Experimental studies of laryngeal articulation

Philip Hoole

Part II: Laryngeal-Oral Coordination in Consonant Sequences

Experimental studies of laryngeal articulation

Habilitationsschrift

zur Erlangung der Venia Legendi für das Fach Phonetik

Fakultät für Sprach- und Literaturwissenschaften
der Ludwig-Maximilians-Universität

vorgelegt von

Dr. phil. Philip Hoole

München, Oktober 2006

Part II: Laryngeal-Oral Coordination in Consonant Sequences

Preface

This study presents two sets of experiments on laryngeal function, entitled "*Electromyographic investigation of laryngeal activity in vowel intrinsic pitch and consonant voicing*" (Part I), and "*Laryngeal-Oral Coordination in Consonant Sequences*" (Part II)[1].

Instrumental studies of articulatory processes are inevitably a collaborative endeavour, and it is now a great pleasure to acknowledge all those involved in these experiments:

The EMG experiments grew out of a set of pilot experiments carried out at ATR Labs, Kyoto, after I had approached Kiyoshi Honda with the idea of performing laryngeal EMG to find out whether German really is a problem for theories of intrinsic pitch. I am very grateful to him for the open ear he lent to my first tentative suggestions. However, it would never have been feasible to carry out a full set of EMG recording sessions in Germany if I had been the only person interested in this kind of data. The crucial elements, in addition to Kiyoshi Honda and Emi Murano who did the needlework, were the interest and support of the phonetics lab of Zentrum für Allgemeine Sprachwissenschaft, Berlin, together with the facilities of Rafael Laboissière's lab for sensorimotor coordination at the Max-Planck-Institute for Psychological Research, Munich. This gave us the opportunity to record data relevant not just for the questions I was originally interested in, but in particular also for lingual and mandibular function. Special recognition is owed here to Christian Kroos who was the first subject for both laryngeal and lingual EMG and thus boldly went - as far as I am aware - where no German subject had gone before, at least in the context of hooked-wire EMG for phonetic research.

The second group of experiments, which used transillumination and fiberoptic filming to study laryngeal kinematics, was carried out in Berlin at the ZAS phonetics lab, using an experimental setup I had implemented some years ago. The specific topic of consonant sequences which is focussed on here represents just one aspect of the many recordings we have carried out together over the last few years. The long-suffering subjects for the present recordings were Ralf Winkler, Suse Fuchs and Christian Geng (the latter two also gave freely of their neck and tongue for the EMG experiments). Jörg Dreyer kept the lab running smoothly, and Dr. Klaus Dahlmeier wielded a mean fiberscope.

Finally, this work is dedicated to Hans-G. Tillmann on the occasion of his retirement. His unflagging enthusiasm for our subject over all the many years that he was head of phonetics at Munich have been a tremendous motivation.

---

[1]Part II includes as an appendix the extended version of a very old conference paper on velar and glottal activity in Icelandic. I prepared this extended version following the conference but never published it. It seemed worth including here because the present experiments in effect apply to German some ideas originally developed in that paper.

Table of Contents for Part II

1 Introduction

Consonant clusters constitute some of the motorically most complex sequences in human speech, and are particularly well represented in many European languages. However, our knowledge of how the component gestures are coordinated with each other is still extremely fragmentary. Consideration of laryngeal-oral coordination as part of this picture is a fruitful area of investigation because the laryngeal component (i.e the so-called devoicing gesture) has, at least superficially, a fairly straightforward movement pattern (compared, for example, to the tongue) and is thus quite amenable to analysis. Moreover, previous investigations have given us a good understanding of basic kinematic properties of these movements.

The present work looks at two specific areas of syllable-initial consonants where the underlying pattern of organization is contentious (the first case), or even basically unknown but potentially revealing (the second case).

The first case is that of purely voiceless sequences such as /**ʃp**/. Here there are differences of opinion as to whether the laryngeal movement is better analyzed as a blending of two underlying gestures corresponding to the two voiceless segments, or whether it reflects a phonological generalization that only one glottal gesture can occur in syllable onsets. In addition, there are also open issues regarding the formulation of the coordination relations, for example whether it is possible to formulate a rather generally applicable rule such as to coordinate peak glottal opening with the midpoint of the fricative if one is present in the syllable onset. For this reason we recorded not only fricative-plosive sequences, but also plosive-fricatives since these have been much less investigated, with German offering a number of possibilities not present in English (e.g /**pf**, **ps**/).

The second area involves syllable onsets consisting of one or more voiceless segments followed by a sonorant (in the present experiments we only consider /**l**/), e.g /**pl, ʃl, ʃpl**/. It is well-known that /**l**/ in such cases can be wholly or partially voiceless. The interesting question is whether this can be seen as a purely passive coarticulatory effect, or whether there are more active adjustments of laryngeal-oral coordination. There are some indications in the literature (mostly interpretation of purely acoustic findings) that there may indeed be an active enhancement of voicelessness when a sonorant is added to the syllable onset. This has intriguing implications for the level of the syllabic hierarchy at which the devoicing gesture is organized, and at least speaks against the assembly of complex onsets from a simple concatenation of the component segments. Thus direct examination of laryngeal activity appeared worthwhile. Moreover, if a seemingly ‘irrelevant’ segment such as a sonorant can result in reorganization of laryngeal-oral coordination, then this has potential implications for how coordination patterns in purely voiceless sequences are interpreted.

The results, briefly, came down in favour of a single-gesture account for purely voiceless onsets, and also gave some evidence for active reorganization in clusters with sonorants. Furthermore, it appeared that the coordination relations may not be formulated most appropriately in terms of the timing of a single time-point in the glottal gesture (generally peak glottal opening) with a specific time-point in the oral segments, but rather in terms of fulfilment of a set of constraints specific to the aerodynamic and functional demands of each syllable onset. Speakers may often have considerable flexibility in how this is achieved.

The presentation of the work is structured as follows:

Chapter 2 reviews the state of our knowledge on laryngeal-oral coordination, focussing on the two areas just outlined above. However, this is preceded by a discussion of single consonants (plosives and fricatives), since the issues raised for the more complex sequences are often only understandable on the background of accounts of single consonants. The chapter concludes by summarizing those issues that have, in my view, not been resolved, thus laying out the motivation for the specific experiments carried out here.

Following Chapter 3, which gives a brief, mainly historical view of appropriate techniques for investigation of laryngeal activity when coordination with supraglottal activity is at the centre of attention, Chapter 4 then presents in detail the procedures followed for the current experiments, including an overview of the speech material recorded (the latter having a somewhat complicated structure because the experiments were designed to collect material for different purposes, of which only part is relevant here). A particular feature of the data collection was the employment not only of photoelectroglottography (aka transillumination) to directly monitor laryngeal activity, but also of electropalatography to capture lingual articulation. Accordingly, Chapter 4 closes with an extensive set of examples examining where EPG is of most benefit for analyzing the details of the oral component of laryngeal-oral coordination.

Perhaps the most important task of this introduction is to give a roadmap to how the presentation of the results can be read. The main purpose of Chapter 5 is to collect at a central location in this volume a tabulation of the complete set of statistical results. Thus it is probably not necessary to read through this section page by page (or, if so, only as an additional summary after reading all of Chapter 6). Its intention is more the following: Regardless of which subsection of the detailed results one is reading in Chapter 6 it should always be possible to quickly find an overview of the results of any other section.

Chapter 6 itself gives a blow-by-blow presentation of the detailed results for each of the categories of syllable onset. Each subdivision (for example the main section on voiceless clusