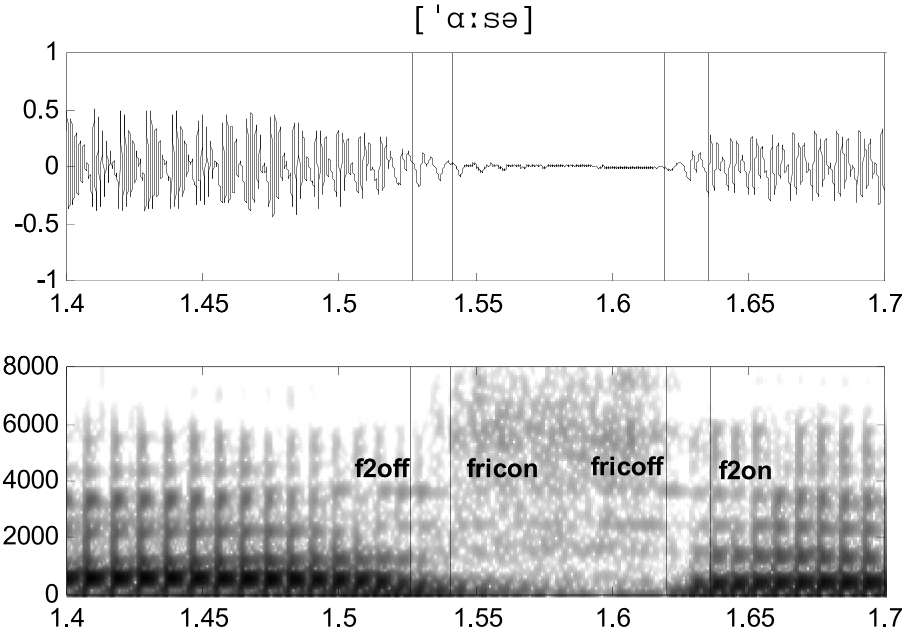

**Figure 3.4:** Example for acoustic time landmarks in fricative production; CG final /s/ in gesahs

Additional comments like glottalisation of the following vowel, persisting second formant through frication are added too. As for stops a student judged the voicing status of the fricatives perceptually.

In voiced fricatives onsets and offsets of the second formant are most difficult to label, since formant patterns often run through the period of frication. If a visible change in intensity was found, a time landmark corresponding to F2 offset was defined. If there was a continuous change in intensity, no time landmark was labelled. Similar difficulties were found defining frication onset, since some

frication often occurs already during the vowel. Therefore the change in intensity was taken as an additional criterion for labelling.
The temporal landmarks selected here are used for further analysis in laryngeal-oral co-ordination.

### *3.6.3. Fiberoptic film labelling*

For stops, an image was taken at about the time of the burst with the highest amount of glottal opening. With reference to this image, two surrounding images were chosen, the preceding and the following one. Since the frame rate of the video film is 25 frames per second, the preceding image occurs 40 ms and the following image 40 ms after the time of the burst.
For fricatives an image was identified for peak glottal opening during the middle part of friction noise together with one preceding and one following image. From the three images the one was chosen, which shows the highest amount of glottal opening (if glottal opening occurs). Fiberoptic film images are discussed in the result chapter in comparison to results from transillumination and acoustics.

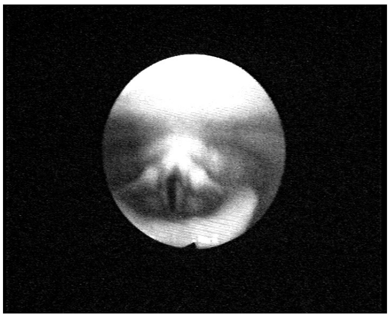

**Figure 3.5:** Example of postprocessed fiberoptic film image for /t/ at the time of the burst CG, C1 in /i/-context

### *3.6.4. EPG time landmarks*

For stops two time landmarks were determined: One at the beginning of the closure **t_clon_epg** and one at the end of closure **t_cloff_epg**. Since closure onset starts already before full closure is made, closure onset was defined as the moment where at least two contacts in the central part of the anterior portion of the palate are „on“ (see Figure 3.6, circle).
For those cases where only one or no contact occurs in the central part of the alveolar region, as it has often been found in word medial position, the moment of most contacts in the anterior region served as an equivalent. Closure offset was defined as the moment before oral release, i.e. one contact in the front two

rows is missing, equivalent to a gap. For the cases of incomplete closure the pattern before one contact is missing served as a landmark.

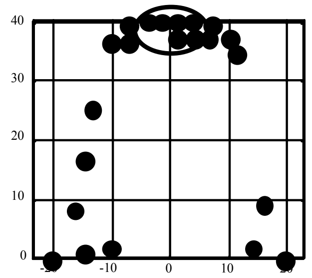

**Figure 3.6:** EPG closure onset pattern with dots as tongue contact patterns. Upper part corresponds to the anterior portion and the lower part (above 0 on the y-axis) to the posterior portion. The area in the circle is relevant for the defined closure onset. X and y axes are in mm.

Additionally, the two EPG time landmarks served as a temporal reference to determine the targets of tongue tip, tongue dorsum and jaw. The vertical position (y) of the relevant EMA coils were taken into account in order to study a possible involvement of the tongue and the jaw in the voicing contrast.

For fricatives the onset and offset of constriction could not be reliably labelled by EPG data. Thus the acoustic landmarks, frication on- and offset, served as a reference for further computations.

Since EPG patterns are binary (contacts can be made or not, i.e. they are 0 or 1), the patterns are easier to discuss when they are reduced and transformed into numerical values. Data reduction methods for different EPG patterns are described in e.g. Hardcastle et al. (1991), Fontdevila et al. (1994), Byrd et al. (1995) or Hardcastle et al. (1999).

All EPG patterns between closure onset and offset were further analysed in calculating several parameters (see Table 3.5) to test whether there would be more contacts in /t / compared to /d/ production.

The percent of contact in the anterior region (ant) increases when more tongue-palate contact is made. A maximum of 100% is possible. The same is true for the percent of contact in the posterior region (post), but the post parameter takes tongue palate contacts patterns in the posterior region into account. The centre of gravity index (cog) is a weighted index in the front-back dimension. The cog is higher, the more fronted tongue palate contact patterns.

**Table 3.5:** EPG parameters with abbreviations and formula

| EPG parameter | hereafter | formula |
|---|---|---|
| % of anterior contact in the 4 front rows | ant | ant = x*100/total number of contacts in the anterior region |
| % of posterior contact in the 4 back rows | post | post = x*100/total number of contacts in the posterior region |
| main concentration of activated electrodes across the palate (higher weight = more fronted) | cog | cog = (0.5R8+1.5R7+2.5R6+3.5R5+4.5R4+5.5R3+6.5R2+7.5R1)[30]/ total number of contacts |

In order to discuss EPG patterns not only with respect to a certain time landmark, but also regarding changes in a certain time interval, the following procedure was applied: the time interval of interest was normalised in time (the duration of the interval of the relevant repetition was set to 1) and the values of the relevant EPG parameter were linearly interpolated and then resampled so that for each parameter 10 values were systematically computed during the time normalised interval. Hereafter the 10 computed EPG parameter values could be averaged over all repetitions.

### *3.6.5. Time landmarks for glottal opening*

The photosensor below the cricoid is the preferred sensor in transillumination studies, and thus only PGG2 was taken into account here. Onsets and offsets of glottal abduction were defined on the velocity signal of PGG2, since labelling on the movement signal itself is rather difficult. There are movements which start slowly, almost like steady states and others which start more quickly. Labelling on the velocity signal is more easy and consistent, because extreme values can be found.

Using the velocity signals for labelling is a common strategy in articulatory studies (e.g. Tasko and Westbury 2002).

To get reliable temporal information on glottal opening on- and offset, a 5 % threshold value was chosen arbitrarily after several measurement with 5, 10 and 20%. The 5 % values are based on the distance between a zero crossing to the next maximum.

---

[30] R corresponds to row, i.e. 0.5R8 would be the number of contacts in row 8 multiplied by 0.5.

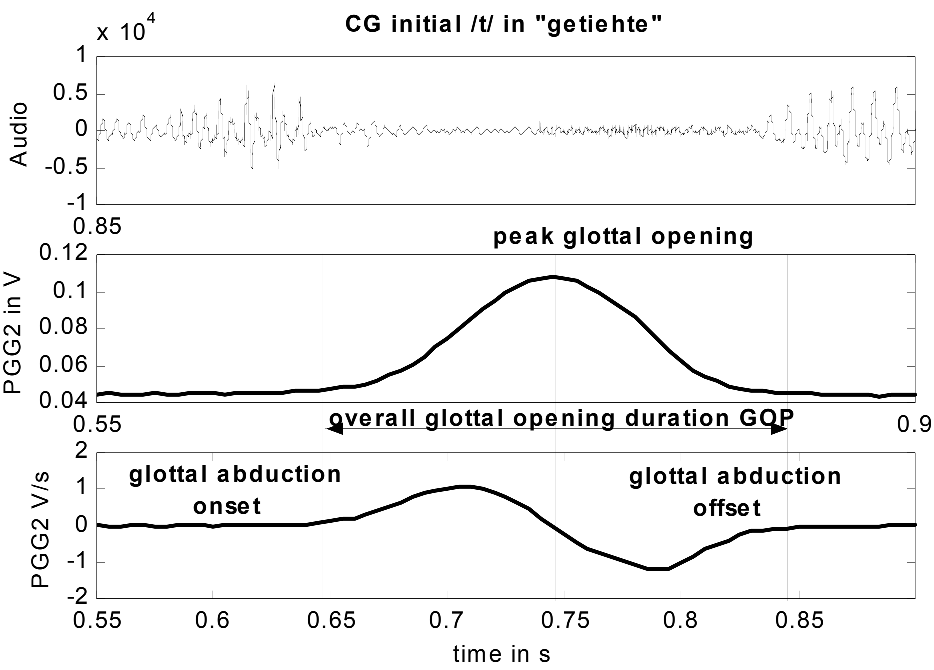

**Figure 3.7**: Example for defined time landmarks concerning glottal opening in /t/, C1, /i/-context

In some cases, where the PGG2 curve was very weak, it was difficult to decide whether the weak amplitude should count as glottal opening or not. The decision is generally based on the velocity signal. If a small glottal opening gesture and a closing gesture were found, and the amplitude shows at least two pixel difference with reference to the baseline, the amplitude counts as glottal abduction.

Figure 3.7 shows the defined time landmarks for cases with glottal opening:

- **t_opg_on_pgg** = glottal abduction onset,
- **t_peak_pgg** = peak glottal opening
- **t_clg_off_pgg** = glottal abduction offset.

As a temporal measurement overall glottal opening duration (hereafter GOP) was calculated as t_clg_off_pgg – t_opg_on_pgg.

To discuss the relations of the amount of glottal opening in different positions without ignoring the possible variation between distance of the tip of the endoscope and glottis, a normalisation procedure was applied. It assumes that changes of the distance between endoscope and glottis can be neglected within one repetition (=3 seconds). All peak amplitudes with respect to the baseline are related to the peaks with the highest amount of glottal opening with respect to the baseline. In stop production the highest amount of glottal opening occurs in the stressed position and thus, the stressed position served as a reference.

The amplitudes between peak glottal opening and the baseline were calculated in each repetition and thereafter set to 100 percent. All other glottal opening peaks within one repetition (C2, C3 for stops) were related to the reference peak following the formula:

x = (PGG2 (t_peak_pgg C2-baseline)*100)/PGG2 (t_peak_pgg C1-baseline)
x = (PGG2(t_peak_pgg C3-baseline)*100/PGG2(t_peak_pgg C1-baseline).

The derived relative glottal opening peak values should give a general impression of the extent to which glottal opening is reduced in the post-stressed positions.

In fricative production the largest glottal opening varies with respect to position in CG's data and thus a normalisation procedure was not carried out. The absolute glottal opening values with reference to the baseline are taken into account.

To investigate laryngeal-oral co-ordination the time of the burst (t_burst) is set to zero (line-up point) and all other glottal time landmarks are related to this point. It would have been possible to take closure offset, defined by EPG as a line-up point, but acoustic data have a higher sampling frequency than the EPG data and hence they are more accurate. For closure onset, the EPG data are taken into account, not the acoustically defined second formant offset. EPG closure onset gives a reliable articulatory criteria of tongue-palate contact, even if it has a lower sampling frequency.

**Table 3.6:** Calculated durations concerning laryngeal-oral co-ordination in stop production

| Calculated durations | hereafter | In | formula |
|---|---|---|---|
| duration between oral closure onset and glottal opening onset | COOR_on | ms | t_clon_epg – t_opg_on_pgg |
| duration between burst and peak glottal aperture | COOR_peak | ms | t_peak_pgg – t_burst |

Concerning laryngeal-oral co-ordination for stops the following two durations were calculated (Table 3.6):

If COOR_on is positive, than oral closure onset started before onset of glottal opening. If it is negative, onset of oral closure started with a delay compared to onset of glottal opening. The duration from COOR_peak corresponds to the timing between oral release and peak glottal opening. Peak glottal opening is known to occur close to oral release for voiceless aspirated stops, whereas for unaspirated stops it is likely to occur before oral release.

Concerning laryngeal-oral co-ordination for fricatives three durations were computed, see Table 3.7.

Again, if COOR_on fric shows positive values, frication onset started later than glottal abduction onset. If COOR_on fric values are negative, frication onset started earlier than the onset of glottal abduction.

COOR_peak fric values give evidence about the relative occurrence of peak glottal opening during the frication interval. Values below 50% correspond to a peak which is produced rather towards the onset frication and values above 50% correspond to peak glottal opening which is realised towards the offset of frication.

When COOR_off fric is a positive value, offset of glottal opening occurs later than offset of frication. If it is a negative value than frication is delayed with respect to offset of glottal abduction.

**Table 3.7:** Calculated durations concerning laryngeal-oral co-ordination in fricative production

| Calculated durations | hereafter | in | formula |
|---|---|---|---|
| duration between onset of frication and onset of glottal opening | COOR_on fric | ms | t_fricon–t_opg_on_pgg |
| relative occurrence of peak glottal opening in the frication interval | COOR_ peak fric | % | ((t_peak_pgg–t_fricon)*100)/ (t_fricoff-t_fricon) |
| duration between offset of glottal opening and offset of frication | COOR_off fric | ms | t_clg_off_pgg-t_fricoff |

### *3.6.6. Time landmarks for tongue tip and jaw movement*

EMA landmarks were labelled on the velocity signal too, i.e. on the tangential velocity signal tv, since tongue tip and jaw involve horizontal and vertical movements and tangential velocity includes both directions. A 20% threshold criterion was used for defining the tongue tip and jaw closing gestures with its onsets and offsets. The advantages of 20% are described in Kroos et al. (1997). Figure 3.8 shows an example of the landmarks at the tongue tip sensor. The same landmarks were also used for jaw movement. The defined time landmarks are:

- **t_clg_left min** = landmark for the vowel target,
- **t_clg_on** = onset of the closing gesture (20% threshold criterion),
- **t_tv_peak** = landmark of the tangential velocity peak of the closing gesture, and

- **t_clg_off** = offset of the closing gesture (20% threshold criterion).

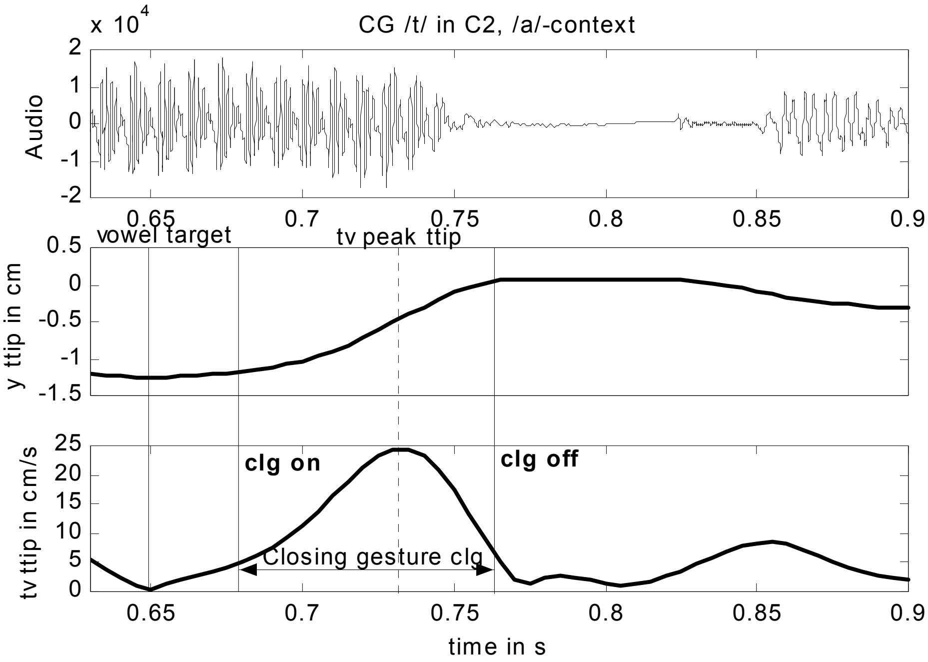

**Figure 3.8**: Example of time landmarks defined on the tongue tip tangential velocity signal of subject CG, producing /t/ in C2, /a/-context

For the tongue tip closing gesture the duration was calculated as t_clg_off-t_clg_on, and the movement amplitude as the integral in this interval. The velocity peak value (tv peak) of the tongue tip movement was also taken into account.

At t_clg_left min the vertical position (y) of tongue tip (y ttip), tongue dorsum (y tdors) and jaw (y jaw) were considered in order to measure positioning of tongue and jaw for the vowel target preceding the relevant voiced or voiceless stop. The horizontal positions are not taken into account here, since the amount of data is already comprehensive.

**Table 3.8:** Calculated durations concerning tongue-jaw co-ordination in stop production

| Calculated durations | in | Formula |
|---|---|---|
| duration between onset of tongue tip closing gesture and onset of jaw closing gesture | ms | ttip clg_on – tjaw clg_on |
| duration between offset of tongue tip closing gesture and offset of jaw closing gesture | ms | ttip clg_off – tjaw clg_off |

In order to study the possible involvement of tongue-jaw co-ordination the durations depicted in Table 3.8 were calculated.

Tongue-jaw co-ordination could only be observed in /a/-context. In all other high vowel contexts labelling the jaw closing gesture was nearly impossible, since jaw movements have only small amplitudes.

### *3.6.7. Statistical procedures*

Due to the techniques used in experiment 1 the possibility of computing statistics is rather poor, since the amount of data is quite small. Generally, it was assumed that speakers would differ significantly due to differences in their vocal tract anatomy, their origin, their sex etc. and thus, they are not averaged.

For the analysis of EPG and acoustic parameters the data from both experiments are taken into account in order to increase the amount of data. It is assumed that individual variations from one to the other experiment would be similar to the ones when a speaker repeats the same corpus at different dates.

For statistical analysis the software packages SPSS 11.0 and Statview 5.0.1. were used. In Statview a 2-way ANOVA was used in order to rule out interaction effects of vowel context (/a/ vs. /i/ vs. /u/) and consonant (/d/ versus /t/). In cases with a strong main effect and no interaction, single comparisons are considered. The confidence interval was set to 95% of the normal distribution.

Different significance levels are described. Values with $p<0.001$ are taken as highly significant, values with $p<0.01$ are taken as moderately significant, and with $p<0.05$ only weakly significant.

Dependent variables are:

- all durations,
- movement amplitudes,
- velocity peaks,
- and y co-ordinates.

Independent variables are:

- consonants CON (/t/vs. /d/) as well as
- vowel contexts VOW( /a/ vs. /i/ vs. /u/).

Data are split by subject (CG, JD and SF) and position (C1, C2, C3). Additionally, mean values and standard deviations were calculated. A complete overview is given in the appendices.

# Chapter 4: Results

The results chapter has the same structure as the corresponding chapter 2, i.e. the trichotomy laryngeal correlates, laryngeal-oral co-ordination and supralaryngeal correlates is maintained.

In the relevant subchapters there will be short descriptions about the origin of the data taken into account. The sections 4.1. (Laryngeal correlates) and 4.2. (Laryngeal-oral co-ordination) consider the production of stops and fricatives, whereas in 4.3. (Supralaryngeal correlates) this work will focus on stops.

## 4.1. Results for laryngeal correlates

### *4.1.1. Introduction*

Results for laryngeal production mechanisms are based on the first experiment, i.e. on data from the transillumination recording and images from fiberoptic films. At the beginning of this section it is outlined which obstruents in different positions were produced with glottal abduction. Besides the first general overview, the amount of glottal opening is discussed in order to enhance insights into laryngeal production mechanisms. Hereafter, temporal phenomena are taken into account since changes in the amount of glottal opening could be related to changes in duration.

Glottal opening was proposed to be an articulatory correlate of the voicing contrast in German (Jessen 1998). Whether this holds true with respect to different stress conditions and to various positions of the obstruent in the syllable is the main research question here.

Glottal abduction is defined on the basis of the velocity signal of the phototransistor below the cricoid cartilage (PGG2). Those gestures are taken into account where at least two pixels difference in amplitude could be observed. Glottal abduction is defined with reference to the baseline (see also methods chapter 3).

Glottal adduction is associated with vocal fold vibrations during the vowels surrounding the relevant obstruents or even with voicelessness, but in the latter case without glottal abduction. Glottal adduction corresponds to the baseline. Since transillumination data in this study are low-pass filtered, no oscillatory patterns could be observed. It should be noted that glottalisation or glottal stops are rather difficult to detect using the transillumination signal since they are often not produced with a considerable amount of glottal abduction.

### *4.1.2. Glottal abduction in stop production*

#### 4.1.2.1. THE OCCURRENCE OF GLOTTAL ABDUCTION IN STOP PRODUCTION

Table 4.1 provides an overview of the number of tokens of /d/ and /t/ which were produced with glottal aperture. Since 5 repetitions were recorded in each condition a total of 5 cases is possible. The two different conditions are vowel context (/a i u/) and position (C1 = stressed, syllable initial position, C2 = post-stressed word medial position, C3 = post-stressed, word final position). Values are split by subject (CG, JD, SF) and vowel environment (/a i u/).

**Table 4.1:** Number of tokens with glottal opening from possible 5, split by subject (from top to bottom CG, JD, SF), by vowel (/a i u/) and by position (C1, C2, C3)

|  | Stressed position | C1 |  | Post-stressed positions | C2 |  | C3 |  |
|---|---|---|---|---|---|---|---|---|
| **CG pgg** |  | **/d/** | **/t/** |  | **/d/** | **/t/** | **/d/** | **/t/** |
| /a/ |  | 0 | 5 |  | 0 | 5 | 0 | 5 |
| /i/ |  | 0 | 5 |  | 0 | 5 | 0 | 5 |
| /u/ |  | 0 | 5 |  | 0 | 5 | 0 | 4 |
| **JD pgg** |  |  |  |  |  |  |  |  |
| /a/ |  | 0 | 5 |  | 0 | 0 | 0 | 0 |
| /i/ |  | 0 | 5 |  | 0 | 0 | 0 | 0 |
| /u/ |  | 0 | 5 |  | 0 | 0 | 0 | 0 |
| **SF pgg** |  |  |  |  |  |  |  |  |
| /a/ |  | 0 | 5 |  | 0 | 4 | 1 | 5 |
| /i/ |  | 0 | 5 |  | 0 | 0 | 0 | 0 |
| /u/ |  | 0 | 5 |  | 0 | 2 | 0 | 0 |

Results provide evidence that in the onset of a stressed syllable /t/ was produced consistently with glottal opening, whereas in both post-stressed positions it varies highly speaker- and vowel-dependently:

- Subject CG produced /t/ with glottal opening in most cases.
- For JD glottal abduction is only found in the stressed position. For all post-stressed conditions the glottis is adducted, at least when filtered transillumination data are taken into account.

- Subject SF results show glottal opening in the post-stressed positions, particularly with respect to the /a/-context.

In /t/ production glottal opening is a compulsory characteristic in the stressed position, whereas it varies with respect to speaker and vowel environment for both post-stressed positions.

In /d/ production only 1 token out of 45 possible cases was realised with glottal opening. Thus, results from this study provide evidence that /d/ surrounded by vowels can be associated with a closed glottis, even though perceptually /d/ is sometimes devoiced as it was proved by an informal perceptual test (see Appendix I).

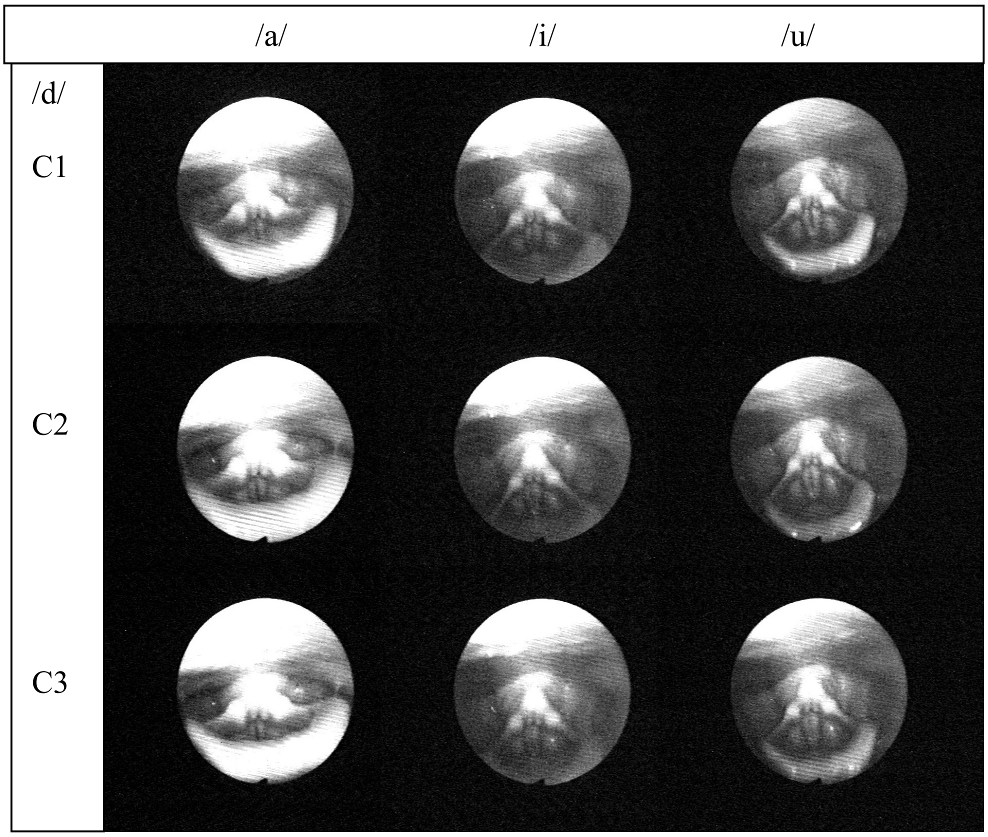

**Figure 4.1:** Example of subject CG's fiberoptic film images for /d/, from C1 to C3 (from top to bottom) and /a, i, u/-context (from left to right)

Images from fiberoptic films for subject CG support these findings for /d/. They exhibit an adducted glottis at the time of the burst (examples are given in Figure 4.1, for comparison with /t/ see Figure 4.3).

One image per condition is plotted in Figure 4.1 (vowel /a i u/-context from left to right; stressed position, post-stressed word medial and final positions from top to bottom). Since no glottal opening is observed for the phonologically voiced stop, the following section will focus on laryngeal production mechanisms in /t/ production.

#### 4.1.2.2. THE AMOUNT OF GLOTTAL OPENING IN /t/ PRODUCTION

In /t/ production the amount of glottal abduction was analysed, but only for cases where a glottal opening amplitude was found. Figure 4.2 exemplifies filtered raw data from CG’s transillumination recording.

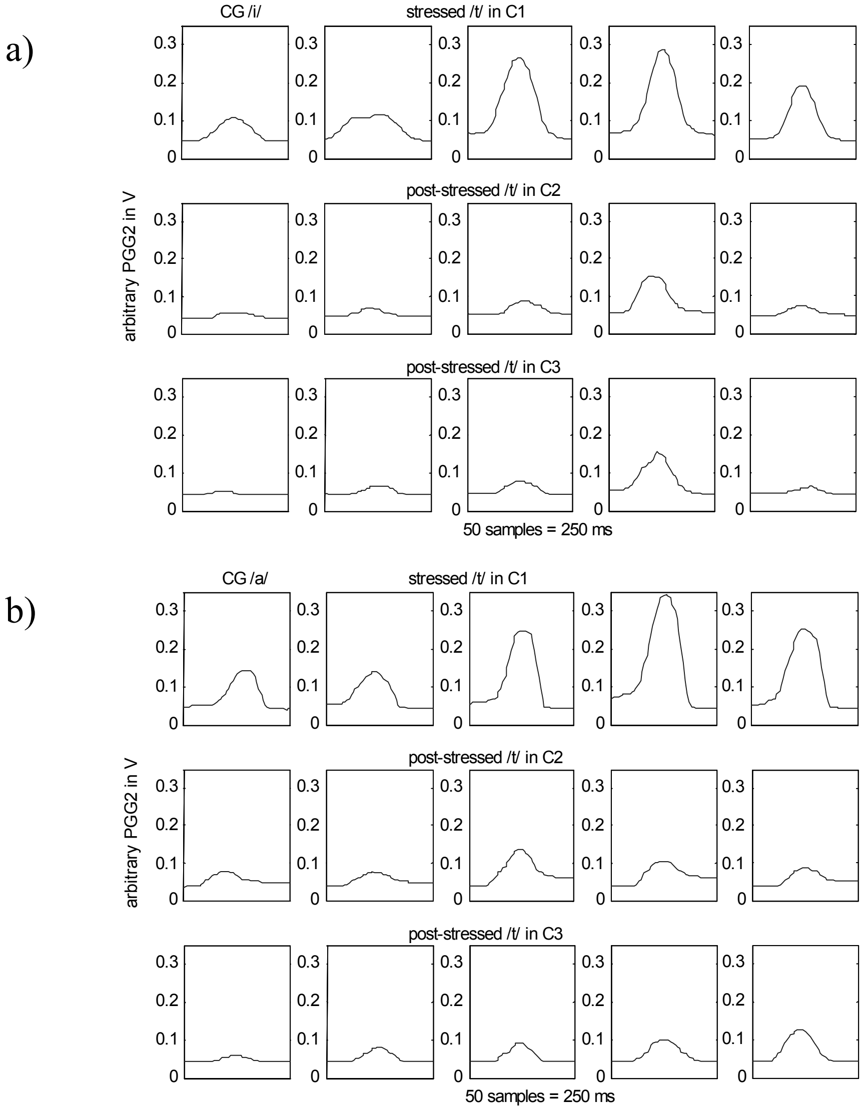

**Figure 4.2:** Glottal opening gestures for /t/ from PGG2, all 5 repetitions (left to right), subject CG, a) /i/-context, b) /a/-context

All five repetitions are shown across the five columns and the corresponding positions within one sentence (C1, C2 and C3) are presented vertically. It is shown that the amount of glottal opening varies with respect to the stressed versus post-stressed positions. The latter two were produced with a considerable decrease in glottal amplitude. The amount of glottal opening in both post-stressed positions is rather similar, i.e. post-stressed word medial and final position exhibit comparable patterns and can not be differentiated for this subject. Hence, deaccentuation from a stressed to an unstressed syllable could be the main factor for the reduction of the glottal opening amplitude.

As Figure 4.2 also displays, the amount of glottal opening within the filtered raw data changes from one repetition to the next (the five repetitions correspond to the five columns). This effect is one of the typical problems of the trans-illumination technique (see also 2.2.3.), since the system can not be calibrated, and the distance between the tip of the endoscope and the glottis varies. A normalisation procedure was applied too, but will be described at a later point.

In Figure 4.2 b) displaying /t/ in /a/-context, a relatively low baseline is observed in some cases on the right side of glottal opening in C1, and on the left side of glottal opening in C2. This artefact is due to tongue retraction in /a/-context which causes epiglottal movement and some shadow onto the glottis. Since the low back vowel /a/ occurs after /t/ in the stressed position, a low baseline can be seen on the right. In both post-stressed positions /a/ preceded /t/ and a lower baseline can be seen on the left side.

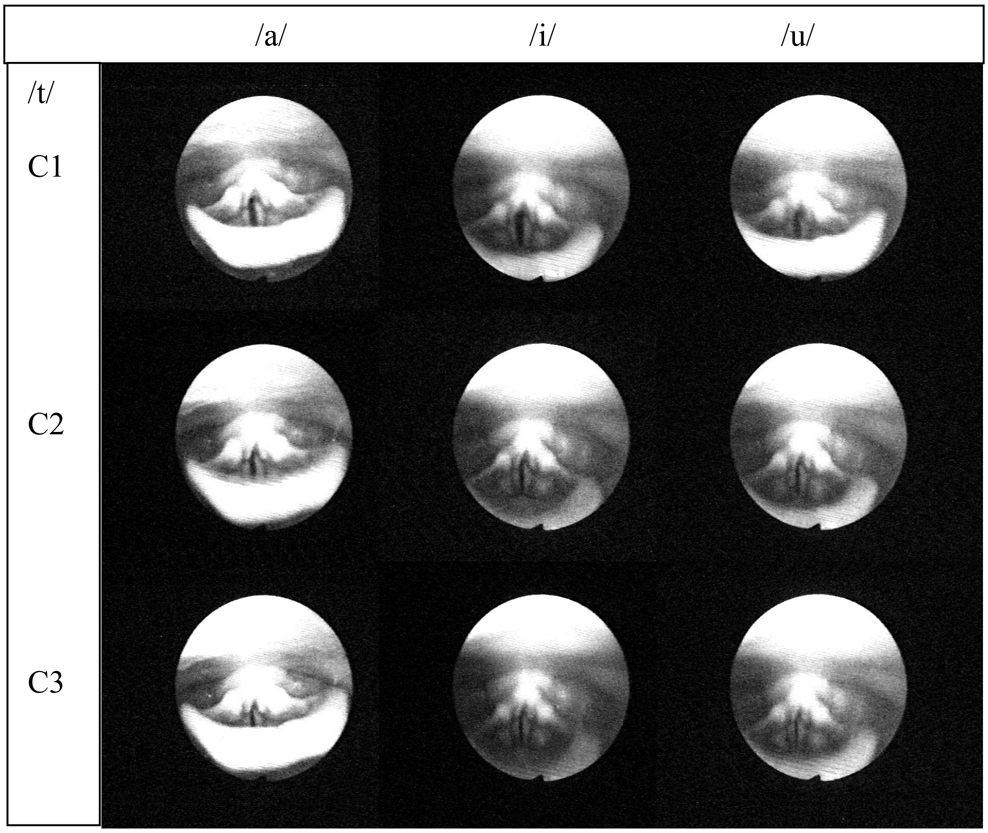

**Figure 4.3:** Example of subject CG's fiberoptic film images for /t/, from C1 to C3 (from top to bottom) and /a, i, u/-context (from left to right).

Images from CG's fiberoptic films support these findings for /t/ more qualitatively. In Figure 4.3 some examples for /t/ in all vowel contexts and conditions are plotted. The highest amount of glottal opening occurs in the stressed position, independent of vowel context. For subject SF glottal abduction occurs most frequently in /a/-context for all positions, although in /u/-context some glottal abduction was found too. Figure 4.4 displays the amount of glottal opening in the different positions in /a/-context. One case in post-stressed syllable initial position is missing and could not be plotted.

Generally, results for subject SF are in agreement with the ones for subject CG, but they exhibit lower glottal opening amplitudes in both post-stressed positions. Since the lower amplitudes in the post-stressed positions are so weak in /a/-context, it seems likely that they disappear in most cases in /i/ and /u/-contexts.

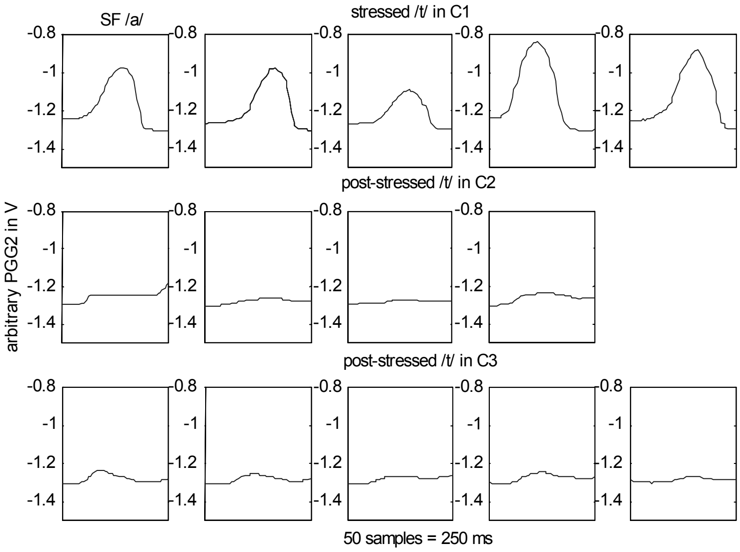

**Figure 4.4:** Glottal opening gestures for /t/ from PGG2, all 5 repetitions (left to right), subject SF, /a/-context

In order to provide more quantitative information about the reduction in amplitude, a normalisation procedure was applied (see 3.6.5.). It is assumed that the distance between the tip of the fiberscope and the glottis is relatively stable within the 3 second interval in which each sentence was recorded. Since glottal abduction amplitude with respect to the baseline is always largest in the stressed position, it is set to 100 percent. All other glottal opening peaks are defined relative to the maximal glottal opening in the stressed position in the relevant sentence.

Figure 4.5 shows barplots with means of relative peak values for the post-stressed positions (C2, C3) with reference to the stressed position (C1). The latter is always 100 percent. Values above the bars represent means.

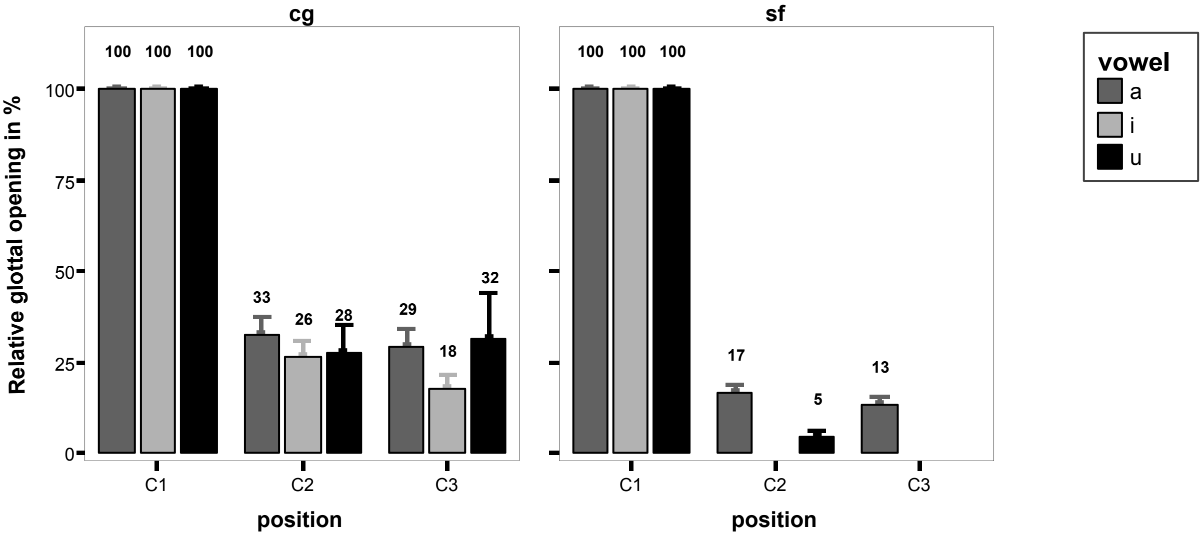

**Figure 4.5:** Barplots with means of relative amount of glottal opening for /t/ in %, error bars with +/-1 standard error; x = different positions (C1, C2, C3); left = CG, right = SF; dark grey bars = /a/ context, light grey bars = /i/-context, black bars = /u/- context

The normalised averaged peak values provide evidence that the amplitude of glottal opening in both post-stressed positions is considerably reduced in comparison to /t/ in the stressed position, if glottal opening occurs:

- For CG the reduction in glottal opening amplitude is approximately a third or below. Lowest relative values are found in /i/-context, highest in /a/ and /u/-contexts.
- For subject SF the reduction in glottal opening amplitude is even below a fifth. In /i/-context no values could be analysed. The largest values are found in /a/-context.

#### 4.1.2.3. THE DURATION OF OVERALL GLOTTAL OPENING IN /t/ PRODUCTION

A reduction in glottal opening amplitude could coincide with a temporal reduction of overall glottal opening duration GOP.

Results (see Figure 4.6) from overall glottal opening duration for CG provide evidence that there is a temporal reduction from the stressed to the post-stressed positions regarding the /i/ and /u/-context, but this temporal reduction is much smaller than the reduction in amplitude. The generally shorter durations in /a/-context might be influenced by measuring with the transillumination technique (tongue retraction in /a/-context causes epiglottal movement and a shadow onto the glottis). Thus, overall glottal opening duration could have similar values in /a/-context as in other vowel contexts. Both post-stressed position (C2 and C3) show similar averaged values.

Results for SF show a temporal reduction regarding overall glottal opening duration in /u/-context and a trend in /a/-context comparing GOP in the stressed position with the post-stressed positions. Again less influence was found for temporal characteristics than for the amplitude of glottal opening.

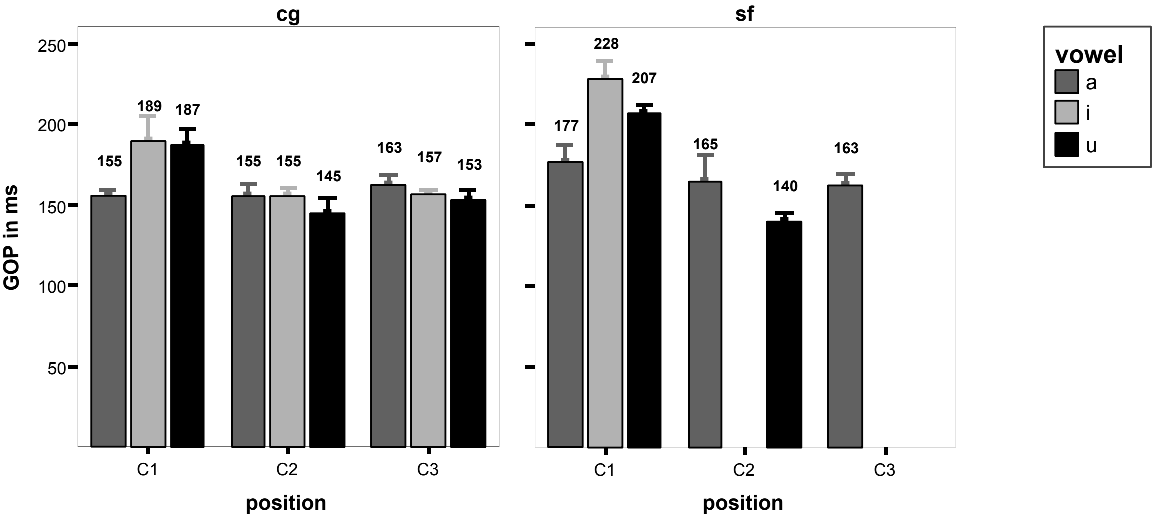

**Figure 4.6:** Barplots with means of overall glottal opening duration (GOP) in ms for /t/, error bars with +/-1 standard error; x = different positions (C1, C2, C3); left = CG, right = SF; dark grey bars = /a/ context, light grey bars = /i/-context, black bars= /u/- context

To summarise the results concerning laryngeal production mechanisms in alveolar stops:

1. Glottal abduction is an obligatory characteristic for /t/ in stressed position, but it occurs speaker- and vowel-dependently in post-stressed position.
2. Glottal abduction is not found for /d/, even in word final position where it could be expected due to the rule of final devoicing in German.
3. The amplitude of glottal opening is considerably reduced in both post-stressed positions compared to /t/ in stressed position, i.e. for CG to approximately 30%, for SF to 10-20% and completely missing for JD. A small trend is observed that glottal abduction for /t/ in /a/-context is reduced less than in /i/-context. This difference could be even larger when epiglottal movements in /a/-context could be neglected, since these artefacts can decrease the values of glottal opening amplitude.
4. The overall glottal opening duration in /t/ shows less influence than glottal opening amplitude comparing stressed with post-stressed positions.
5. The hypothesis that glottal opening would be the relevant articulatory correlate of the voicing contrast for alveolar stops fails, if all conditions are taken into account, but it could be confirmed when only the stressed position is considered.

### *4.1.3. Glottal abduction in fricative production*

The reader is reminded that alveolar fricatives in German do not show the same distributions as stops concerning the voicing contrast (see chapter 1). Alveolar fricatives investigated here are phonologically voiced in all positions with a following or preceding stressed **tense** vowel. Phonologically voiceless fricatives considered here are accompanied with a preceding **lax** vowel in the post-stressed word medial (= ambisyllabic) position. This work focuses on phonologically voiced fricatives, in all positions. Voiceless fricatives in lax vowel context are also taken into account, but only in the post-stressed ambisyllabic position (C2).

#### 4.1.3.1. THE OCCURRENCE OF GLOTTAL ABDUCTION IN FRICATIVE PRODUCTION

Table 4.2 shows findings regarding the occurrence of glottal abduction in alveolar fricative production.

**Table 4.2:** Number of tokens with glottal opening for /s, z/ from possible 5; split by subject (from top to bottom CG, JD, SF), by vowel (/a i u/) and by position (C1, C2, C3)

| | Stressed position | C1 | Post-stressed positions | C2 | | C3 |
|---|---|---|---|---|---|---|
| **CG pgg** | | /z/ | | /z/ | /s/ | /z/ |
| /a/ - /a/ | | 4 | | 5 | 5 | 5 |
| /i/ - /ɪ | | 5 | | 5 | 5 | 5 |
| /u/ - /ʊ/ | | 5 | | 5 | 5 | 5 |
| **JD pgg** | | | | | | |
| /a/ - /a/ | | 1 | | 2 | 5 | 2 |
| /i/ - /ɪ/ | | 0 | | 0 | 5 | 5 |
| /u/ - /ʊ/ | | 0 | | 0 | 5 | 4 |
| **SF pgg** | | | | | | |
| /a/ - /a/ | | 1 | | 5 | 5 | 5 |
| /i/ - /ɪ/ | | 5 | | 5 | 5 | 5 |
| /u/ - /ʊ/ | | 3 | | 5 | 5 | 5 |

Results for the occurrence of glottal abduction in alveolar fricatives provide evidence that:

- Most consistently glottal abduction was produced for /s/ in the post-stressed ambisyllabic position.
- For /z/ in none of the positions a consistent picture can be seen independent of speaker and vowel context. Most frequently glottal abduction occurs in word final position, in all cases for subjects CG and SF, but with some missing cases in /a/ and /u/ context for subject JD.

- Speaker CG realised 44 out of 45 possible cases with glottal opening. So far the data provide no evidence of a difference between phonologically voiced and voiceless fricatives.
- Speaker JD's results for /z/ are rather variable. In stressed position he produced /z/ with an adducted glottis, except from 1 token in /a/-context and he also realised the alveolar fricative in post-stressed word medial position without glottal opening, again with 2 exceptions in /a/-context. In word final position a weak glottal abduction for /z/ was found.
- Speaker SF realised all fricatives in both post-stressed conditions with glottal abduction. Some vowel dependent variations were found in the stressed position. In /i/-context glottal opening occurs in all cases, whereas in /a/ and /u/-context it occurs less frequently.

### 4.1.3.2. THE AMOUNT OF GLOTTAL OPENING IN FRICATIVE PRODUCTION

**Phonologically voiced fricatives in tense vowel context**: Figure 4.7 displays transillumination data from subject CG in /i/-context. The amount of glottal opening varies from token to token:

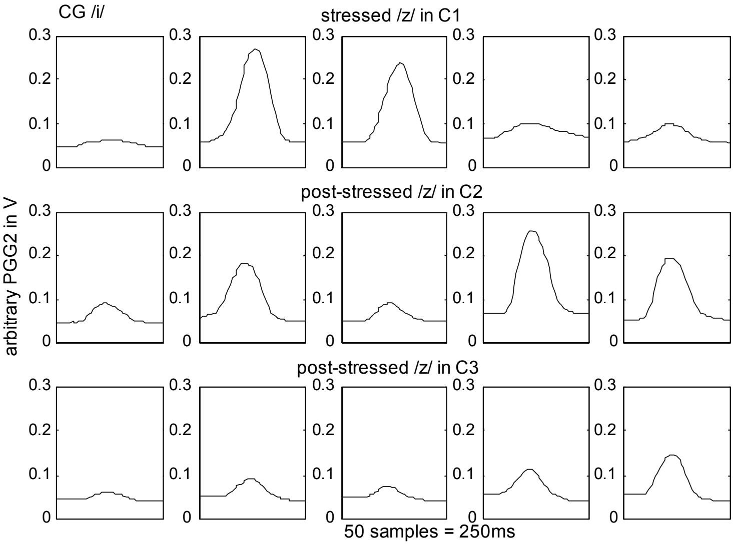

**Figure 4.7:** Glottal opening gestures for /z/ from PGG2, subject CG, all 5 repetitions, /i/ context, C1 to C3 (from top to bottom)

For the stressed position repetitions 1, 4 and 5 (first row) show the smallest amplitudes of glottal opening, and repetitions 2 and 3 the largest amplitudes. In the post-stressed word medial position, repetitions 2, 4 and 5 are the ones with the largest glottal abduction.

Variation was also found in the fiberoptic film images. Figure 4.7 shows some typical images of the glottis of CG's /z/ production. Except for the first upper image where the epiglottis masks the glottis in /a/-context (transillumination artefact due to tongue retraction), glottal abduction was found in all cases. It is rather difficult to make any further consistent conclusions concerning the amplitude of glottal opening in different positions based on these images. So far the main characteristics for CG's /z/ production is glottal abduction, but free variations concerning the degree of glottal opening are found.

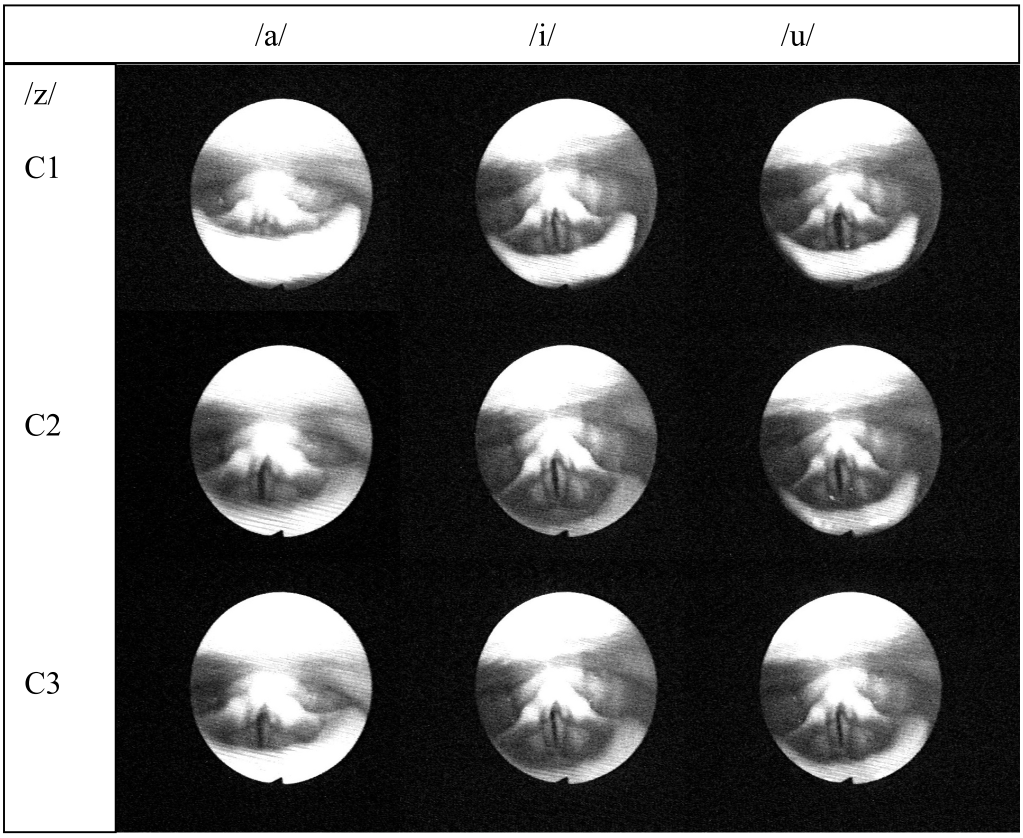

**Figure 4.8:** Example of subject CG's fiberoptic film images for /z/, from C1 to C3 (from top to bottom) and /a i u/-context (from left to right)

For subject JD the transillumination signal shows a weak glottal abduction for all /z/ in tense vowel environment. Figure 4.9 provides evidence that even though glottal opening in word final position is produced more frequently, it does not differ considerably from the amplitudes in the stressed and post-stressed positions. Note that in Figure 4.9 all possible tokens with some glottal abduction are plotted and the relevant positions and vowel-contexts are written in the left corners of each track. Baseline shifts can be seen, especially in word final position on the right side of peak glottal opening.

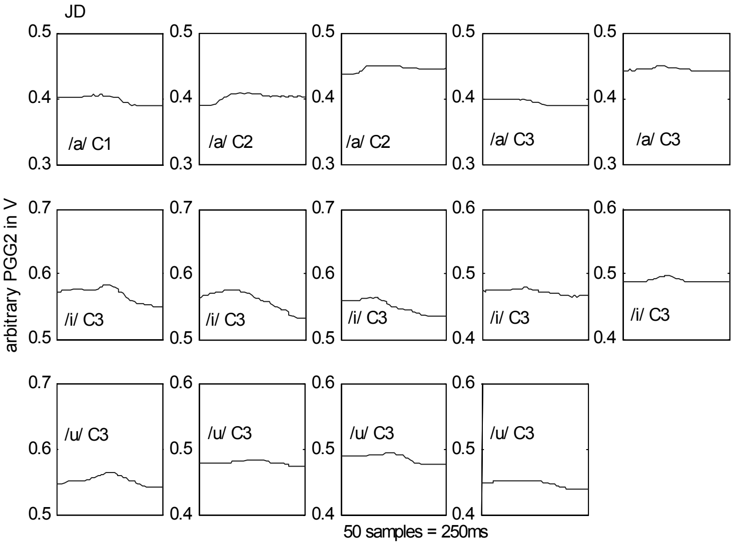

**Figure 4.9:** Glottal opening gestures for /z/ from PGG2, subject JD, /a i u/-context (from top to bottom), positions are written in subplots

Jessen (1999) explained baseline shifts in word final position as being caused by glottalisation processes associated with the vowel following word final obstruents. In the current study, approximately 95% of all vowels following word final obstruents /d t/ and /s/ are glottalised (which has been observed in the acoustic data), and support Jessen's findings.

Another explanation for the small rather asymmetric glottal opening is related to the specific aerodynamic conditions during fricative production. The voiced fricative /z/ might be partially voiceless since the high airflow rate which is necessary for fricative production could be responsible for the small glottal opening. At the end of the fricative the vocal folds are adducted again and probably tensed in order to produce glottalisation.

For speaker SF Figure 4.10 is plotted, reflecting the glottal opening for /z/ in /i/-context, again with some different characteristics than the ones found for CG and JD. Results are relatively consistent within the five repetitions of the relevant positions. The largest amount of glottal opening in /z/ was found in the post-stressed word medial position. Smaller amplitudes occur in word final position, and weak amplitudes in the stressed position. A similar relationship between the three positions is also observed in /a/ and /u/-contexts.

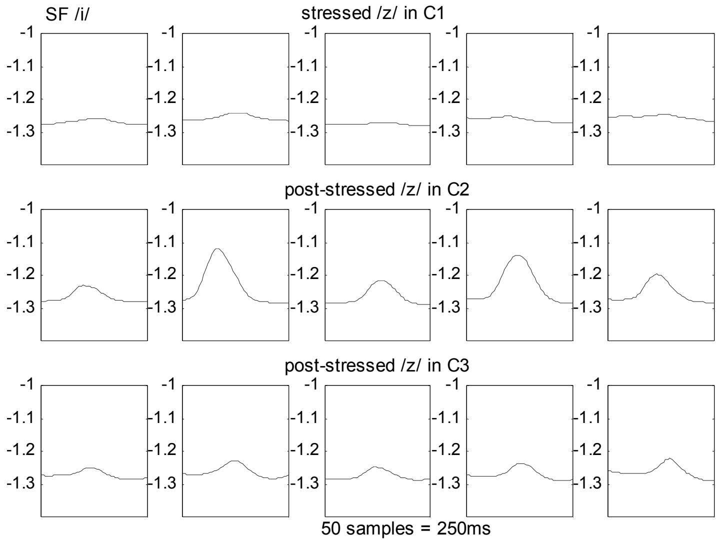

**Figure 4.10:** Glottal opening for /z/ from PGG2, subject SF, all 5 repetitions, /i/ context, C1 to C3 (from top to bottom)

Generally, the degree of glottal abduction in phonologically voiced alveolar fricative production shows a relatively high intra-speaker and inter-speaker variability.

## Phonologically voiceless fricatives in lax vowel context

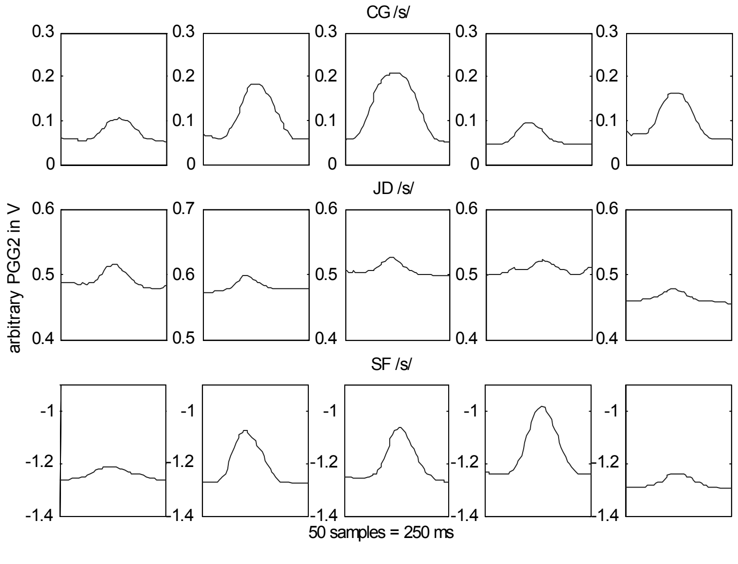

**Figure 4.11:** Example for glottal abduction in post-stressed ambisyllabic /s/ production in /ɪ/-context; all 5 repetitions (from left to right), speaker CG, JD, SF (from top to bottom); y-axis = PGG2 in V

Figure 4.11 exemplifies glottal abduction for /s/ in /ɪ/-context, post-stressed ambisyllabic position. Similar patterns are also found in /a/ and /u/-context. Speaker CG and SF show similar glottal opening amplitudes to /z/. Speaker JD shows some difference, but rather in the sense that glottal abduction occurs more often in /s/, but less frequently in /z/.

Fiberoptic film images (see Figure 4.12) from CG confirm his transillumination findings.

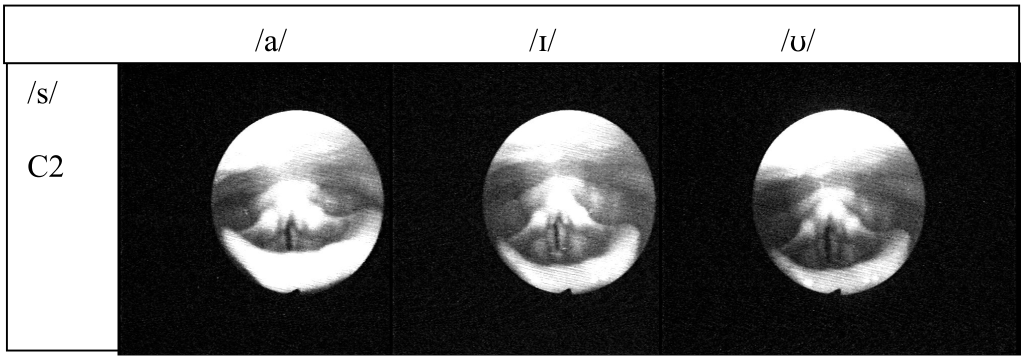

**Figure 4.12**: Example of subject CG's fiberoptic film images for /s/, for C2 in /a ɪ ʊ/-context (from left to right).

**Quantitative values for the amount of glottal opening in /s/ and /z/**: In order to get more detailed insights into speaker specific laryngeal production mechanisms the amount of glottal opening in several positions is considered. Again, it is mainly focused on the phonologically voiced fricatives, since they show several variations in their production and they are comparable to stops in tense vowel environment.

To study the amount of glottal opening quantitatively absolute glottal opening values with reference to the baseline are calculated. A similar normalisation procedure as for /t/ could not be applied here, since there was no general trend for all speakers that glottal opening in one position would be the largest in comparison to all other positions. Subject CG's glottal abduction amplitudes are largest in the stressed position for a few cases and for a few other cases they are largest in the post-stressed word medial position. Therefore absolute values of glottal abduction rather than relative values were taken into account.

The following Figure 4.13 displays glottal opening amplitudes with reference to the baseline. For the post-stressed C2 position /s/ and /z/ are taken into account whereas in the other positions only /z/ is considered.

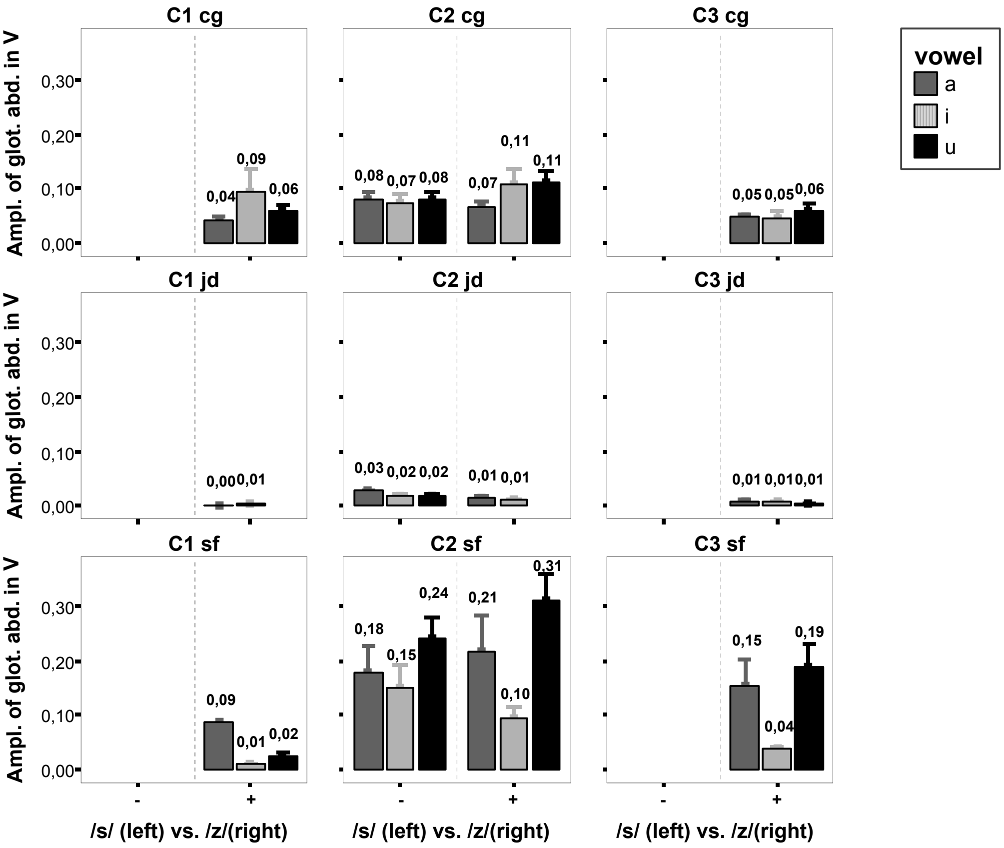

**Figure 4.13:** Barplots with means of amplitude of glottal abduction with reference to the baseline in V for /s/ and /z/, error bars with +/-1 standard error; - = lax vowel & /s/; + = tense vowel & /z/; different positions C1, C2, C3 (from left to right track); CG, JD, SF (from top to bottom); dark grey bars = /a/ context, light grey bars = /i ɪ/-context, black bars = /u ʊ/-context

Glottal abduction amplitudes are discussed with caution, because during the transillumination recording the distance between endoscope tip and glottis varies. Comparing /s/ and /z/ in the post-stressed C2 position with each other it can be seen that no relevant differences occur for the cases where glottal opening is produced.

Speaker SF’s results show a smaller amount of glottal opening for /s/ and /z/ in /i ɪ/-context compared to /a/ and /u ʊ/-contexts.

Considering the amount of glottal opening for phonological /z/ in different positions, a tendency for a smaller glottal opening amplitude can be seen in the stressed position and in the word final position for CG compared to the post-stressed word medial position.

For speaker SF the weakest glottal abduction amplitude for /z/ was produced in stressed position, the largest abduction in post-stressed word medial position and some intermediate glottal opening amplitude can be reported in word final position.

For speaker JD no reliable differences were found between the three positions, since all amplitudes in /z/ are weak.

#### 4.1.3.3. THE OVERALL DURATION OF GLOTTAL OPENING IN /z/ AND /s/ PRODUCTION

Temporal correlates of glottal abduction could explain differences in the amount of glottal opening, i.e. a larger glottal abduction could be associated with a longer overall glottal opening duration. Barplots with overall glottal opening duration (GOP) for /z/ in different positions are displayed in Figure 4.14.

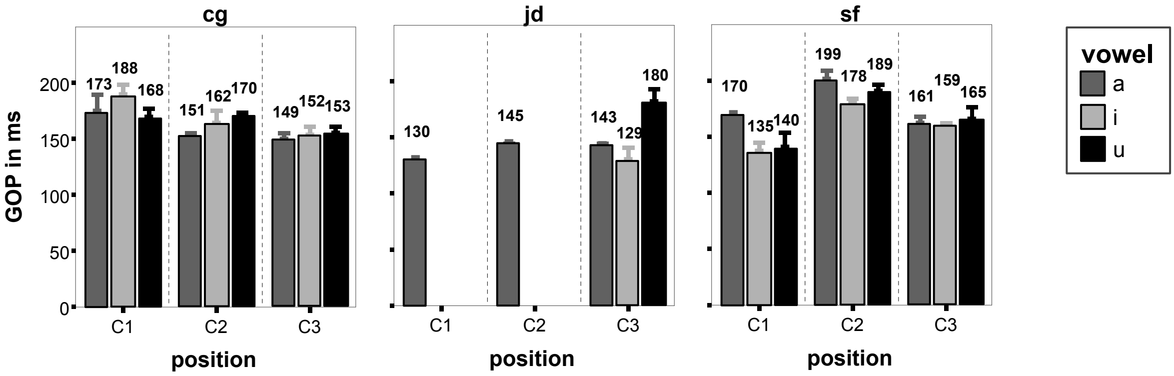

**Figure 4.14:** Barplots with means of GOP in ms for /z/, error bars with +/-1 standard error; x = different positions (C1, C2, C3); left = CG, middle = JD, right = SF; dark grey bars = /a/ context, light grey bars = /i/-context, black bars= /u/- context

Figure 4.14 exhibits a weak reduction of overall glottal opening duration in /z/ for speaker CG from the stressed to both post-stressed positions. For speaker JD GOP is of little use, since only a few tokens are included here, except from the word final position. Speaker SF produced the relatively high glottal opening amplitude in the post-stressed word medial position together with a longer overall glottal opening duration compared to the shorter duration and smaller glottal abduction amplitude in C3, and even shorter and smaller glottal abducttion in C1.

Comparing overall glottal opening duration for /z/ with values for /s/ in post-stressed word medial position similar values can be seen for CG and SF (Figure 4.15). For JD differences are unclear, since glottal opening was less frequently produced in /z/.

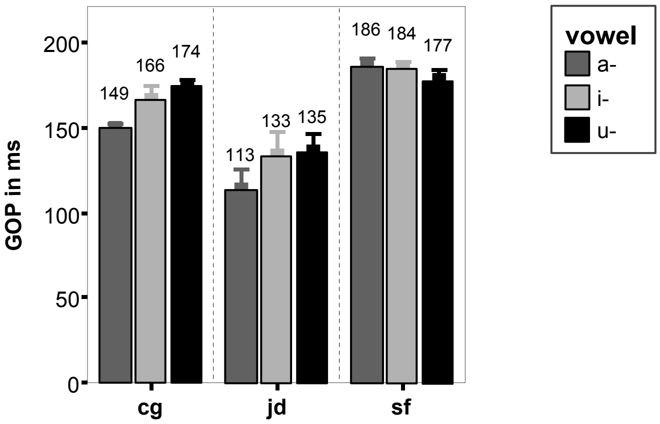

**Figure 4.15:** Barplots with means of overall glottal opening duration in ms for /s/, error bars with +/-1 standard error; x = different speaker (CG, JD, SF); dark grey bars = /a/ context, light grey bars = /ɪ/-context, black bars= /ʊ/- context

## Summary

The following results concerning laryngeal production mechanisms in alveolar fricatives can be summarised:

1. Glottal abduction occurs always for phonologically voiceless fricatives in lax vowel context (here they are analysed in the post-stressed ambi-syllabic position).
2. For phonologically voiced fricatives glottal abduction occurs most frequently in word final position, but with a rather weak amplitude for all the speakers. The rule of final devoicing applies for German concerning this position.
3. A trend is found that variations in amplitude of glottal abduction for /z/ depend rather on speaker than on position:
   - Surprisingly, speaker CG produced a considerable glottal aperture in the stressed position. Glottal opening was also found for CG in all other positions. These results can be interpreted in terms of CG’s south German heritage, a region were differences between /z/ and /s/ are often not realised.
   - JD’s transillumination data for the fricatives are generally very weak. Most consistently he produced a small amount of glottal opening in word final /z/.
   - SF realised large glottal openings in the post-stressed word medial position compared to smaller amplitudes in word final position and weak amplitudes in the stressed position.

4. Temporal aspects play a role in particular in SF's /z/ production: For this subject larger glottal apertures coincided with a longer glottal opening duration and vice versa.
5. The hypothesis that glottal abduction could be a reliable articulatory correlate of the voicing contrast in German fails for most of the results reported here, since the phonologically voiced alveolar fricatives are often produced with some amount of glottal opening.

Stops and fricatives differ from each other with respect to occurrence of glottal opening in different position and to reduction phenomena:

- In stop production only the voiceless show glottal abduction, whereas in fricatives glottal abduction was frequently produced in the voiced ones.
- Most consistent differences in stop production occur in the stressed position, whereas this is not the case for fricatives, since phonologically only voiced /z/ occurs in German vocabulary. In the post-stressed word medial position speaker dependent variations were found. Speakers CG and SF do not show reliable differences in this position regarding the amount of glottal opening and the duration of overall glottal opening, whereas for JD the amount of glottal abduction is more distinctive.

Besides glottal opening, the articulatory timing between laryngeal and oral gestures should be relevant for the voicing contrast. This will be discussed in the following section.

## 4.2. Results for laryngeal-oral co-ordination

### *4.2.1. Laryngeal-oral co-ordination in stop production*

In this section data from different experiments are considered:

- acoustic data (noise duration) from both experiments in order to increase the number of tokens for further statistical analysis, and
- transillumination data to investigate laryngeal gestures in relation to oral gestures, derived from EPG and acoustics.

In a first step noise duration was computed in order to discuss laryngeal-oral co-ordination. Noise duration is equated with VOT or aspiration duration. The two last measurements have been frequently used in other studies (e.g. Lisker and Abramson 1964, Docherty 1992, Cho and Ladefoged 1999, Scobbie in press). VOT or aspiration duration are relatively easy to label in the acoustic signal compared to articulatory measurements. In this study noise duration is defined as the duration between oral release (the burst in the acoustic signal) and the offset of high frequency noise. When VOT is longer than approximately 20ms (Stevens and Klatt 1974) then it is likely that high frequency noise was produced with an open glottis. Such stops are likely to be perceived as being aspirated. When noise duration is rather short, stops are often perceived as unaspirated.

It should be noted here that noise duration is adopted, a measure very similar to VOT, but it differs with respect to the offset (see 3.6.1.). Using the high frequency criterion proved to be an accurate measure to label the offset for noise duration even in noisy acoustic speech material as it occurs in this study. The offsets in VOT computations are typically either the beginning of periodicity (VOT after Lisker and Abramson) or the second formant onset (VOT after Klatt, for review see Sock 1998). By definition VOT characterises laryngeal-oral co-ordination, since it is the duration between the burst (corresponding to the supralaryngeal articulation) and the onset of vocal fold vibration (corresponding to the laryngeal level). The high frequency noise adopted here, however, does not necessarily reveal laryngeal-oral timing. The noise could be produced by a vocal tract constriction (supralaryngeal) as well as by a glottal constriction (laryngeal).

Results in Figure 4.16 show barplots with means of noise duration, split by consonant (/d/ versus /t/), speaker (from top to bottom), vowel (different colours and barfillings), and position (from left to right). Regarding different positions, there is evidence that /d/ and /t/ differ considerably in the stressed position, whereas differences become weak or do not occur in both post-stressed positions. Differences between /d/ and /t/ become less distinctive in terms of noise duration mainly due to a decrease for /t/ from C1 to C2 to C3[31]. In word final position the rule of final devoicing applies in German. Complete neutralisation of noise duration was found for JD and SF, but not for CG.

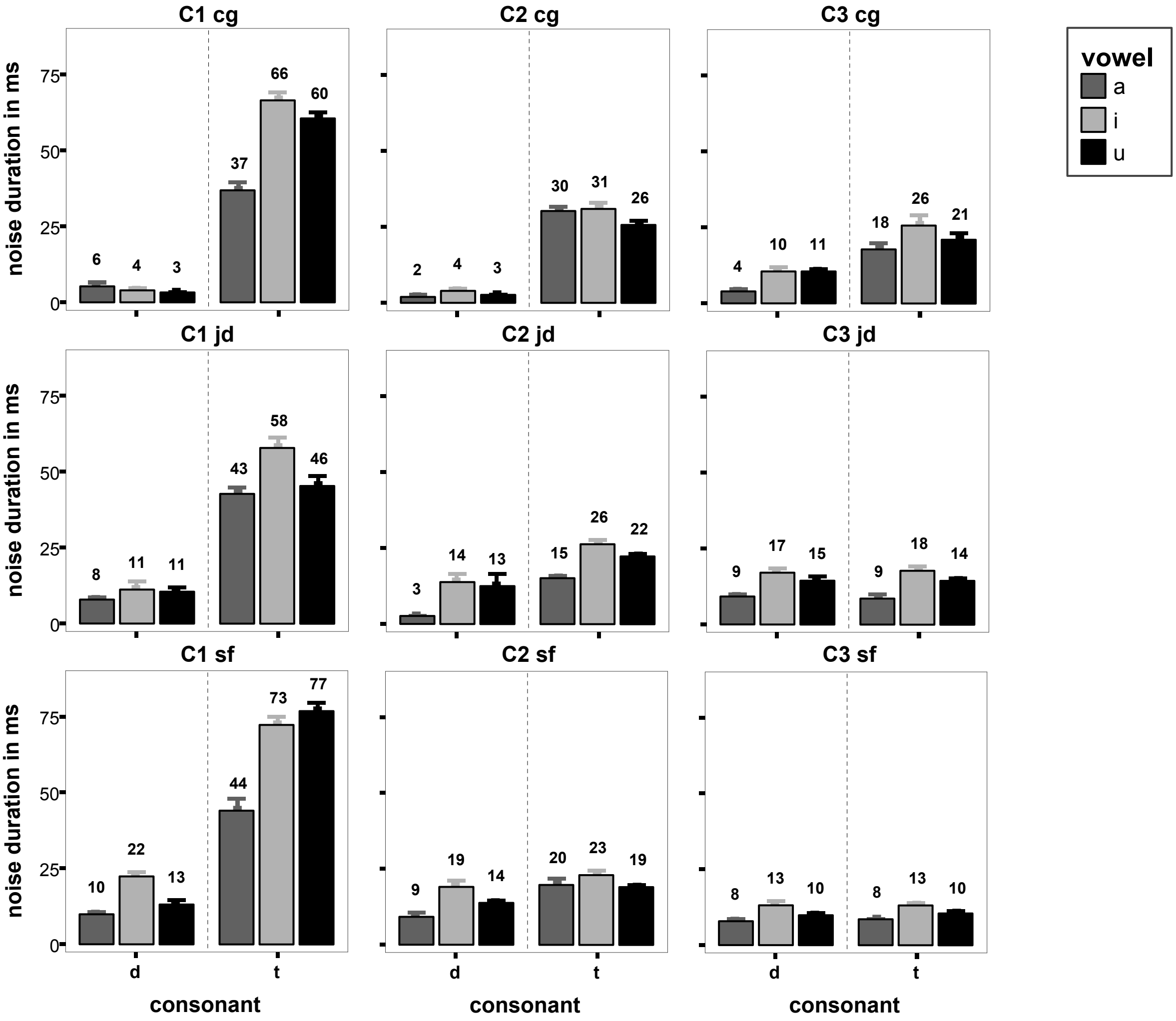

**Figure 4.16:** Bar plots showing means of noise duration in ms; error bars = +/- 1 standard error; /d/ = left 3 bars on the x-axis, /t/ = right 3 bars; subject CG, JD, SF (from top to bottom track), positions C1, C2, C3 (from left to right track); different vowel contexts (/a/ = dark grey bars, /i/ = light grey bars, /u/ = black bars)

Another finding is that vowels surrounding the consonant influence noise duration. Noise duration for /t/ is often longer in /i/-context than in /a/-context. A two-way ANOVA was computed in order to rule out the possibility of significant interaction effects between vowel and consonant (see Table 4.3). Interactions between vowel and consonant are found for the stressed position in CG's and SF's data. Results for noise duration are significantly shorter in /d/

---

[31] If only noise duration is taken into account one could ask why the phonological /d/ in word final position would become /t/ rather than phonological /t/ becomes /d/, since the productions of /t/ shows more variations and the /d/'s are relatively stable.

compared to /t/. Differences are also found for noise duration in /a/- vs. /u/-context and /a/- versus /i/-context, but no differences or weak differences occur in /i/ vs. /u/-context (SF, CG). The significant interaction effect is caused by vowel height.

**Table 4.3:** Two-way ANOVAs for noise duration comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 972.1*** | 387.8*** | 644.5*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 26.4*** | 7.7*** | 43.6*** |
| | **CON*VOW** | 2 | 36.1*** | 3.8 | 19.0*** |
| **C2** | **CON** (/d/-/t/) | 1 | 677.5*** | 38.9*** | 42.5*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 4.3 | 11.3*** | 14.4*** |
| | **CON*VOW** | 2 | 2.3 | 0.3 | 4.7* |
| **C3** | **CON** (/d/-/t/) | 1 | 80.6*** | 0.0 | 0.4 |
| | **VOW** (/a/-/i/-/u/) | 2 | 8.6*** | 36.4*** | 29.0*** |
| | **CON*VOW** | 2 | 1.1 | 0.1 | 0.1 |

In the post-stressed word medial position no significant interactions are found, but two main effects for JD and SF (VOW and CON) and one main effect for CG (CON).

In word final position vowel environment has a significant main effect for all subjects (except from CG, who exhibits a main effect of CON). Looking at single contrasts, vowel height turns out again as a significant factor for JD and CG. Concerning the /d/ vs. /t/ contrast, noise duration for /d/ in JD’s and SF’s results is completely neutralised to /t/, whereas in CG’s results no neutralisation was found. Data for CG are in agreement with results from Piroth and Janker (2004), who describe incomplete neutralisation for subjects from Bavaria. It is assumed that similar findings are likely to occur in a region next to Bavaria (CG grew up at Lake Constance). Thus, findings for CG would support Piroth and Janker’s results.

In a next step, articulatory data demonstrating laryngeal-oral co-ordination are analysed. All cases with glottal abduction are averaged in amplitude and plotted together with:

- closure onset defined on EPG data (see 3.6.4.) since EPG data provide an adequate articulatory information of the moment where the tongue tip starts to touch the palate,
- and oral release defined as the onset of the acoustic burst signal. Since the acoustic signal has a much higher sampling frequency than the EPG data, the acoustic signal was taken for determining oral release. The acoustic burst is used as a line-up point in Figure 4.17.

Figure 4.17 displays laryngeal-oral timing for all speakers' /t/ production in the stressed position. Closure onsets and offsets are marked by vertical lines and glottal aperture can be identified by the rather triangular shaped curves. The y-axis corresponds to glottal opening values at PGG2 in voltage (negative estimates are possible, since it is an arbitrary measure of glottal opening) and the x-axis corresponds to time. The burst in real time is set to zero and all other landmarks are calculated with reference to the burst.

Results for laryngeal-oral timing for stressed /t/ provide evidence that the beginning of glottal aperture is tightly coupled with the beginning of closure onset as tongue-alveolar contact.

For speaker SF both temporal events are coupled too, but glottal abduction starts slightly earlier relative to oral closure onset. Peak glottal opening was produced shortly after the burst (all subjects, stressed position). The second articulatory co-ordination characteristic, the timing between peak glottal opening and oral release (burst) shows a close coupling between laryngeal and supralaryngeal events too. Since the amount of glottal opening at oral release is close to its peak with a large amplitude, it can be concluded that all /t/ in stressed position are aspirated[32]. The shorter glottal closing gesture during /a/-context is likely a mixture of an artefact of the transillumination recording (epiglottal movement) and a shorter overall glottal opening duration (see Figure 4.17).

---

[32] It should be noted following Dixit (1987, p. 87): "In fact, the glottis is said to be a necessary condition for the production of aspiration (Chomsky and Halle, 1968). However, it is not a sufficient condition; in addition the supraglottal vocal tract must be unobstructed. Unless the supraglottal vocal tract is unobstructed while the glottis is open, higher than normal airflow rate cannot be generated, which is so important in the generation of aspiration."

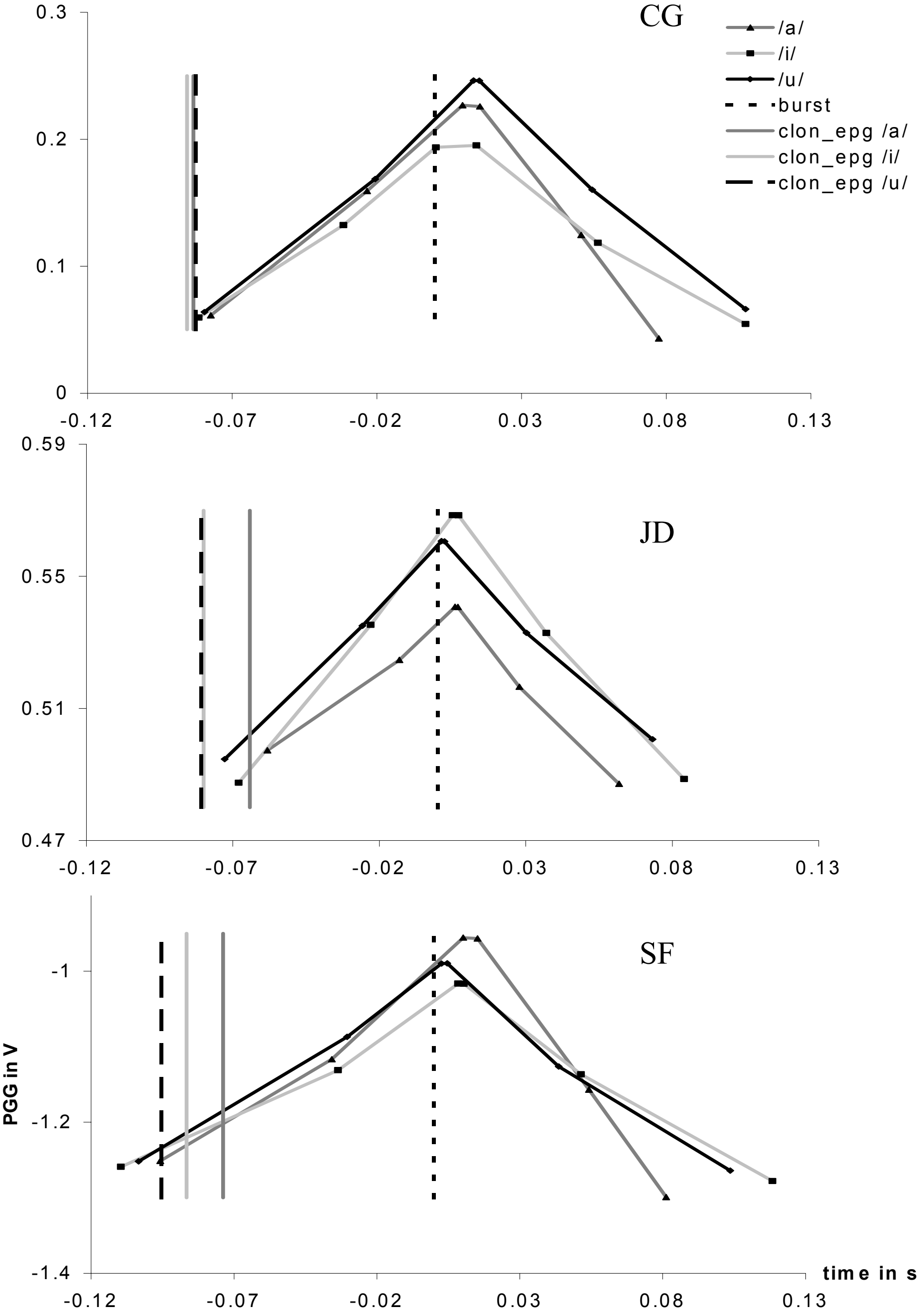

**Figure 4.17:** Means of glottal opening for /t/ in C1 (dark grey curve = /a/- context, light grey curve = /i/- context, black curve = /u/- context) in relation to closure onset (clon_epg; vertical lines: dark grey /a/- context, light grey = /i/- context, dashed line = /u/- context ) and burst as line up point set to 0 in time (dotted vertical line); from top to bottom: CG, JD, SF; x-axis = time in s; y-axis = PGG2 in V

**Quantitative values for laryngeal-oral co-ordination**: For a general overview about laryngeal-oral timing and for a more quantitative evaluation several time intervals were computed. The duration between onset of laryngeal opening and oral closure COOR_on as well as the duration between oral release and peak glottal opening COOR_peak were calculated:

COOR_on = t_clon_epg - t_opg_pgg
COOR_peak = t_peak_pgg - t_burst.

Values for COOR_on are shown in Figure 4.18. Negative values correspond to an earlier closure onset relative to glottal opening onset and positive values to a later closure onset.

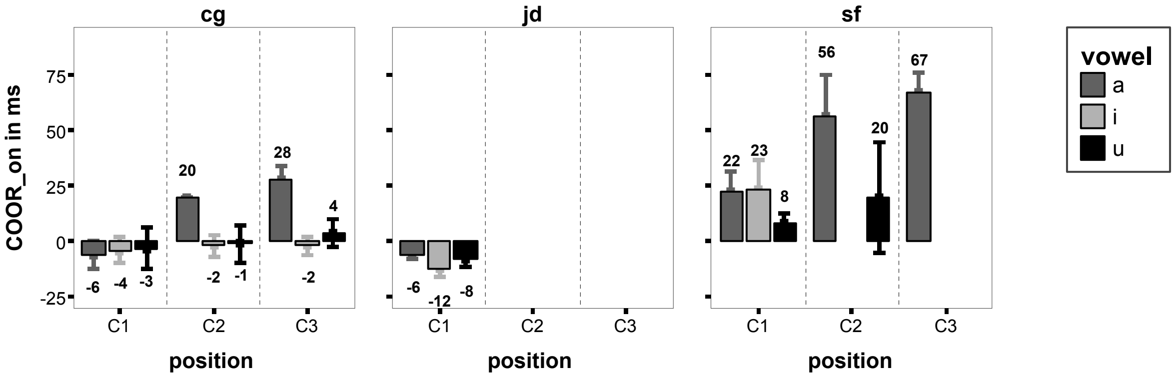

**Figure 4.18 :** Bar plots with means of COOR_on in ms for /t/ with +/- 1 standard error; split by subjects (CG, JD, SF from left to right track), vowel contexts (dark grey bars = /a/-context, light grey bars = /i/-context, black bars =/u/-context) and position (C1, C2, C3 on the x-axis)

Generally, results in Figure 4.18 provide evidence that for stressed /t/ (C1) oral closure onset and glottal opening onset are co-ordinated within an averaged range of -6 to 23ms. Speakers CG and JD show comparable values, but for SF closure onset started later with respect to glottal abduction.

In the post-stressed positions results become more speaker- and vowel-dependent, i.e. for CG laryngeal-oral co-ordination varies with respect to vowel height. He realised post-stressed /t/ in /i/ and /u/-context tightly coupled as in stressed position, but with a delay of oral closure onset in /a/-context.

For speaker SF a delay of oral closure onset with respect to laryngeal abduction onset can be seen in Figure 4.18. Latencies show a weak difference between the post-stressed word medial and word final position. The delay of closure onset is longer in word final position. Vowel context has a similar influence as for subject CG, with a longer delay of oral closure onset in /a/-context compared to /u/-context.

Figure 4.19 depicts laryngeal-oral co-ordination at COOR_peak. Positive values hold for peak glottal opening after the burst and negative values for peak glottal opening before oral release. The first is associated with aspirated stops, and the second with unaspirated stops. Generally the distinction between aspirated and unaspirated is rather continuous than abrupt, i.e. some tokens show a co-ordination which is associated with an aspirated stop, others with an unaspirated stops and a few which could be both, aspirated or unaspirated (e.g. SF, C2, /a/-context). However, for all /t/ in the stressed position peak glottal opening was realised between 5 to 16ms after oral release. Summarising these results together with results from the amount of glottal opening in the stressed position there is no doubt that these tokens are aspirated.

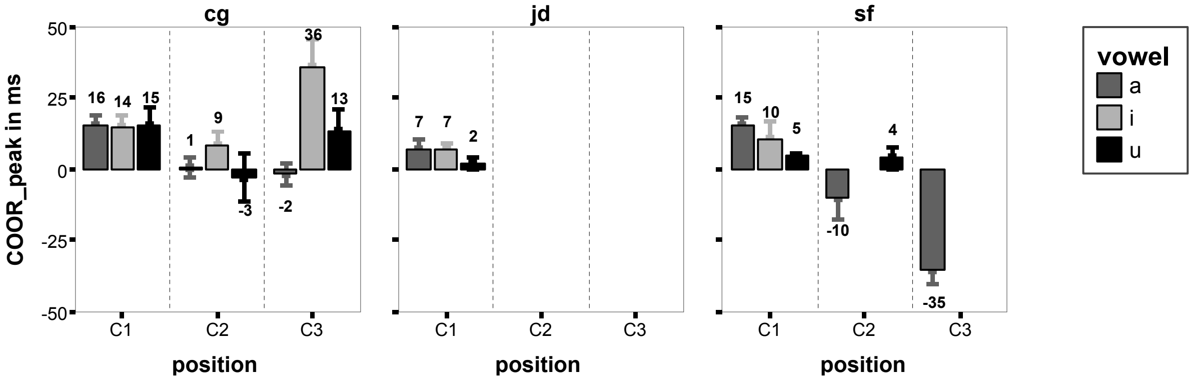

**Figure 4.19**: Bar plots with means of COOR_peak in ms for /t/ with +/- 1 standard error; split by subjects (CG, JD, SF from left to right track), vowel contexts (dark grey bars = /a/-context, light grey bars = /i/-context, black bars =/u/-context) and position (C1, C2, C3 on the x-axis)

Speaker dependent variations can be seen in both post-stressed positions. CG produced /t/ in post-stressed word medial position with a close co-ordination between peak glottal opening and oral release too, but this does not hold for word final position. In word final position a late peak glottal opening was found at least in high vowel context. Such co-ordination is rather known for fricatives. However, some effects with respect to the glottalisation of the following word initial vowel could be involved too.

Subject SF's results exhibit positive values in /u/-context, but negative values in /a/-context. The small amount of glottal abduction and the early peak glottal opening with respect to oral release provide evidence that /t/ in /u/-context was realised as an unaspirated stop. As it was pointed out earlier, tokens for /i/-context have a status in between, i.e. between aspirated and unaspirated. Their laryngeal-oral co-ordination tends to be more like aspirated stops, but on the other hand there is not a sufficient amount of glottal abduction.

## Summary

To summarise laryngeal-oral co-ordination in /t/ production (/d/ did not show glottal abduction and hence laryngeal-oral co-ordination could not be observed):

1. The beginning of laryngeal abduction is tightly coupled with the onset of oral closure. Furthermore, glottal opening reached its peak shortly after oral release (burst) in the stressed position. These results of close co-

ordination between laryngeal and supralaryngeal articulators - together with the large glottal abduction in this position - should be organised in order to produce the relevant perceptual output, i.e. aspiration noise after oral release. Laryngeal-oral co-ordination is most consistent in the stressed position and exhibits similar patterns for all subjects.

2. In the post-stressed positions laryngeal-oral timing varies across speakers:
    - Speaker CG shows timing patterns comparable with /t/ in stressed position. However a few tokens were realised with a late glottal opening which is more common in fricative production.
    - Speaker SF's patterns vary with position. Comparing stressed with both post-stressed positions, results differ with respect to both COOR_on and COOR_peak. In both post-stressed positions the laryngeal opening gesture starts much earlier in /a/-context than the onset of oral closure and the same is true for the small glottal peak with reference to oral release. The glottis is almost closed at oral release which coincides with the idea of laryngeal-oral co-ordination of unaspirated stops. No long aspiration noise can be produced at the laryngeal level with an almost closed glottis.
    - For subject JD no glottal aperture was found in the post-stressed condition and hence, no comparison of different positions is possible.

### *4.2.2. Laryngeal-oral co-ordination in fricative production*

For fricative production transillumination data with their corresponding acoustic data are considered. A similar measurement to the acoustic noise duration in stop production was difficult to compute, since it is rather challenging to separate noise produced due to the constriction at the glottis or due to the vocal tract in the acoustic signal.

Thus, transillumination data are used to plot averaged glottal opening per subject, position and vowel context in relation to acoustically defined frication on- and offsets (t_fricon and t_fricoff). The onset of glottal opening is taken as a line-up point.

#### 4.2.2.1. LARYNGEAL-ORAL CO-ORDINATION IN PHONOLOGICALLY VOICED FRICATIVES

Since glottal aperture varies speaker dependently, the next section will be divided in results for different speakers. Again, only tokens with at least some amount of glottal opening are included.

1. CG: Figure 4.20 displays laryngeal-oral timing for subject CG in all positions. Independent of the relevant position in which the fricative was produced, laryngeal-oral co-ordination shows the following patterns:
    - Glottal opening onset starts earlier than acoustically defined frication onset. Note that in Figure 4.20 frication onsets in /a/ and /u/-context (word final position) overlap and only one vertical line can be seen.
    - The glottis reaches its maximal width before the middle of the way through the frication.
    - Afterwards a delay of glottal abduction offset with respect to frication offset was found, i.e. the glottis is still slightly open at frication offset. These results are relatively constant through all the tokens for this subject. In word final position frication onset (t_fricon) occurs slightly before glottal opening onset in /i/-context. This result could be due to the high tongue and jaw position for /i/, which could have caused some noise already during the vowel and hence, the t_fricon was labelled before glottal abduction onset.

2. SF: In Figure 4.21 results from speaker SF are shown. The findings are generally in agreement with the ones from CG:
    - Glottal opening onset occurs before frication onset.
    - The glottis reaches its peak amplitude approximately in two thirds of the way through frication.
    - The glottis closes with a delay compared to frication offset. This is true for all positions, even for those with a weak glottal opening.
    - The plotted curve in /a/-context stressed position exhibits a relatively long increasing period until peak glottal abduction is reached. This result can be explained with the beginning of glottal opening, which starts already in the devoiced velar stop during the prefix ge.

3. JD: Data from this subject are not averaged and plotted, since the amplitude of glottal aperture is so weak. However, looking at single tokens a similar oral-laryngeal co-ordination was found as described for the previous speakers.

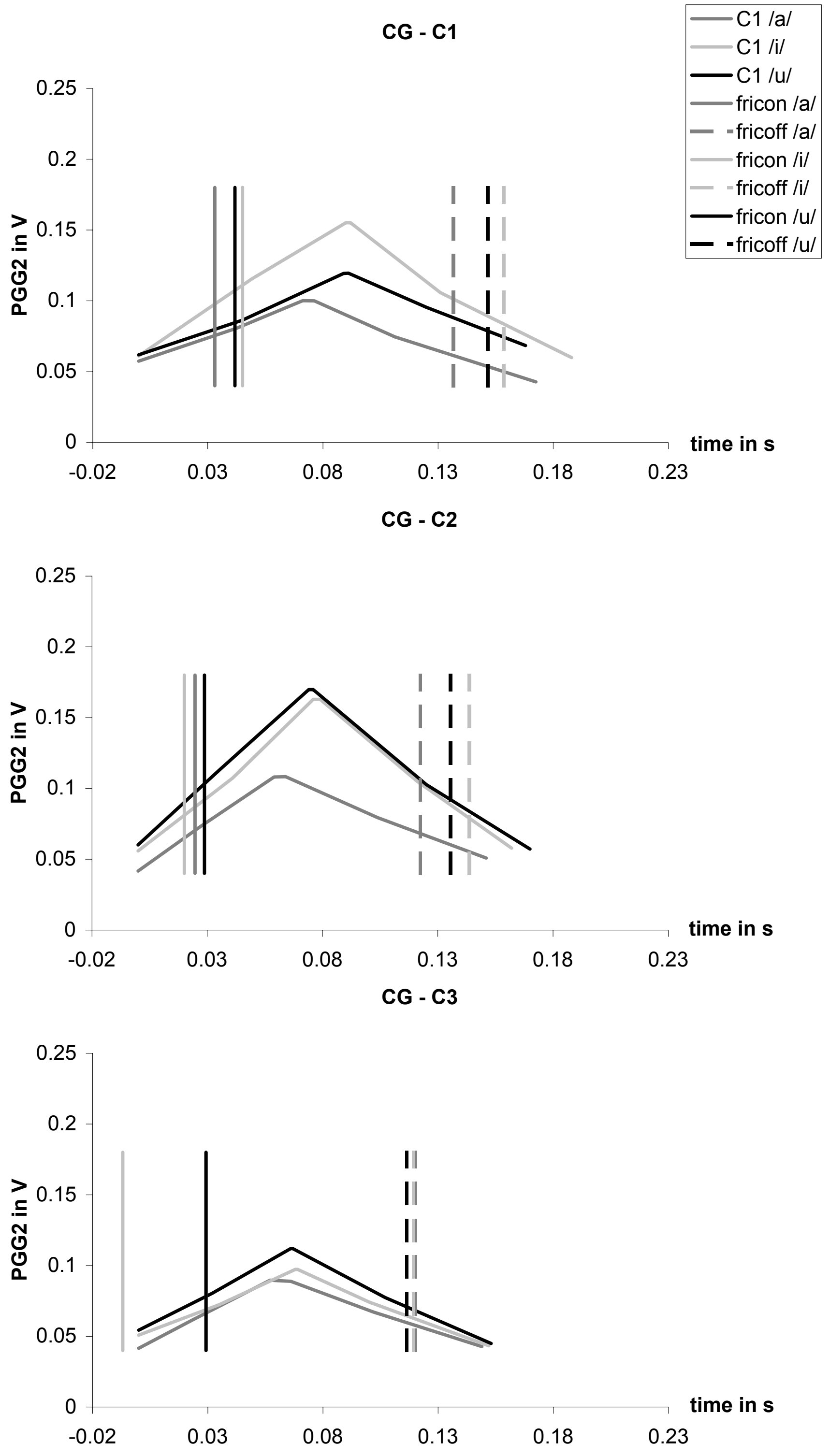

**Figure 4.20:** Means of glottal opening (dark grey curve = /a/- context, light grey curve = /i/- context, black curve = /u/- context) in C1, C2, C3 (from top to bottom) in relation to frication onset (t_fricon; vertical lines: dark grey /a/- context, light grey = /i/- context, black line = /u/- context ) and frication offset (t_fricoff dashed lines), subject CG; x-axis = time in s; y-axis = PGG2 in V.

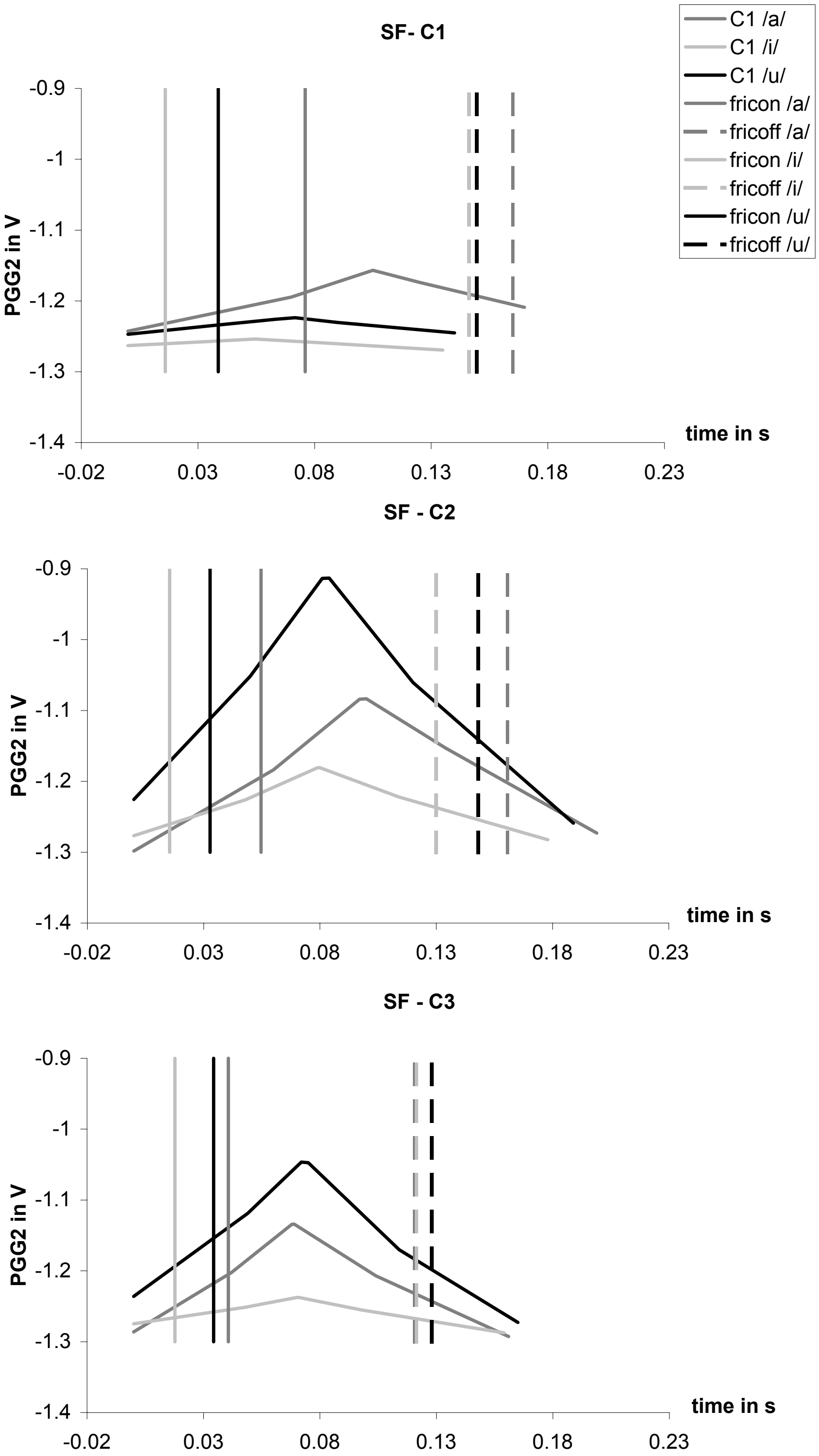

**Figure 4.21:** Means of glottal opening for /z/ (dark grey curve = /a/-context, light grey curve = /i/-context, black curve = /u/-context) in C1, C2, C3 (from top to bottom) in relation to frication onset (t_fricon; vertical lines: dark grey /a/-context, light grey = /i/- context, black line = /u/- context ) and frication offset (t_fricoff dashed lines), subject SF; x-axis = time in s; y-axis = PGG2 in V.

### 4.2.2.2. LARYNGEAL-ORAL CO-ORDINATION IN PHONOLOGICALLY VOICELESS FRICATIVES

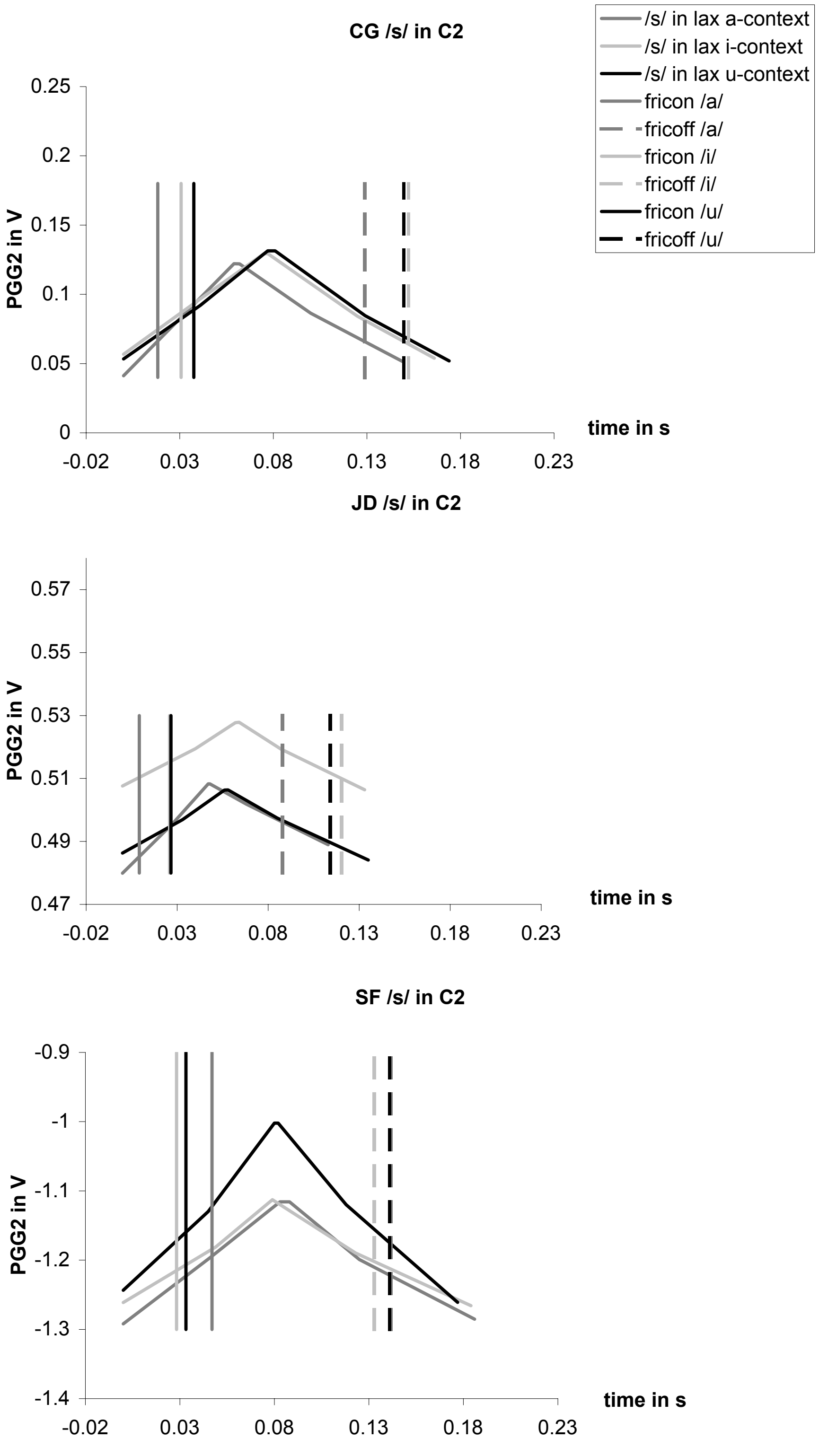

**Figure 4.22:** Means of glottal opening for /s/ (dark grey curve = /a/-context, light grey curve = /i/-context, black curve = /u/-context) in C2 in relation to frication onset (t_fricon; vertical lines: dark grey /a/-context, light grey = /i/- context, black line = /u/- context ) and frication offset (t_fricoff dashed lines), subjects CG, JD, SF (top to bottom); x-axis = time in s; y-axis = PGG2 in V

Similar to the findings in the previous section, laryngeal-oral co-ordination is plotted for /s/ in lax vowel context in the relevant post-stressed ambisyllabic position taken into account here.

Figure 4.22 provides evidence that in all speakers' /s/ production the onset of glottal abduction occurs earlier than the onset of frication noise. Glottal abduction offset was produced with a delay compared to offset of frication. So far no differences concerning laryngeal-oral co-ordination in /s/ and /z/ productions can be seen as long as glottal abduction is produced.

### 4.2.2.3. QUANTITATIVE VALUES FOR LARYNGEAL-ORAL CO-ORDINATION

In order to provide more quantitative information about laryngeal-oral timing in fricative production, the following durations were computed:

- COOR_on fric = t_fricon - t_opg_pgg as the duration between frication and glottal opening onset,
- COOR_peak fric = ((t_peak_pgg - t_fricon)*100)/(t_fricoff - t_fricon) as the relative duration of the glottal peak occurrence during the whole frication interval, which was set to 100%,
- and COOR_off fric= t_clg_off_pgg – t_fricoff.

First the phonologically voiced fricatives with glottal abduction are considered and thereafter phonologically voiceless fricatives.

**COOR_on fric in /z/**: Figure 4.23 provides an overview about the delay of frication onset with respect to glottal opening onset in /z/. Positive values provide evidence about a later onset of supralaryngeal constriction. Since it is likely that the results are influenced by acoustic labelling and frication on- and offsets are rather difficult segmentation criteria (±10ms might be a good estimate for the error due to subjective labelling), these findings are described in a broader sense.

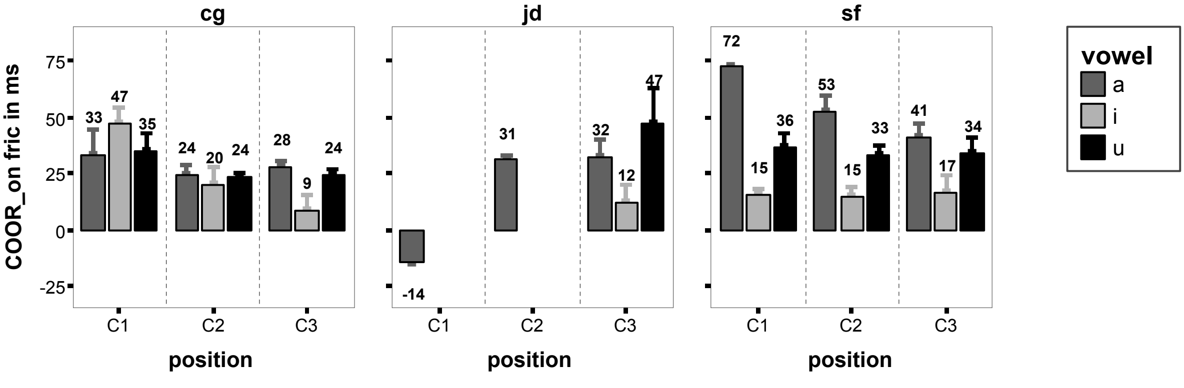

**Figure 4.23:** Bar plots with means of COOR_on fric in ms for /z/ with +/- 1 standard error; split by subjects (CG, JD, SF from left to right track), vowel contexts (dark grey bars = /a/-context, light grey bars = /i/-context, black bars =/u/-context) and position (C1, C2, C3 on the x-axis)

The most consistent result in Figure 4.23 is that frication onset starts approximately between 9 and 53ms after the onset of glottal abduction.

In speaker CG's laryngeal-oral co-ordination a small trend can be seen regarding stressed vs. post-stressed position in high vowel context. The first shows a longer delay of oral closure onset relative to glottal abduction onset than the latter.

JD's data are most variable which could be caused by the limits of labelling weak glottal opening amplitudes on the velocity signal.

Speaker SF's barplots show a clear vowel effect, i.e. frication onset was produced with a delay relative to glottal abduction onset in the following order: /a/>/u/>/i/-context.

**COOR_on fric in /s/**: Similar values are also found for laryngeal-oral co-ordination onset during /s/ production in the lax vowel environment (see Figure 4.24) as in /z/.

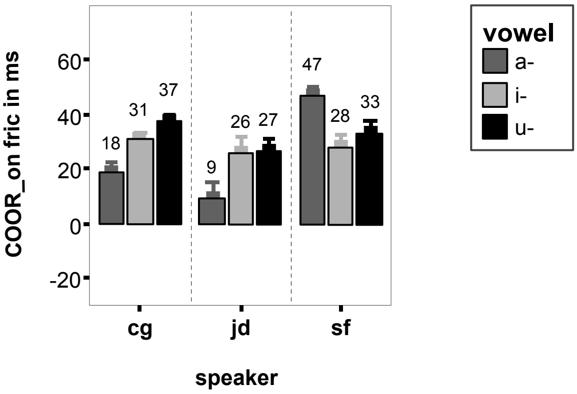

**Figure 4.24:** Bar plots with means of COOR_on fric in ms for /s/ at C2 with +/- 1 standard error; split by subjects (CG, JD, SF from left to right x-axis), vowel contexts (dark grey bars = /a/-context, light grey bars = /ɪ/-context, black bars =/ʊ/-context)

They range on average between 9 - 47ms, i.e. a delay of frication onset occurs. Vowel context shows comparable influences/trends as described previously, i.e. for speaker SF the delay is long in /a/-context and it is shorter in /ɪ/ and /ʊ/-contexts. For speakers CG and JD it seems to be vice versa, /s/ in /a/-context exhibits the shortest delay compared to /ɪ/ and /ʊ/-context.

**COOR_peak fric in /z/**: In a next step COOR_peak fric was calculated, an estimate where peak glottal opening is related to the frication interval. Figure 4.25 displays the equivalent barplots for the analysis.

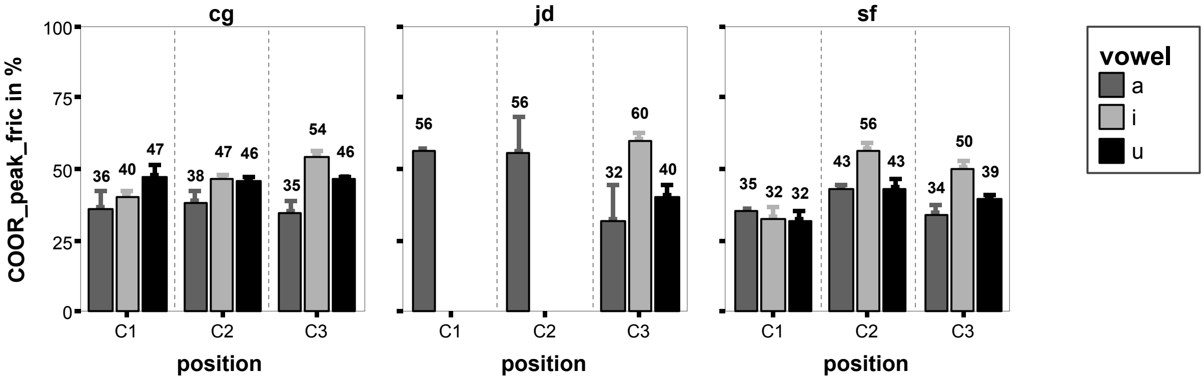

**Figure 4.25:** Barplots for means of COOR_peak fric in ms (y-axis) with +/- 1 standard error; subject CG, JD, SF (left to right track), vowel context (dark grey bars = /a/-context, light grey bars = /i/-context, black bars = /u/ context), positions (C1, C2, C3) at the x-axis

Bar plots provide evidence that peak glottal abduction occurs approximately between 30 to 60 % of the way through relative frication period (= 100%). In speaker CG and SF's averaged bar plots a tendency was found that peak glottal aperture occurs more to frication onset, which corresponds to values below 50 percent. Findings for JD exhibit more variation, i.e. means above and below 50% are found. Again, this could be a consequence of the difficulty in labelling on- and offsets of the weak glottal abduction.

**COOR_peak fric in /s/:** Laryngeal-oral co-ordination in /s/ production shows comparable values for COOR_peak in /s/ (Figure 4.26) as in /z/ (Figure 4.25). Glottal abduction is realised during the first half of the frication period. COOR_peak values are most of the time below 50%. No matter to which extent the glottis is open in phonologically voiced or voiceless fricatives, laryngeal-oral timing consistently ensured that peak glottal abduction is produced in the middle part of the frication interval.

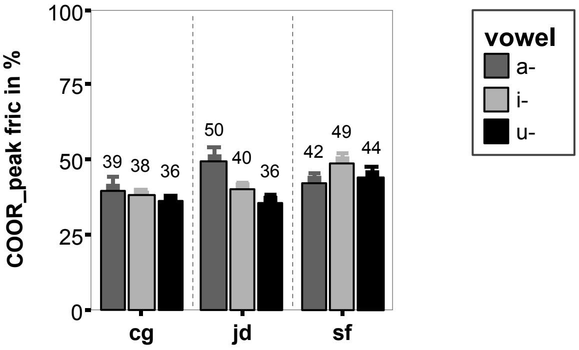

**Figure 4.26:** Bar plots with means of COOR_peak fric in % for /s/ at C2 with +/- 1 standard error; split by subjects (CG, JD, SF from left to right x-axis), vowel contexts (black grey bars = /a/-context, light grey bars = /ɪ/-context, black bars =/ʊ/-context)

**COOR_off fric in /z/**: As Figure 4.27 displays, in most cases the glottis is still open at frication offset, except for subject SF's stressed /z/ where frication and glottal abduction offsets are realised close to each other. Findings in Figure 4.27 show comparable durations to COOR_on fric, but in COOR_off fric glottal abduction offset delays frication offset and in COOR_on fric glottal abduction onset precedes frication onset.

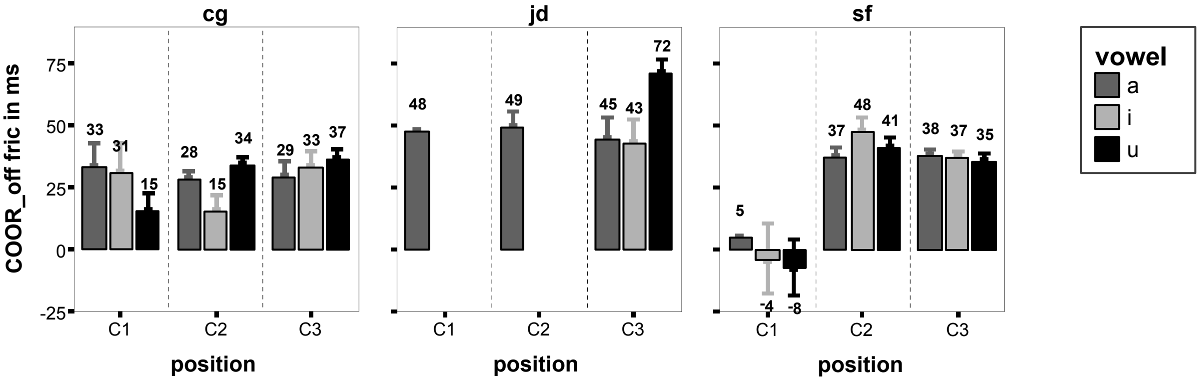

**Figure 4.27:** Barplots for means of COOR_peak fric for /z/ in ms (y-axis) with +/- 1 standard error; subject CG, JD, SF (left to right track) and vowel contexts vowel contexts (dark grey bars = /a/-context, light grey bars = /i/-context, black bars =/u/-context), x-axis = positions (C1, C2, C3)

**COOR_off fric in /s/:** In /s/ production comparable pattern are found as in /z/ regarding COOR_off fric (both in C2).

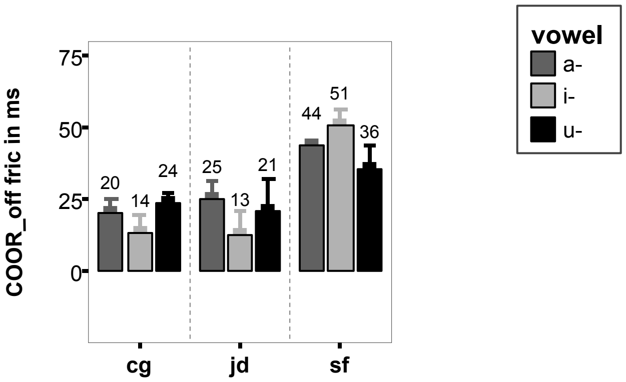

**Figure 4.28:** Bar plots with means of COOR_off fric in ms for /s/ at C2 with +/- 1 standard error; split by subjects (CG, JD, SF from left to right x-axis), vowel contexts (dark grey bars = /a/-context, light grey bars = /ɪ/-context, black bars =/ʊ/-context)

## Summary

1. Laryngeal-oral co-ordination is relatively stable for all phonologically voiced and voiceless fricatives that show at least some amount of glottal opening:
    - Glottal opening onset occurs on average between 9-53 ms earlier with respect to frication onset and it closes 15-50 ms later with respect to frication offset. This co-ordination is independent of the amount of glottal opening, which was realised with a weak glottal opening amplitude for stressed /z/ in SF's data and also in all data concerning word final position.
    - Peak glottal opening was produced during the first half of the frication interval.
2. Laryngeal-oral timing is less affected by position than in stop production. No matter what amount of glottal opening was produced, the co-ordination kept relatively constant.
3. Phonologically voiceless and voiced fricatives resemble each other quite closely in the post-stressed word medial position.

## 4.3. Results for supralaryngeal correlates

### *4.3.1. Introduction*

This section focuses on results for supralaryngeal correlates of alveolar stops in the same positions (C1, C2, C3) as discussed previously. Data from the second experiment (simultaneous EMA, EPG and acoustic recordings) are taken into account, except for acoustic measurements where data from both experiments are included. The results are presented in the following order: First, acoustic results for vowel duration preceding the voiced or voiceless stop are discussed. Second, results are considered for the acoustically defined closure duration, followed by results for the articulatory defined closing gesture duration, its movement amplitudes, its velocity peaks, and its tongue-jaw co-ordination. Third, the articulatory alveolar closure is considered with respect to changes of tongue palate contact patterns during the closure interval. Finally, articulatory target positions for tongue tip, tongue dorsum and jaw are discussed at the vowel target preceding the consonant, at EPG defined closure onset and offset.

### *4.3.2. Acoustic vowel duration*

Vowel duration is one cue which can be attributed to the voicing distinction (e.g. Chen 1970, Lisker 1978, Luce and Charles-Luce 1985, Esposito 2002). For a review on previous work see Luce and Charles-Luce (1985). Vowel duration is generally known to be longer before voiced stops and shorter preceding voiceless stops. However, it does not necessarily imply that it is a universal characteristic.

Vowel duration in both post-stressed positions is considered. The vowel preceding stressed /d/ or /t/ (the unstressed schwa in the prefix ge) was not taken into account, since the unstressed syllable is often very short and the consonant in the stressed syllable belongs rather to the following stressed vowel. Table 4.4 shows results from the 2-way ANOVA and Figure 4.29 barplots with the relevant averaged vowel duration. For speakers CG and JD two main effects are displayed: consonant CON (/d/ vs. /t/) and vowel VOW (/a/ vs. /i/ vs. /u/).

For speaker SF the two main effects are also significant, but with strong interactions. Looking at single contrasts in SF's data it was found that vowels differ in their intrinsic duration in both post-stressed positions. Additionally, the voicing status (/d/ vs. /t/) affected vowel duration in the reported way, i.e. it is longer in voiced stop environment.

For subject CG the vowel effect is primarily an effect of vowel height, since /i/ and /u/ do not differ significantly in word final position and differ weakly in post-stressed syllable initial position. All other vowel duration show significant differences in single comparisons. And again, vowel duration is shorter in /t/ than in /d/ (see also Figure 4.29). The small significant interaction between CON*VOW in C2 is not taken into account.

Subject JD's findings show two main effects too, again with no considerable interaction. Vowel duration is significantly shorter in /t/ than in /d/ and all vowels differ in their intrinsic duration (/a/>/u/>/i/).

However, the first strong differences on vowel duration and voicing distinction could also be due to the consonant in stressed position. This corpus does not include /tVd/ or /dVt/-sequences, but always /dVd/ or /tVt/. Both voiced stops in /dVd/ or voiceless stops in /tVt/ can affect vowel duration. Comparable results have been presented on Italian stops in Esposito (2002), and vowel duration showed a strong effect in the expected direction. However, Esposito's findings and the effects described here might become weaker, if a non-symmetrical environment (/tVd/, /dVt/ etc.) would be taken into account.

**Table 4.4:** Two-way ANOVAs comparing vowel duration for /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C2** | **CON** (/d/-/t/) | 1 | 253.9*** | 291.0*** | 376.4*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 204.9*** | 272.7*** | 226.9*** |
| | **CON*VOW** | 2 | 3.8* | 4.5* | 8.5*** |
| **C3** | **CON** (/d/-/t/) | 1 | 132.8*** | 67.5*** | 283.0*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 103.4*** | 179*** | 215.3*** |
| | **CON*VOW** | 2 | 0.4 | 0.2 | 7.7*** |

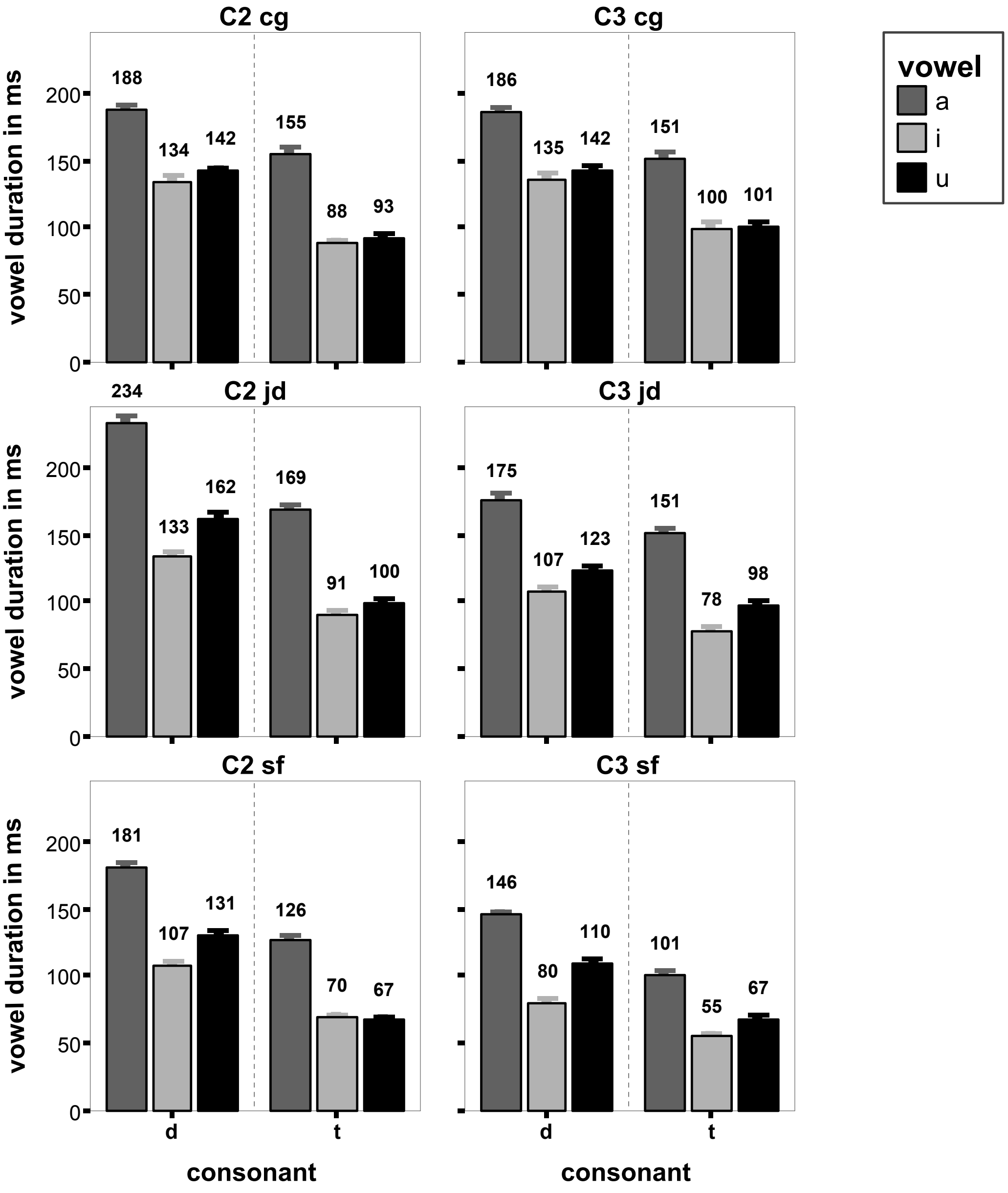

**Figure 4.29:** Bar plots showing means of vowel duration in ms; error bars = +/- 1 standard error; /d/ = left 3 bars on the x-axis, /t/ = right 3 bars; subject CG, JD, SF (from top to bottom), positions C2, C3 (left and right track); different vowel contexts (/a/ = dark grey bars, /i/ = light grey bars squares, /u/ = black bars)

## *4.3.3. Closing gestures*

### 4.3.3.1. ACOUSTIC CLOSURE DURATION

Acoustic closure has been frequently attributed to the voicing contrast in different languages with a two-way voicing contrast (Lisker 1957, Stathopoulos and Weismer 1983, Lisker 1986, Docherty 1992, Jessen 1998, Esposito 2002). The longer articulatory closure is held, the greater the likelihood that voicing will disappear (Ohala 1983), since intraoral pressure rises during closure and equalises the transglottal pressure difference. Hence, a longer acoustic closure duration is associated with voiceless stops whereas a shorter closure duration with voiced stops. Most consistent evidence was found for a longer closure duration regarding the post-stressed intervocalic position (Lisker 1957).

In order to prove the dependency of closure duration on the voicing status and the surrounding vowel environment a two-way ANOVA was calculated. Findings are presented in Table 4.5 and the corresponding barplots with means of closure duration are given in Figure 4.30 (see also Appendix IV).

**Table 4.5:** Two-way ANOVAs comparing acoustic closure duration for /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 40.1*** | 3.9* | 18.7*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 0.0 | 39.4*** | 13.9*** |
| | **CON*VOW** | 2 | 0.1 | 0.0 | 1.5 |
| **C2** | **CON** (/d/-/t/) | 1 | 46.1*** | 194.2*** | 47.1*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 6.8** | 1.1 | 3.7* |
| | **CON*VOW** | 2 | 3.0* | 2.4 | 1.4 |
| **C3** | **CON** (/d/-/t/) | 1 | 3.9* | 2.9 | 0.6 |
| | **VOW** (/a/-/i/-/u/) | 2 | 9.8*** | 19.7*** | 16.0*** |
| | **CON*VOW** | 2 | 1.4 | 2.4 | 1.1 |

Most variable findings can be seen in the stressed position. For JD and SF two significant main effects are found, the vowel environment (VOW) and the voicing status (CON), whereas for CG only voicing status shows a main effect. Closure duration is longer for stressed /d/ compared to /t/ for all subjects. The long closure duration in /d/ could be one explanation why phonologically voiced stops are often produced without or with only a short period of vocal fold oscillation during oral closure. Vowel environment influenced JD's and SF's closure duration in dependency on vowel height. Closure duration is signifycantly longer in high vowels (/i/ and /u/) than in /a/- context.

In the post-stressed word medial position (C2) closure duration is consistently longer for /t/ compared to /d/ (when a rather weak CON*VOW interaction for

CG is not taken into account). It is the same position which Lisker (1957) mentioned. Vowel context is another main effect onto closure duration, in particular for subject CG, and weaker in SF's result. In single comparisons both subjects show no differences between the high vowels /i/ and /u/, but significant differences for closure duration in /a/ vs. /u/ and for /a/ vs. /i/-contexts.

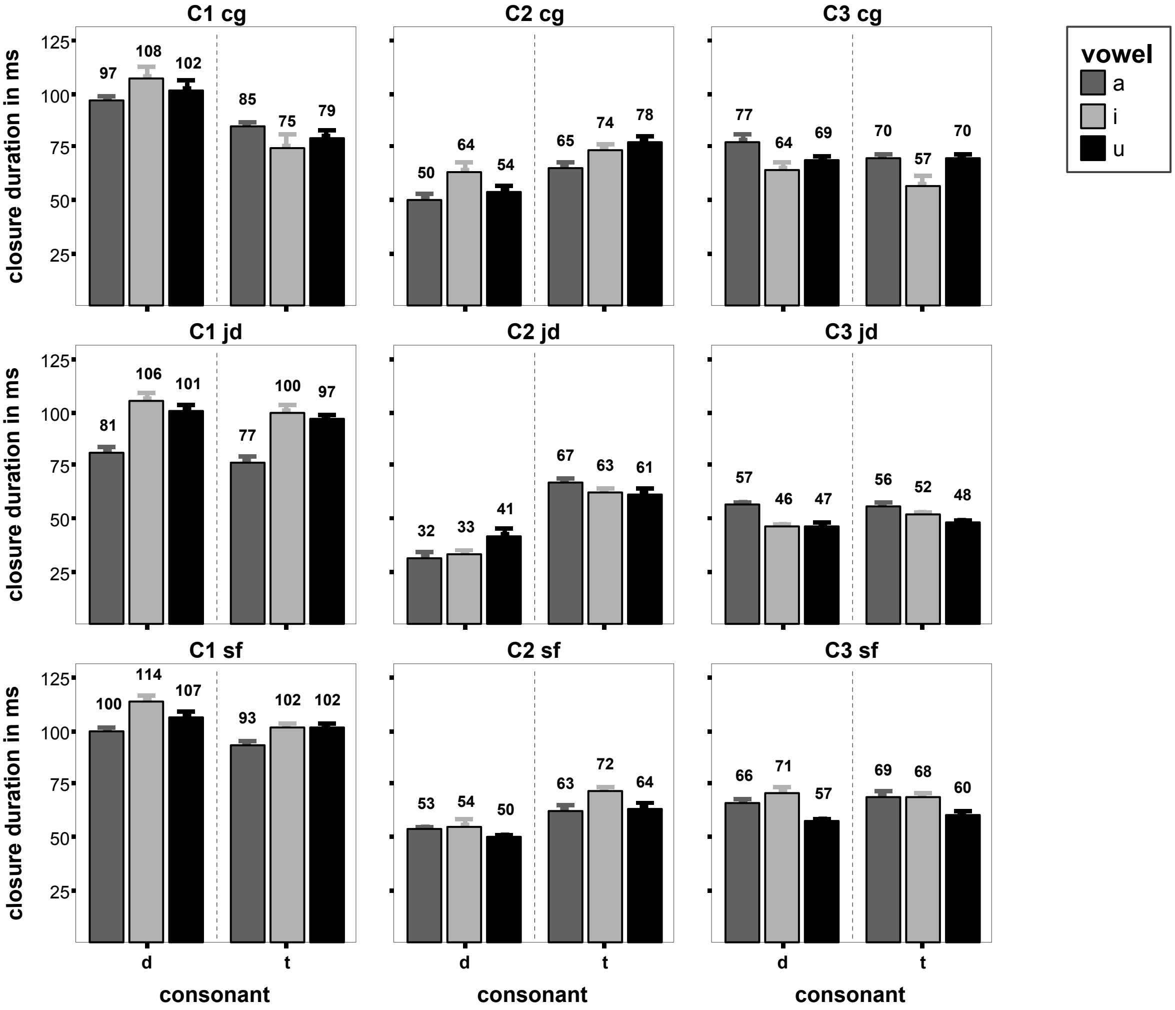

**Figure 4.30:** Bar plots showing means of acoustic closure duration in ms; error bars = +/- 1 standard error; /d/ = left 3 bars on the x-axis, /t/ = right 3 bars; subject CG, JD, SF (from top to bottom), positions C1, C2, C3 (from left to right track); different vowel contexts (/a/ = dark grey bars, /i/ = light grey bars, /u/ = black bars)

In word final position closure duration does not differ significantly, except from a weak effect in CG's data. To remind the reader, the word final position is the one where final devoicing applies in German. Speaker CG's aspiration duration does not show neutralisation from voiced to voiceless and some residue in closure duration can be seen too.

Closure duration is affected by vowel environment, but not by the voicing status (/t/ vs. /d/).

Acoustically defined closure duration is not necessarily similar to temporal characteristics of the articulatory closing gesture. For instance the articulatory closing gesture probably starts during the acoustic vowel duration. The articulatory closing gesture includes those actions of the tongue tip where it moves with a certain speed, but it does not include steady state positioning, which is produced at the alveolar ridge. In the next paragraph results from the articulatory tongue tip closing gesture are provided: the duration of the articulatory closing gesture, the corresponding movement amplitude and velocity peak.

#### 4.3.3.2. ARTICULATORY CLOSING GESTURE

**Duration of the closing gesture**: Figure 4.31 and Table 4.6 exhibit a slightly different picture comparing acoustically defined closure duration with articulatory defined closure duration. The latter is more affected by vowel environment than the acoustically defined closure duration.

In stressed position differences between closing gesture duration for /d/ and /t/ disappear. Consequently, differences found for acoustically defined closure duration should be based on a longer steady state positioning of the tongue tip at the alveolars, but not on the duration of the movement towards closure (articulatory closing gesture). For subject JD and SF the main effect is significant for vowel height. If the following vowel is low, a shorter articulatory closing gesture duration was found compared to a longer duration for a following high vowel context. This effect provides evidence for an anticipatory coarticulation, at least for SF and JD.

In the post-stressed word medial position two significant main effects can be seen for CG and JD, but only one for SF. For CG /d/ was produced with a shorter closing gesture duration than /t/, but this is the opposite for speaker JD. However, all subjects show a longer articulatory closing gesture duration in /a/-context compared to /i/ and /u/-contexts. For SF and JD the articulatory closing gesture duration in /i/ vs. /u/-context is not significant.

In word final position a comparable strong effect of vowel environment onto articulatory defined closing gesture duration was found for all speakers (together with a main effect concerning the voicing status for SF). The articulatory closing gesture is longer when /a/ preceded the consonant and it is shortest in /i/-context. For subject SF a longer closing gesture can be seen (Figure 4.31) in /t/ production.

Generally, the duration of the articulatory closing gesture does not show consistent results with respect to the voicing contrast. It is more affected by the surrounding vowel environment than by the voicing status.

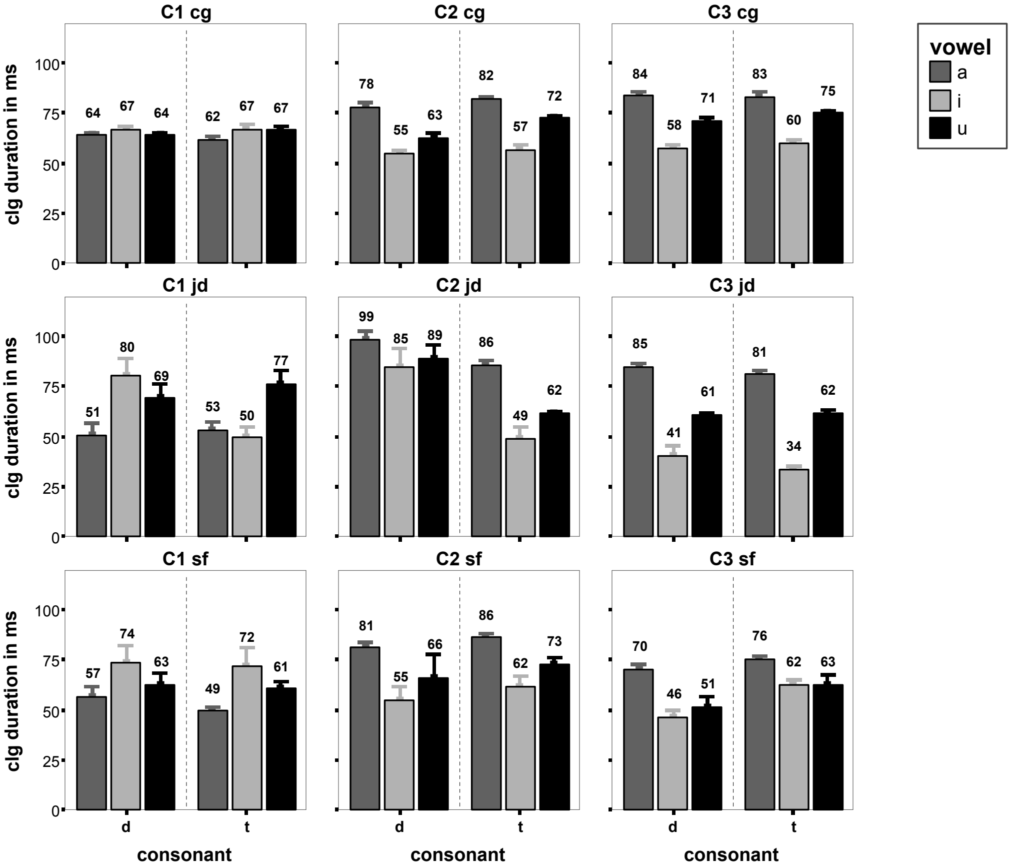

**Figure 4.31:** Bar plots showing means of articulatory tongue tip closing gesture duration in ms; error bars = +/- 1 standard error; /d/ = left 3 bars on the x-axis, /t/ = right 3 bars; subject CG, JD, SF (from top to bottom), positions C1, C2, C3 (from left to right track); different vowel contexts (/a/ = dark grey bars, /i/ = light grey bars, /u/ = black bars)

**Table 4.6:** Two-way ANOVAs for tongue tip closing gesture duration in ms comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 0.0 | 1.7 | 0.6 |
| | **VOW** (/a/-/i/-/u/) | 2 | 3.1 | 5.2** | 5.0** |
| | **CON*VOW** | 2 | 1.0 | 5.1** | 0.1 |
| **C2** | **CON** (/d/-/t/) | 1 | 11.8** | 33.3*** | 1.4 |
| | **VOW** (/a/-/i/-/u/) | 2 | 79.6*** | 11.1*** | 7.8** |
| | **CON*VOW** | 2 | 2.4 | 2.3 | 0.0 |
| **C3** | **CON** (/d/-/t/) | 1 | 1.5 | 2.1 | 15.4*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 87.3*** | 145.6*** | 16.2*** |
| | **CON*VOW** | 2 | 0.6 | 1.1 | 1.1 |

**Movement amplitude of the closing gesture**: Figure 4.32 provides evidence that movement amplitudes of the closing gesture are affected by vowel environment in all post-stressed positions, but not by the voicing contrast. In both post-stressed positions vowel environment plays a major role for the amplitude of the closing gesture (see Table 4.7). Vowel environment also shows some interactions with the voicing status (/d/ vs. /t/), even though movement amplitudes do not show differences for /d/ and /t/ in single comparisons, except from JD in C2 and SF in C3.

In both post-stressed conditions movement amplitude of the closing gesture are largest in /a/-context and smallest in /i/-context. Thus, the previously described long closing gesture duration for low vowels could be caused by larger movement amplitudes.

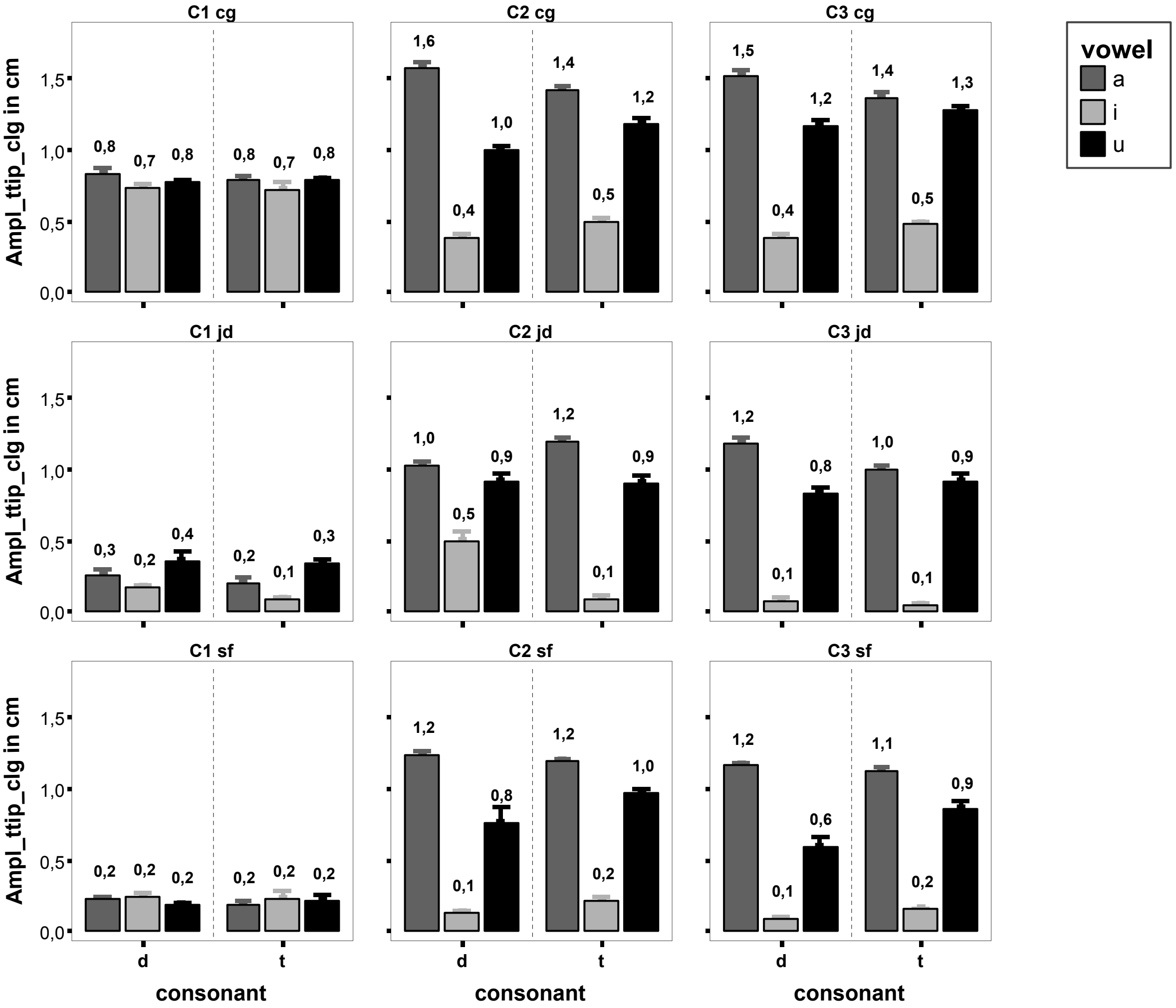

**Figure 4.32:** Bar plots showing means of the movement amplitude for the tongue tip closing gesture in cm; error bars = +/- 1 standard error; /d/ = left 3 bars on the x-axis, /t/ = right 3 bars; subject CG, JD, SF (from top to bottom), positions C1, C2, C3 (from left to right track); different vowel contexts (/a/ = dark grey bars, /i/ = light grey bars, /u/ = black bars)

Regarding the interaction between vowel environment and voicing status for both post-stressed positions the following trend can be seen (Table 4.7):

- Movement amplitudes are larger for /t/ compared to /d/ in /i/ or /u/-context (CG, SF, both post-stressed positions).
- Movement amplitudes are larger for /d/ compared to /t/ in /a/-context (all post-stressed results for CG, and word final position for SF and JD).

**Table 4.7:** Two-way ANOVAs for movement amplitudes of the tongue tip closing gesture in cm comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 0.4 | 2.3 | 0.0 |
| | **VOW** (/a/-/i/-/u/) | 2 | 3.6* | 14.6*** | 0.7 |
| | **CON*VOW** | 2 | 0.7 | 0.3 | 0.5 |
| **C2** | **CON** (/d/-/t/) | 1 | 2.7 | 4.9* | 3.8 |
| | **VOW** (/a/-/i/-/u/) | 2 | 428.1*** | 182.3*** | 201.4*** |
| | **CON*VOW** | 2 | 12.2*** | 22.0*** | 2.8 |
| **C3** | **CON** (/d/-/t/) | 1 | 0.4 | 1.7 | 9.2** |
| | **VOW** (/a/-/i/-/u/) | 2 | 509.0*** | 380.2*** | 342.6*** |
| | **CON*VOW** | 2 | 9.1*** | 5.5** | 7.2** |

**Velocity peak of the closing gesture**: The velocity peak of the closing gesture could be another characteristic possibly involved in the voicing contrast. Generally, findings provide evidence that the velocity peak coincides with vowel environment rather than with voicing status (Table 4.8 and Figure 4.33). However, some differences regarding the voicing contrast occur in C2.

**Table 4.8:** Two-way ANOVAs for tongue tip velocity peak of the closing gesture comparing /d/ versus /t/ (CON), /a i u/ (VOW), the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 1.6 | 2.6 | 0.2 |
| | **VOW** (/a/-/i/-/u/) | 2 | 6.6** | 22.3*** | 0.8 |
| | **CON*VOW** | 2 | 0.1 | 0.1 | 0.3 |
| **C2** | **CON** (/d/-/t/) | 1 | 0.1 | 53.3*** | 8.9** |
| | **VOW** (/a/-/i/-/u/) | 2 | 167.6*** | 436.6*** | 987.8*** |
| | **CON*VOW** | 2 | 10.7*** | 89.4*** | 16.3*** |
| **C3** | **CON** (/d/-/t/) | 1 | 0.2 | 0.7 | 5.3* |
| | **VOW** (/a/-/i/-/u/) | 2 | 1324.4*** | 247.4*** | 1369.6*** |
| | **CON*VOW** | 2 | 3.5* | 1.9 | 28.9*** |

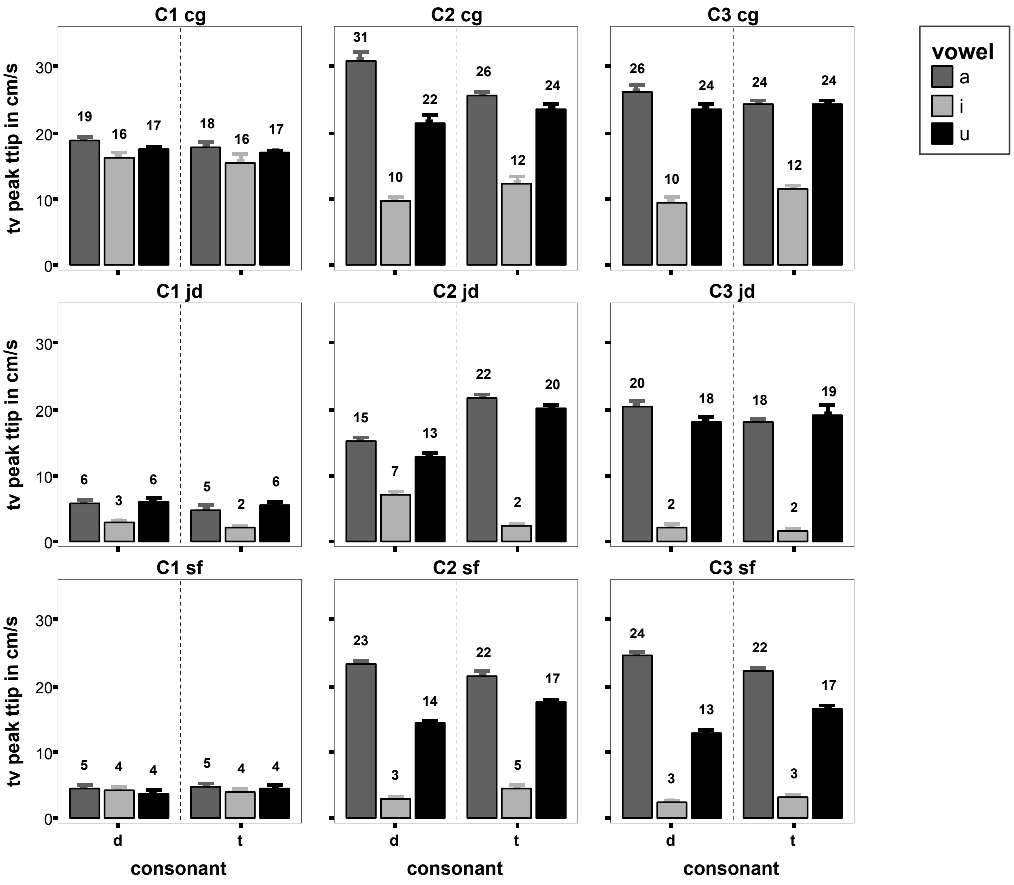

**Figure 4.33:** Bar plots showing means of the tongue tip velocity peak of the closing gesture in cm/s; error bars = +/- 1 standard error; /d/ = left 3 bars on the x-axis, /t/ = right 3 bars; subject CG, JD, SF (from top to bottom), positions C1, C2, C3 (from left to right track); different vowel contexts (/a/ = dark grey bars, /i/ = light grey bars, /u/ = black bars)

In the post-stressed positions word medial position an interaction effect was found that tongue tip velocity peak is higher for /t/ in high vowel context (/i/ and /u/) compared to /d/ (except from JD /i/-context). The opposite trend occurs in /a/-context, where /d/ was more frequently produced with a higher velocity peak compared to /t/ (except from JD).

In word final position, no considerable significant differences are found when velocity peaks are taken into account. It can be suggested that the longer the distance of the vowel from the consonantal target the greater the likelihood for a higher velocity peak and a longer closing gesture duration.

Generally, velocity peaks for the closing gesture do not affect the voicing status in the stressed position and in the word final position. However, an interaction with vowel environment can be seen in the C2 position, which is similar to the results for the movement amplitudes. In high vowel context movement amplitudes are larger for /t/ than /d/ and the velocity peak is higher as well. In low vowel context the opposite occurs, i.e. the movement amplitude is larger for the voiced stop compared to the voiceless and it is also faster.

**Tongue jaw co-ordination in /a/-context**: It is generally agreed that the jaw can contribute to tongue tip movement. For the current study the tongue tip sensor was not decomposed of jaw movement since decomposition is not a straight forward process. In order to get an idea of possible tongue-jaw interactions, the latencies between tongue tip and jaw closing gesture onsets were computed and the latencies between the two articulators at closing gesture offsets. Closing gestures with preceding low vowels were taken into account since in /a/-context there is a considerable amount of jaw movement. Negative values correspond to a jaw delay, i.e. the jaw reaches its target later than the tongue tip.

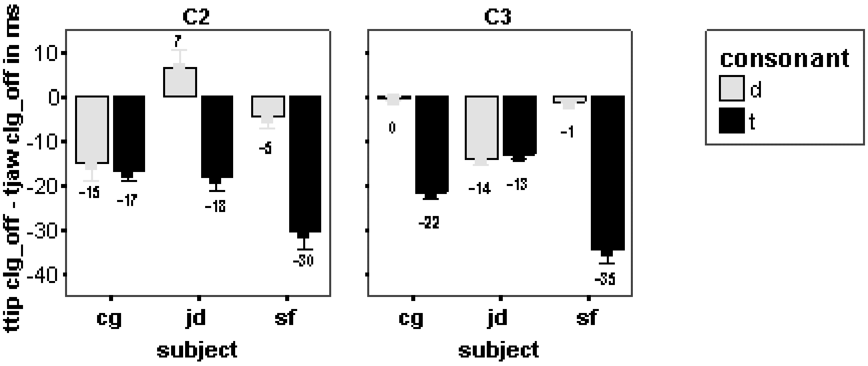

**Figure 4.34:** Barplots showing means of tongue tip – jaw closing gesture offset in ms with +/- 1 std. error; bright bars = /d/, black bars = /t/; left track = C2, right track = C3, subject (CG, JD, SF) on the x-axis

The timing between tongue and jaw at closing gesture onset is not taken into account here, even though it shows some differences. However the sampling rate of the EMA recording is 200 Hz and thus small timing differences should be close to its reliability range (for estimates see Appendix IV).

The timing between tongue tip closing gesture and jaw at closing gesture offset is more relevant. In Figure 4.34 it can be seen that tongue tip and jaw are highly synchronised in /d/ production (JD, SF in C2 left track and CG, SF in C3 right track), whereas in /t/ production the jaw reaches its target with a considerable delay compared to tongue tip movement. Subject dependent differences can be seen: jaw delay is longest for SF in both post-stressed positions. The continuing jaw movement could be involved in the production of a sufficient salient burst in /t/ (see Mooshammer et al. 2003 and Fuchs and Perrier 2003).

#### 4.3.3.3. ARTICULATORY CLOSURE

**Duration of articulatory defined oral closure**: Articulatory defined closure duration using EPG data is very similar to the acoustical defined closure duration. Since these results have already been described here, they are included in the Appendix IV.

**Tongue palate contact pattern during oral closure**: Tongue palate contact pattern differences between /d/ and /t/ are discussed here, since previous studies have found some differences with respect to the voicing contrast (see 2.4.3.). The /d/ often showed less anterior contact than /t/ when taking maximum percent of contact during oral closure into account.

However changes of tongue palate contact patterns can still occur during alveolar closure, similar to the huge changes known for velars (the so called loops represent forward movements of the tongue dorsum along the palate). Hence, looking at changes of tongue palate contact patterns during the whole closure interval should enhance the insights into supralaryngeal production mechanisms.

The amount of contact in the anterior region (ant), in the posterior region (post) and the centre of gravity index (cog) were calculated and are plotted in the EPG defined closure interval (for procedure see 3.6.4.). All closure intervals are normalised in time, i.e. temporal differences are not taken into account at this point.

Results of tongue palate contact patterns provide evidence that most significant differences between /d/ and /t/ occur in the post-stressed word medial position. This position is described here and the other positions are reported in Appendix IV. The stressed position and the word final position show relatively similar patterns between /d/ and /t/.

Figure 4.35 exhibits the three EPG parameters (ant, post, cog) for subject CG from top to the bottom, and the different vowel contexts from left to right (/a i u/). As it is evident from Figure 4.35, most changes occur in the anterior portion of the palate where the tongue tip touches the alveolar region. In speaker CG's data /d/ was produced with less contact after closure onset. Another trend can be seen that the parameter for percent of contact in the anterior region exhibits a plateau phase in /t/, i.e. it kept relatively stable at one level. This is somewhat different in /d/ production, where the interpolated line changes more, in particular in /a/ and /u/-contexts.

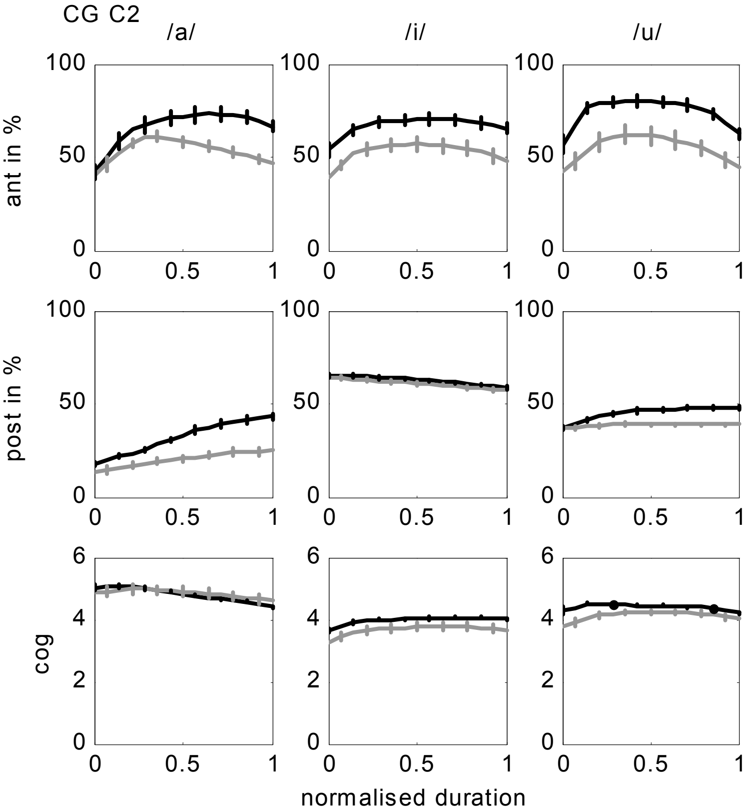

**Figure 4.35:** Means of interpolated tongue palate contact patterns (anterior contact in %= ant, posterior contact in %= post, centre of gravity = cog from top to bottom) with standard errors (vertical lines); x-axis = normalised time duration; grey lines = /d/ and black lines = /t/; vowel contexts /a i u/ from left to right track; subject CG; C2

Coming to the percent of contact in the posterior region differences between /d/ and /t/ are most substantial at the end of oral closure, close to the burst (again in /a/ and /u/-context). The posterior contact increases in /t/ production rather than decreases in /d/ production. For /d/ the percent of contact in the posterior region does not change much during oral closure. In /i/-context voiced and voiceless stops show similar patterns which is probably caused by the high amount of palatal contact already during the vowel /i/.

The centre of gravity index does not reveal any particular influence which could be associated with the voicing contrast.

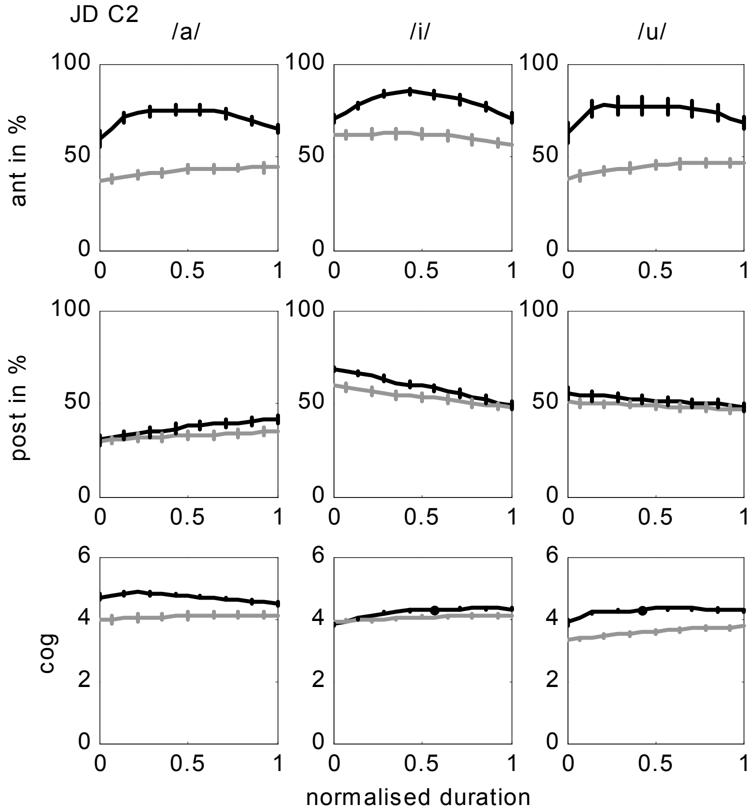

Figure 4.36: same as Figure 4.35, but for JD

Results for speaker JD (Figure 4.36) are similar to CG in terms of the considerable differences in ant between voiced and voiceless cognates. A similar plateau trend as in CG's data is also found for /t/, but tongue palate contacts do not change much either in /d/ production.

Results from the percent of contact in the posterior region are relatively similar for /d/ and /t/ in /a u/-contexts, but there are some small differences the beginning of alveolar closure in /i/-context.

However, differences in the centre of gravity (cog) are found. JD produced /t/ with a more fronted tongue position. The cog gives a weighted estimate about the front-back dimension.

Speaker SF (Figure 4.37) does not show particular differences in tongue palate contact patterns in most of the cases. Surprisingly, the only exceptions are /t/ vs. /d/ in /i/-context and at the beginning of oral closure in /u/-context.

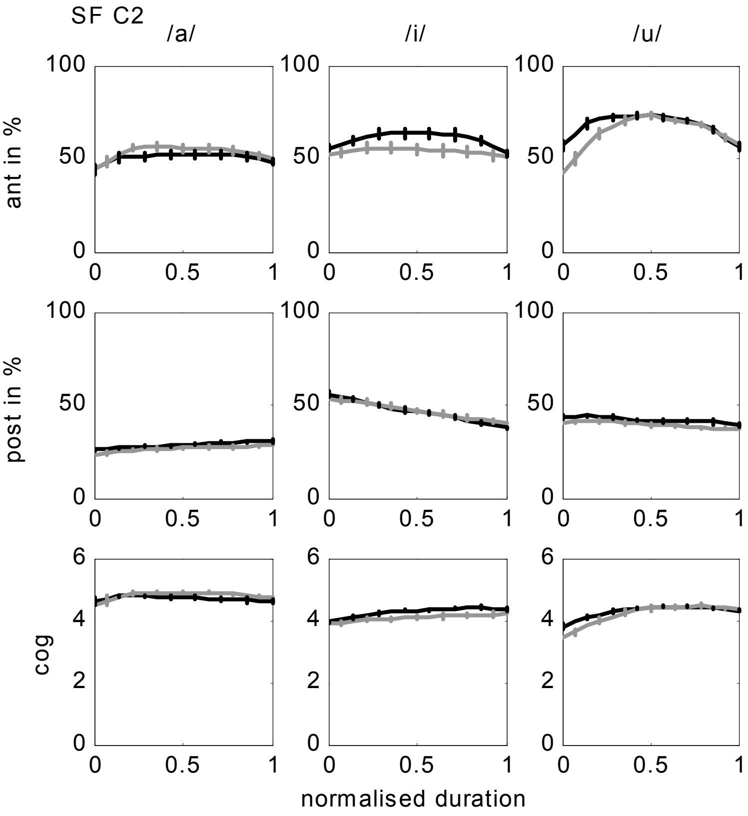

**Figure 4.37:** same as Figure 4.35, but for SF

Since these plots were all time normalised, temporal differences are not represented here.

### 4.3.3.4. ARTICULATORY TARGET POSITIONS

In this paragraph results from three time landmarks are reported, one landmark at the velocity minimum of the closing gesture (preceding vowel target[33]), another landmark at EPG defined alveolar closure onset and the third at EPG defined closure offset. For all three landmarks vertical positions (y) of the jaw, the tongue tip and the tongue dorsum are taken into account. The tongue tip positioning is assumed to realise the consonantal gesture during alveolar closure and tongue dorsum as well as jaw are assumed to participate in the production of the preceding vowel and they could help to support increase (for voiced stops) or decrease the oral cavity (for voiceless stops) as well.

**Tongue tip, tongue dorsum and jaw positions at the vowel target**: Vowel targets of vertical jaw positions are not influenced by the voicing status of the following consonant, except from CG's results in word final position, where /a/

---

[33] Note that the vowel target before C1 is always schwa (and hence unstressed) and for both post-stressed positions it varies between stressed /a i u/.

and /i/ are realised with a lower jaw position when followed by /d/ (see Table 4.9 and Appendix IV). However, the main effect onto jaw position is vowel height, i.e. the jaw has the lowest position for /a/ and the highest for /i/ and /u/.

**Table 4.9:** Two-way ANOVAs for jaw position (y jaw at left min vowel target) comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 0.9 | 0.1 | 0.1 |
| | **VOW** (/a/-/i/-/u/) | 2 | 0.7 | 15.8*** | 2.2 |
| | **CON*VOW** | 2 | 0.6 | 3.1 | 0.2 |
| **C2** | **CON** (/d/-/t/) | 1 | 1.8 | 6.2 * | 0.1 |
| | **VOW** (/a/-/i/-/u/) | 2 | 455.3*** | 576.9*** | 966.2*** |
| | **CON*VOW** | 2 | 2.9 | 0.8 | 0.9 |
| **C3** | **CON** (/d/-/t/) | 1 | 11.1** | 1.7 | 0.1 |
| | **VOW** (/a/-/i/-/u/) | 2 | 439.9*** | 439.8*** | 1215.1*** |
| | **CON*VOW** | 2 | 7.6** | 4.7* | 1.0 |

The tongue tip position exhibits a similar relationship (see Table 4.10), but with strong interaction effects between voicing status and vowel in both post-stressed positions. Tongue tip has a lower position during vowel /a/ when the following consonant is /d/ (all post-stressed positions, all subject, but for SF C3 no difference was found). In /i/ and /u/ tongue tip has most frequently a lower position when /t/ follows compared to a higher positioning when /d/ follows.

For the post-stressed positions these results are in agreement with results concerning the movement amplitude of the closing gesture and the velocity peak. It can be summarised that in high vowel context /t/ has a lower tongue position already at the vowel target, it moves faster and moves over a larger distance in comparison to /d/. In low vowel context the tongue tip has already a lower tongue tip position at the vowel target when a voiced stop follows. The following movement amplitude is larger and additionally the velocity peak.

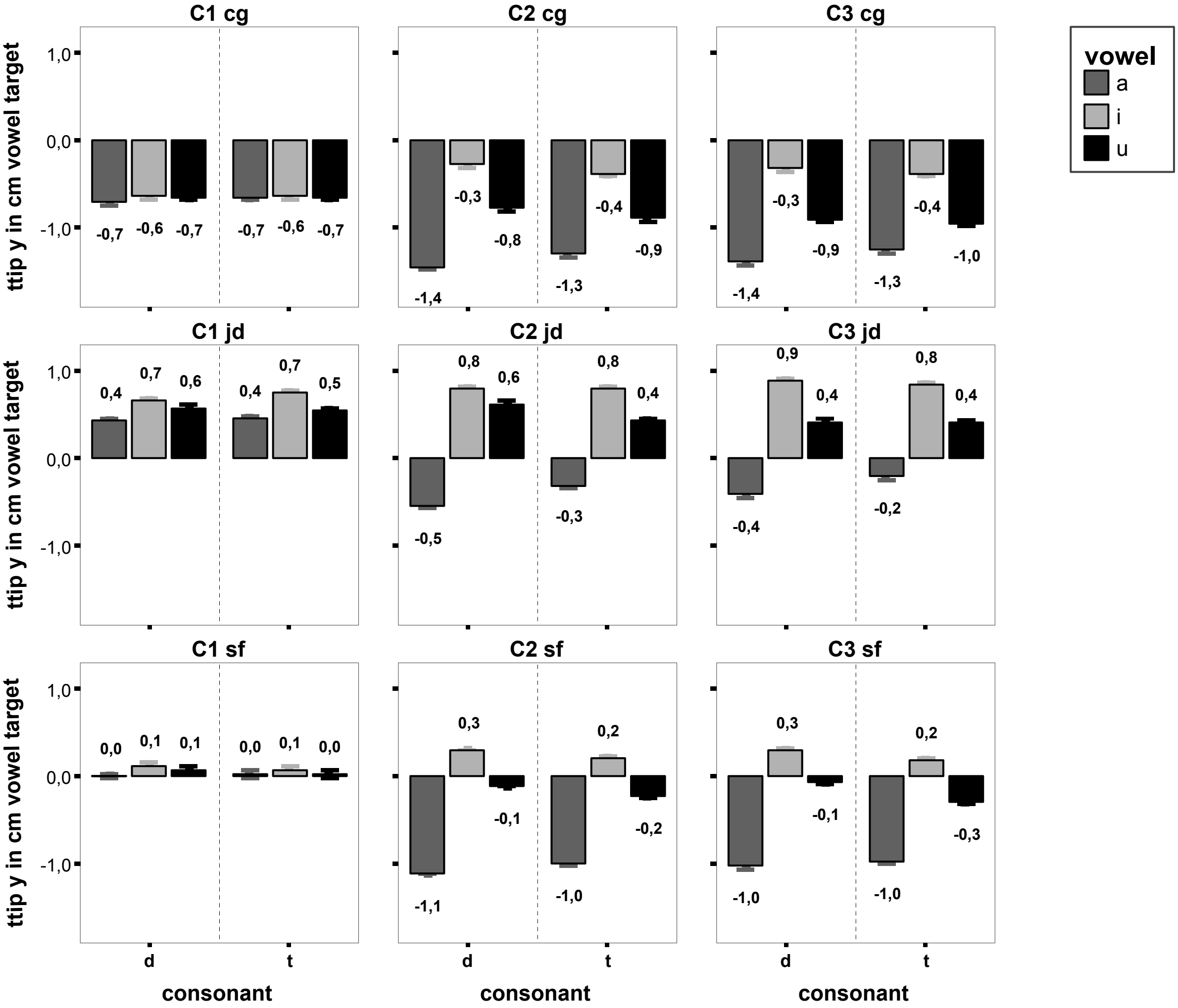

**Figure 4.38:** Bar plots showing means of the tongue tip position at vowel target in cm; error bars = +/- 1 standard error; /d/ = left 3 bars on the x-axis, /t/ = right 3 bars; subject CG, JD, SF (from top to bottom), positions C1, C2, C3 (from left to right track); different vowel contexts (/a/ = dark grey bars, /i/ = light grey bars, /u/ = black bars)

**Table 4.10:** Two-way ANOVAs for tip position (y ttip at left min vowel target) comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 0.7 | 1.1 | 0.4 |
| | **VOW** (/a/-/i/-/u/) | 2 | 1.6 | 46.6*** | 1.9 |
| | **CON*VOW** | 2 | 0.4 | 2.1 | 0.4 |
| **C2** | **CON** (/d/-/t/) | 1 | 0.6 | 0.9 | 6.1* |
| | **VOW** (/a/-/i/-/u/) | 2 | 435.4*** | 1106.1*** | 2605.0*** |
| | **CON*VOW** | 2 | 8.8*** | 28.3*** | 20.2*** |
| **C3** | **CON** (/d/-/t/) | 1 | 0.2 | 4.8* | 25.6*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 460.6*** | 774.4*** | 1672.2*** |
| | **CON*VOW** | 2 | 6.8** | 10.4*** | 18.2*** |

**Table 4.11:** Two-way ANOVAs for tongue dorsum position (y tdors at vowel target, left min) comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 5.5* | 0.7 | 0.3 |
| | **VOW** (/a/-/i/-/u/) | 2 | 1.7 | 16.3*** | 3.9* |
| | **CON*VOW** | 2 | 0.0 | 1.0 | 0.2 |
| **C2** | **CON** (/d/-/t/) | 1 | 1.0 | 15.4*** | 8.6** |
| | **VOW** (/a/-/i/-/u/) | 2 | 3467.5*** | 1650.0*** | 4137.6*** |
| | **CON*VOW** | 2 | 12.7*** | 7,1** | 32.7*** |
| **C3** | **CON** (/d/-/t/) | 1 | 9.6** | 29.5*** | 15.2*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 2791.0*** | 2210.6*** | 1627.4*** |
| | **CON*VOW** | 2 | 21.8*** | 17.9*** | 17.1*** |

Looking at tongue dorsum vertical positioning (see Table 4.11) the voicing status as well as vowel environment are main significant effects in the post-stressed positions, and show an interaction too. At the vowel target a higher tongue dorsum position was found when /t/ follows (all subjects), but this holds only for /a/-context. In /i/ and /u/ context vertical positions do not differ with respect to the voicing contrast (CG, JD). However, SF's findings exhibit a lower tongue dorsum position at the vowel target (/i/ or /u/) when /t/ follows.

It can be summarised that in the post-stressed positions the vowel position shows already some anticipation for the following consonant. If a voiced stop follows the tongue dorsum has a lower position, in particular when the vowel is /a/. Tongue tip and jaw positions do show a similar anticipation.

**Tongue tip, tongue dorsum and jaw positions at the beginning of oral closure**: At closure onset (defined on EPG data) jaw position shows some significant differences with respect to vowel (all subjects) and to voicing contrast (CG, SF; with an interaction for CG, JD in C2) in both post-stressed positions (Table 4.12). There is no difference regarding the stressed position.

The jaw exhibits a higher position at closure onset in /t/ production, in particular when the preceding vowel was /a/ or /u/ (SF, CG).

**Table 4.12:** Two-way ANOVAs for jaw position (y jaw at t_clon_EPG) comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 0.2 | 0.1 | 0.2 |
| | **VOW** (/a/-/i/-/u/) | 2 | 1.4 | 3.9* | 2.4 |
| | **CON*VOW** | 2 | 0.6 | 0.6 | 0.1 |
| **C2** | **CON** (/d/-/t/) | 1 | 107.1*** | 0.1 | 19.5*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 72.4*** | 32.9*** | 176.1*** |
| | **CON*VOW** | 2 | 18.8*** | 13.8*** | 0.7 |
| **C3** | **CON** (/d/-/t/) | 1 | 6.0* | 0.9 | 9.2*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 50.5*** | 74.2*** | 122.8*** |
| | **CON*VOW** | 2 | 0.7 | 2.2 | 1.5 |

The vertical position of the tongue tip at closure onset seems to play a minor role in terms of the voicing contrast, but some significant differences are found (see Table 4.13) in particular for JD (C2, C3) and CG (C3). For JD tongue tip has a higher position in /t/ realisation compared to /d/. This is true for /a/-context in post-stressed word medial position and for /i/- and /u/-context in word final position. However, differences are often only 1mm.

The vertical position of tongue dorsum seems to be more influenced by vowel environment than by voicing. Occurring differences (see Table 4.14) are rather inconsistent and speaker dependent and they are not taken into account here (for overview see Appendix IV).

**Table 4.13:** Two-way ANOVAs for tip position (y ttip at t_clon_EPG) comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 0.0 | 5.8* | 1.4 |
| | **VOW** (/a/-/i/-/u/) | 2 | 1.6 | 42.2*** | 6.3** |
| | **CON*VOW** | 2 | 0.1 | 0.7 | 0.6 |
| **C2** | **CON** (/d/-/t/) | 1 | 0.5 | 12.1*** | 0.5 |
| | **VOW** (/a/-/i/-/u/) | 2 | 3.6* | 242.3*** | 141.6*** |
| | **CON*VOW** | 2 | 0.1 | 7.6** | 0.0 |
| **C3** | **CON** (/d/-/t/) | 1 | 7.4** | 14.3*** | 0.3 |
| | **VOW** (/a/-/i/-/u/) | 2 | 5.6** | 306.1*** | 163*** |
| | **CON*VOW** | 2 | 4.5* | 1.2 | 1.4 |

**Table 4.14:**: Two-way ANOVAs for tongue dorsum position (y tdors at t_clon_EPG) comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 5.5* | 6.0* | 4.7* |
| | **VOW** (/a/-/i/-/u/) | 2 | 40.5*** | 61.2*** | 7.3** |
| | **CON*VOW** | 2 | 1.8 | 0.1 | 1.6 |
| **C2** | **CON** (/d/-/t/) | 1 | 7.4** | 1.0 | 0.4 |
| | **VOW** (/a/-/i/-/u/) | 2 | 1190.2*** | 717.4*** | 1631.3*** |
| | **CON*VOW** | 2 | 8.5*** | 5.6** | 8.2*** |
| **C3** | **CON** (/d/-/t/) | 1 | 0.0 | 10.1** | 11.1** |
| | **VOW** (/a/-/i/-/u/) | 2 | 375.6*** | 996.6*** | 1028.3*** |
| | **CON*VOW** | 2 | 2.0 | 0.3 | 8.2*** |

Findings from tongue tip and jaw position provide evidence that /t/ can be associated with higher jaw and tongue tip position and /d/ with a lower position of the two articulators at the beginning of articulatory defined closure onset. However results vary with respect to the relevant speaker and to the surrounding vowel context.

**Tongue tip, tongue dorsum and jaw positions at the end of oral closure:** Jaw position at oral release shows most consistent differences for /d/ vs. /t/ (see Table 4.15) comparing it with tongue tip and tongue dorsum positions. The voiceless stop was produced with a relatively high jaw position compared to a lower jaw position for /d/. However, JD's highly significant results in stressed and post-stressed positions are often below 0.5mm which is close to the reliability of the EMA-system. For CG and SF differences in jaw position are more reliable.

**Table 4.15**: Two-way ANOVAs for jaw position (y jaw at t_cloff_EPG) comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 1.4 | 38.6*** | 7.7** |
| | **VOW** (/a/-/i/-/u/) | 2 | 0.6 | 64.9*** | 25.6*** |
| | **CON*VOW** | 2 | 6.1 | 12.4*** | 2.5 |
| **C2** | **CON** (/d/-/t/) | 1 | 223.8*** | 12.2*** | 38.2*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 2.6 | 1.5 | 18.3*** |
| | **CON*VOW** | 2 | 14.2*** | 4.8 | 0.0 |
| **C3** | **CON** (/d/-/t/) | 1 | 34.0*** | 0.9 | 17.8*** |
| | **VOW** (/a/-/i/-/u/) | 2 | 1.0 | 1.8 | 21.1*** |
| | **CON*VOW** | 2 | 0.9 | 0.4 | 1.9 |

Again, it is supposed that a high jaw position could be realised for the production of a salient burst or coarticulatory effects due to lower jaw positions in the vowels.

Results for tongue tip position at oral release in Table 4.16 show two main effects in the stressed and post-stressed word medial position with an interaction (except SF C2). In stressed position tongue tip position is higher in /t/

production in /a/ and /i/ context for JD, whereas it is lower in /t/ production (/i/-context) for SF.

In the post-stressed word medial position tongue tip exhibits a significantly higher position in /t/ realisation compared to /d/ (JD /a/ and /i/ context, CG /a/ and /u/ context). In the word final position effects are weaker.

**Table 4.16**: Two-way ANOVAs for tip position (y ttip at t_cloff_EPG) comparing /d/ versus /t/ (CON), /a i u/ (VOW) and the interaction between CON and VOW (CON*VOW); split by position (Pos.) C1, C2, C3; by speaker CG, JD, SF; Given are the degrees of freedom (Df), the F-values, and the significance level (*** p<0.001, ** p<0.01, * p<0.05)

| | | | **CG** | **JD** | **SF** |
|---|---|---|---|---|---|
| **Pos.** | | **Df** | **F-value** | **F-value** | **F-value** |
| **C1** | **CON** (/d/-/t/) | 1 | 36.0*** | 12.1*** | 4,7* |
| | **VOW** (/a/-/i/-/u/) | 2 | 6.2** | 192.9*** | 262.7*** |
| | **CON*VOW** | 2 | 3.4* | 3.6* | 11.6*** |
| **C2** | **CON** (/d/-/t/) | 1 | 41.8*** | 15.0*** | 2.3 |
| | **VOW** (/a/-/i/-/u/) | 2 | 5.0** | 37.5*** | 11.6*** |
| | **CON*VOW** | 2 | 11.0*** | 12.2*** | 2.7 |
| **C3** | **CON** (/d/-/t/) | 1 | 6.6** | 4.1* | 0.1 |
| | **VOW** (/a/-/i/-/u/) | 2 | 1.4 | 27.1*** | 4.6* |
| | **CON*VOW** | 2 | 4.3* | 1.2 | 0.8 |

Tongue dorsum position is not discussed here since it does not reveal strong differences with respect to the voicing contrast.

The main results are summarised again in the discussion section.

# Chapter 5: Discussion

Chapter 5 is structured as follows: Section 5.1 provides an overview of the laryngeal and supralaryngeal findings of the current study and relates these to the hypotheses drawn in chapter 1. It is shown that for the voicing contrast in German different articulatory mechanisms are favoured depending on the position of the sequence. Stress has a major impact on the production mechanism used in the relevant position. Section 5.2. presents some limitations of the current work and considers how future studies of the voicing contrast in German could be improved.

Section 5.3. opens up a general discussion of what can be drawn from the current findings. It is first discussed which phonological feature terms should be used for the voicing contrast in German, if features are meant to be representative of phonetic reality. Secondly, different models of stress are described and evaluated in light of the current findings, and the extent to which the results confirm or disagree with those models is addressed. In particular, it seems de Jong's (1995) suggestion of stress as 'localized hyperarticulation' fits the current results best. Finally, in section 5.4. the results of this study are reviewed in light of the theoretical concepts introduced in chapter 1.

## 5.1. Linking the results to the hypotheses

Since in the current work alveolar stops have been studied more extensively, and since they behave differently in comparison to fricatives (see also Fuchs et al. 2003), the two classes are discussed separately.

### *5.1.1. The voicing contrast in alveolar stop production*

#### 5.1.1.1. TABULATION OF RESULTS

Table 5.1 provides an overview of all the results found in this study. Columns are split by position and subject. Findings are sorted with respect to laryngeal correlates, laryngeal-oral co-ordination and supralaryngeal correlates. The following abbreviations have been used:

**Abbrevations**

X = Significant effect
X(+) = Significant effect with /t/ being longer in duration, larger in amplitude, having more tongue- palate contact, and a higher target position in comparison to /d/
X(-) = Significant effect with /d/ being longer in duration, larger in amplitude, having more tongue-palate contact, and a higher target position in comparison to /t/
empty = No significant effect
INT = Interaction effects with vowel context
Part = Partial effects depending on the relevant vowel context
\# = No measurements possible

“X” holds not only for a significant effect based on statistical analysis, but also for those cases where a statistical analysis was not possible due to the limited dataset and a clear distinction between the minimal pair could be described.

**Table 5.1**: Summary of the results concerning the voicing contrast in alveolar stop production, columns are split by position (C1, C2, C3), subject (CG, JD, SF)

| **Position** | **C1** | | | **C2** | | | **C3** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Subject** | **CG** | **JD** | **SF** | **CG** | **JD** | **SF** | **CG** | **JD** | **SF** |
| **Laryngeal correlates** | | | | | | | | | |
| Glottal abduction | X (+) | X(+) | X(+) | X(+) | | Part(+) | X(+) | | Part(+) |
| **Laryngeal-oral co-ordination** | | | | | | | | | |
| Noise duration | X(+) | X(+) | X(+) | X(+) | X(+) | X(+)<br>INT | X(+) | | |
| COOR_on (+/-20ms) | X | X | X | X | | Part | Part | | |
| COOR_peak (+/-20ms) | X | X | X | X | | X | Part | | |
| **Supralaryngeal correlates** | | | | | | | | | |
| Vowel duration | # | # | # | X(-) | X(-) | X(-) | X(-) | X(-) | X(-) |
| Closure duration | X (-) | X (-) | X (-) | X(+) | X(+) | X(+) | X(+) | | |
| Ttip closing gesture duration | | Part(+)<br>INT | | X(+) | X(-) | | | | X(+) |
| Ttip movement amplitudes | | | | Part<br>INT | Part<br>INT | | INT | INT | Part<br>INT |
| Ttip velocity peaks | | | | INT | Part<br>INT | Part<br>INT | INT | | Part<br>INT |
| Tongue-jaw co-ord. at closing gesture offset /a/-context X(+)= earlier ttip in /t/ | # | # | # | | X(+) | X(+) | X(+) | | X(+) |
| Tongue palate contact patterns: | | | | | | | | | |
| ant | | | | X(+) | X(+) | Part(+) | | | |
| post | | | | Part(+) | Part(+) | | | | |
| cog | | | | Part(+) | Part(+) | | | | |

| Position | C1 | | | C2 | | | C3 | | |
|---|---|---|---|---|---|---|---|---|---|
| **Subject** | **CG** | **JD** | **SF** | **CG** | **JD** | **SF** | **CG** | **JD** | **SF** |
| Target position at preceding vowel: | | | | | | | | | |
| Jaw | | | | | Part | | Part<br>INT | INT | |
| Ttip | | | | INT | INT | Part<br>INT | INT | Part<br>INT | Part<br>INT |
| Tdors | | | | INT | Part<br>INT | Part<br>INT | Part<br>INT | Part<br>INT | Part<br>INT |
| Target position at oral closure onset: | | | | | | | | | |
| Jaw | | | | Part<br>INT | Part<br>INT | Part | Part | | Part |
| Ttip | | Part | | | Part<br>INT | Part | Part | | |
| Tdors | Part | Part | Part | Part<br>INT | INT | INT | | Part | Part<br>INT |
| Target position at oral closure offset | | | | | | | | | |
| Jaw | | X(+)<br>INT | Part(+) | X(+)<br>INT | X(+) | X(+) | X(+) | | X(+) |
| Ttip | X(+)<br>INT | X(+)<br>INT | Part(+)<br>INT | Part(+)<br>INT | Part(+)<br>INT | | Part<br>INT | Part | |
| Tdors | | Part<br>INT | Part<br>INT | Part<br>INT | INT | INT | | | |

### 5.1.1.2. GLOTTAL OPENING: A LARYNGEAL CORRELATE OF THE VOICING CONTRAST IN GERMAN?

Generally, glottal opening is **not** a consistent correlate of the voicing contrast in German alveolar stop production. The occurrence of glottal abduction varies with respect to position as follows:

**Glottal opening in stressed position**: All the speakers produced consistently large glottal abduction for /t/ in the stressed position, but did not realise glottal opening for /d/. It is inferred from EMG studies on other languages (see Hirose and Gay 1972, Lisker and Baer 1984, Hutters 1984, Fischer-Jørgensen and Hirose 1974) that glottal abduction in stressed position is produced due to neural activity, i.e. activity of the posterior cricoarytenoid muscle (PCA) with simultaneous suppression of the interarytenoid muscle (INT). In principle, it is possible, on the other hand that a high intraoral pressure could have driven the vocal folds apart, but it is unlikely that the large glottal abduction in /t/ can be explained by aerodynamics alone.

**Glottal opening in post-stressed positions**: If glottal abduction does still participate in the voicing contrast, then its amplitude is reduced considerably compared to the stressed position. The occurrence of glottal abduction varies speaker- and vowel-dependently: CG realised glottal opening in most of the cases, JD did not realise glottal abduction, and for SF some very weak glottal opening peaks occur, particularly in the /a/-context. Interestingly, the reduction of glottal opening amplitude does not coincide linearly with a reduction in overall glottal opening duration. The glottal opening amplitude for /t/ is more affected than its duration, comparing the stressed with the post-stressed positions.

This result confirms earlier empirical work from Sawashima and Miyazaki (1973) and Hutters (1984). Similar amplitude reduction due to stress has also been observed by Cooper (1991), Munhall (1984), and Löfqvist (1980) for English and Swedish, and by Hoole et al. (1983) for German. The disappearance of glottal abduction in phonologically voiceless stops seems surprising, but it has also been reported for unstressed positions by Hoole et al. (1983) and Fuchs et al. (2004b) for German, by Löfqvist (1980) for Swedish, and by Lisker & Baer (1984) for American English. A nearly closed glottis can be seen in Iwata and Hirose's data (1976) for Mandarin Chinese. Most of these authors mention an absence of glottal abduction for alveolar stops. These results are interpreted with respect to the fact that alveolars are generally most often affected by articulatory reduction phenomena (e.g. assimilation, coarticulation). It is unclear to what extent the small amount of glottal opening that was found is caused by neural activity or whether it is due to aerodynamics. However, since small abductions occur with a great intra-speaker and inter-speaker variability it is supposed that they are rather a by-product of intraoral pressure rise than due to neural activity.

The results here concerning the role of glottal abduction are partially in disagreement with Jessen's work (1998) on German, since he reported glottal abduction for ***all*** phonologically voiceless stops in utterance initial position, and furthermore in post-stressed intervocalic position. More surprisingly, he even found some weak glottal abduction for the phonologically voiced stops. These differences could be due to the fact that articulatory data are generally variable, particularly between speakers - similarly to the results for post-stressed positions in the current study. Jessen (1998) presented just one subject, whereas the current study is based on three subjects. The different findings might also be due to the materials. The amount of glottal abduction could be influenced by the number of syllables of the target word (with a smaller glottal abduction in words containing more syllables), the length of the whole phrase (longer phrases involve more frequently reduction phenomena), and the relevant speaking style (hypospeech or hyperspeech, slow versus faster speech rate - again, fast speech often involves an amplitude reduction of articulatory movements similarly to hypospeech). In Jessen's study just the target word was produced (except for the #_V form) whereas in the current study target words were embedded in a carrier phrase. Jessen recorded mostly bisyllabic real words whereas here bisyllabic and trisyllabic nonsense words were taken into account.

The current findings show that stress plays a major role in determining the phonetics of the voicing contrast. Stops in stressed syllables show a clear voicing distinction in terms of presence versus absence of glottal opening. This distinction becomes smaller or disappears when the contrast appears in the post-stressed position. The differences in glottal abduction found in stressed versus post-stressed position support the concept of ‘strong’ and ‘weak’ syllables (for an overview see Krakow 1999). Basically, strong syllables are generally associated with larger movement amplitudes and longer overall duration in comparison to weak syllables. Our results provide evidence that this is also true for laryngeal articulation. A large glottal opening is conditioned by a syllable being stressed (a strong syllable) whereas a reduced amount of glottal opening or even no opening reflects post-stressed positions (a weak syllable). The amplitude of glottal opening is more affected than its overall duration by stress. There are presumably some lower temporal limits for the production of glottal opening. If this temporal limit cannot be achieved (in the current study overall glottal opening duration is below 140ms for stops) glottal aperture disappears.

#### 5.1.1.3. SUPRALARYNGEAL CORRELATES OF THE VOICING CONTRAST IN GERMAN

Let us turn now to the role of supralaryngeal articulators in the voicing contrast. Effects can be approached from three different perspectives:

1. Laryngeal-oral co-ordination: Supralaryngeal correlates are well synchronised with laryngeal correlates, particularly in the production of aspiration (see chapter 2.3.1.). In addition, aspiration noise could neither be produced nor perceived when the constriction area of the vocal tract is larger than the constriction area at the glottis (Dixit 1987, Maeda 1996). Hence, constriction degrees of supralaryngeal and laryngeal articulations have to be adjusted in the production of phonologically voiceless obstruents.
2. Vocal tract-source interaction: The realisation of the size (and duration) of vocal tract constrictions has an effect onto the maintenance of vocal fold vibrations (source), i.e. a vocal tract-source interaction takes place. This interaction can be used in the production of the voicing contrast.
3. Pure supralaryngeal correlates: Supralaryngeal correlates such as preceding vowel duration or closure duration can be used to distinguish between phonologically voiced and voiceless cognates of a following stop.

The present results provide evidence for the three perspectives as follows:
**Laryngeal-oral co-ordination**: For the STRESSED POSITION the results provide clear evidence that all phonologically voiceless stops are aspirated. Laryngeal events are well synchronised with oral closure. When the tongue tip reaches the alveolars ridges, glottal abduction starts, in the current study within a range from –6 to 23 ms. Peak glottal opening and oral release are co-ordinated within a range from 5-15 ms with a slightly later peak. The latter co-ordination is produced in order to guarantee a sufficient amount of glottal abduction at oral release necessary for the production of aspiration noise (see also Kim 1970). The long acoustic noise duration found for stressed /t/ supports the transillumination results and additionally, Dixit's and Maeda's notion of a smaller constriction degree at the glottis in comparison to oral articulation. Moreover, in /d/ production no glottal abduction could be observed, making the contrast clear and consistently cued by aspiration. It was of course not possible to find any oral-laryngeal co-ordination for /d/, but note that the acoustic data for /d/ in the stressed position show some noise duration values (burst) below 20ms and thus provide confirmatory acoustic evidence for /d/ being a voiceless unaspirated stop.

In the POST-STRESSED POSITIONS the current study shows continuous variation in laryngeal-oral co-ordination with respect to the surrounding vowel context, and also inter-speaker variation. For CG the synchronisation of laryngeal and supralaryngeal movements is similar to that found in stressed position for the high vowel context (/i u/). In the /a/-context, however, glottal abduction onset appears before the oral closure is made (20-28ms) and glottal opening reaches its maximum either at oral release (in the post-stressed word medial position) or before oral release (in the word final position). The latter co-ordination together with a reduced amount of glottal opening is known for unaspirated stops and results in a shorter acoustic noise duration (approximately below 20ms). Comparable results for unaspirated /t/ are also found for subject SF.

To summarise, the co-ordination of glottal opening to supralaryngeal events takes place in both stress positions for /t/. The timing is more stable however, in the stressed syllable with fewer speaker-specific effects compared to the post-stressed syllable.

The results from the analysis of the duration of post-release acoustic noise are in general agreement with Jessen (1998), who found that aspiration duration is the main correlate of the voicing contrast in German. There is a categorical distinction for noise duration between the stressed and the post-stressed word medial positions (the post-stressed final position will be discussed in 5.1.1.5.).

The articulatory results for laryngeal-oral co-ordination, however, vary more continuously from a reduced amount of glottal opening but with a good synchronisation with supralaryngeal events to a rather variable synchronisation with an early peak glottal opening with respect to oral release. The small amount of noise generated for /t/ in the post-stressed positions is assumed to be produced in many cases due to vocal tract constriction rather than by an open glottis, at least for two speakers.

Both phonemes /d/ and /t/ differ with respect to acoustic noise duration, but both are most of the time unaspirated (assuming aspiration refers to a realisation of above approximately 20ms) in the post-stressed positions. It can be concluded that phonemic differences concerning noise duration are relative differences rather than absolute values. A value of 20ms noise duration, for instance, does not necessarily provide information about the voicing status of the phoneme concerned. It could be either phonologically voiced or voiceless. The question arises as to whether the small differences in noise duration concerning /d/ and /t/ in the post-stressed position is of any perceptual relevance to the detection of the voicing contrast in this position, a question that cannot be addressed here. However, looking at all acoustic differences, preceding vowel and closure duration seem to have more perceptual relevance than noise duration in these positions.

**Vocal tract-source interaction**: In the STRESSED POSITION the longer acoustic closure duration for /d/ (81-114ms) compared to /t/ (75-102ms) is responsible for devoicing of /d/, a common phenomenon in initial/stressed position in German (see Pape et al. 2003). It should be noted at this point that a quantitative analysis of the duration of voicing during closure was not made here due to noisy acoustic data. However, inspections of the acoustic signals showed devoicing in the word initial stressed position as well as in the post-stressed word final position. It is assumed that the long oral closure coincides with an intraoral pressure rise, which decreases the likelihood of maintaining voicing. In both cases (for /d/ and /t/) the relatively long oral closure should also be of perceptual importance, i.e. the acoustically silent period improves the strength of the perception of the following burst and the aspiration noise.

In the POST-STRESSED WORD MEDIAL POSITION the opposite is found: a longer acoustic closure duration for /t/ (63-78ms) in comparison to /d/ (32-64ms). Since /d/ is often produced with voicing during oral closure in this position and /t/ with a small amount of voicing or without voicing, a clear threshold seems to exist between the two. In /d/ the transglottal pressure difference is kept high, which allows the maintenance of voicing and in /t/ the diminishing of the transglottal pressure difference causes devoicing. The idea of a threshold can be strengthened since the POST-STRESSED WORD FINAL

POSITION exhibits comparable closure duration values for /d/ and /t/ as the /t/ (but not /d/) in word medial position. Moreover, both /t/ and /d/ are phonetically voiceless in word final position. Assuming a threshold in closure duration for or against the maintenance of voicing, the voicing during closure characteristic is produced by differences in oral closure duration with everything else being equal.

Closure duration, like aspiration duration, is a relatively consistent correlate of the voicing contrast in German. However, if aspiration (noise) duration and closure duration were ranked in a hierarchy of correlates (Jessen 2001, Brunner et al. 2003), aspiration/noise duration would be a stronger correlate to describe the contrast for two reasons: First, significant differences are more pronounced for aspiration/noise duration and second, the silent period in word initial closure cannot always be detected in language use - if following silence, for example. In the current study a prefix preceded the morpheme initial stressed syllable so that word boundary effects were excluded, and it should be borne in mind that the behaviour of absolute initial stops may be different to the phrase-internal ones examined here, as perhaps exemplified by the different results of Jessen (see above).

**Pure supralaryngeal correlates**: An example of what might be called a pure supralaryngeal correlate of the voicing contrast is the duration of a preceding vowel. It can also be responsible to regulate the transglottal pressure differences, reported previously. A differentiation between both processes is sometimes hard to define. Pure supralaryngeal correlates of the voicing contrast do most frequently occur in the POST-STRESSED POSITIONS. They are highly speaker-dependent, significant interactions with the surrounding vowel environment are often found, and significant effects are often weak. Since they are weak, it is unclear to what extent listeners can use them for discrimination. The strongest and most consistent effects which have been found in the current study are vowel duration (with a longer duration before /d/ in word medial as well as in word final position), jaw position at oral release (vertically higher for /t/ in word medial position and speaker-dependently in word final position) as well as tongue-jaw co-ordination at closing gesture offset (with synchronised movements for /d/, but with a jaw delay reaching its target for /t/; 2 speakers in both post-stressed positions show the effect; only /a/-context could be taken into account). Considering vowel duration, the clear distinction may not only be due to the voicing status of the following consonant, but also due to contributory effects from the preceding /d/ or /t/: recall that the materials matched voicing in C1 and C2. Consequently, if the post-vocalic consonant were to be varied while keeping the prevocalic consonant fixed (either as /d/ or /t/) the size of this effect may be less.

The production of a higher jaw position in /t/ is interpreted in accordance with Mooshammer et al. (2003) in order to explain the production of a salient burst, i.e. the high jaw position decreases the oral cavity and increases intraoral pressure - a prerequisite for the production of noise. Additional evidence could be found by means of tongue-palate contact patterns (2 speakers). The phonologically voiceless stop in post-stressed word medial position shows a higher percentage of contact in the anterior region than /d/, and the percentage of posterior contact increases during the oral closure interval.

From the above, we can see that oral (supralaryngeal) correlates play a non-neglible role in the realisation of the voicing contrast, particularly in both post-stressed positions (see also Fuchs & Perrier 2003). In conclusion, in STRESSED syllables the voicing contrast is based on an interaction of laryngeal and supralaryngeal articulators, whereas in POST-STRESSED POSITIONS supra-laryngeal correlates dominate.

#### 5.1.1.4. CAN SUPRALARYNGEAL CORRELATES COMPENSATE FOR THE LACK OF LARYNGEAL ADJUSTMENT?

Let us consider in more detail those cases where glottal opening disappears: /t/ in post-stressed intervocalic position (subjects JD and SF). It would seem that supralaryngeal correlates compensate for the lack of laryngeal adjustment, i.e. they might be able to produce an equivalent to the noise generated at the glottal level: frication noise - realised with supralaryngeal correlates - compensates for aspiration noise - realised with an open glottis. A relatively high tongue and jaw position and a long oral closure were considered as potential possibilities for the production of noise, since a high tongue and jaw position with a long closure phase could increase the intraoral pressure. At the time of oral release a salient burst could be produced due to the high intraoral pressure and hence compensate for the missing glottal abduction.

High tongue and jaw positions together with a long closure duration were found in the current study, but their use as a possible compensation for the lack of laryngeal adjustment can only be confirmed for one subject (SF). For JD, differences in jaw position between /d/ and /t/ are only minor. For CG, there is a significant difference with respect to tongue and jaw position, but this subject produced glottal opening too. This suggests that tongue and jaw movements might be readily available as a compensatory mechanism for at least some speakers who lack glottal opening. Such compensation appears to be a speaker dependent strategy. Further experimental investigations or simulations with a physical model of the larynx and the vocal tract are necessary to expand on these observations.

#### 5.1.1.5. FINAL DEVOICING IN GERMAN

Final devoicing has been a major issue in the discussion of the phonetics-phonology interface (see 1.3.3.). The main question around this issue is whether phonologically voiced stop behave in a way indistinguishable from their voiceless counterparts in word final position, i.e. whether neutralisation is "complete", or whether on the contrary some residue occur which suggests the contrast is suppressed or diminished, rendering it "incomplete". Results of this study suggest different perspectives depending on whether we look at acoustic or articulatory data:

**Acoustic results**: Noise duration as well as closure duration exhibit generally no differences between /d/ becoming devoiced and /t/. However, speaker CG did realise a significant difference in acoustic noise duration, plus some weak effects for closure duration, suggesting incomplete neutralisation. These findings are interpreted in agreement with a recent acoustic investigation of Piroth and Janker (2004) who studied neutralisation phenomena with respect to three dialectal regions in Germany. Their results showed incomplete neutralisation for Bavarian speakers. CG's spend his childhood near Bavaria, at Lake Constance, so the possibility arises that his natural accent is one with partial devoicing. Perceptual analysis of this speaker does not give any overt impression of hyperarticulation towards standard Northern German. Additionally, CG realised phonologically voiced fricatives most of the time without voicing: a typical South German feature. Thus this incomplete neutralisation is probably due to his regional origin.

Note that differences in vowel duration are not taken into account, since it is unclear whether the strong differences observed are due to final /d/ versus /t/ or to preceding /t/ or /d/ or a combination of both.

**Articulatory results**: Articulatory residue occurred more frequently than acoustic residue and they were highly speaker dependent. From an articulatory point of view neutralisation is clearly not complete.

CG still produced a difference in glottal abduction, but also differences in supralaryngeal articulation such as tongue-jaw latency and jaw position at closing gesture offset. SF's findings exhibit significant differences at the supra-laryngeal level for the articulatory closing gesture duration, tongue-jaw latency, and jaw position at closing gesture offset. All other differences are vowel dependent or interact with the surrounding vowel environment. For JD a few supralaryngeal production mechanisms exhibit differences, but none of them are consistent for all the surrounding vowels.

The neutralisation process in the acoustics shows generally a trend towards being complete whereas articulation is always incomplete (at least for JD and

SF). For CG neutralisation was incomplete on both grounds. Based on the differences between the acoustic and the articulatory results (see also Fuchs & Perrier 2003, Fuchs et al. 2004c) it can be concluded:

1. Complete neutralisation can neither be considered in isolation from facts of acoustics and ultimately perception, nor just on a speech production basis alone. Speaker-specific articulatory mechanisms can be used in order to produce a particular acoustic/perceptual target. Given a phonetic model of neutralisation via, for example, articulatory reduction rather than a phonological model of abstract category deletion, there is a high probability that speaker specific articulatory residues of the contrast will be maintained, even though strong, general and reliable acoustic cues to the contrast do not occur. The perceptual importance of such articulatory residues may be minimal, especially if speakers' purpose is to neutralise. Additionally, the word final position is perceptually less salient than for instance the word initial position (see Trubetzkoy 1936). The interpretation of final devoicing according to the speech production-perception link brings a new spirit into the old debate whether neutralisation of phonologically voiced obstruents in word final position would be complete or not.
2. Articulatory residue of phonologically voiced stops in word final position occur most frequently at the supralaryngeal level than at the laryngeal level. Thus, there is empirical evidence that glottal opening is not generally an articulatory residue of the voicing contrast though it may be present in some dialects.
3. Final devoicing in German varies with respect to dialectal regions as found out by Piroth and Janker (2004). South German speakers preserve the voicing contrast. The current findings support Piroth and Janker's suggestion.

### *5.1.2. Voicing contrast in alveolar fricatives*

#### 5.1.2.1. GLOTTAL OPENING IN PHONOLOGICALLY VOICED VERSUS VOICELESS FRICATIVES

To remind the reader: In the current corpus voiced and voiceless fricatives not only differ with respect to their claimed phonological voicing status, but also according to the tenseness of the preceding vowel. Voiced fricatives occur following a tense vowel and voiceless fricatives following a lax vowel, both in the post-stressed[34] word medial position. The fricatives preceding the stressed

---

[34] A fricative occurring after a lax vowel is ambisyllabic and belongs to the stressed as well as to the post-stressed position.

vowel („stressed position“) is phonologically voiced. The post-stressed intervocalic position is the only position of these taken into account here, involving a /z/-/s/ contrast.

Results of this study provide evidence that glottal opening as well as overall glottal opening duration is not a consistent correlate of the voicing contrast in alveolar fricatives in German in post-stressed intervocalic position. Looking at glottal opening in detail: In none of the results would presence versus absence of glottal abduction explain the voicing contrast, with the exception of subject JD’s results in /i ɪ u ʊ/-context. Interestingly, two subjects (CG, SF) realise the phonologically voiced fricatives with a significant glottal opening, similar to the voiceless fricatives. So far it is unclear whether the two subjects contrast /s/ and /z/ by means of the tense versus lax vowel pairs or whether supralaryngeal production mechanisms are involved too. For the third subject (JD) glottal abduction is produced in /s/, but not in /z/, at least for the high vowel context.

Unlike the stops, glottal abduction is frequently found in phonologically voiced fricatives, but varies in its amount speaker dependently. In particular, a large amount of glottal opening is found for subject CG but only a weak opening for SF, for both in the stressed position. For all subjects weak glottal abduction was observed for /z/ in word final position. Three possible interpretations are presented for this heterogeneous picture:

1. **The role of aerodynamics**: The specific characteristic of fricatives is that they are produced with turbulent airflow, which can be perceived as frication noise. The main factors to generate turbulent airflow are: the size of the channel the air is moving through and the volume velocity, i.e. the number of air particles passing an observed point. The smaller the size of the channel and the higher the volume velocity, the higher the frequency and the amplitude of the produced noise. In fricatives turbulent airflow is created at different stages: at the laryngeal level (only for phonologically voiceless fricatives) as well as at the supralaryngeal level due to a downstream obstacle (here at the teeth) and when the air jet escapes the narrow channel.

   Johnson (1997) notes that:

   > *“Voiced fricatives are relatively unusual in the languages of the world, undergo a variety of phonetically motivated alternations, and are surprisingly difficult to produce. The difficulty, which may underlie the cross-linguistic and phonological patterns, arises because high volume velocity is needed to produce the turbulent noise characteristic of*

> *fricatives, and the vibrating vocal cords impede the flow of air through the vocal tract"* (Johnson 1997, p.115).

He also reports that the volume velocity is much lower when the glottis is shut as in a phonatory state. Since the creation of turbulence is one of the main characteristics of fricatives, strategies for the realisation of turbulence seem to dominate over those for the production of voicing. Consequently, glottal opening occurs even in phonologically voiced fricatives, but in turn, it increases the likelihood of devoicing too.

2. **Regional variations**: Similar to the interpretations of the results for stops (speaker CG), variation is found for the fricatives. Glottal abduction for this subject is observed in all positions for phonologically voiced /z/. This result is probably not only a consequence of the aerodynamic conditions, but also due to his South German heritage. South German speakers, specifically Bavarian speakers, are known to realise phonologically voiced fricatives as voiceless (see, Mangold 1978).
3. **The role of the lexicon**: The realisation of /z/ as a voiceless fricative could be a consequence of the nonsense speech material used in this study. Nonsense words, even though they follow German phonotactical rules, have no representation in the lexicon. Since there are only a few minimal pairs for the alveolar fricatives in the German lexicon, the nonsense speech material might induce free variation of /z/.

Out of the three interpretations the first, the role of aerodynamics is the strongest concerning the results of this study. Devoicing of phonologically voiced fricatives has also been observed in other studies (e.g. Smith 1997, Shih and Möbius 1999). Thus, it is concluded that the creation of turbulent noise has a major impact on phonologically voiced fricatives, which are often half or fully devoiced.

#### 5.1.2.2. SUPRALARYNGEAL CORRELATES IN PHONOLOGICALLY VOICED VERSUS VOICELESS FRICATIVES

As with stops, laryngeal-oral co-ordination, tract-source-interaction as well as purely supralaryngeal correlates of the voicing contrast are possible in fricatives. The first sort of effects are found, but there are no clear difference with respect to the voicing contrast. For the second sort a tract-source-interaction takes place as described previously, but it seems to be independent of the voicing contrast. For the third type of effect, no data have been analysed yet. Hence, the following short section is dedicated to laryngeal-oral co-ordination and the voicing contrast in German alveolar fricatives.

**Laryngeal-oral co-ordination**: In both cases (/s/ and /z/) the onset of glottal abduction is synchronised with the onset of supralaryngeal constriction, i.e. the oral constriction starts on average around 30ms later than it does for stops, and the synchronisation does not differ with respect to the voicing contrast. Peak glottal opening is located around the centre of the frication interval with a trend towards the beginning of the interval, again with no distinction between voiced and voiceless fricatives. Glottal abduction ends after frication offset.

## 5.2. Limitations of the current work

### *5.2.1. Technical limitations*

Studying laryngeal and supralaryngeal correlates of the voicing contrast by means of a combination of transillumination/fiberoptic films, EPG and EMA forms one of the unique aspects of this work. These techniques are relatively rarely used compared to acoustic analyses, and even more rarely combined, despite their scientific power. Here, they were used to reveal aspects of the underlying articulation of the voicing contrast in German, despite there being a number of methodological difficulties in using and combining these techniques. In this section are also some suggestions for further improvements to the use and combination of these techniques.

1. **Acoustic and transillumination recordings**: The quality of the acoustic and the transillumination data suffered from electromagnetic interference from the main power supply. Using an adapter could reduce this interference. The low frequency periodic interference on the acoustic signal limited the analysis of voicing during oral closure. The interference on the transillumination signal also masked the phonatory patterns, which are normally superimposed on the transillumination waveform. This meant that in the current work the transillumination signal needed to be filtered and information about the extent of phonation at the beginning and end of the glottal opening gesture was lost.
2. **Limited statistical analysis due to a small number of repetitions**: The transillumination technique limits the number of data being recorded to approximately 30min for one session due to the physical discomfort. Consequently, two separate sessions were recorded for each subject, each with different speech material. Methodologically, the choice was either to record a smaller corpus with lots of repetitions or to record target words within various contexts. Since the second option was chosen, only 5

repetitions of each target word could be recorded limiting statistical analysis.

3. **Frame rates**: Frame rates of EPG and fiberoptic films are limited to 100Hz for the first and approximately 40Hz for the second. It would be particularly desirable to increase the video frame rate by using a high-speed camera instead of the standard format, since stops in the post-stressed position are relatively short.

### *5.2.2. Limitations due to the chosen speech material*

1. **Influence of /a/-context**: Through the course of experimentation it became evident that epiglottal movement sometimes blocked the transillumination signal in /a/-vowel context. The strength of this effect varies according to the speaker. For current purposes, the back vowel context was recorded to make the materials consistent with the EMA recordings, but /i/-context is favoured. Contrary, with EMA data the /i/-context is most difficult to label since movements are so small in amplitude. If future work does not aim at recording EMA, only high front vowel context should be chosen for the transillumination recording.
2. **Nonsense words**: Although the current speech material was constructed following German phonotactical rules and nonsense target word were embedded in a real word carrier phrase, the current speech material is far from being ‘natural speech’. More real speech material may reveal weaker effects since natural materials are more influenced by coarticulation processes. A subsequent study (Fuchs et al 2004c) considered acoustic and articulatory results for final devoicing in nonsense words and compared them with real words. It turned out that articulatory differences between /d/ and /t/ were even smaller and more speaker-dependent in the real word material. However, articulatorily /d/ was still not completely neutralised to /t/. The nonsense speech material was chosen to investigate the underlying articulatory movements and their control as much as possible. Using real lexical items in such experiments would increase the naturalness of the speech process. Future work should consider both: nonsense words and real words in order to study the underlying kinematics as well as the specifics of more real speech.

### *5.2.3. Other limits*

1. **Awareness of the task of this study**: Since the author of the current study was one of the subjects (SF), it may be that the results were

influenced by an awareness of the task. However, it is generally difficult to find subjects prepared to volunteer for a transillumination recording and with the necessary experience in speaking in a relaxed and naturalistic way when subjected to such an extended experimental procedure. In addition, the speech processes addressed are highly automatic, so that it is nearly impossible to control the relevant articulatory movements consciously (within a range of 1-3mm) given a relaxed speaker and normal speech rate. It is always possible that a subject may hypo- or hyperarticulate, but the results from the author do not differ to a great extent from the other subjects.

2. **Missing data of factors which are also involved in the production of the voicing contrast**: This study looked at glottal opening as a laryngeal correlate of the voicing contrast. It does not take into account variations of vocal fold tension. Vocal fold tension, particularly an increase of fundamental frequency due to CT activity is known to occur during the production of voiceless stops (for our preliminary results of CT activity in German see Hoole et al. 2004). It would therefore be useful to examine such data if a more complete picture of muscular activity were required in addition to its effects. Additionally, aerodynamic factors are only inferred from the measured articulatory kinematics. However, the assumptions are in agreement with a number of publications on this topic. Future work should also include a perceptual evaluation of the observed variations in order to study the potential role of the variability of speech production and their perception.

In summary, future work should improve the technical equipment of the transillumination recording so that phonatory oscillations can be detected in the transillumination data as well as voicing during closure can be segmented in the acoustics. I would also suggest to record more repetitions of the same speech material in order to do a reliable statistic analysis instead of recording a variety of material and providing more or less a descriptive analysis. For a more comprehensive study, simultaneous aerodynamic and EMA-recordings should be targeted in order to get quantitative evidence about the correlation between tongue/jaw movements and the development of intraoral pressure with respect to vocal fold vibrations and glottal abduction. Additionally, a perceptual evaluation of the small articulatory changes in the word final position are necessary and would provide insights into the question whether the incomplete neutralisation concerning articulatory data is of perceptual relevance.

## 5.3. Understanding the voicing contrast in German

### *5.3.1. Some general methodological remarks*

1. **Investigating a contrast in different prosodic conditions**: For languages like German where the realisation of the voicing contrast changes with respect to position, methodologically comprehensive investigations are necessary and important. However, focusing on the contrast in one particular position may be worth studying, but conclusions about the nature of the contrast should be derived with caution. Here results differ quite markedly in stressed vs. post-stressed positions.
2. **Terminology of theoretical concepts needs empirical support**: Theoretical concepts in phonetics and phonology need an empirical foundation, calling the need for comprehensive studies. The apparently unremarkable proposal that the German voicing contrast could be based on the phonological feature spread versus non-spread glottis, for example, is not supported by the results presented here. Indeed, the very idea that articulatory terms used in phonology must take into account that articulatory processes are very often speaker-specific and variable, and henceforth not suitable as general terms for a language. Thus, if generally applicable feature names are desirable, then it is worth considering an acoustic or perceptual terminology as being more consistent or to use terminology which is inherently more abstract in combination with detailed empirical studies which provide a quantitative and scientific meaning for such abstract terminology (for further details see next section).
3. **Evaluation of speaker's heritage**: A detailed evaluation of the speaker's heritage and possible dialectal influences should be a common procedure prior to the experiment. Every language is characterised by variation, and such an evaluation may tease apart variability due to dialectal influences and speaker-specific morphology. In the current study subject CG does not show remarkable dialectal colours in his daily speech. However, his articulatory patterns showed clearly typical pronunciation forms with respect to his South German heritage and therefore morphological differences are assumed to play a minor role.

### *5.3.2. Which phonological feature terms reflect the realisation of the voicing contrast in German properly?*

Phonological feature terms describing a contrast are often named with respect to phonetic characteristics (see also 1.1.), and of course the phonological contrasts themselves are correlated with specific phonetic correlates. Since there ought to be some kind of relationship between these equivalences but this is complicated by the subtle positional and inter-speaker differences which can be found, the question as to what is the ‘correct’ (or more weakly, the ‘best’) feature term for the voicing contrast in German is addressed here.

**[±voice]**: Chomsky and Halle (1968) subsume four different phonetic characteristics under the [±voice] feature. One of those phonetic characteristics which is investigated here is glottal abduction, called ‘constricted glottis’ in Chomsky and Halle’s terms. Although it is only one of four phonetic characteristics the constricted versus not constricted glottis mechanism would not account for the results of alveolar stops in the post-stressed positions (because glottal abduction is often missing). Additionally the feature term is problematic for the opposite reason for alveolar fricatives in the post-stressed word medial position (because glottal abduction is realised most of the time for both voiced and voiceless fricatives).

**[±aspirated]**: The findings of this study do not support a description of the contrast as aspirated versus unaspirated if by “aspiration” we require a delay of voicing of no less than 20ms (a reasonable perceptual perspective and one in keeping with studies on other languages). In both post-stressed positions, only a small amount of noise is produced in /t/, often below 20ms/23ms (JD, SF) and therefore, /t/ should be classified as unaspirated. If the feature [±aspirated] is associated with temporal differences in aspiration (here it is called noise duration with no reference to the source of noise - it can be laryngeal or supralaryngeal), the results of the current study can support it. Temporal differences were found with longer noise duration for /t/ and shorter noise duration for /d/. Therefore a more appropriate term for the temporal differences would be “noise duration”, since it is more neutral concerning the source of the noise. Aspiration is commonly associated with noise produced due to an open glottis, but glottal opening can disappear and a certain amount of noise can be produced due to the vocal tract.

**[±spread glottis]**: No matter whether the spread glottis feature is interpreted as the occurrence of glottal opening or more specifically with respect to an active opening due to neural activity as opposed to merely a passive opening due to a high intraoral pressure (see 1.1.), results of this study do not support the [±spread glottis] feature for the post-stressed position (at least not for JD and SF). Thus, using this phonological term does not reflect the phonetic realisation in the way that the use of such a phonetically-specific feature would imply .

**[tense] vs. [lax]**: The definition of [tense] versus [lax] can only be tested in line with the proposals of Jessen (1998) rather than Perkell (1969), since for the latter no data on muscular tension of the tongue or the pharyngeal walls were carried out. Jessen relates the tenseness feature to the temporal domain, with a longer duration for tense and a shorter for the lax phonemes. Results of this study reveal that acoustically, duration is one of the major correlates of the voicing contrast in German, particularly concerning noise duration, closure duration and, rather differently, vowel duration. However, temporal differences do not always point in the same direction. Noise duration (all positions, except JD's and SF's word final position) and closure duration (post-stressed word medial position) are longer for /t/ compared to /d/ whereas closure duration (stressed position) behaves in the opposite direction, i.e. /t/ is shorter than /d/. If we consider vowel duration, then there is a compensatory relationship: a post-stressed tense stop is longer, and the preceding vowel is shorter. If the tense-lax feature could be established as a temporal feature (not as a correlate of muscular tension) it would properly describe the voicing contrast in German.

**[fortis] vs. [lenis]**: No direct conclusion can be drawn concerning the fortis-lenis feature as a proprioceptive impression based on intraoral pressure variations (see Malécot's definition in 1.1.). It is supposed that intraoral pressure differences occur with higher pressure for /t/ compared to /d/, especially in the post-stressed word medial position. Whether or not intraoral pressure differences are still significant in the stressed position can be questioned, since closure duration was longer for /d/ than /t/ and /d/ was most of the time devoiced.

Kohler's (1984) definition of the fortis-lenis distinction involves several factors: power in the supraglottal movements, air stream and tension in the larynx. Additionally, he incorporates articulatory timing, for instance in that the speed of stricture formation is greater for tense consonants. Insofar as conclusive remarks can be derived from the current study, speed of closure, i.e. the velocity peak of the closing gesture did not provide evidence for reliable differences between /d/ and /t/ (see Table 4.8). The maximal speed of the movement towards closure was much more dependent on the surrounding vowel context and all the significant differences found, interact with vowel environment. Consequently, speed of stricture formation cannot count as a reliable phonetic correlate of the contrast. However, Kohler's fortis - lenis feature terms are rather broad and involve more than one particular characteristic so, in common with all abstract labels, they cannot be rejected straightforwardly on the basis of empirical evidence.

To summarise: If phonological feature terms of the voicing contrast represent a contrast on the basis of phonetic correlates, a relatively abstract temporal feature such as tense vs. lax (proposed by Jessen 1998) would fit the results of the current study best. More superficially concrete feature terms corresponding to one particular articulatory characteristic such as [±spread glottis] should be avoided since articulatory patterns were variable and showed high interspeaker variations. Broader feature terms (fortis versus lenis or tense versus lax) involving more than one particular correlate have, of course, more flexibility and thus can be used to characterise the voicing contrast in German properly. However, other possibilities consist of either defining a phonological feature as an even more abstract term such as "voicing" in the current study or avoiding a particular name and calling it the /d/ versus /t/ distinction while simultaneously providing enough quantitative information to allow a detailed understanding of the contrast, a position espoused earlier by Docherty (1992).

### *5.3.3. Voicing contrast in German and stress effects: Implications for models of stress*

In this study stress has been seen similar to the definition of de Jong, Beckman and Edwards (1993):

> *"Within this framework, we define stress as a set of prosodic categories with relationships of relative prominence between syllables. In contrast with the traditional view, we view stress not as a phonetic content feature, but as an organizational property. A syllable is stressed if, at some level in the metrical tree, that syllable occupies a strong position relative to other syllables"* (de Jong, Beckman and Edwards 1993, p.200).

Stress is a property of the whole syllable, not only of the nuclear vowel. Within the current study stress had a MAJOR IMPACT particularly on the production of phonologically voiceless alveolar stops in German. In the STRESSED syllable the voiceless alveolar was consistently realised with glottal abduction and glottal abduction was tightly synchronised with supralaryngeal events. In the POST-STRESSED positions /t/ exhibited a considerably reduced amplitude of glottal opening or no glottal abduction at all, a small amount of temporal shortening for glottal opening as well as variable synchronisation patterns to supralaryngeal events. Since differences between /d/ and /t/ in laryngeal adjustment are diminished in post-stressed position, supralaryngeal events were more frequently involved in the realisation of the contrast, although they were highly speaker-dependent and interact with the preceding vowel environment. So, how do some of different models of stress proposed in the literature deal with these findings?

Stress has been modelled as ‘localized hyperarticulation’ by de Jong (1995), ‘jaw expansion’ by Macchi (1985), ‘sonority expansion’ by Beckman, Edwards and Fletcher (1992), or an ‘increase of global articulatory effort’ by Fowler (1995). Most of these studies focus on supralaryngeal events for the production of the nucleus of the stressed syllable, the vowel.

Since the jaw exhibits a lower position in the stressed syllable, Macchi (1985) suggested that stress coincided with jaw expansion. Looking at the results of the current study, jaw position at closure offset does not show consistently the distinctive patterns between stressed and post-stressed syllables, neither for /d/ nor for /t/. Thus, the jaw expansion model is not supported by the current data.

Similarly, Beckman, Edwards and Fletcher (1992) proposed a model of stress as ‘sonority expansion’, focusing not on the jaw, but on the opening of the vocal tract. Stressed syllables would be at a higher level in a sonority scale since more energy can radiate from the mouth, which is mainly based on an open vocal tract and a longer hold open vocal tract configuration. Results of this study, particularly tongue dorsum position at closure offset for JD, do not confirm Beckman, Edwards and Fletcher’s proposal of a ‘sonority expansion’.

So far results fit best with the proposal from de Jong (1995). De Jong’s (1995) model of stress as localised hyperarticulation predicts:

> “*that all phonemically distinctive contrasts will be directly affected by stress”* (de Jong 1995, p. 493).

His terminology goes back to Lindblom (1990) who describes speech along a continuum between ‘hypospeech’ and ‘hyperspeech’ with principles of motor economy on the one hand and listeners’ demands concerning communicative effort on the other hand. Hypospeech would be strongly affected by principles of motor economy such as reduction processes, assimilation, coarticulation etc. whereas hyperspeech can be associated with a relatively high amount of communicative effort, i.e. the speaker pronounces carefully with greater phonemic contrast in order to be understood by the listener. Thus, according to de Jong’s model of stress as local hyperarticulation, stress enhances perceptual clarity due to more pronounced differences in production within a particular interval, the syllable. De Jong considered only supralaryngeal articulation, but we need to apply his suggestion to laryngeal adjustment as well as to laryngeal–oral coordination, since these factors are involved in the production of the voicing contrast. The clearest distinctions of the voicing contrast are produced in the stressed position, in terms of laryngeal abduction and of laryngeal-oral co-ordination at the articulatory level and of noise duration at the acoustic level. In both post-stressed positions economical principles of the motor system apply with respect to temporal characteristics (shorter durations in general), an enormous reduction of glottal opening amplitude, and a greater amount of coarticulatory influence due to the preceding vowel environment (interaction effects). Additionally, laryngeal-oral co-ordination for /t/ in the post-stressed positions shows quite variable patterns, suggesting either that the small laryngeal abduction is not controlled by the neural nervous system, but a result of intraoral air pressure changes or that articulatory coupling in functional units is less strong in weak/post-stressed positions compared to strong/stressed positions.

Finally, Fowler's model of stress as an increase of global articulatory effort for stressed syllables could also hold true for the current findings. However, this would only be on the basis of the previous work on EMG data, not on kinematic data as observed here. Previous EMG work showed a greater amount of neuronal activity for phonemes in stressed compared to post-stressed position (e.g. Harris 1978, Hirose and Gay 1972, Hirose and Ushijima 1978).

Future work on stress could investigate the question whether stress affects different articulators to a different degree since in the current dataset laryngeal adjustment seems to be more sensitive to changes in stress as for instance jaw movement in stop production.

## 5.4. The voicing contrast and its links to different theoretical concepts

While the findings here arise purely from an experimental study of German, nevertheless it is possible to extrapolate to other languages' voicing contrasts and to consider some of the broader issues that were originally raised in chapter 1.

### *5.4.1. The phonetics-phonology interface*

1. Results of this study provide evidence that positional factors cannot be separated from the realisation of a segmental contrast. There are intrinsic

properties for the stressed versus post-stressed position (e.g. stressed being longer in overall duration) which do affect the realisation of the voicing contrast. If variability causes the need for abstraction (1.2.1, p.11 citation to Pierrehumbert, Beckman and Ladd 2001), the abstraction would probably be affected by the frequency of occurrence of a certain phenomenon. For German, relatively consistently occurring phonetic phenomena associated with the voicing contrast are duration (generally longer for phonologically voiceless) and aspiration/noise duration (again longer for phonologically voiceless). The high jaw position at closure offset in /t/ responsible for the production of a salient burst could be involved too, but the strength of the salient burst was not investigated here.

2. Not only one, but several acoustic and articulatory correlates are involved in the voicing contrast in German. Thus, a phonological contrast like the voicing contrast in German corresponds to more than one difference at the phonetic level, an observation in agreement with Lisker (1957, 1978, 1986), Jessen (1999, 2001), Kohler (1984), Luce and Charles-Luce (1985). It could be that these various phonetic differences are mutually interdependent and unable to vary independently, thus indicating the redundancy of speech production. But in fact, inter-speaker differences show that many are not unavoidably co-variant. Depending on the frequency with which different independent phonetic correlates occur in the different positions, a hierarchy could be drawn between stronger phonetic correlates of the contrast (Jessen refers to 'primary cues', but the term 'cue' is skewed towards the perception of a correlate, not to its production.). Based on the results of this study, noise duration, closure duration and jaw position at the oral release are stronger correlates of the contrast and weaker correlates are for instance movement amplitude or velocity peak. So far it is unknown whether in the perception of the contrast several phonetic correlates are integrated into one unit (as Malécot proposed, p.10) or whether listeners focus on a particular correlate and compensate if it was not produced. Further work is necessary to investigate this issue.

### *5.4.2. The role of timing and its possible control*

With respect to the acoustic results, timing in general is a clear correlate involved in the voicing contrast (see 1.2.2.) whereas articulatory results do not show the same trend that frequently. However this is partly because some of the more robust acoustic measures which are relevant to the voicing contrast are timing related, though non-timing related correlates do also exist, and because to allow a broader range of measurements to be made in this study fewer articulatory measurements were made which specifically target timing. The results of this study taken as a whole therefore support the incorporation of timing into phonology. The voicing contrast in German can be modelled with longer acoustic noise duration for /t/ compared to /d/ in the relevant position. Closure and vowel duration are other relatively stable acoustic characteristics to distinguish between both cognates. Additionally, prosodic factors such as position and stress influence temporal characteristics of the contrast.

Looking at the differences between the acoustic and articulatory results, the question arises how and whether the timing of particular acoustic events themselves are controlled, whether timing is a consequence of the articulatory movement from a static target position to the next (Perrier 2003) or whether timing is a consequence of articulatory stiffness (Browman and Goldstein 1989). In the second case it is not timing, but target positions that would be controlled, as summarised by Perrier (2003) and experimentally tested by our work on the voicing contrast in Korean velar stops (Brunner et al. 2004). Perrier (2003, in press) assumes a simple internal representation including the knowledge about mass, stiffness, and inertia of articulators. During speech acquisition the speaker learns about these physical properties.

Such a model supposes target positions without a precise control of timing, and can hold for the results of the voicing contrast in the post-stressed word medial position. A lower vertical target position for tongue and jaw position at the alveolars produced for /d/ would have X effect with everything else being equal. X is:

- tongue palatal contact patterns (with less anterior contact for /d/),
- closure duration as well as overall duration (shorter for /d/), and
- intraoral pressure - a shorter oral closure would coincide with a lower intraoral pressure and a greater likelihood to produce voicing.

A lower target for /d/ with a smaller amount of tongue palatal contact patterns and a shorter closure duration (+ longer voicing during closure without quantitative evidence) has been confirmed in this study for the post-stressed word medial position. Hence, a very economic and simple explanation - a lower target position for /d/ - can describe the manifold results at different levels. In addition, it supports de Jong’s concept on stress as localised hyperarticulation in a sense that in a non-stressed position (here post-stressed) motor economy principles apply.

However, in stressed position a model based on different target positions for /d/ and /t/ would not explain the temporal findings as well as the articulatory results (longer closure duration for /d/, shorter noise duration for /d/, glottal abduction for /t/, no differences in tongue-palatal contact patterns etc.). It can be hypothesised therefore that timing is a consequence of the movement from one target to the next, **but only in those positions where economical principles apply** (hypospeech in Lindblom's terms). In strong position, where communicative effort (listeners' demands) plays a major role, timing or interarticulatory co-ordination seems to be under specific control.

In the third case, it is not timing, but stiffness of the articulators that would be controlled with a stiffer articulator corresponding to faster movements and smaller movement amplitudes in comparison to a less stiff articulator. If the results of this study support such a model then changes in stiffness affect articulators to a different extent. Changes in stiffness would have to be greater for laryngeal adjustment (glottal opening is reduced considerably or even diminishes) than for tongue movement.

So far Perrier's proposal seems to fit the results of the current study best.

### *5.4.3. The stability of interarticulatory co-ordination*

As was pointed out in section 1.2.3. Kelso, Saltzman and Tuller (1986) proposed a concept of flexible and task-dependent 'coordinative structures'. The authors assume that although kinematic trajectories change due to variations of stress (or speech rate) the interarticulatory timing relations would be stable. In opposition, Alfonso and van Lieshout (1999) provided evidence for considerable variations for the relative timing of different articulators.

The results of the current study exhibit stable laryngeal-oral co-ordination for /t/ in the STRESSED position, but variable patterns for the same phoneme in the POST-STRESSED positions (from a similar timing as in the stressed position to an remarkably earlier glottal opening onset and glottal opening peak with respect to supralaryngeal events). Thus, the stability of laryngeal-oral timing (laryngeal abduction supposed) depends to a great extent on the relevant position. Based on the current findings Kelso, Saltzman and Tuller's concept of invariant interarticulatory co-ordination, i.e. relative timing, cannot be supported. However, the differences in co-ordination may also be explained by the fact that /t/ in stressed position is clearly aspirated (noise duration 37-77ms), but in the post-stressed position it can be either aspirated or unaspirated (noise duration 15-31ms). If the realisation of the same phoneme changes from aspirated to unaspirated then laryngeal-oral co-ordination changes too. In that sense Kelso, Saltzman and Tuller's suggestion may still hold true for the same allophonic variation of a phoneme - here it would be an aspirated stop. If the authors assume stability of interarticulatory co-ordination for all variations of one phoneme, the results of the current study do clearly exhibit a counter-example with a large variability.

# Chapter 6: Summary, conclusion, and perspectives

The present study was dedicated to the voicing contrast in German alveolars and its articulatory variations across several positions. To summarise, it extends previous work on mainly acoustic characteristics of the contrast (Jessen 1998) to **articulatory correlates** and relates these to the corresponding acoustic output. It was hypothesized that the voicing contrast is in fact based not only on differences at the laryngeal level, but is rather a **complex phenomenon involving laryngeal and supralaryngeal production mechanisms** as well as **laryngeal-supralaryngeal co-ordination**. Specific positions of the alveolars in the word were considered which are known to show acoustic differences in the realisation of the voicing contrast:

- The intervocalic syllable/morpheme initial stressed position where the contrast is mainly due to aspiration duration.
- The intervocalic syllable initial post-stressed position where the contrast is realised as a ‘real’ voicing distinction.
- The intervocalic post-stressed word final position. Particular attention was given

to this position, since it is ruled by **final devoicing** where phonologically voiced obstruents should be neutralised to voiceless.

Three native speakers of German were analysed by means of simultaneous transillumination, fiberoptic films, EPG recordings (experiment 1) and an EMA, EPG recording (experiment 2).

1. One of the major outcomes of this study is a **missing glottal abduction for /t/ as well as for /d/** in post-stressed positions (found for two of the three subjects concerning /t/). Some small amount of noise was however measured after oral release in the acoustics for /t/. The question arises immediately, how in the case of /t/ can noise be produced without an open glottis? It has been hypothesized in the current study that supralaryngeal mechanisms can compensate for the lack of glottal adjustment by means of a high tongue and jaw position increasing the intraoral pressure and therefore the density of air particles. In general, noise is realised by means of an obstruction where air particles pass through and get a higher density. The obstruction and the density of air particles cause an increase in velocity and thus, in airflow. Turbulent airflow (acoustically noise) is produced when the air jet leaves the obstruction. Mooshammer, Hoole and Geumann (submitted) suggest that a high jaw position ensures the production of a salient burst. They point to the role of the lower teeth as

an obstacle. Since the teeth are close to the constriction location, the noise of the explosion might be enhanced. This is not the case for voiceless bilabial and velar stops. Thus, it is concluded here that the missing glottal abduction in the post-stressed positions is a potential characteristic of voiceless **alveolar** stops, but may not pertain for bilabials and velars. A high jaw position at closure offset for /t/ is the most stable articulatory supralaryngeal correlate in all positions (but it is more pronounced in the post-stressed positions) in this study. It was found for all speakers (except CG in stressed position and JD in word final position). To further investigate the potential role of the teeth in alveolar stop production, future work should involve vocal tract modelling with and without teeth or perturbation studies including, for instance, subjects with a dental prostheses who can put them in and take them out.

Additionally, the fact that (some) speakers use supralaryngeal production mechanisms to compensate for the lack of glottal abduction provides empirical evidence that speech is planned towards acoustic goals rather than articulatory targets (e.g. Perkell et al. 1997, Perrier 2005). This does not mean that articulatory modalities are not of any relevance, but that acoustic goals are dominant in the speech production task. Regarding the voicing contrast in **post-stressed positions**, noise can be realised by means of laryngeal-supralaryngeal co-ordination or supralaryngeal mechanisms. However, the aspiration noise in the stressed position is too pronounced in duration and intensity to be produced by a high tongue and jaw position; glottal abduction is required as well. Hence, the strength of the acoustically required noise leads to a laryngeal-oral co-ordination in the stressed position, but to a potential (speaker dependent) supralaryngeal compensation in the post-stressed positions. Another fact supporting the importance of acoustic goals becomes apparent when comparing acoustic and articulatory inter-speaker variations. In general, acoustic results between speakers are more coherent, less ambiguous and less variable than articulatory results.

2. Another main achievement of the current study is the **variety of laryngeal and supralaryngeal correlates** found during the realisation of the voicing contrast. Indeed, those were linked to the different positions. With respect to the **phonetics-phonology interface** the findings challenge the idea that a phonemic minimal pair such as the voicing contrast can be separated by **one** specific production mechanism only (as implied, for example, in the IPA chart). However, we may have one particular representative or abstraction of a phonemic contrast in mind which has been learnt during speech acquisition. This representative would be based on the perception of variable phonemes in their relevant context. As

discussed in section 1.2.1., Pierrehumbert, Beckman and Ladd (2001) have argued that variability causes the need for abstraction. Following up on this idea, what could be an appropriate abstraction taking the results of this study into account? It should be noted that the abstraction is primarily an acoustic correlate since the main representatives of speech production are auditory goals as concluded above. If the abstraction is acquired due to the frequency of occurrence of a particular acoustic characteristic then noise duration would be the characteristic on top of a hierarchy. If prominence of a syllable (here stressed versus post-stressed) is more important for building an abstraction then aspiration duration (assuming it is above 20ms) - not noise duration would be the appropriate cue, since aspiration duration coincides with the stressed position which is less frequent than any unstressed position (German words contain on average 2-3 syllables and one of them is stressed). Furthermore the dominance of noise or aspiration duration as a potential abstract cue of the voicing contrast in German should be supplemented by information on their perceptual strength. It seems difficult to imagine that a cue like voicing into closure with a relatively weak intensity could be the most important perceptual cue when ordinary speech, **including background noise,** is considered. Future work investigating the perception of single and multiple cues of the voicing contrast with and without background noise would enhance our understanding of the voicing contrast.

3. Another contribution of the current study is that it extends previous work on **final devoicing** in German from an **acoustic to an articulatory level.** To my knowledge no data were available at least for laryngeal production mechanisms, although there have been speculations that an incomplete neutralisation of German phonologically voiced obstruents is realised due to a missing glottal abduction (e.g. Gafos submitted). As was found in two of the three speakers, glottal abduction is missing not only for /d/ but, interestingly, also for /t/. Additionally, small articulatory differences occur generally more frequently at the supralaryngeal level. A comparison of acoustic and articulatory results revealed more stable patterns in the acoustics suggesting a complete neutralisation. With respect to articulatory results, however, neutralisation is clearly not complete. Consequently, the discussion on complete versus incomplete neutralisation can neither be considered in isolation from facts of acoustics, nor just on speech production mechanisms alone. It is highly probable that speaker specific articulatory residues of the contrast will be maintained, even though strong, general and reliable acoustic cues of the contrast do not occur. The perceptual importance of such articulatory residues may be minimal,

especially if speakers' purpose is to neutralise. One **exception** in the current dataset is the speaker who grew up in **Southern Germany,** close to Lake Constance. His results did **not show any clear acoustic or articulatory neutralisation:** a small amount of glottal abduction was found for /t/, but not for /d/, resulting in differences in noise duration. These findings are in agreement with Piroth and Janker (2004) who reported incomplete neutralisation for Bavarian speakers. In the current study subject CG does not show obvious dialectal variations in his daily speech. Additionally, he has lived in Berlin for more than 10 years. Thus, it can be concluded that regional variations of a speaker's origin who migrated to another area a long time ago are still present, although perceptually they are very weak.

4. The implication of the current work was further discussed with respect to the following theoretical concepts: **Stress effects** and different **models of stress, the role of timing** and its **possible control,** and the **stability of interarticulatory co-ordination.**

   **Stress effects and different models of stress:** Stress has a major impact on articulatory production mechanisms of the voicing contrast in alveolar stop production in German. More specifically, the amplitude of glottal abduction in /t/ following the stressed syllable (post-stressed position) was reduced considerably or did even disappear. Supralaryngeal production mechanisms are not affected with a similar strength. Comparing these findings with different models of stress (the jaw expansion model (Macchi 1985), the sonority expansion model (Beckman, Edwards and Flechter 1992), stress as 'localized hyperarticulation' model (de Jong 1995) and stress as 'an increase of global articulatory effort' model (Fowler 1995)) it is evident that de Jong's proposal corresponds best to the results. It is concluded that the **stressed position is 'hyperarticulated'** when the voicing contrast is primarily based on **glottal abduction tightly linked with oral events,** resulting acoustically in a long aspiration duration for /t/. For phonologically **voiced stops no glottal abduction** is produced and therefore a comparatively small amount of noise is found after the burst. Additionally, the voicing contrast is perceptually enhanced by means of the silent closure period (similar in duration for /d/ and /t/) preceding the aspiration noise and the burst. The mechanisms described for the **stressed position** are realised with respect to the **listener's demands.** On the contrary, the results for the **post-stressed positions** follow **economical principles of the motor system -** glottal abduction is reduced, or supralaryngeal mechanisms can compensate for a missing glottal abduction. Furthermore, temporal differences of the contrast become smaller.

The importance of stress on articulatory production mechanisms of the voicing contrast may be further investigated by comparing the current dataset with non-stressed languages such as French (with a phrasal accent, but no word accent) or Japanese (a mora timed language). It is predicted that positional differences should be smaller than in the study discussed here.

**The role of timing and its possible control:** The results of this study provide further evidence (see e.g. also Port and Leary 2003) of the importance of timing in order to understand the voicing contrast. Additionally, the results allow some predictions about the question of whether timing itself is controlled or not. For the post-stressed positions, differences in tongue tip and jaw target position (lower target for /d/ than for /t/) can explain the articulatory and acoustic differences found. Thus, it is concluded that timing is a consequence of the movement from one target to the next, **but only in those positions where economical principles apply** (hypospeech in Lindblom's terms). In strong position, where communicative effort (listeners' demands) plays a crucial role, timing or interarticulatory co-ordination seems to be the main object of the control since the differences between /t/ and /d/ can not be explained by a movement from one target to the next.

**The stability of interarticulatory co-ordination:** The outcome of this work sheds some light on the 'coordinative structure' concept (see Kelso, Saltzman and Tuller 1986) where the authors assume stability in interarticulatory timing. The current findings provide an excellent example that when considering interarticulatory co-ordination at a phoneme level, it is clearly not stable. However, when taking allophonic variations of one phoneme such as [$t^h$] and [t] into account (assuming some small amount of glottal abduction occurs for the unaspirated stop), laryngeal-supralaryngeal co-ordination is more stable. The greatest stability of interarticulatory co-ordination can be found with respect to the prominent (stressed) syllable position and therefore it is concluded that one characteristic of prominence is a strong cohesion between the involved articulators. Since prominence often coincides with a longer syllable duration and probably with a greater articulatory effort, future work regarding speech motor control could consider duration and effort as two effects which may be responsible for the interarticulatory linkage.

Although more work is needed, particularly on the aerodynamics and perception of the voicing contrast, I hope this work has provided a comprehensive picture on articulatory correlates of the voicing contrast in alveolar obstruent production in German.

# References

Abramson, A.S. and Lisker, L. (1967) Discrimination along the voicing continuum: Cross-language tests. *Haskins Laboratories: Status Report on Speech Research*: 17-22.

Andersen, P. (1981) The effect of increased speaking rate and (intended) loudness on the glottal behaviour in stop consonant production as exemplified by Danish p. *Annual Report of the Institute of Phonetics University of Copenhagen* 15:103-146.

Alfonso, P.J. and van Lieshout, P. (1999) Flexibility and stability in bilabial gestures: 1) Evidence from discrete syllable production. *In Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco,* USA: 49-52.

Allen, J.S., Miller, J.L. and DeSteno, D. (2003): Individual talker differeneces in voice-onset-time. *Journal of the Acoustical Society of America* 113(1): 544- 552.

Baer, T., Löfqvist, A. and McGarr, N.S. (1983) Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques. *Journal of the Acoustical Society of America* 73(4): 1304-1308.

Baken, R.J. (1987) *Clinical measurement of speech and voice.* Little Brown Company, London.

Baum, S.R. and McFarland, D.H. (1997) The development of speech adaptation to an artificial palate. *Journal of the Acoustical Society of America* 102(4): 2353-2359.

Beckman, M., Edwards, J. and Fletcher, J. (1992) Prosodic structure and tempo in a sonority model of articulatory dynamics. In Docherty, G.J. & Ladd, D.R. (Eds.) *Papers in laboratory phonology II: gesture, segment, prosody.* Cambridge University Press, Cambridge: 68–86.

Beckman, M.E.and Pierrehumbert, J. (2000) Positions, probabilities, and level of categorization, Keynote address at the 8th Australian International Conference on Speech Science and Technology, Canberra, Australia. (see http://www.ling.nwu.edu/~jbp/conference.html)

Bell-Berti, F. and Hirose, H. (1971) Velopharyngeal function in oral/nasal articulation and voicing gestures. *Haskins Laboratories: Status Report on Speech Research* SR-28: 143-156.

Bell-Berti, F. and Hirose, H. (1972): Velar activity in voicing distinctions: a simultaneous fiberoptic and electromyographic study. *Haskins Laboratories: Status Report on Speech Research* SR-31/32: 223-229.

Bell-Berti (1975) Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of the Acoustical Society of America* 57(2): 456-461.

Benguerel, A.P., Hirose, H., Sawashima, M. and Ushijima, T. (1978) Laryngeal control in French stop production: a fiberscopic, acoustic and electromyographic study. *Folia phoniatrica* 30: 175-198.

Benguerel, A.P. and Bhatia, T.K. (1980) Hindi stop consonants: an acoustic and fiberscopic study. *Phonetica* 37: 149-158.

Brockhaus, W. (1995) Final devoicing in the phonolgy of German. *Linguistische Arbeiten 337*, Niemeyer: Tübingen.

Browman, C.P. and Goldstein, L. (1986) Towards an articulatory phonology. *Phonology 3*: 219-252.

Browman, C.P. and Goldstein, L. (1989) Articulatory gestures as phonological units. *Phonology 6*: 201-252.

Brunner, J., Fuchs, S., Perrier, P. and Kim, H.Z. (2003) Mechanisms of contrasting Korean velar stops: A catalogue of acoustic and articulatory parameters. *ZAS-papers in Linguistics* 32: 15-30.

Brunner, J., Fuchs, S., Perrier, P. and Kim, H.-Z. (2004) Correlation between angle of incidence and sliding patterns of the tongue along the palate in Korean velar stops. *147th ASA-meeting, New York* [Abstract & Poster]

Butcher, A. (1977) Coarticulation in intervocalic voiceless plosives and fricatives in connected speech. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel* 8: 154-212.

Byrd, D., Flemming, E., Mueller, C.A. and Tan, C.C. (1995) Using regions and indices in data reduction. *Journal of Speech and Hearing Research* 38, 821-827.

Chen, M. (1970) Vowel length variation as a function of the voicing of the consonant environement. *Phonetica* 22: 129-159.

Cho, T. and Ladefoged, P. (1999): Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics* 2: 207-229.

Cho, T., Jun, S.A. and Ladefoged, P. (2002) Acoustis and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30: 193-228.

Collier, R., Lisker, L., Hirose, H. and Ushijima, T. (1979) Voicing in intervocalic stops and fricatives in Dutch. *Journal of Phonetics* 7: 357-373.

Chomsky, N. and Halle, M. (1968) *The Sound Pattern of English*. Harper & Row Publishers, New York/Evanston/London.

Cooper, A.M. (1991) Laryngeal and oral gestures in English /p, t, k/. *In Proceedings of the 12th International Congress of Phonetic Sciences Aix-en-Provence*, Vol. 2: 50-53.

Dagenais, P.A., Lorendo, L.C. and McCutcheon; M.J. (1994) A study of voicing and context effects upon consonant linguapalatal contact patterns. *Journal of Phonetics* 22: 225-238.

Debrock, M. (1977) An acoustic correlate of force of articulation. *Journal of Phonetics* 5: 61-80.

de Jong, K.J., Beckman, M.E. and Edwards, J. (1993) The interplay between prosodic structure and coarticulation. *Language and Speech* 36(2,3): 197-212.

de Jong, K.J. (1995) The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *Journal of the Acoustical Societyof America* 97: 491–504.

de Jong, K.J. and Zawyadeh, B.A. (2002) Comparing stress, lexical focus, and segmental focus: Patterns of variation in Arabic vowel duration. *Journal of Phonetics* 30: 53-75.

Dixit, R.P. and MacNeilage, P. F. (1980) Cricothyroid activity and control of voicing in Hindi stops and affricates. *Phonetica* 37: 397-406.

Dixit, R.P. (1987) Mechanisms for the control of voicing and aspiration in plosive production. *UCLA Working Papers in Phonetics* 67: 49-102.

Dixit, R.P. (1989) Glottal gestures in Hindi plosives. *Journal of Phonetics* 17: 213-237.

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

Docherty, G. J. (1992) *The timing of voicing in British English obstruents*. Foris Publications, Berlin/New York.

Earnest, M. and Max, L. (2001) Maxillary and mandibular morphometry and the relation to articulatory kinematics. In *Proceedings of the 4th International Speech Motor Conference, Nijmegen*, The Netherlands: 134-137.

Eimas, P.D., Siqueland, E.R., Jusczyk, P. and Vigorito, J. (1971) Speech perception in infant. *Science* 171: 303-306.

Eisenberg, P. (1991) Syllabische Struktur und Wortakzent; Prinzipien der Prosodik deutscher Wörter. *Zeitschrift für Sprachwissenschaft* 10 (1): 37-64.

## References

Elgendy, A. M. (2001) *Aspects of pharyngeal coarticulation*. ACLC: LOT Netherlands Graduate School of Linguistics.

Esposito, A. (2002) On vowel height and consonantal voicing effects: data from Italian. *Phonetica* 69:197-231.

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

Fischer-Jørgensen, E. and Hirose, H. (1974) A preliminary electromyographic study of labial and laryngeal muscles in Danish stop consonant production. *Haskins Laboratories: Status Report on Speech Research* SR-39/40: 231-254.

Fischer- Jørgensen, E. (1976) Some data on North German stops and affricates. *Annual Report of the Institute of Phonetics, University of Copenhagen* 10: 149-200.

Fiukowski, H. (1992) *Sprecherzieherisches Elementarbuch*. Max Niemeyer Verlag: Tübingen.

Fletcher, S. (1989) Palatometric specification of stop, affricate, and sibilant sounds. *Journal of Speech and Hearing Research* 32: 736-748.

Folkins, J. and Abbs, J. (1975) Lip and jaw motor control during speech: Responses to resistive loading of the jaw. *Journal of Speech and Hearing Research* 18: 207-220.

Fontdevila, J., Pallarès, M.D. and Recasens, D. (1994) The contact index method of electropalatographic data reduction. *Journal of Phonetics* 22: 141-154.

Fourakis, M. and Iverson, G. (1984) On the 'incomplete neutralization' of German final obstruents. *Phonetica* 41: 140-149.

Fowler, C.A. (1980) Coarticulation and theories of extrinsic timing. *Journal of Phonetics 8*: 113-133.

Fowler, C.A. and Turvey, M.T. (1980) Immediate compensation in bite-block speech. *Phonetica* 37: 306-326.

Fowler, C. A. (1995) Acoustic and kinematic correlates of contrastive stress accent in spoken English. In Bell-Berti, F. and Raphael, L.J.(eds.): *Producing speech: contemporary issues for Katherine Saffort Harris.* American Institute of Physics, Woodbury/New York: 355–373.

Francis, A.L., Ciocco, V. and Yu, J.M.C. (2003): Accuracy and variability of acoustic measures of voicing onset. *Journal of the Acoustical Society of America* 113(2): 1025-1032.

Frøkjær-Jensen (1967): A photo-electric glottograph. *Annual Report of the Institute of Phonetics, University of Copenhagen* 2: 5-19.

Frøkjær-Jensen, B., Ludvigsen, C. and Rischel, J. (1971): A glottographic study of some Danish consonants. 123-140. In Hammerich, L.L., Jakobson, R. and Zwirner, E. (eds.): *Form and Substance.* Akademisk, Copenhagen.

Fuchs, S., Perrier, P. and Mooshammer, C. (2001) The role of the palate in tongue kinematics: an experimental assessment in VC sequences from EPG and EMMA data. *In Proceedings of Eurospeech. Aalborg, Danemark*, Vol. 3:1487-1490.

Fuchs, S., Mooshammer, C. and Hoole, P. (2002) The timing of laryngeal and lingual gestures of apical stops in dependence on syllable position: an EPG and transillumination study. *Conference on Motor Speech: Motor Speech Disorders & Speech Motor Control, Williamsburg* U.S. [Abstract & Poster]

Fuchs, S. and Perrier, P. (2003): An EMMA/EPG study of voicing contrast correlates in German. In *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*: 1057-1060.

Fuchs, S., Perrier, P. Geng, C. and Mooshammer, C. (2003) What role does the palate play in speech motor control? Insights from tongue kinematics for German alveolar obstruent production. *CD-ROM Speech Production Seminar Sydney*: 61-66.

Fuchs, S., Hoole, P, Brunner, J. and Inoue, M. (2004a) The trough effect: an aerodynamic phenomenon? *CD-ROM MIT-meeting From Sound to Sense.* C25-C30.

Fuchs, S., Hoole, P., Pelorson, X., Van Hirtum, A., Perrier, P., Dahlmeier, K. and Creutzburg, J. (2004b) Laryngeal adjustment in voiceless consonant production: I. An experimental study of glottal abduction in loud versus normal speech. In *Proceedings of the conference on Voice physiology and Biomechanics*: 39-43.

Fuchs, S., Scobbie, J., Perrier, P. and Wrench, A. (2004c) Final obstruent devoicing and neutralisation in German: an articulatory perspective. *Manchester Phonology Meeting* [poster].

Fujimura, O. and Sawashima, M. (1971) Consonant sequences and laryngeal control. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 5: 1-6.

Fujimura, O., Tatsumi, I.F. and Kagaya, R. (1972) Computational processing of palatographic patterns. *Journal of Phonetics* 1: 47-54.

Fujimura, O. (1977) Control of the Larynx in Speech. *Phonetica* 34: 280-288.

Fujimura, O., Baer, T. and Niimi, S. (1979) A stereo-fiberscope with a magnetic interlens bridge for larnygeal observation. *Journal of the Acoustical Society of America* 65(2): 478-480.

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

Fujino, A., Kaburagi, T., Honda, M., Murando, E. Z. and Niimi, S. (2000) Relative timing between articulatory and glottal motions in voiceless consonant production. *In Proceedings of the 5th Seminar on Speech Production: Models and Data, Kloster Seon, Germany*: 89-92.

Fukui, N. and Hirose, H. (1983) Laryngeal adjustments in Danish voiceless obstruent production. *Annual Report of the Institute of Phonetics, University of Copenhagen* 17: 61-71.

Gafos, A.I. (submitted) Dynamics: the non-derivational alternative to modeling phonetics-phonology. *Labphon 8*, New Haven.

Gandour, J. and Maddieson, I. (1976): Measuring larynx movement in Standard Thai using the cricothyrometer. *Phonetica* 33:241-267.

Gay, T. and Hirose, H. (1972) Electromyography of the intrinsic laryngeal muscles during phonation. *Annals of Otology, Rhinology and Laryngology* 81(3): 401-409.

Gibbon, F. and Nicolaidis, K. (1999) Palatography. In: Hardcastle, W.J. and Hewlett N. (eds.) *Coarticulation in Speech Production: Theory, Data, and Techniques.* Cambridge University Press, Cambridge: 229-245

Giegerich, H.J. (1989) *Syllable structure and lexical derivation in German*. Indiana University Linguistics Club.

Glück, H. (ed.) (1993) *Metzler Lexikon Sprache*. Metzlersche Verlagsbuchhandlung und C.E. Poeschel Verlag, Stuttgart/Weimar.

Gracco, V.L. and Abbs, J.H. (1986) Variant and invariant characteristics of speech movement. *Experimental Brain Research* 65: 156-166.

Gracco, V. L. (1988) Timing factors in the coordination of speech movements. *The Journal of Neuroscience* 8(12): 4628-4639.

Gracco, V.L. and Löfqvist, A. (1994) Speech motor coordination and control: evidence from lip, jaw, and laryngeal movements. *The Journal of Neuroscience* 14(11): 6585-6597.

Gracco, V.L. (2001) Perturbations of speech motor output: history, interpretation and future directions. *Crest Workshop on Speech Motor Control and Modeling. Sapporo*, Japan. [Abstract]

# References

Habermann, G. (1986) *Stimme und Sprache. Eine Einführung in ihre Physiologie und Hygiene. Für Ärzte, Sänger, Pädagogen und alle Sprechberufe*. (2nd ed.) Thieme Verlag, Stuttgart/New York.

Hall, T.A. (2001) Introduction: Phonological representations and phonetic implementation of distinctive features. In Hall, T.A. (ed.) *Distinctive feature theory.* Mouton DeGruyter, Berlin: 1-40.

Halle, M. and Stevens, K. (1971) A note on laryngeal features. *MIT Reserach Laboratory Electronics Quaterly Progress Report* 101: 198-213.

Hamlet, S. and Stone, M. (1978) Compensatory alveolar consonant production induced by wearing a dental prothesis. *Journal of Phonetics* 6: 227-248.

*Handbook of the International Phonetic Association* (1999). Cambridge University Press, Cambridge.

Hanson, H. M. (1997) Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America* 101**:** 466–481.

Hardcastle, W.J. (1970) *Electropalatography in the investigation of some physiological aspects of speech production*. Thesis at University of Edinburgh.

Hardcastle, W.J. (1976) *Physiology of speech production*. Academic Press, London/New York/San Francisco.

Hardcastle, W.J., Gibbon, F. and Nicolaidis, K. (1991): EPG data reduction and their implications for studies of lingual coarticulation. *Journal of Phonetics* 19, 251-266.

Hardcastle, W.J. (1999) Electromyography. In Hardcastle, W.J. and Hewlett, N.(eds.): *Coarticulation: theory, data and techniques*. University Press, Cambridge: 270-283.

Harrington, J., Fletcher, J. and Roberts, C. (1995) Coarticulation and the accented/unaccented distinction: evidence from jaw movement data. *Journal of Phonetics* 23: 305-322.

Hauger, W., Schnell, W and Gross, D. (1999) *Technische Mechanik.* Volume 3: Kinetik. (6th ed.) Springer Verlag, New York/Berlin/Heidelberg.

Hawkins, S. (1999a) Auditory capacities and phonological development: animal, baby, and foreign listeners. In Pickett, J.M. (1999) *The acoustics of speech communication. Fundamentals, speech perception theory and technology.* Allyn and Bacon, Boston/London/Toronto/Sydney/Tokyo/Singapore: 183-197.

Hawkins, S. (1999b) Looking for invariant correlates of linguistic units: two classical theories of speech perception. In Pickett, J.M. (1999) *The acoustics of speech communication. Fundamentals, speech perception theory and technology.* Allyn and Bacon, Boston/London/Toronto/Sydney/Tokyo/Singapore: 198- 232.

Hawkins, S. (1999c) Reevaluating assumptions about speech perception: interactive and integrative theories. In Pickett, J.M. (1999) *The acoustics of speech communication. Fundamentals, speech perception theory and technology.* Allyn and Bacon, Boston/London/Toronto/Sydney/Tokyo/Singapore: 233-288.

Hirano, M. and Ohala, J. (1969) Use of hooked-wire electrodes for electromyography of the intrinsic laryngeal muscles. *Journal of Speech and Hearing Research* 12: 362-373.

Hirose, H. and Gay, T. (1972) The activity of the intrinsic laryngeal muscles in voicing control. *Phonetica* 25: 140-164.

Hirose, H. Lisker, L. and Abramson, A. (1972) Physiological aspects of certain features in stop production. *Haskins Laboratories: Status Report on Speech Research* SR-31/32: 183-191.

Hirose, H. Lee, C.Y. and Ushijima, T. (1974) Laryngeal control of Korean stop production. *Journal of Phonetics* 2: 145-152.

Hirose, H. (1975) The posterior cricoarytenoid as a speech muscle. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 9: 47-66.

Hirose, H. Yoshioka, H. and Niimi, S. (1978) A cross language study of laryngeal adjustment in consonant production. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 12: 61-71.

Hirose, H. and Ushijima, T. (1978) Laryngeal control for voicing distinction in Japanese consonant production. *Phonetica* 35: 1-10.

Hirose, H. (1979) Laryngeal EMG. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 13: 13-22.

Hirose, H. and Sawashima, M. (1981) Functions of the laryngeal muscles in speech. In Stevens, K.N. and Hirano, M. (eds.): *Vocal Fold Physiology*. University of Tokyo Press: 137-154.

Hirose, H. and Niimi, S. (1987) The relationship between glottal opening and the transglottal pressure differences during consonant production. 381-390. In Baer, T. Sasaki, C. and Harris, K.S. (eds.): *Laryngeal function in phonation and respiration.* College-Hill Press, Boston.

Hoole, P., Pompino-Marschall, B. and Dames, M. (1983) Glottal timing in German voiceless occlusives. In Van den Broecke, M.P.R and Cohen, A.(eds.): *Proceedings of the 10th International Congress of Phonetic Sciences, Utrecht*, The Netherlands: 309-403.

Hoole P. (1987) Velar and glottal activity in a speaker of Icelandic. *Proceedings of the 11th International Congress on Phonetic Sciences, Tallinn,* Estonia 3: 31-34.

Hoole, P. (1993) Methodological considerations in the use of Electromagnetic Articulography in phonetic research. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 31: 43-64.

Hoole, P. (1996) Theoretische und methodische Grundlagen der Artikulationsanalyse in der experimentellen Phonetik. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 34: 3-173.

Hoole, P. (1996) Issues in the aquisition, processing, reduction and parameterisation of articulographic data. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 34, 158-173.

Hoole, P. and Kroos, C. (1998) Control of larynx height in vowel production. *Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Australia*, 2: 531-534.

Hoole, P. (1999) Laryngeal coarticulation. Section A: Coarticulatory investigations of the devoicing gesture. In Hardcastle, W. and Hewlett, N.(eds.): *Coarticulation: theory, data and techniques.* Cambridge: University Press: 105-121.

Hoole, P. (1999) Techniques for investigating laryngeal articulation. Section A: Investigation of the devoicing gesture. In Hardcastle, W. and Hewlett, N.(eds.): *Coarticulation: Theory, data and techniques.* Cambridge: University Press: 294-300.

Hoole, P. and Nguyen, N. (1999) Electromagnetic articulography in coarticulation research. In Hardcastle, W. and Hewlett, N.(eds.): *Coarticulation: Theory, data and techniques.* Cambridge: University Press: 260-269.

Hoole, P., Fuchs, S. and Dahlmeier, K. (2003) Interarticulatory timing in initial consonant clusters. *In Proceedings of the 6th Speech Production Seminar, Sydney*: 101-106.

Hoole, P., Honda, K., Murano, E., Fuchs, S. and Pape, D. (2004) Cricothyroid activity in consonant voicing and vowel intrinsic pitch. *Proceedings of the Conference on Voice Physiology and Biomechanics, Marseille.*

Honda, K., Maeda, S., Hashi, M., Dembowski, J.S., and Westbury J.R. (1996) Human palate and related structures: their articulatory consequences. *In Proceedings of the 4th International Conference on Speech and Language Processing, Philadelphia*, USA: 784-787.

# References

Honda, M. and Kaburagi, T. (2000) Speech compensation to dynamical structural perturbation of the palate shape. *In Proceedings of the 5th Seminar on Speech Production: Models and Data, Kloster Seeon*, Germany: 21-24.

Honda, M., Fujino, A. and Kaburagi, T. (2002) Compensatory responses of speech articulators to unexpected perturbation of the palate shape. *Journal of Phonetics* 30, 281-302.

Hutters, B. (1976) Problems in the use of the photoelectric-glottograph. *Annual Report of the Institute of Phonetics, University of Copenhagen* 10: 274-312.

Hutters, B. (1978) The glottal gesture in some Danish consonants – preliminary observations. *Annual Report of the Institute of Phonetics, University of Copenhagen* 12: 87-101.

Hutters, B. (1984) Vocal fold adjustments in Danish voiceless obstruent production. *Annual Report of the Institute of Phonetics, University of Copenhagen* 18: 293-385.

Hutters, B. (1985) Vocal fold adjustments in aspirated and unaspirated stops in Danish. *Phonetica* 42: 1-24.

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

Iwata, R. and Hirose, H. (1976) Fiberoptic acoustic studies of Mandarin stops and affricates. *Annual Bulletin Research Institute of Logopedics and Phoniatrics* 10: 47-60.

Iwata, R., Sawashima, M., Hirose, H. and Niimi, S. (1979) Laryngeal adjustments of Fukienese stops. Initial plosivs and final applosives. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 13: 61-81.

Iwata, R., Sawashima, M. and Hirose, H. (1981) Laryngeal movements for syllable-final stops in Cantonese. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 15: 45-54.

Jakobson, R. (1929) Remarques sur l'evolution phonologique du russe. *Travaux du cercle linguistique de Prague* 2.

Jakobson, R. and Waugh, L. (1979) *The sound shape of language*. Harvester Press: Brighton.

Jessen, M. (1995) Glottal opening in German obstruents. *In Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm*, Sweden, Vol. 3: 429-431.

Jessen, M. (1998) *Phonetics and phonology of tense and lax obstruents in German.* John Benjamins Publishing Company, Amsterdam/Philadelphia.

Jessen, M. (1998) Word boundary marking at the glottal level in the production of German obstruents. In Alexiadou, A. Fuhrop, N., Kleinhenz, U. and Law, P. (eds.) Papers on morphology and phonetics. *ZAS Papers in Linguistics* 11 (9): 147-166.

Jessen, M. (1999) Redundant aspiration in German is primarily controlled by closure duration. *In Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco,* USA: 993-996.

Jessen, M. (2001) Phonetic implementation of the distinctive auditory features [voice] and [tense] in stop consonants. In Hall, T.A. (ed.) *Distinctive feature theory.* Mouton DeGruyter, Berlin: 237-294.

Jessen, M. and Ringen, C. (2002) Laryngeal features in German. *Phonology* 12: 189-218.

Johnson, K. (1997) *Acoustic and Auditory Phonetics*. Blackwell, Oxford.

Jones, J. and Munhall, K.G. (2003) Learning to produce speech with an altered vocal tract: The role of auditory feedback. *Journal of the Acoustical Society of America* 113(1): 532-543.

Kagaya, R. (1971) Laryngeal gestures in Korean stop consonants. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 5: 15-23.

Kagaya, R. (1974) A fiberscopic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics* 2: 161-180.

Kagaya, R. and Hirose, H. (1975) Fiberoptic electromyographic and acoustic analysis of Hindi stop consonants. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 9: 27-46.

Kelso, J.A.S., Saltzman, E.L. and Tuller, B. (1986) The dynamical perspective on speech production: data and theory. *Journal of Phonetics* 14: 29-59.

Kent, R.D. and Moll, K.L. (1969) Vocal-tract characteristics of stop cognates. *Journal of the Acoustical Society of America* 46(6):1549-1555.

Kim, C.W. (1970) A theory of aspiration. *Phonetica* 21: 107-116.

Kinnear, P.R. and Gray, C.D. (2001) *SPSS for Windows made simple: Release 10.* Psychology Press Ltd, East Sussex.

Kiritani, S. (1971) X-ray monitoring of the position of the fiberscope by means of computer controlled radiography. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 5: 35-39.

Kiritani, S. and Niimi, S. (1997) High-speed digital image analysis of temporal changes in vocal fold vibration in tremor. p. 53-67. In Kiritani, S., Hirose, H. and Fujisaki, H.: *Speech production and language. In Honour of Osamu Fujimura.* Mouton DeGryuter, Berlin/New York.

Kjellin, O. (1977) Observations on consonant types and “tone” in Tibetan. *Journal of Phonetics* 3: 317-338.

Klatt, D. (1975): Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18: 686-706.

Kohler, K. (1975) Die Instabilität wortfinaler Alveolarplosive im Deutschen: eine elektropalatographische Untersuchung. *Arbeitsberichte des Instituts für Phonetik der Universität Kiel* 5: 47-79.

Kohler, K. (1984) The phonetic explanation in phonology: The feature fortis/lenis. *Phonetica 41*: 150-174.

Kohler, K. (1990) German. *Journal of the International Phonetic Association* 20: 48-50.

Kohler, K. (1992) Gestural reorganization in connected speech: a functional viewpoint on ‘articulatory phonology’. *Phonetica* 49: 205-211.

Kohler, K. (1994) Glottal stops and glottalisation in German. *Phonetica* 51: 38-51.

Kohler, K. (1995) *Einführung in die Phonetik des Deutschen*. (2nd edition) Erich Schmidt Verlag: Berlin.

König, W. (1978) *dtv-Atlas zur deutschen Sprache*. Deutscher Taschenbuchverlag: München.

Krakow, R.A. (1999) Physiological organization of syllables: a review. *Journal of Phonetics* 27: 23-54.

Kröger, B.J. (1998) *Ein phonetisches Modell der Sprachproduktion*. Niemeyer Verlag, Tübingen.

Kroos, C., Hoole, P., Kühnert, B, and Tillmann, H. (1997) Phonetic evidence for the phonological status of the tense-lax distinction. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 35: 17-25.

Kuhl, P.K. and Miller, J.D. (1975) Speech perception by the chinchilla: Voice-voiceless distinction in alveolar plosive consonants. *Science* 190: 69-72.

Kuhl, P.K. (1991) Human adults and human babies show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics* 50: 93-107.

Kühnert, B. (1996) Die alveolar-velare Assimilation bei Sprechern des deutschen und Englischen: Kinematische und perzeptive Grundlagen. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)* 34: 175-392.

## References

Laboissière, R., Ostry, D.J. and Feldman, A.G. (1996) The control of multi-muscle systems: human jaw and hyoid movements. *Biological Cybernetics* 74: 373-384.

Ladefoged, P. (1973) The features of the larynx. *Journal of Phonetics* 1: 73-83.

Ladefoged, P. and Maddieson, I. (1996) *The sounds of the world's languages.* Blackwell Publishers, Oxford.

Latash, M.L. (1993) *Control of human movement*. Human Kinetics Publishers.

Laver, J. (1980) *The phonetic description of voice quality*. University Press, Cambridge.

Lavoie, L. (2001) *Consonant strength: Phonological pattern and phonetic manifestations.* Garland Publications, New York.

Leonhardt, H. (1991) *dtv-Atlas der Anatomie. Innere Organe.* Volume 2. Thieme Verlag, Stuttgart/ New York.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. (1967) Perception of the speech code. *Psychological Review* 74:431-461.

Liberman, A.M. and Mattingly, I.G. (1985) The motor theory of speech perception revised. *Cognition* 21: 1-36.

Liberman, A. M., and Whalen, D.H. (2000) On the relation of speech to language. *Trends in Cognitive Sciences* 4: 187-196.

Lieberman, P., Klatt, D.H. and Wilson, W.A. (1969) Vocal tract limitations on the vowel repertoires of rhesus monkey and other nonhuman primates. *Science* 164: 1185-1187.

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

Lindblom, B., Sussman, H.M., Modarresi, G. and Burlingame, E. (2002) The through effect: implications for speech motor programming. *Phonetica* 59(4): 245-262.

Lindqvist, J. (1972) B. Laryngeal articulation studied on Swedish subjects. *STL-QPSR* 2-3: 10-26.

Lisker, L. (1957) Closure duration and the intervocalic voiced-voiceless distinction in English. *Language* 33(1): 42-49.

Lisker, L. and Abramson, A.S. (1964) A cross-language study of voicing in initial stops. *Word* 20: 384-422.

Lisker, L., Abramson, A.S., Cooper, F.S. and Schvey, M.H. (1969) Transillumination of the larynx in running speech. *Journal of the Acoustical Society of America* 45(6): 1544-1546.

Lisker, L., Sawashima, M., Abramson, A.S. and Cooper, F.S. (1970) Cineradiographic observations of the larynx during voiced and voiceless stops. *Haskins Laboratories: Status Report on Speech Research* SR 21/22: 201-211.

Lisker, L. and Abramson, A.(1971) Distinctive features and laryngeal control. *Language* 47: 767-784.

Lisker (1977) Factors in maintenance and cessation of voicing. *Phonetica* 34: 304-306.

Lisker, L. (1978) Rapid vs. rabid: a catalogue of acoustic features that may cue the distinction. *Haskins Laboratories: Status Report on Speech Research* SR-54: 127-132.

Lisker, L. and Baer, T. (1984) Laryngeal management at utterance-internal word boundary in American English. *Language and Speech* 27 (2): 163-171.

Lisker, L. (1986) "Voicing" in English: a catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech* 29(1): 3-11.

Löfqvist, A. (1975) A study of subglottal pressure during the production of Swedish stops. *Journal of Phonetics* 3: 175-189.

Löfqvist, A. and Pétursson, M. (1978) Swedish and Icelandic stops - A glottographic investigation. In Weinstock J. (ed.) *The Nordic Languages and Modern Linguistics* 3: 454-461. The University of Texas at Austin Press: Austin.

Löfqvist, A. (1980) Interarticulator programming in stop production. *Journal of Phonetics* 8: 475-490.

Löfqvist, A. and Yoshioka, H. (1980) Laryngeal activity in Swedish obstruent clusters. *Journal of the Acoustical Society of America* 68(3): 792-799.

Löfqvist, A., Baer, T. and Yoshioka, H. (1981) Scaling of glottal opening. *Phonetica* 38: 265-276.

Löfqvist, A., McGarr, N.S. and Honda. K. (1984) Laryngeal muscles and articulatory control. *Journal of the Acoustical Society of America* 76(3): 951-954.

Löfqvist, A. and Yoshioka, H. (1984) Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production. *Speech Communication* 3: 279-289.

Löfqvist, A. (1986) Stability and change. *Journal of Phonetics* 14: 139-144.

Löfqvist, A. and McGarr, N. (1987) Laryngeal dynamics in voiceless consonant production. p. 391-402 In Baer, T. Sasaki, C. and Harris, K.S. (eds.): *Laryngeal function in phonation and respiration.* College-Hill Press, Boston.

Löfqvist, A., Baer, T., McGarr, N. and Story, R.S. (1989) The cricothyroid muscle in voicing control. *Journal of the Acoustical Society of America* 85(3): 1314-1321.

Löfqvist, A. (1990) Speech as audible gestures. p. 289-322 In Hardcastle, W., and Marchal, A. (eds.): *Speech production and speech modelling.* Kluwer, Dordrecht.

Löfqvist, A. (1991) Proportional timing in speech motor control. *Journal of Phonetics* 19: 343-350.

Löfqvist, A. and McGowan, R.S. (1992) Influence of consonantal environment on voice source aerodynamics. *Journal of Phonetics* 20: 93-110.

Löfqvist, A. and Gracco, V.L. (1994) Tongue body kinematics in velar stop production: Influences of consonant voicing and vowel context. *Phonetica* 51: 52-67.

Löfqvist, A. and Lindblom, B. (1994) Speech motor control. *Current Opinion in Neurobiology* 4: 823-826.

Löfqvist, A. (1995) Laryngeal mechanisms and interarticulator timing in voiceless consonant production. In Bell-Berti F. and Raphael, L. (eds.) *Producing speech: Contemporary issues for Katherine Safford Harris.* 99-116. American Institute of Physics,Woodbury.

Löfqvist, A. (1997) Theories and models of speech production. p. 405-426 In Hardcastle, W. and Laver, J. (eds.): *The Handbook of Phonetic Sciences*. Blackwell Press, Oxford.

Löfqvist, A. and Gracco, V.L. (1997) Lip and jaw kinematics in bilabial stop consonant production. *Journal of Speech, Language, and Hearing Research* 40(4): 877-893.

Lombardi, L. (1991) *Laryngeal features and laryngeal neutralization*. PhD dissertation Amherst, MA, University of Massachusetts.

Luce, P.A. and Charles-Luce, J. (1985) Contextual effects on vowel duration, closure duration, and consonant vowel ratio in speech production. *Journal of the Acoustical Society of America* 78: 1949-1957.

Macchi, M. (1985) *Segmental and suprasegmental features of lip and jaw articulators*. PhD dissertation, New York University.

Maddieson, I. (2003) Phonological typology in geographical perspective. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*: 719-722.

Malécot, A. and Peebles, K. (1965) An optical device for recording glottal adduction-abduction during normal speech. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikation* 18: 545-550.

Malécot, A. (1966) The effectiveness of intra-oral air-pressure-pulse parameters in distinguishing between stop cognates. *Phonetica* 14, 65-81.

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

# References

Mangold, M. (1978) Zur Frage der Opposition /s/:/z/ im Deutschen. *Frankfurter Phonetische Beiträge 11 = Forum Phoneticum* 16: 23-41.

Martinet, A. (1936) Neutralisation et archiphonème. In: *Ètudes dediées au quatrième Congrès de Linguistes. Travaux du Cercle Linguistique de Prague* 6 : 46-57.

Mascaró, J. (1987) Underlying voicing recoverability of finally devoiced obstruents in Catalan. *Journal of Phonetics* 15: 183-186.

Matsumura, M., Kimura, K., Yoshino, K., Tachimura, T. and Wada, T. (1994) Measurement of palatolingual contact pressure during consonant productions using strain gauge transducer mounted palate plate. *In Proceedings of the 3rd International Conference on Spoken Language Processing, Yokohama*, Japan, Vol. 2: 655-658.

McFarland, D.H., Baum, S.R., Chabot, C. (1996) Speech compensation to articulatory perturbations. *Journal of the Acoustical Society of America* 97: 1865-1873.

Maeda, S. (1996) Phonemes as concatenable units: VCV synthesis using a vocal-tract synthesizer. *Arbeitsberichte des Instituts für Phonetik und Digitale Sprachverarbeitung (ARIPUC)* 31: 145-164.

Minifie, F.D., Abbs, J.H., Tarlow, A. and Kwaterski, M. (1974) EMG activity within the pharynx during speech production. *Journal of Speech and Hearing Research* 17: 489-496.

Moen, I. and Simonsen (1997) Effects of voicing on /t,d/ tongue/palate contact in English and Norwegian. *ESCA Eurospeech Rhodes*, Greece: 2399-2402.

Moen, I., Simonsen, H.G., Huseby, M. and Grue (2001) The relationsship between intraoral air pressure and tongue/palate contact during the articulation of Norwegian /t/ and /d/. *CD-ROM Eurospeech Aalborg*, Danemark: article 265.

Mooshammer, C., Hoole, P. and Kühnert, B. (1995): On loops. *Journal of Phonetics* 23: 3-21.

Mooshammer, C. (1998) Experimentalphonetische Untersuchung zur artikulatorischen Modellierung der Gespanntheitsopposition im Deutschen. *Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München (FIPKM)*: 3-191.

Mooshammer, C. Fuchs, S. and Fischer, D. (1999) Effects of stress and tenseness on the production of CVC syllables in German. In *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco*, USA: 409-412.

Mooshammer, C. and Fuchs, S. (2000) Stress Distinction in German: Modeling kinematic parameters of tongue tip gestures. In *Proceedings of the 5th Seminar on Speech Production: Models and Data, Kloster Seeon*, Germany: 85-88.

Mooshammer, C., Perrier, P., Fuchs, S., Geng, C. and Payan, P. (2001) The Control of Token-to-Token Variability: an Experimental and Modeling Study. In *Proceedings of the 4th International Speech Motor Conference*, Nijmegen, The Netherlands: 78-81.

Mooshammer, C., Geumann, A., Hoole, P., Alfonso, P., van Lieshout, P. and Fuchs, S. (2003) Coordination of lingual and mandibular gestures for different manners of articulation. In *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*: 81-84.

Mooshammer, C., Hoole, P. and Geumann, A. (submitted) The role of the jaw in the production of coronal consonants. *Language and Speech.*

Munhall, K. (1984) *Temporal adjustment in speech motor control: Evidence from laryngeal kinematics.* PhD Thesis Montreal: Department of Psychology.

Munhall, K., Löfqvist, A. (1992) Gestural aggregation in speech: laryngeal gestures. *Journal of Phonetics* 20: 111-126.

Munhall, K.G., Löfqvist, A. and Kelso, J.A. (1994): Lip-larynx coordination in speech: effects of mechanical perturbations to the lower lip. *Journal of the Acoustical Society of America* 95(6): 3605-3616.

Narayanan, S. (2000) Noise source models for fricative consonants. *IEEE Transactions on Speech and Audio Processing* 8 (2): 328-344.

Netsell, R. (1969) Subglottal and intraoral air pressures during the intervocalic contrast of /t/ and /d/. *Phonetica* 20: 68-73.

Ni Chasaide, A. (1987) Glottal control of aspiration and of voicelessness. *Proceedings of the 11th International Congress of Phonetic Sciences, Tallinn*, Estonia, Vol. 6: 28-31.

Ohala, J.J. and Riordan, C.J. (1980): Passive vocal tract enlargement during voiced stops. *Report of the Phonology Laboratory* No.5, University of Berkeley: 78-88.

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

Paeschke, A., Kienast, M. and Sendlmeier, W.F. (1999) F0-contours in emotional speech. In *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco*, USA: 929-932.

Pape, D., Mooshammer, C., Hoole, P. and Fuchs, S. (2003) Devoicing of word-initial stops: A consequence of the following vowel? *Proceedings of the Speech Production Seminar, Sydney*: 207-212.

Payan, Y. and Perrier, P. (1997) Synthesis of V-V sequences with a 2D biomechanical tongue model controlled by the Equilibrium Point Hypothesis. *Speech Communication* 22 (2/3): 185-205.

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

Perkell, J. and Klatt, D.H. (1986) *Invariance and variability in speech processes*. LEA: Hillsdale, New York.

Perkell, J., Matthies, M.L., Svirsky, M.A. and Jordan, M.I. (1993) Trading relations between tongue-body raising and lip rounding in the production of the vowel /u/: A pilot motor equivalence study. *Journal of the Acoustical Society of America* 93: 2948-2961.

Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J. and Guiod, P. (1997) Speech motor control: acoustic goals, saturation effects, auditory feedback and internal models. *Speech Communication* 22: 227-250.

Perrier, P. (2003) About speech motor complexity. In *Proceedings of the 6th Speech Production Seminar, Sydney*: 225-230.

Perrier, P. (in press) About speech motor complexity. In Harrington, J. & Tabain M. (Eds.): *Speech Production: Models, Phonetic Processes and Techniques*. Psychology Press: New York.

Perrier, P. (2005) Control and representations in speech production. Fuchs, S., Perrier, P. and Pompino-Marschall, B. S. (Eds.): Speech production and perception: Experimental analyses and models. *ZAS papers in Linguistics 40*:109-132.

Petersen, N.R. (1983) The effect of consonant type on fundamental frequency and larynx height in Danish. *Annual Report of the Institute of Phonetics, University of Copenhagen* 17: 55-86.

Pétursson, M. (1977) Timing of glottal events in the production of aspiration after [s]. *Journal of Phonetics* 5: 205-212.

Pickett, J.M. (1999) *The acoustics of speech communication. Fundamentals, speech perception theory and technology.* Allyn and Bacon, Boston/London/Toronto/Sydney/-Tokyo/Singapore.

Pierrehumbert, J., Beckman, M.E. and Ladd, B. (2001) Conceptual foundation of phonology as a laboratory science. In Burton-Roberts, N., Carr, P. and Docherty, G. (eds.): *Phonological knowledge.* Oxford University Press, Oxford:273-304. (see http://www.ling.nwu.edu/~jbp/publications.html)

Pierrehumbert, J. (2003) Probabilistic Phonology: Discrimination and robustness. In Bod, R., Hay, J. and Jannedy, S. (eds.) *Probability theory in linguistics*. The MIT Press.

Piroth, H.P. and Janker, P.M. (2004) Speaker-dependent differences in voicing and devoicing of German obstruents. *Journal of Phonetics* 32(1): 81-109.

Pompino-Marschall, B. (1995) *Einführung in die Phonetik*. Walter de Gruyter, Berlin/New York.

Port, R.F. and O'Dell, M.L. (1985) Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13: 455-471.

Port, R.F. and Crawford, P. (1989) Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17: 257-282.

Port, R.F. and Leary, A. P. (2003) Speech timing and linguistic theory. In Drake, C. and Lorenzi, C. (eds.): *Hearing in Time/ Temp et Audition*. De Boeck University Press, Louvin-la-Neuve.

Repp, B.H. (1982) Phonetic trading relations and context effects : New experimental evidence for a speech mode of perception. *Psychological Bulletin* 92:81-110.

Ridouane, R. (2003) Geminates vs. singleton stops in Berber: An acoustic, fiberscopic and photoglottographic study. In *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona*: 1753-1746.

Ridouane, R., Fuchs, S. and Hoole, P. (2003) Laryngeal adjustments in the production of voiceless obstruent clusters in Berber. In *Proceedings of the 6th Speech Production Seminar, Sydney*: 237-242.

Ridouane, R., Fuchs, S. and Hoole, P. (in press) Laryngeal adjustments in the production of voiceless obstruent clusters in Berber. In Harrington, J. & Tabain M. (Eds.): Harrington, J. & Tabain, M. (eds) (to appear) *Speech Production: Models, Phonetic Processes and Techniques* Psychology Press: New York.

Riordan, C.J. (1980) Larynx height during English stop consonants. *Journal of Phonetics* 8: 353-360.

Romero, J.(1999) The effect of voicing assimilation on gestural coordination. In *Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco*, USA, 1793-1796.

Rouco, A. and Recasens, D. (1996) Reliability of electromagnetic midsagittal articulometry and electropalatography data aquired simultaneously. *Journal of the Acoustical Society of America* 100(5): 3384-3389.

Saltzman, E. and Munhall, K.G. (1989) A dynamical approach to gestural patterning in speech production. *Ecological Psychology 1(4)*: 333-382.

Saltzman, E., Löfqvist, A., Kay, B., Kinsella-Shaw, J. and Rubin, P. (1998) Dynamics of intergestural timing: a pertubation study of lip-larynx coordination. *Experimental Brain Research* 123: 412-424.

Saltzman, E. and Byrd, D. (2000) Task-dynamics of gestural timing: Phase windows and multifrequency rhythms. *Human Movement Science* 19: 499-526.

Sanders, I., Wu, B.L., Mu, L., Li, Y. and Biller H. F. (1994) The innervation of the human posterior cricoarytenoid muscle: evidence for at least two neuromuscular compartments. *Laryngoscope* 104(7): 880-884.

Savariaux, C., Perrier, P. and Orliaguet, J.P. (1995) Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control space in speech production. *Journal of the Acoustical Society of America* 98(5): 2428-2442.

Sawashima, M. (1968) Movements of the larynx in articulation of Japanese consonants. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 2:11-20.

Sawashima, M. and Hirose, H. (1968) New laryngoscopic technique by use of fiberoptics. *Journal of the Acoustical Society of America* 43(1): 168-169.

Sawashima, M., Abramson, A.S., Cooper, F.S. and Lisker, L. (1970) Observing laryngeal adjustments during running speech by use of a fiberoptics system. *Phonetica* 22: 193-201.

Sawashima, M. (1970) Glottal adjustments for English obstruents. *Haskins Laboratories: Status Report on Speech Research* SR-21/22: 186-200.

Sawashima, M. and Ushijima, T. (1971) Use of the fiberscope in speech research. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 5: 25-34.

Sawashima, M. and Miyazaki, S. (1973) Glottal opening for Japanese voiceless consonants. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 7: 1-9.

Sawashima, M. and Miyazaki, S. (1974) Stereo-fiberscopic measurement of the larynx: a preliminary experiment by use of ordinary laryngeal fiberscopes. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 8: 7-12.

Sawashima, M. and Niimi, S. (1974) Laryngeal conditions in articulations of Japanese voiceless consonants. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 8: 13-18.

Sawashima, M., Hirose, H., Ushijima, T. and Niimi, S. (1975) Laryngeal control in Japanese consonants with special reference to those in utterance initial position. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 9: 21-26.

Sawashima, M., Hirose, H. and Yoshioka, H. (1978) Abductor (PCA) and adductor (INT) muscles of the larynx in voiceless sound production. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 12: 53-60.

Sawashima, M. and Park, H.S. (1979) Laryngeal adjustments for syllable final stops in Korean: Some preliminary results for fiberoptic observation. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 13: 83-89.

Sawashima, M. and Hirose, H: (1983) Laryngeal gestures in speech production. In MacNeilage, P.F. (ed.): *The Production of Speech.* New York, Heidelberg, Berlin: Springer Verlag. 11-38.

Sawashima, M. (1997) Phonetic control of the glottal opening. In Kiritani, S., Hirose, H. and Fujisaki, H.: *Speech production and language.* In Honor of Osamu Fujimura. Berlin, New York: Mouton DeGryuter. 69-75.

Scherer, K. (1986) Vocal affect expression: A review and a model for future research. *Psychological Bulletin* 99, 2: 143-165.

Schwartz, J.L., Abry, C., Boë, J.L. and Cathiard, M. (2002) Phonology in a theory of perception-for-action-control. In Durand, J. & Laks B. (eds.) *Phonology: from Phonetics to Cognition.* Oxford, Oxford University Press: 255-280.

Scobbie, J. (in press) Flexibility in the face of incompatible English VOT systems. Labphon VI.

Shaiman, S. and Abbs, J. (1987) Sensorimotor contributions to the temporal coordination of oral and laryngeal movements for speech. *Speech Motor Control Laboratories University of Maddision*, Spring-Summer: 185-202.

Shih, C. and Möbius, B. (1999) Contextual effects on consonantal voicing profiles: A cross-linguistic study. *In Proceedings of the 14th International Congress of Phonetic Sciences in San Francisco* 2: 989-992.

Shockey, L. (1991) Electropalatography of conversational speech. *Proceedings of the 12th International Congress of Phonetic Sciences, Aix-en-Provence*, France, Vol. 3:10-13.

Shockey, L. and Gibbon, F. (1993) “Stopless Stops” in connected English. *Speech Research Laboratory, University of Reading Work in Progress* 7: 1-7.

Slis, I.H. (1967) What causes the voice-voiceless distinction?. *I.P.O. Annual Progress Report* 2: 71-76.

Slis, I.H. and Damsté (1967) Transillumination of the glottis during voiced and voiceless consonants. *I.P.O. Annual Progress Report* 2: 103-109.

Slis, I.H. (1970): Articulatory measurements on voiced, voiceless and nasal consonants. *Phonetica* 21: 193-210.

Smith, C.L. (1997) The devoicing of /z/ in American English: effects of local and prosodic context. *Journal of Phonetics* 25: 471-500.

Sock, R. (1998): *Organisation temporelle en production de la parole émergence de catégories sensori-motorices phonétiques*. Presses universitaires du septentrion: Villeneuve.

Sonesson, B. (1960) On the Anatomy and vibratory pattern of the human vocal folds. With special reference to a photo-electrical method for studying the vibratory movements. *Acta Oto-Laryngologica Supplementum* 156: 7-76.

Stathopoulos, E.T. and Weismer, G. (1983) Closure duration of stop consonants. *Journal of Phonetics* 11: 395-500.

Stevens, K.N. (1972) The quantal nature of speech: Evidence from articulatory-acoustic data. In David, E.E. and Deenes, P.B. (eds.): *Human communication: a unified view.* McGraw-Hill, New York: 51-66.

Stevens, K.N. and Klatt, D.H. (1974) Role of formant transistions in the voice-voiceless distinction for stops. *Journal of the Acoustical Society of America* 55 (3): 653-659.

Stevens, K.N. (1977) Physics of laryngeal behavior and larynx mode. *Phonetica* 34: 264-279.

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

Stevens, K. N. (1998) *Acoustic phonetics*. The MIT Press, Cambridge, MA.

Stone, M. (1995) How the tongue takes advantage of the palate during speech. p. 143-153 In Bell-Berti F. and Lawrence J. R. (eds.): *Producing Speech: Contemporary Issues for Katherine Safford Harris.* AIP Press, New York.

Stone, M. (1997) Instrumentation for the study of speech physiology. p. 495-524 In Hardcastle, W.J. and Laver, J. (eds.): *The Handbook of Phonetic Scienes.* Blackwell Publishers, Oxford.

Summerfield, Q. and Haggard, M.P. (1975) Vocal tract normalization as demonstrated by reaction times. In Fant, G. and Tatham, M. (eds.) *Auditory signals and perception of speech*. Academic, New York: 115-141.

Svirsky, M., Stevens, K., Matthies, M., Manzella, J., Perkell, J. and Wilhelms-Tricarico, R. (1997) Tongue surface displacement during bilabial stops. *Journal of the Acoustical Society of America* 102: 562-571.

Tabain, M. (2002) Voiceless consonants and locus equations: a comparison with electropalatographic data on coarticulation. *Phonetica* 59: 20-37.

Tasko, S.T. and Westbury, J.R. (2002) Defining and Measuring speech movement events. *Journal of Speech, Language, and Hearing Research* 45: 127-142.

Tatam, M.A.A. and Morton, K. (1973) Electromyographic and intraoral air-pressure studies of bi-labial stops. *Language and Speech* 16, 336-350.

Tiede, M. (1998) Palate shape effects on characteristic vowel tongue postures. [Abstract] *Journal of the Acoustical Society of America* 103: 2893.

Tiede, M. (2003) A new approach to pressure-sensitive palatography using a capacitive sensing device. *Proceedings of the 15th International Congress of Phonetic Sciences, Barcelona.*

Trubetzkoy, N.S. (1936) Die Aufhebung der phonologischen Gegensätze. In: *Ètudes dediées au quatrième Congrès de Linguistes. Travaux du Cercle Linguistique de Prague* 6 : 29-45.

Ushijima, T. and Sawashima, M. (1972) Fiberscopic observation of velar movements during speech. *Annual Bulletin Research Institute of Logopedics and Phoniatrics, University of Tokyo* 6, 25-38.

van den Berg, J. (1958) Myoelastic-aerodynamic theory of voice production. *Journal of Speech and Hearing Research* 1(3): 227-244.

Wakumoto, M., Masaki, S., Honda, K. and Ouhe, T. (1998) A pressure sensitive palatography: application of the new pressure sensitive sheet for measuring tongue-palatal contact pressure. In *Proceedings of the 5rd International Conference on Spoken Language Processing, Sydney,* Australia, Vol. 7: 3151-3154.

Weismer, G. and Bunton, K. (1999) Influences of pellet markers on speech production behavior: Acoustic and perceptual measures. *Journal of the Acoustical Society of America* 105(5): 2882-2894.

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

Wiese, R. (1996) *The phonology of German*. Clarendon Press: Oxford.

Yoshioka, H., Löfqvist, A. and Hirose H. (1980) Laryngeal adjustments in Japanese voiceless sound production. *Haskins Laboratories: Status Report on Speech Research* SR-63/64: 293-308.

Yoshioka, H. (1981) Laryngeal adjustments in the production of the fricative consonants and devoiced vowels in Japanese. *Phonetica* 38: 236-251.

Yoshioka, H., Löfqvist, A. and Hirose H. (1981) Laryngeal adjustments in the production of consonant clusters and geminates in American English. *Journal of the Acoustical Society of America* 70(6): 1615-1623.

Zemlin, W.R. (1988) *Speech and Hearing Science: Anatomy and Physiology*. (3rd ed), New Jersey: Prentice Hall.

# Appendices

## Appendix I: Perceptual rating

**Table 1:** Perceptual rating of /d t/ with respect to the different positions (C1, C2, C3), speakers (CG, JD, SF), vowel contexts (/a i u/), 1= voiceless, 0= voiced

| | **C1** | | **C2** | | **C3** | |
|---|---|---|---|---|---|---|
| | /d/ | /t/ | /d/ | /t/ | /d/ | /t/ |
| **CG** | | | | | | |
| /a/-context | 1 | 15 | 0 | 15 | 15 | 15 |
| /i/-context | 0 | 15 | 0 | 15 | 14 | 15 |
| /u/-context | 4 | 15 | 0 | 15 | 15 | 15 |
| **JD** | | | | | | |
| /a/-context | 4 | 15 | 0 | 15 | 15 | 15 |
| /i/-context | 5 | 15 | 0 | 15 | 15 | 15 |
| /u/-context | 6 | 15 | 0 | 15 | 15 | 15 |
| **SF** | | | | | | |
| /a/-context | 5 | 15 | 0 | 15 | 15 | 15 |
| /i/-context | 6 | 15 | 1 | 15 | 15 | 15 |
| /u/-context | 6 | 15 | 0 | 15 | 15 | 15 |

**Table 2:** Perceptual rating of /s z/ with respect to the different positions (C2) speakers (CG, JD, SF), vowel contexts (/a i u/), 1= voiceless, 0= voiced

| | **C2** | |
|---|---|---|
| | /z/ | /s/ |
| **CG** | | |
| /a a/-context | 5 | 5 |
| /i ɪ/-context | 5 | 5 |
| /u ʊ/-context | 5 | 5 |
| **JD** | | |
| /a a/-context | 0 | 5 |
| /i ɪ/-context | 0 | 5 |
| /u ʊ/-context | 0 | 5 |
| **SF** | | |
| /a a/-context | 5 | 5 |
| /i ɪ/-context | 5 | 5 |
| /u ʊ/-context | 5 | 5 |

## Appendix II: Laryngeal correlates

**Table 3:** Number of tokens, means and standard deviations for the relative amount of glottal opening in post-stressed positions (C2, C3) with respect to C1 (set to 100%) /t/ production, speaker (CG, SF), vowel contexts (/a i u/)

**Relative amount of glottal opening in % in /t/ production**

| SPEAKER | POSITION | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| cg | C1 | a | 5 | 100.00 | .00 |
| | | i | 5 | 100.00 | .00 |
| | | u | 5 | 100.00 | .00 |
| | C2 | a | 5 | 32.60 | 11.01 |
| | | i | 5 | 26.40 | 9.71 |
| | | u | 5 | 27.60 | 16.95 |
| | C3 | a | 5 | 29.20 | 11.52 |
| | | i | 5 | 17.80 | 8.90 |
| | | u | 4 | 31.75 | 25.26 |
| sf | C1 | a | 5 | 100.00 | .00 |
| | | i | 5 | 100.00 | .00 |
| | | u | 5 | 100.00 | .00 |
| | C2 | a | 4 | 16.75 | 3.77 |
| | | u | 2 | 4.50 | 2.12 |
| | C3 | a | 5 | 13.20 | 5.67 |

**Table 4:** Number of tokens, means and standard deviations for overall glottal opening duration in ms, /t/ production, speaker (CG, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Overall glottal opening duration for /t/ in ms**

| SPEAKER | POSITION | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| cg | C1 | a | 5 | 155.00 | 9.35 |
| | | i | 5 | 189.00 | 35.07 |
| | | u | 5 | 187.00 | 22.25 |
| | C2 | a | 5 | 155.00 | 16.96 |
| | | i | 5 | 155.00 | 10.61 |
| | | u | 5 | 145.00 | 19.69 |
| | C3 | a | 5 | 163.00 | 12.04 |
| | | i | 5 | 157.00 | 4.47 |
| | | u | 4 | 152.50 | 11.90 |
| sf | C1 | a | 5 | 177.00 | 22.53 |
| | | i | 5 | 228.00 | 24.14 |
| | | u | 5 | 207.00 | 9.75 |
| | C2 | a | 4 | 165.00 | 32.91 |
| | | u | 2 | 140.00 | 7.07 |
| | C3 | a | 5 | 163.00 | 15.25 |

**Table 5:** Number of tokens, means and standard deviations for the amount of glottal opening in /s/ and /z/ production; /s/ in lax vowel context (-) position (C2), /z/ in tense vowel context (+) positions (C1, C2, C3), subjects (CG, JD, SF)

**Amplitude of glottal abduction for /s/ (lax vowel context -) and /z/ (tense vowel context +) with reference to the baseline**

| SUBJECT | POSITION | TENSENES | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|---|
| cg | C1 | + | a | 4 | .043 | .012 |
| | | | i | 5 | .094 | .091 |
| | | | u | 5 | .058 | .023 |
| | C2 | + | a | 5 | .067 | .019 |
| | | | i | 5 | .107 | .066 |
| | | | u | 5 | .110 | .049 |
| | | - | a | 5 | .081 | .025 |
| | | | i | 5 | .073 | .036 |
| | | | u | 5 | .078 | .033 |
| | C3 | + | a | 5 | .048 | .009 |
| | | | i | 5 | .046 | .031 |
| | | | u | 5 | .058 | .036 |
| jd | C1 | + | a | 1 | .002 | . |
| | | | i | 1 | .006 | . |
| | C2 | + | a | 2 | .015 | .003 |
| | | | i | 2 | .011 | .003 |
| | | - | a | 5 | .028 | .005 |
| | | | i | 5 | .020 | .006 |
| | | | u | 5 | .020 | .008 |
| | C3 | + | a | 4 | .008 | .008 |
| | | | i | 5 | .007 | .004 |
| | | | u | 4 | .006 | .002 |
| sf | C1 | + | a | 1 | .086 | . |
| | | | i | 5 | .009 | .004 |
| | | | u | 3 | .023 | .012 |
| | C2 | + | a | 5 | .215 | .148 |
| | | | i | 5 | .096 | .046 |
| | | | u | 5 | .312 | .107 |
| | | - | a | 5 | .176 | .114 |
| | | | i | 5 | .148 | .094 |
| | | | u | 5 | .242 | .087 |
| | C3 | + | a | 5 | .152 | .114 |
| | | | i | 5 | .037 | .009 |
| | | | u | 5 | .190 | .088 |

**Table 6:** Number of tokens, means and standard deviations for overall glottal opening duration in ms, /z/ production, speaker (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Overall glottal opening duration for /z/ in ms**

| SUBJECT | POSITION | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| cg | C1 | a | 4 | 172.50 | 31.75 |
| | | i | 5 | 188.00 | 20.19 |
| | | u | 5 | 168.00 | 20.80 |
| | C2 | a | 5 | 151.00 | 10.84 |
| | | i | 5 | 162.00 | 29.07 |
| | | u | 5 | 170.00 | 7.91 |
| | C3 | a | 5 | 149.00 | 14.32 |
| | | i | 5 | 152.00 | 19.87 |
| | | u | 5 | 153.00 | 15.25 |
| jd | C1 | a | 1 | 130.00 | . |
| | C2 | a | 2 | 145.00 | .00 |
| | C3 | a | 2 | 142.50 | 3.54 |
| | | i | 5 | 129.00 | 28.37 |
| | | u | 4 | 180.00 | 24.83 |
| sf | C1 | a | 1 | 170.00 | . |
| | | i | 5 | 135.00 | 20.31 |
| | | u | 3 | 140.00 | 22.91 |
| | C2 | a | 5 | 199.00 | 19.81 |
| | | i | 5 | 178.00 | 12.55 |
| | | u | 5 | 189.00 | 16.73 |
| | C3 | a | 5 | 161.00 | 13.87 |
| | | i | 5 | 159.00 | 8.94 |
| | | u | 5 | 165.00 | 23.98 |

**Table 7:** Number of tokens, means and standard deviations for overall glottal opening duration in ms, /s/ production, speaker (CG, JD, SF), position (C2), vowel contexts (lax /a i u/)

**Overall glottal opening duration for /s/ in ms**

| SPEAKER | POSITION | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| cg | C2 | a | 5 | 149.00 | 7.42 |
| | | i | 5 | 166.00 | 17.82 |
| | | u | 5 | 174.00 | 7.42 |
| jd | C2 | a | 5 | 113.00 | 27.75 |
| | | i | 5 | 133.00 | 30.94 |
| | | u | 5 | 135.00 | 25.25 |
| sf | C2 | a | 5 | 186.00 | 8.94 |
| | | i | 5 | 184.00 | 8.94 |
| | | u | 5 | 177.00 | 13.51 |

# Appendix III: Laryngeal-oral co-ordination

**Table 8**: Number of tokens, means and standard deviations for noise duration in ms, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a, i, u/)

**Noise duration in ms for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 15 | 5.63 | 2.64 | 15 | 37.01 | 9.17 |
| | | i | 12 | 3.75 | 2.56 | 15 | 66.33 | 11.16 |
| | | u | 14 | 3.06 | 1.94 | 15 | 60.30 | 9.73 |
| | C2 | a | 12 | 2.33 | .70 | 15 | 30.08 | 5.66 |
| | | i | 12 | 4.10 | 1.81 | 15 | 31.23 | 6.52 |
| | | u | 14 | 2.67 | .93 | 15 | 25.62 | 5.82 |
| | C3 | a | 15 | 3.75 | 2.69 | 15 | 17.79 | 7.68 |
| | | i | 15 | 10.45 | 5.63 | 13 | 25.87 | 11.56 |
| | | u | 15 | 10.56 | 2.51 | 15 | 20.81 | 7.78 |
| jd | C1 | a | 15 | 7.66 | 3.35 | 15 | 42.54 | 9.22 |
| | | i | 13 | 11.19 | 9.39 | 15 | 57.81 | 12.41 |
| | | u | 15 | 10.58 | 5.42 | 15 | 45.68 | 12.06 |
| | C2 | a | 2 | 2.50 | .71 | 15 | 14.93 | 3.82 |
| | | i | 10 | 13.99 | 8.36 | 15 | 26.16 | 4.95 |
| | | u | 6 | 12.67 | 9.20 | 15 | 22.31 | 3.55 |
| | C3 | a | 15 | 9.00 | 3.31 | 15 | 8.88 | 3.41 |
| | | i | 15 | 17.37 | 4.55 | 15 | 17.94 | 4.88 |
| | | u | 15 | 14.65 | 3.37 | 15 | 14.33 | 4.19 |
| sf | C1 | a | 15 | 9.99 | 1.88 | 15 | 44.39 | 15.56 |
| | | i | 15 | 22.46 | 5.20 | 15 | 72.76 | 9.90 |
| | | u | 15 | 13.24 | 4.43 | 15 | 77.07 | 11.10 |
| | C2 | a | 15 | 9.07 | 4.76 | 15 | 20.08 | 6.31 |
| | | i | 13 | 19.36 | 5.81 | 15 | 22.99 | 5.19 |
| | | u | 15 | 13.72 | 2.33 | 15 | 19.14 | 3.38 |
| | C3 | a | 15 | 7.71 | 2.29 | 15 | 8.38 | 2.10 |
| | | i | 14 | 13.24 | 3.68 | 15 | 13.35 | 2.85 |
| | | u | 15 | 10.07 | 1.33 | 15 | 10.31 | 3.13 |

**Table 9:** Number of tokens, means and standard deviations for COOR_on in ms, /t/-production, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**COOR_on in ms for /t/**

| SUBJECT | POSITION | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| cg | C1 | a | 5 | -6.00 | 13.87 |
| | | i | 5 | -4.00 | 12.94 |
| | | u | 5 | -3.00 | 21.39 |
| | C2 | a | 5 | 20.00 | .00 |
| | | i | 5 | -2.00 | 11.51 |
| | | u | 5 | -1.00 | 19.17 |
| | C3 | a | 5 | 28.00 | 12.55 |
| | | i | 5 | -2.00 | 9.08 |
| | | u | 4 | 3.75 | 11.81 |
| jd | C1 | a | 5 | -6.00 | 4.18 |
| | | i | 5 | -12.00 | 7.58 |
| | | u | 5 | -8.00 | 7.58 |
| sf | C1 | a | 5 | 22.00 | 20.19 |
| | | i | 5 | 23.00 | 30.74 |
| | | u | 5 | 8.00 | 9.75 |
| | C2 | a | 4 | 56.25 | 38.16 |
| | | u | 2 | 20.00 | 35.36 |
| | C3 | a | 5 | 67.00 | 19.24 |

**Table 10:** Number of tokens, means and standard deviations for COOR_peak in ms, /t/-production, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**COOR_peak in ms for /t/**

| SUBJECT | POSITION | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| cg | C1 | a | 5 | 15.60 | 7.44 |
| | | i | 5 | 14.40 | 10.62 |
| | | u | 5 | 15.40 | 13.46 |
| | C2 | a | 5 | .60 | 7.77 |
| | | i | 5 | 8.60 | 9.76 |
| | | u | 5 | -2.60 | 18.93 |
| | C3 | a | 5 | -1.80 | 8.81 |
| | | i | 3 | 35.67 | 17.21 |
| | | u | 4 | 13.25 | 14.93 |
| jd | C1 | a | 5 | 6.80 | 8.98 |
| | | i | 5 | 7.00 | 4.64 |
| | | u | 5 | 2.20 | 5.07 |
| sf | C1 | a | 5 | 15.20 | 6.53 |
| | | i | 5 | 10.40 | 14.10 |
| | | u | 5 | 4.60 | 2.07 |
| | C2 | a | 4 | -9.75 | 15.35 |
| | | u | 2 | 4.00 | 5.66 |
| | C3 | a | 5 | -35.40 | 10.67 |

**Table 11:** Number of tokens, means and standard deviations for COOR_on fric in ms, /z/-production, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**COOR_on fric in ms for /z/**

| SUBJECT | POSITION | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| cg | C1 | a | 4 | 33.08 | 22.53 |
| | | i | 5 | 46.74 | 17.28 |
| | | u | 5 | 35.18 | 17.75 |
| | C2 | a | 5 | 24.37 | 9.22 |
| | | i | 5 | 20.12 | 17.37 |
| | | u | 5 | 23.73 | 3.91 |
| | C3 | a | 5 | 28.12 | 5.52 |
| | | i | 5 | 8.52 | 15.03 |
| | | u | 5 | 24.02 | 5.94 |
| jd | C1 | a | 1 | -14.46 | . |
| | C2 | a | 2 | 30.90 | 2.98 |
| | C3 | a | 2 | 32.41 | 10.49 |
| | | i | 5 | 12.46 | 15.83 |
| | | u | 4 | 47.28 | 31.65 |
| sf | C1 | a | 1 | 72.00 | . |
| | | i | 5 | 15.40 | 6.88 |
| | | u | 3 | 36.35 | 11.32 |
| | C2 | a | 5 | 52.69 | 14.63 |
| | | i | 5 | 15.00 | 8.60 |
| | | u | 5 | 32.99 | 10.57 |
| | C3 | a | 5 | 40.63 | 13.65 |
| | | i | 5 | 16.87 | 17.37 |
| | | u | 5 | 34.34 | 13.88 |

**Table 12:** Number of tokens, means and standard deviations for COOR_on fric in ms, /s/-production, subjects (CG, JD, SF), position (C2), vowel contexts (lax /a i u/)

**COOR_on fric in ms for /s/**

| SPEAKER | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|
| cg | a | 5 | 18.28 | 9.43 |
| | i | 5 | 30.72 | 5.27 |
| | u | 5 | 37.48 | 4.58 |
| jd | a | 5 | 9.24 | 12.64 |
| | i | 5 | 25.86 | 12.69 |
| | u | 5 | 26.56 | 9.21 |
| sf | a | 5 | 46.98 | 6.06 |
| | i | 5 | 28.26 | 9.54 |
| | u | 5 | 33.18 | 9.02 |

**Table 13:** Number of tokens, means and standard deviations for COOR_peak fric in %, /z/-production, subjects (CG, JD, SF), positions (C1,C2,C3), vowel contexts (/a i u/)

**COOR_peak fric in % for /z/**

| SUBJECT | POSITION | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| cg | C1 | a | 4 | 36.12 | 12.58 |
| | | i | 5 | 40.36 | 4.37 |
| | | u | 5 | 47.47 | 8.13 |
| | C2 | a | 5 | 37.85 | 10.52 |
| | | i | 5 | 46.60 | 2.15 |
| | | u | 5 | 45.54 | 2.93 |
| | C3 | a | 5 | 34.75 | 9.47 |
| | | i | 5 | 54.49 | 4.73 |
| | | u | 5 | 46.50 | 1.89 |
| jd | C1 | a | 1 | 56.46 | . |
| | C2 | a | 2 | 55.66 | 17.62 |
| | C3 | a | 2 | 32.03 | 17.36 |
| | | i | 5 | 59.80 | 6.02 |
| | | u | 4 | 40.13 | 8.03 |
| sf | C1 | a | 1 | 35.48 | . |
| | | i | 5 | 32.17 | 9.25 |
| | | u | 3 | 31.78 | 5.88 |
| | C2 | a | 5 | 43.31 | 1.95 |
| | | i | 5 | 56.28 | 6.81 |
| | | u | 5 | 42.73 | 8.46 |
| | C3 | a | 5 | 34.00 | 6.70 |
| | | i | 5 | 50.07 | 6.69 |
| | | u | 5 | 39.47 | 3.54 |

**Table 14:** Number of tokens, means and standard deviations for COOR_peak fric in %, /s/-production, subjects (CG, JD, SF), position (C2), vowel contexts (lax /a i u/)

**COOR_peak fric in % for /s/**

| SPEAKER | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|
| cg | a | 5 | 39.37 | 10.46 |
| | i | 5 | 38.10 | 2.03 |
| | u | 5 | 36.01 | 2.30 |
| jd | a | 5 | 49.70 | 9.70 |
| | i | 5 | 40.25 | 3.90 |
| | u | 5 | 35.81 | 4.84 |
| sf | a | 5 | 42.19 | 6.85 |
| | i | 5 | 49.01 | 6.94 |
| | u | 5 | 44.08 | 7.61 |

**Table 15:** Number of tokens, means and standard deviations for COOR_off fric in ms, /z/-production, subjects (CG, JD, SF), positions (C1,C2,C3), vowel contexts (/a i u/)

**COOR_off fric in ms for /z/**

| SUBJECT | POSITION | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|---|
| cg | C1 | a | 4 | 33.00 | 18.85 |
| | | i | 5 | 30.60 | 28.05 |
| | | u | 5 | 14.90 | 16.53 |
| | C2 | a | 5 | 27.88 | 7.39 |
| | | i | 5 | 15.48 | 13.89 |
| | | u | 5 | 33.72 | 7.13 |
| | C3 | a | 5 | 29.05 | 14.02 |
| | | i | 5 | 33.40 | 13.70 |
| | | u | 5 | 36.58 | 8.06 |
| jd | C1 | a | 1 | 48.00 | . |
| | C2 | a | 2 | 49.27 | 8.87 |
| | C3 | a | 2 | 44.84 | 12.49 |
| | | i | 5 | 43.00 | 21.70 |
| | | u | 4 | 71.50 | 11.56 |
| sf | C1 | a | 1 | 5.00 | . |
| | | i | 5 | -4.00 | 31.66 |
| | | u | 3 | -7.67 | 19.86 |
| | C2 | a | 5 | 36.98 | 8.55 |
| | | i | 5 | 47.99 | 12.37 |
| | | u | 5 | 40.96 | 9.40 |
| | C3 | a | 5 | 37.82 | 5.13 |
| | | i | 5 | 36.80 | 6.65 |
| | | u | 5 | 35.50 | 6.87 |

**Table 16:** Number of tokens, means and standard deviations for COOR_off fric in ms, /s/-production, subjects (CG, JD, SF), position (C2), vowel contexts (lax /a i u/)

**COOR_off in ms for /s/**

| SPEAKER | VOWEL | N | Mean | Std. Deviation |
|---|---|---|---|---|
| cg | a | 5 | 20.10 | 11.57 |
| | i | 5 | 13.78 | 13.17 |
| | u | 5 | 24.20 | 6.49 |
| jd | a | 5 | 25.24 | 14.06 |
| | i | 5 | 12.56 | 18.93 |
| | u | 5 | 20.80 | 25.03 |
| sf | a | 5 | 44.34 | 2.25 |
| | i | 5 | 51.24 | 12.50 |
| | u | 5 | 35.86 | 18.06 |

# Appendix IV: Supralaryngeal correlates

**Table 17:** Number of tokens, means and standard deviations for vowel duration in ms, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Vowel duration in ms for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C2 | a | 14 | 187.67 | 11.35 | 15 | 155.39 | 14.52 |
| | | i | 14 | 134.01 | 16.65 | 15 | 88.47 | 8.64 |
| | | u | 15 | 142.20 | 9.11 | 15 | 92.60 | 13.01 |
| | C3 | a | 15 | 185.91 | 10.97 | 15 | 151.36 | 19.54 |
| | | i | 15 | 135.41 | 18.82 | 15 | 99.74 | 15.74 |
| | | u | 15 | 142.39 | 11.62 | 15 | 101.47 | 12.48 |
| jd | C2 | a | 15 | 233.77 | 20.79 | 15 | 168.66 | 15.29 |
| | | i | 14 | 133.04 | 15.22 | 15 | 90.51 | 11.95 |
| | | u | 14 | 161.66 | 18.62 | 15 | 99.72 | 8.21 |
| | C3 | a | 15 | 174.80 | 20.34 | 15 | 151.29 | 13.98 |
| | | i | 15 | 107.01 | 15.12 | 15 | 78.44 | 14.34 |
| | | u | 15 | 123.31 | 10.51 | 15 | 97.61 | 13.83 |
| sf | C2 | a | 15 | 181.28 | 11.37 | 15 | 125.98 | 15.62 |
| | | i | 15 | 107.04 | 17.01 | 15 | 69.80 | 7.66 |
| | | u | 15 | 130.67 | 10.23 | 15 | 66.77 | 12.22 |
| | C3 | a | 15 | 145.95 | 7.35 | 15 | 100.92 | 12.95 |
| | | i | 15 | 79.91 | 10.76 | 15 | 54.66 | 7.72 |
| | | u | 15 | 109.57 | 10.32 | 15 | 67.20 | 13.00 |

**Table 18:** Number of tokens, means and standard deviations for acoustic closure duration in ms, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Acoustic closure duration in ms for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 15 | 96.68 | 7.22 | 15 | 84.54 | 10.02 |
| | | i | 12 | 107.51 | 18.83 | 15 | 74.68 | 26.35 |
| | | u | 15 | 101.80 | 17.49 | 15 | 79.44 | 11.69 |
| | C2 | a | 12 | 50.44 | 8.84 | 15 | 65.44 | 7.84 |
| | | i | 12 | 63.54 | 14.55 | 15 | 73.50 | 12.11 |
| | | u | 14 | 53.76 | 10.74 | 15 | 77.74 | 10.42 |
| | C3 | a | 15 | 77.40 | 13.08 | 15 | 69.72 | 6.34 |
| | | i | 15 | 64.34 | 13.98 | 13 | 56.96 | 15.14 |
| | | u | 15 | 68.65 | 7.19 | 15 | 69.59 | 9.09 |
| jd | C1 | a | 15 | 81.00 | 11.05 | 15 | 76.68 | 9.10 |
| | | i | 13 | 105.78 | 11.28 | 15 | 100.17 | 14.57 |
| | | u | 15 | 100.94 | 10.14 | 15 | 96.70 | 9.94 |
| | C2 | a | 7 | 31.60 | 5.92 | 15 | 66.92 | 6.28 |
| | | i | 11 | 32.97 | 7.95 | 15 | 62.65 | 7.06 |
| | | u | 6 | 41.50 | 9.91 | 15 | 61.43 | 9.68 |
| | C3 | a | 15 | 56.59 | 3.14 | 15 | 55.77 | 7.36 |
| | | i | 15 | 46.09 | 6.04 | 15 | 51.68 | 6.58 |
| | | u | 15 | 46.77 | 5.27 | 15 | 48.20 | 5.28 |
| sf | C1 | a | 15 | 99.62 | 6.93 | 15 | 92.97 | 7.33 |
| | | i | 15 | 114.19 | 10.83 | 15 | 101.79 | 7.96 |
| | | u | 15 | 106.64 | 11.12 | 15 | 101.71 | 7.46 |
| | C2 | a | 15 | 53.42 | 6.19 | 15 | 62.74 | 7.88 |
| | | i | 13 | 54.41 | 16.23 | 15 | 71.66 | 7.18 |
| | | u | 15 | 50.01 | 5.55 | 15 | 63.56 | 8.88 |
| | C3 | a | 15 | 66.14 | 6.32 | 15 | 69.29 | 8.02 |
| | | i | 14 | 70.64 | 10.16 | 15 | 68.45 | 8.98 |
| | | u | 15 | 57.30 | 6.29 | 15 | 60.29 | 6.64 |

**Table 19:** Number of tokens, means and standard deviations for articulatory closing gesture duration in ms, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Articulatory closing gesture duration in ms for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | 64.00 | 3.94 | 10 | 62.00 | 4.22 |
| | | i | 10 | 67.00 | 4.22 | 10 | 67.00 | 7.89 |
| | | u | 10 | 64.00 | 3.94 | 10 | 66.50 | 5.30 |
| | C2 | a | 10 | 78.00 | 7.53 | 10 | 82.00 | 3.50 |
| | | i | 10 | 55.00 | 4.08 | 10 | 57.00 | 5.87 |
| | | u | 10 | 62.50 | 9.20 | 10 | 72.50 | 3.54 |
| | C3 | a | 10 | 83.50 | 7.47 | 10 | 83.00 | 7.15 |
| | | i | 10 | 57.50 | 4.86 | 10 | 60.00 | 6.24 |
| | | u | 10 | 71.50 | 4.74 | 10 | 75.00 | 4.08 |
| jd | C1 | a | 10 | 50.50 | 18.63 | 10 | 53.00 | 13.58 |
| | | i | 10 | 80.50 | 28.23 | 10 | 49.50 | 16.57 |
| | | u | 10 | 69.00 | 22.58 | 10 | 76.50 | 21.74 |
| | C2 | a | 10 | 98.50 | 11.80 | 10 | 85.50 | 7.62 |
| | | i | 10 | 85.00 | 28.38 | 10 | 49.00 | 17.45 |
| | | u | 10 | 89.00 | 20.66 | 10 | 61.50 | 4.74 |
| | C3 | a | 10 | 84.50 | 4.97 | 10 | 81.00 | 7.75 |
| | | i | 10 | 40.50 | 17.07 | 10 | 33.50 | 5.80 |
| | | u | 10 | 61.00 | 2.11 | 10 | 62.00 | 4.22 |
| sf | C1 | a | 10 | 57.00 | 13.78 | 9 | 49.44 | 7.26 |
| | | i | 10 | 74.00 | 26.75 | 10 | 72.00 | 28.30 |
| | | u | 10 | 62.50 | 18.45 | 10 | 60.50 | 12.57 |
| | C2 | a | 10 | 81.00 | 8.10 | 9 | 86.11 | 6.97 |
| | | i | 10 | 55.00 | 22.24 | 10 | 61.50 | 15.64 |
| | | u | 10 | 66.00 | 37.77 | 10 | 73.00 | 9.19 |
| | C3 | a | 10 | 70.00 | 8.16 | 9 | 75.56 | 3.91 |
| | | i | 10 | 46.50 | 10.01 | 10 | 62.50 | 7.17 |
| | | u | 10 | 51.50 | 16.34 | 10 | 63.00 | 13.58 |

**Table 20:** Number of tokens, means and standard deviations for movement amplitude of the tongue tip closing gesture in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Tongue tip movement amplitude of the artiuculatory closing gesture in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | .836 | .096 | 10 | .784 | .085 |
| | | i | 10 | .735 | .096 | 10 | .722 | .163 |
| | | u | 10 | .772 | .046 | 10 | .791 | .059 |
| | C2 | a | 10 | 1.569 | .138 | 10 | 1.415 | .099 |
| | | i | 10 | .382 | .092 | 10 | .493 | .111 |
| | | u | 10 | .994 | .114 | 10 | 1.183 | .129 |
| | C3 | a | 10 | 1.510 | .135 | 10 | 1.364 | .127 |
| | | i | 10 | .385 | .098 | 10 | .475 | .072 |
| | | u | 10 | 1.170 | .113 | 10 | 1.279 | .068 |
| jd | C1 | a | 10 | .252 | .140 | 10 | .208 | .111 |
| | | i | 10 | .173 | .057 | 10 | .089 | .031 |
| | | u | 10 | .358 | .215 | 10 | .338 | .121 |
| | C2 | a | 10 | 1.026 | .112 | 10 | 1.192 | .089 |
| | | i | 10 | .500 | .195 | 10 | .095 | .053 |
| | | u | 10 | .908 | .193 | 10 | .908 | .138 |
| | C3 | a | 10 | 1.180 | .136 | 10 | 1.002 | .091 |
| | | i | 10 | .080 | .072 | 10 | .052 | .025 |
| | | u | 10 | .829 | .120 | 10 | .909 | .212 |
| sf | C1 | a | 10 | .226 | .079 | 9 | .192 | .077 |
| | | i | 10 | .242 | .095 | 10 | .236 | .163 |
| | | u | 10 | .187 | .060 | 10 | .221 | .116 |
| | C2 | a | 10 | 1.235 | .070 | 9 | 1.196 | .057 |
| | | i | 10 | .130 | .065 | 10 | .211 | .084 |
| | | u | 10 | .759 | .360 | 10 | .969 | .113 |
| | C3 | a | 10 | 1.160 | .074 | 9 | 1.130 | .075 |
| | | i | 10 | .096 | .028 | 10 | .158 | .016 |
| | | u | 10 | .595 | .214 | 10 | .854 | .175 |

**Table 21:** Number of tokens, means and standard deviations for tangential velocity of the tongue tip closing gesture in cm/s, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Tongue tip velocity peak of the closing gesture in cm/s for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | 18.92 | 1.87 | 10 | 17.95 | 1.83 |
| | | i | 10 | 16.27 | 2.25 | 10 | 15.55 | 3.71 |
| | | u | 10 | 17.50 | 1.27 | 10 | 17.01 | 1.33 |
| | C2 | a | 10 | 30.85 | 4.49 | 10 | 25.60 | 2.07 |
| | | i | 10 | 9.71 | 2.16 | 10 | 12.46 | 2.60 |
| | | u | 10 | 21.62 | 3.65 | 10 | 23.57 | 2.34 |
| | C3 | a | 10 | 26.16 | 3.26 | 10 | 24.25 | 1.58 |
| | | i | 10 | 9.57 | 2.51 | 10 | 11.50 | 2.03 |
| | | u | 10 | 23.54 | 2.68 | 10 | 24.40 | 1.71 |
| jd | C1 | a | 10 | 5.79 | 1.94 | 10 | 4.84 | 2.06 |
| | | i | 10 | 2.88 | .81 | 10 | 2.12 | .68 |
| | | u | 10 | 5.99 | 2.27 | 10 | 5.59 | 1.66 |
| | C2 | a | 10 | 15.24 | 1.67 | 10 | 21.78 | 1.44 |
| | | i | 10 | 7.20 | 1.22 | 10 | 2.47 | .98 |
| | | u | 10 | 12.97 | 1.75 | 10 | 20.12 | 2.15 |
| | C3 | a | 10 | 20.42 | 2.85 | 10 | 18.00 | 1.95 |
| | | i | 10 | 2.11 | 1.75 | 10 | 1.70 | .58 |
| | | u | 10 | 18.06 | 2.91 | 10 | 19.07 | 4.78 |
| sf | C1 | a | 10 | 4.63 | 1.16 | 9 | 4.75 | 1.55 |
| | | i | 10 | 4.22 | 1.87 | 10 | 4.03 | 1.63 |
| | | u | 10 | 3.80 | 1.41 | 10 | 4.44 | 1.86 |
| | C2 | a | 10 | 23.24 | 1.54 | 9 | 21.60 | 1.69 |
| | | i | 10 | 2.91 | .68 | 10 | 4.51 | 1.41 |
| | | u | 10 | 14.33 | 1.44 | 10 | 17.50 | 1.05 |
| | C3 | a | 10 | 24.49 | 2.06 | 9 | 22.22 | 1.36 |
| | | i | 10 | 2.53 | .60 | 10 | 3.30 | .48 |
| | | u | 10 | 12.93 | 1.17 | 10 | 16.63 | 1.00 |

**Table 22:** Number of tokens, means and standard deviations for tongue tip–jaw latency at closing gesture onset in ms, /d/ vs. /t/, subjects (CG, JD,SF), positions (C2, C3), vowel contexts (/a/)

**ttip clg_on - tjaw clg_off in ms for /d/ vs. /t/**

| | | consonant | | | | | |
|---|---|---|---|---|---|---|---|
| | | d | | | t | | |
| SUBJECT | POSITION | N | Mean | Std. Deviation | N | Mean | Std. Deviation |
| cg | C2 | 10 | -1.34 | 2.23 | 10 | -4.05 | 2.79 |
| | C3 | 10 | -1.23 | 1.94 | 10 | -8.16 | 5.39 |
| jd | C2 | 10 | -8.50 | 8.51 | 10 | -17.50 | 7.17 |
| | C3 | 10 | -22.62 | 7.34 | 10 | -14.12 | 7.02 |
| sf | C2 | 10 | -2.43 | 2.98 | 9 | 1.20 | 3.07 |
| | C3 | 10 | -1.04 | 2.48 | 9 | 4.81 | 2.13 |

**Table 23:** Number of tokens, means and standard deviations for tongue tip–jaw latency at closing gesture offset in ms, /d/ vs. /t/, subjects (CG, JD, SF), positions (C2, C3), vowel contexts (/a/)

**ttip clg_off -tjaw clg_off in ms for /d/ vs. /t/**

| | | consonant | | | | | |
|---|---|---|---|---|---|---|---|
| | | d | | | t | | |
| SUBJECT | POSITION | N | Mean | Std. Deviation | N | Mean | Std. Deviation |
| cg | C2 | 10 | -14.94 | 12.65 | 10 | -16.80 | 6.83 |
| | C3 | 10 | -.41 | 1.81 | 10 | -21.63 | 4.95 |
| jd | C2 | 10 | 6.50 | 12.48 | 10 | -18.00 | 11.11 |
| | C3 | 10 | -14.05 | 4.41 | 10 | -13.00 | 4.01 |
| sf | C2 | 10 | -4.65 | 8.95 | 9 | -30.50 | 12.55 |
| | C3 | 10 | -1.12 | 1.12 | 9 | -34.53 | 10.00 |

**Table 24:** Number of tokens, means and standard deviations for closure duration (EPG defined) in ms, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Closure duration (EPG defined) in ms for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | 86.50 | 10.01 | 10 | 72.50 | 6.35 |
| | | i | 10 | 103.50 | 16.67 | 10 | 60.00 | 22.61 |
| | | u | 10 | 101.00 | 14.30 | 10 | 66.00 | 10.22 |
| | C2 | a | 10 | 39.50 | 8.96 | 10 | 52.50 | 10.07 |
| | | i | 10 | 44.50 | 16.06 | 10 | 65.50 | 15.36 |
| | | u | 10 | 42.50 | 16.20 | 10 | 74.00 | 8.43 |
| | C3 | a | 10 | 64.00 | 11.97 | 10 | 66.00 | 7.38 |
| | | i | 10 | 50.00 | 17.16 | 10 | 55.00 | 14.14 |
| | | u | 10 | 65.00 | 9.43 | 10 | 69.00 | 5.16 |
| jd | C1 | a | 10 | 73.50 | 14.35 | 10 | 74.50 | 10.12 |
| | | i | 10 | 108.00 | 13.37 | 10 | 103.50 | 22.86 |
| | | u | 10 | 96.50 | 10.55 | 10 | 97.50 | 11.84 |
| | C2 | a | 10 | 23.00 | 10.33 | 10 | 63.00 | 7.15 |
| | | i | 10 | 47.50 | 14.95 | 10 | 74.50 | 9.85 |
| | | u | 10 | 27.00 | 12.74 | 10 | 65.00 | 7.07 |
| | C3 | a | 10 | 48.00 | 5.87 | 10 | 48.50 | 9.73 |
| | | i | 10 | 51.50 | 8.83 | 10 | 59.50 | 6.85 |
| | | u | 10 | 45.00 | 5.27 | 10 | 45.00 | 7.07 |
| sf | C1 | a | 10 | 78.00 | 13.37 | 9 | 76.11 | 9.28 |
| | | i | 10 | 108.00 | 11.60 | 10 | 91.00 | 13.50 |
| | | u | 10 | 101.00 | 10.75 | 10 | 93.50 | 10.01 |
| | C2 | a | 10 | 44.00 | 3.94 | 9 | 55.56 | 12.10 |
| | | i | 10 | 37.50 | 13.59 | 10 | 67.50 | 10.34 |
| | | u | 10 | 53.50 | 8.18 | 10 | 65.50 | 7.98 |
| | C3 | a | 10 | 62.00 | 8.23 | 9 | 63.89 | 7.82 |
| | | i | 10 | 63.00 | 13.17 | 10 | 57.00 | 7.89 |
| | | u | 10 | 54.00 | 5.68 | 10 | 58.00 | 9.49 |

**Table 24:** Time normalised interpolated EPG parameters (ant, post, cog) during oral closure, /d/ (grey) vs. /t/ (black), subject (CG), position (C1), vowel contexts (/a i u/)

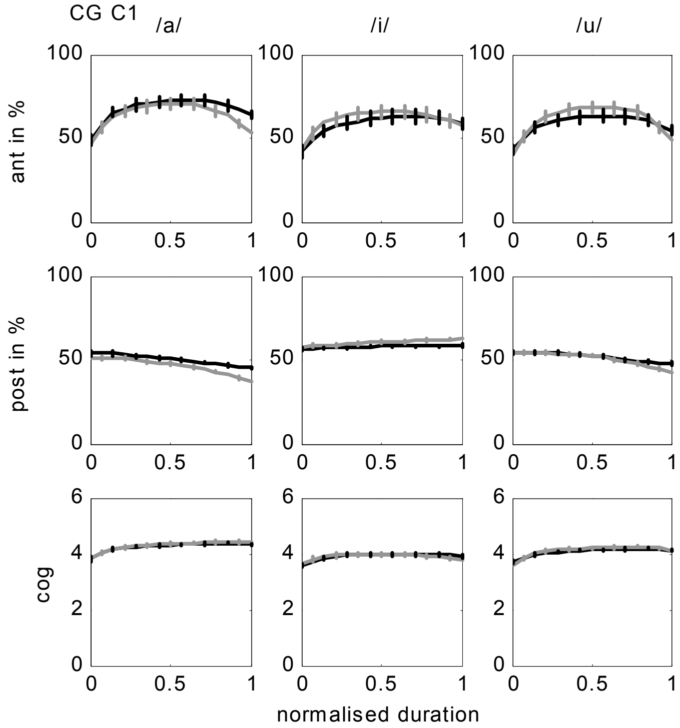

**Table 25:** Time normalised interpolated EPG parameters (ant, post, cog) during oral closure, /d/ (grey) vs. /t/ (black), subject (CG), position (C3), vowel contexts (/a i u/)

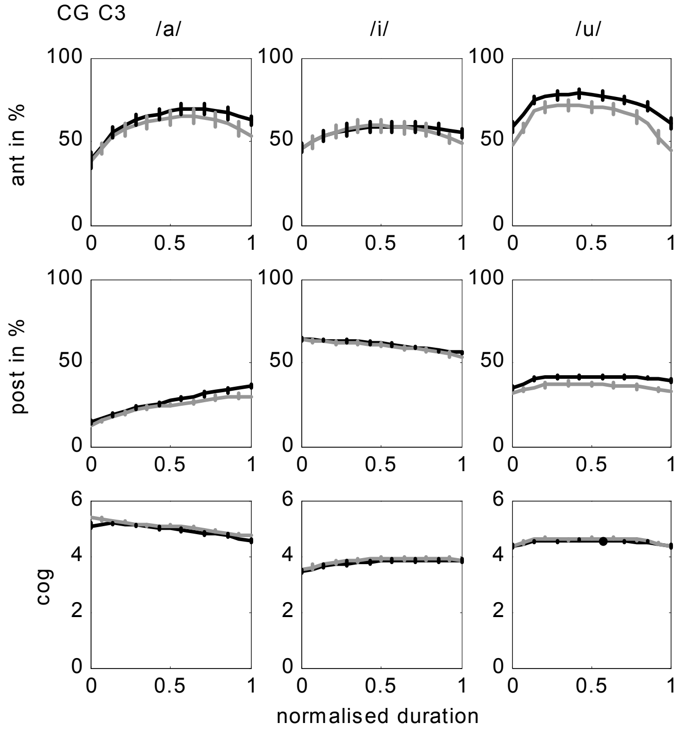

**Table 26:** Time normalised interpolated EPG parameters (ant, post, cog) during oral closure, /d/ (grey) vs. /t/ (black), subject (JD), position (C1), vowel contexts (/a i u/)

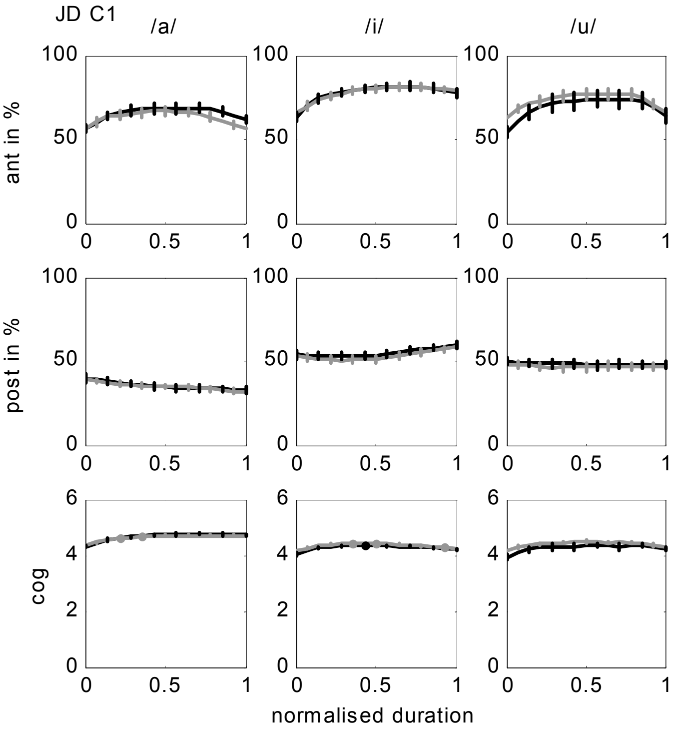

**Table 27:** Time normalised interpolated EPG parameters (ant, post, cog) during oral closure, /d/ (grey) vs. /t/ (black), subject (JD), position (C3), vowel contexts (/a i u/)

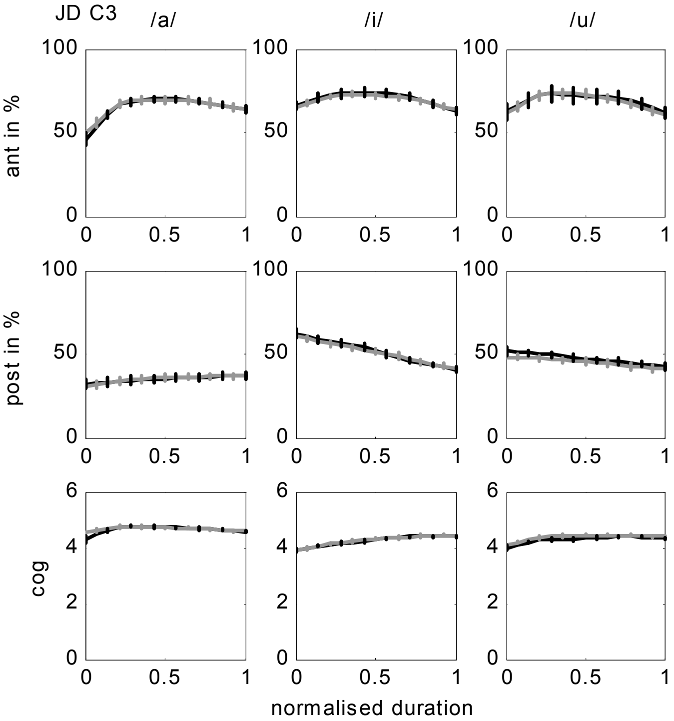

**Table 28:** Time normalised interpolated EPG parameters (ant, post, cog) during oral closure, /d/ (grey) vs. /t/ (black), subject (SF), position (C1), vowel contexts (/a i u/)

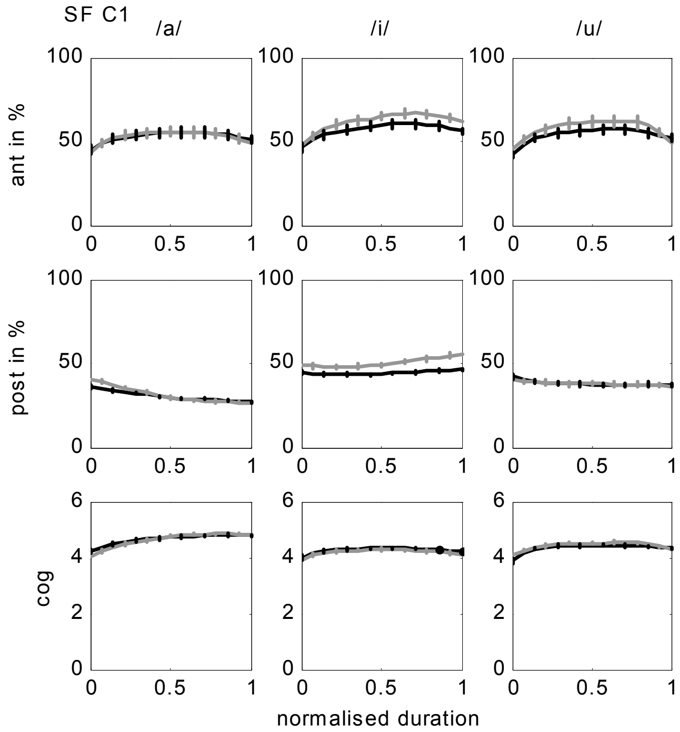

**Table 28:** Time normalised interpolated EPG parameters (ant, post, cog) during oral closure, /d/ (grey) vs. /t/ (black), subject (SF), position (C3), vowel contexts (/a i u/)

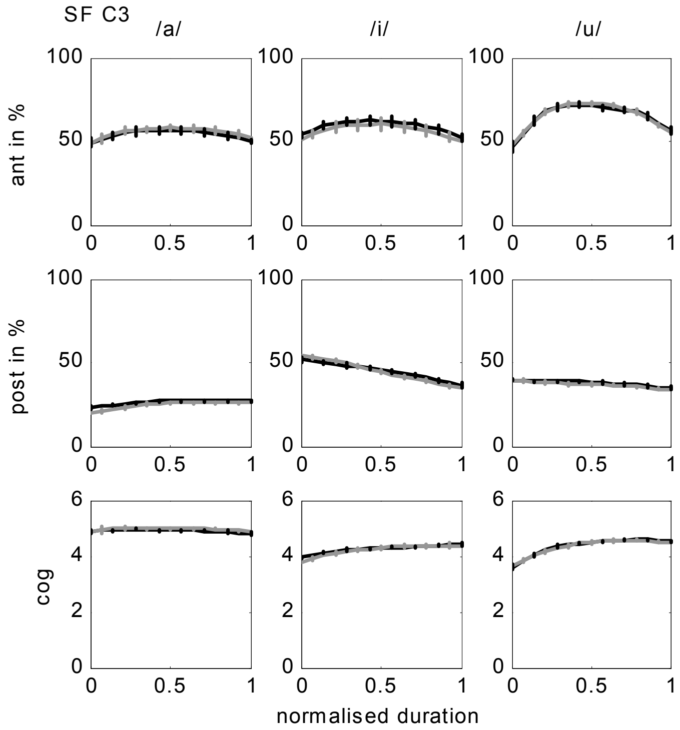

**Table 29:** Number of tokens, means and standard deviations for tongue tip y position at vowel target in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Tongue tip y position at vowel target in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | -.715 | .097 | 10 | -.664 | .086 |
| | | i | 10 | -.645 | .092 | 10 | -.628 | .146 |
| | | u | 10 | -.666 | .041 | 10 | -.670 | .079 |
| | C2 | a | 10 | -1.446 | .139 | 10 | -1.298 | .121 |
| | | i | 10 | -.285 | .094 | 10 | -.390 | .087 |
| | | u | 10 | -.786 | .109 | 10 | -.894 | .108 |
| | C3 | a | 10 | -1.396 | .121 | 10 | -1.250 | .141 |
| | | i | 10 | -.328 | .094 | 10 | -.394 | .061 |
| | | u | 10 | -.903 | .096 | 10 | -.951 | .067 |
| jd | C1 | a | 10 | .430 | .086 | 10 | .445 | .083 |
| | | i | 10 | .663 | .075 | 10 | .749 | .051 |
| | | u | 10 | .574 | .122 | 10 | .546 | .097 |
| | C2 | a | 10 | -.544 | .093 | 10 | -.311 | .079 |
| | | i | 10 | .789 | .101 | 10 | .795 | .068 |
| | | u | 10 | .616 | .097 | 10 | .440 | .073 |
| | C3 | a | 10 | -.421 | .116 | 10 | -.210 | .107 |
| | | i | 10 | .876 | .061 | 10 | .833 | .069 |
| | | u | 10 | .418 | .114 | 10 | .411 | .087 |
| sf | C1 | a | 10 | .003 | .088 | 9 | .021 | .112 |
| | | i | 10 | .101 | .157 | 10 | .075 | .123 |
| | | u | 10 | .075 | .101 | 10 | .021 | .141 |
| | C2 | a | 10 | -1.108 | .060 | 9 | -1.011 | .067 |
| | | i | 10 | .287 | .037 | 10 | .201 | .068 |
| | | u | 10 | -.109 | .050 | 10 | -.230 | .057 |
| | C3 | a | 10 | -1.033 | .094 | 9 | -.982 | .072 |
| | | i | 10 | .293 | .059 | 10 | .184 | .045 |
| | | u | 10 | -.078 | .056 | 10 | -.290 | .073 |

**Table 30:** Number of tokens, means and standard deviations for jaw y position at vowel target in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Jaw y position at vowel target in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | -1.528 | .067 | 10 | -1.515 | .089 |
| | | i | 10 | -1.463 | .101 | 10 | -1.508 | .080 |
| | | u | 10 | -1.486 | .079 | 10 | -1.526 | .146 |
| | C2 | a | 10 | -2.339 | .158 | 10 | -2.196 | .155 |
| | | i | 10 | -1.357 | .070 | 10 | -1.368 | .082 |
| | | u | 10 | -1.249 | .108 | 10 | -1.260 | .096 |
| | C3 | a | 10 | -2.277 | .113 | 10 | -2.032 | .146 |
| | | i | 10 | -1.361 | .084 | 10 | -1.343 | .092 |
| | | u | 10 | -1.231 | .112 | 10 | -1.217 | .085 |
| jd | C1 | a | 10 | -1.006 | .026 | 10 | -1.022 | .043 |
| | | i | 10 | -.967 | .035 | 10 | -.938 | .035 |
| | | u | 10 | -.978 | .023 | 10 | -.998 | .040 |
| | C2 | a | 10 | -1.219 | .026 | 10 | -1.192 | .039 |
| | | i | 10 | -.971 | .043 | 10 | -.945 | .032 |
| | | u | 10 | -.907 | .008 | 10 | -.902 | .013 |
| | C3 | a | 10 | -1.216 | .041 | 10 | -1.169 | .031 |
| | | i | 10 | -.934 | .037 | 10 | -.947 | .045 |
| | | u | 10 | -.910 | .012 | 10 | -.912 | .019 |
| sf | C1 | a | 10 | -1.109 | .048 | 9 | -1.120 | .033 |
| | | i | 10 | -1.108 | .084 | 10 | -1.094 | .057 |
| | | u | 10 | -1.080 | .076 | 10 | -1.071 | .044 |
| | C2 | a | 10 | -1.613 | .080 | 9 | -1.629 | .063 |
| | | i | 10 | -1.108 | .043 | 10 | -1.088 | .036 |
| | | u | 10 | -.958 | .022 | 10 | -.973 | .026 |
| | C3 | a | 10 | -1.621 | .054 | 9 | -1.605 | .052 |
| | | i | 10 | -1.100 | .054 | 10 | -1.086 | .032 |
| | | u | 10 | -.942 | .041 | 10 | -.962 | .020 |

**Table 31:** Number of tokens, means and standard deviations for tongue dorsum y position at vowel target in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Tongue dorsum y position at vowel target in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | 1.610 | .050 | 10 | 1.580 | .042 |
| | | i | 10 | 1.644 | .046 | 10 | 1.609 | .039 |
| | | u | 10 | 1.620 | .038 | 10 | 1.578 | .106 |
| | C2 | a | 10 | .147 | .060 | 10 | .262 | .087 |
| | | i | 10 | 1.717 | .013 | 10 | 1.711 | .020 |
| | | u | 10 | 1.118 | .069 | 10 | 1.052 | .059 |
| | C3 | a | 10 | .041 | .065 | 10 | .252 | .105 |
| | | i | 10 | 1.718 | .019 | 10 | 1.720 | .016 |
| | | u | 10 | 1.046 | .063 | 10 | .994 | .083 |
| jd | C1 | a | 10 | 1.409 | .157 | 10 | 1.377 | .129 |
| | | i | 10 | 1.316 | .119 | 10 | 1.386 | .123 |
| | | u | 10 | 1.532 | .091 | 10 | 1.566 | .036 |
| | C2 | a | 10 | .405 | .079 | 10 | .569 | .044 |
| | | i | 10 | 1.625 | .101 | 10 | 1.644 | .082 |
| | | u | 10 | 1.471 | .048 | 10 | 1.497 | .031 |
| | C3 | a | 10 | .469 | .088 | 10 | .666 | .060 |
| | | i | 10 | 1.637 | .058 | 10 | 1.677 | .031 |
| | | u | 10 | 1.479 | .032 | 10 | 1.476 | .043 |
| sf | C1 | a | 10 | 1.103 | .057 | 9 | 1.091 | .054 |
| | | i | 10 | 1.128 | .064 | 10 | 1.134 | .041 |
| | | u | 10 | 1.087 | .061 | 10 | 1.069 | .079 |
| | C2 | a | 10 | -.545 | .068 | 9 | -.421 | .072 |
| | | i | 10 | 1.175 | .031 | 10 | 1.009 | .084 |
| | | u | 10 | .883 | .034 | 10 | .794 | .035 |
| | C3 | a | 10 | -.510 | .072 | 9 | -.420 | .150 |
| | | i | 10 | 1.117 | .058 | 10 | .886 | .070 |
| | | u | 10 | .872 | .063 | 10 | .748 | .084 |

**Table 32:** Number of tokens, means and standard deviations for tongue tip y position at closure onset in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Tongue tip y position at closure onset (EPG defined) in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | .011 | .031 | 10 | .012 | .034 |
| | | i | 10 | .002 | .035 | 10 | -.003 | .024 |
| | | u | 10 | -.006 | .028 | 10 | -.005 | .027 |
| | C2 | a | 10 | .038 | .103 | 10 | .033 | .043 |
| | | i | 10 | .031 | .058 | 10 | .014 | .034 |
| | | u | 10 | .095 | .129 | 10 | .075 | .053 |
| | C3 | a | 10 | -.043 | .071 | 10 | -.023 | .060 |
| | | i | 10 | -.013 | .036 | 10 | -.014 | .061 |
| | | u | 10 | -.024 | .057 | 10 | .079 | .057 |
| jd | C1 | a | 10 | .476 | .077 | 10 | .515 | .073 |
| | | i | 10 | .646 | .065 | 10 | .722 | .057 |
| | | u | 10 | .668 | .099 | 10 | .691 | .065 |
| | C2 | a | 10 | .327 | .033 | 10 | .462 | .084 |
| | | i | 10 | .791 | .081 | 10 | .845 | .057 |
| | | u | 10 | .700 | .062 | 10 | .680 | .038 |
| | C3 | a | 10 | .365 | .048 | 10 | .426 | .075 |
| | | i | 10 | .846 | .046 | 10 | .874 | .070 |
| | | u | 10 | .568 | .070 | 10 | .654 | .038 |
| sf | C1 | a | 10 | .234 | .064 | 9 | .213 | .044 |
| | | i | 10 | .297 | .071 | 10 | .261 | .028 |
| | | u | 10 | .219 | .079 | 10 | .223 | .045 |
| | C2 | a | 10 | .124 | .024 | 9 | .131 | .045 |
| | | i | 10 | .373 | .047 | 10 | .389 | .072 |
| | | u | 10 | .408 | .077 | 10 | .417 | .062 |
| | C3 | a | 10 | .096 | .043 | 9 | .116 | .062 |
| | | i | 10 | .371 | .047 | 10 | .336 | .044 |
| | | u | 10 | .378 | .040 | 10 | .371 | .068 |

**Table 33:** Number of tokens, means and standard deviations for jaw y position at closure onset in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Jaw y position at closure onset (EPG defined) in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | -1.361 | .057 | 10 | -1.332 | .055 |
| | | i | 10 | -1.321 | .094 | 10 | -1.343 | .079 |
| | | u | 10 | -1.380 | .073 | 10 | -1.364 | .081 |
| | C2 | a | 10 | -1.741 | .107 | 10 | -1.352 | .070 |
| | | i | 10 | -1.358 | .057 | 10 | -1.290 | .067 |
| | | u | 10 | -1.345 | .105 | 10 | -1.138 | .079 |
| | C3 | a | 10 | -1.463 | .101 | 10 | -1.416 | .056 |
| | | i | 10 | -1.307 | .082 | 10 | -1.282 | .091 |
| | | u | 10 | -1.219 | .086 | 10 | -1.133 | .076 |
| jd | C1 | a | 10 | -.988 | .031 | 10 | -.979 | .029 |
| | | i | 10 | -.981 | .038 | 10 | -.973 | .035 |
| | | u | 10 | -.953 | .013 | 10 | -.963 | .029 |
| | C2 | a | 10 | -.956 | .016 | 10 | -1.008 | .026 |
| | | i | 10 | -.969 | .040 | 10 | -.944 | .029 |
| | | u | 10 | -.927 | .021 | 10 | -.907 | .012 |
| | C3 | a | 10 | -1.054 | .021 | 10 | -1.023 | .029 |
| | | i | 10 | -.947 | .036 | 10 | -.955 | .052 |
| | | u | 10 | -.921 | .015 | 10 | -.921 | .020 |
| sf | C1 | a | 10 | -1.055 | .058 | 9 | -1.057 | .024 |
| | | i | 10 | -1.032 | .047 | 10 | -1.025 | .043 |
| | | u | 10 | -1.028 | .074 | 10 | -1.017 | .041 |
| | C2 | a | 10 | -1.200 | .050 | 9 | -1.143 | .049 |
| | | i | 10 | -1.094 | .027 | 10 | -1.060 | .036 |
| | | u | 10 | -.975 | .021 | 10 | -.943 | .016 |
| | C3 | a | 10 | -1.211 | .038 | 9 | -1.149 | .053 |
| | | i | 10 | -1.103 | .061 | 10 | -1.074 | .034 |
| | | u | 10 | -.966 | .049 | 10 | -.952 | .018 |

**Table 34:** Number of tokens, means and standard deviations for tongue dorsum y position at closure onset in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Tongue dorsum y position at closure onset (EPG defined) in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | 1.391 | .064 | 10 | 1.396 | .053 |
| | | i | 10 | 1.593 | .039 | 10 | 1.527 | .045 |
| | | u | 10 | 1.458 | .052 | 10 | 1.409 | .097 |
| | C2 | a | 10 | .466 | .109 | 10 | .632 | .101 |
| | | i | 10 | 1.694 | .041 | 10 | 1.703 | .013 |
| | | u | 10 | 1.275 | .072 | 10 | 1.258 | .070 |
| | C3 | a | 10 | .402 | .126 | 10 | .492 | .148 |
| | | i | 10 | 1.685 | .037 | 10 | 1.602 | .254 |
| | | u | 10 | 1.152 | .083 | 10 | 1.151 | .070 |
| jd | C1 | a | 10 | 1.025 | .161 | 10 | 1.110 | .097 |
| | | i | 10 | 1.395 | .070 | 10 | 1.457 | .102 |
| | | u | 10 | 1.348 | .131 | 10 | 1.415 | .085 |
| | C2 | a | 10 | .811 | .040 | 10 | .750 | .094 |
| | | i | 10 | 1.560 | .090 | 10 | 1.587 | .060 |
| | | u | 10 | 1.369 | .073 | 10 | 1.456 | .045 |
| | C3 | a | 10 | .724 | .050 | 10 | .786 | .042 |
| | | i | 10 | 1.523 | .081 | 10 | 1.556 | .065 |
| | | u | 10 | 1.338 | .066 | 10 | 1.386 | .031 |
| sf | C1 | a | 10 | .845 | .079 | 9 | .742 | .122 |
| | | i | 10 | .951 | .097 | 10 | .877 | .092 |
| | | u | 10 | .833 | .090 | 10 | .841 | .119 |
| | C2 | a | 10 | .040 | .042 | 9 | .093 | .040 |
| | | i | 10 | 1.030 | .093 | 10 | 1.037 | .029 |
| | | u | 10 | .846 | .055 | 10 | .759 | .045 |
| | C3 | a | 10 | -.038 | .049 | 9 | .006 | .124 |
| | | i | 10 | 1.041 | .049 | 10 | .909 | .073 |
| | | u | 10 | .827 | .058 | 10 | .727 | .056 |

**Table 35:** Number of tokens, means and standard deviations for tongue tip y position at closure offet in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Tongue tip y position at closure offset (EPG defined) in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | -.040 | .034 | 10 | .029 | .019 |
| | | i | 10 | -.019 | .043 | 10 | .001 | .016 |
| | | u | 10 | -.069 | .038 | 10 | -.007 | .036 |
| | C2 | a | 10 | .016 | .054 | 10 | .077 | .034 |
| | | i | 10 | .000 | .052 | 10 | .017 | .028 |
| | | u | 10 | -.025 | .065 | 10 | .124 | .024 |
| | C3 | a | 10 | -.059 | .055 | 10 | .035 | .043 |
| | | i | 10 | -.039 | .029 | 10 | -.076 | .225 |
| | | u | 10 | -.082 | .042 | 10 | .063 | .060 |
| jd | C1 | a | 10 | .265 | .076 | 10 | .407 | .084 |
| | | i | 10 | .789 | .071 | 10 | .828 | .060 |
| | | u | 10 | .577 | .083 | 10 | .600 | .080 |
| | C2 | a | 10 | .357 | .041 | 10 | .537 | .037 |
| | | i | 10 | .603 | .081 | 10 | .673 | .091 |
| | | u | 10 | .564 | .081 | 10 | .525 | .070 |
| | C3 | a | 10 | .421 | .045 | 10 | .469 | .036 |
| | | i | 10 | .589 | .107 | 10 | .591 | .114 |
| | | u | 10 | .359 | .097 | 10 | .443 | .078 |
| sf | C1 | a | 10 | .111 | .043 | 9 | .131 | .035 |
| | | i | 10 | .418 | .034 | 10 | .334 | .025 |
| | | u | 10 | .185 | .045 | 10 | .188 | .032 |
| | C2 | a | 10 | .162 | .033 | 9 | .196 | .040 |
| | | i | 10 | .247 | .039 | 10 | .226 | .056 |
| | | u | 10 | .223 | .048 | 10 | .265 | .053 |
| | C3 | a | 10 | .107 | .047 | 9 | .129 | .048 |
| | | i | 10 | .176 | .044 | 10 | .156 | .060 |
| | | u | 10 | .154 | .060 | 10 | .168 | .063 |

**Table 36:** Number of tokens, means and standard deviations for jaw y position at closure offet in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Jaw y position at closure offset (EPG defined) in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | -1.333 | .083 | 10 | -1.211 | .058 |
| | | i | 10 | -1.236 | .079 | 10 | -1.257 | .078 |
| | | u | 10 | -1.243 | .075 | 10 | -1.273 | .088 |
| | C2 | a | 10 | -1.551 | .082 | 10 | -1.156 | .058 |
| | | i | 10 | -1.385 | .052 | 10 | -1.228 | .065 |
| | | u | 10 | -1.482 | .074 | 10 | -1.214 | .086 |
| | C3 | a | 10 | -1.333 | .074 | 10 | -1.188 | .065 |
| | | i | 10 | -1.313 | .090 | 10 | -1.232 | .083 |
| | | u | 10 | -1.348 | .066 | 10 | -1.238 | .064 |
| jd | C1 | a | 10 | -1.016 | .028 | 10 | -.947 | .023 |
| | | i | 10 | -.937 | .022 | 10 | -.922 | .015 |
| | | u | 10 | -.918 | .011 | 10 | -.906 | .016 |
| | C2 | a | 10 | -.933 | .010 | 10 | -.931 | .007 |
| | | i | 10 | -.969 | .048 | 10 | -.920 | .016 |
| | | u | 10 | -.941 | .029 | 10 | -.926 | .012 |
| | C3 | a | 10 | -.988 | .021 | 10 | -.967 | .015 |
| | | i | 10 | -.968 | .042 | 10 | -.962 | .062 |
| | | u | 10 | -.955 | .032 | 10 | -.954 | .040 |
| sf | C1 | a | 10 | -1.077 | .055 | 9 | -1.016 | .034 |
| | | i | 10 | -1.040 | .054 | 10 | -1.012 | .032 |
| | | u | 10 | -.957 | .038 | 10 | -.956 | .026 |
| | C2 | a | 10 | -1.054 | .043 | 9 | -.997 | .051 |
| | | i | 10 | -1.082 | .035 | 10 | -1.024 | .032 |
| | | u | 10 | -1.014 | .026 | 10 | -.958 | .016 |
| | C3 | a | 10 | -1.010 | .036 | 9 | -.988 | .020 |
| | | i | 10 | -1.102 | .060 | 10 | -1.062 | .025 |
| | | u | 10 | -1.065 | .053 | 10 | -.994 | .030 |

**Table 37:** Number of tokens, means and standard deviations for tongue dorsum y position at closure offet in cm, /d/ vs. /t/, subjects (CG, JD, SF), positions (C1, C2, C3), vowel contexts (/a i u/)

**Tongue dorsum y position at closure offset (EPG defined) in cm for /d/ vs. /t/**

| | | | consonant | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | d | | | t | | |
| SUBJECT | POSITION | VOWEL | N | Mean | Std. Dev. | N | Mean | Std. Dev. |
| cg | C1 | a | 10 | .861 | .046 | 10 | .979 | .066 |
| | | i | 10 | 1.685 | .028 | 10 | 1.511 | .062 |
| | | u | 10 | 1.175 | .033 | 10 | 1.176 | .064 |
| | C2 | a | 10 | .771 | .127 | 10 | .933 | .108 |
| | | i | 10 | 1.489 | .099 | 10 | 1.476 | .067 |
| | | u | 10 | 1.288 | .076 | 10 | 1.300 | .088 |
| | C3 | a | 10 | .712 | .088 | 10 | .842 | .139 |
| | | i | 10 | 1.394 | .085 | 10 | 1.105 | .669 |
| | | u | 10 | 1.011 | .049 | 10 | 1.021 | .067 |
| jd | C1 | a | 10 | .619 | .107 | 10 | .787 | .106 |
| | | i | 10 | 1.486 | .081 | 10 | 1.456 | .114 |
| | | u | 10 | 1.258 | .096 | 10 | 1.274 | .072 |
| | C2 | a | 10 | .866 | .045 | 10 | 1.040 | .066 |
| | | i | 10 | 1.281 | .074 | 10 | 1.223 | .100 |
| | | u | 10 | 1.265 | .064 | 10 | 1.219 | .058 |
| | C3 | a | 10 | .896 | .079 | 10 | .942 | .079 |
| | | i | 10 | 1.096 | .123 | 10 | 1.126 | .121 |
| | | u | 10 | 1.013 | .100 | 10 | 1.091 | .098 |
| sf | C1 | a | 10 | .205 | .073 | 9 | .228 | .061 |
| | | i | 10 | 1.018 | .085 | 10 | .809 | .050 |
| | | u | 10 | .663 | .127 | 10 | .607 | .052 |
| | C2 | a | 10 | .149 | .062 | 9 | .251 | .069 |
| | | i | 10 | .746 | .110 | 10 | .641 | .056 |
| | | u | 10 | .596 | .067 | 10 | .489 | .076 |
| | C3 | a | 10 | .053 | .069 | 9 | .106 | .142 |
| | | i | 10 | .596 | .122 | 10 | .487 | .110 |
| | | u | 10 | .514 | .113 | 10 | .427 | .086 |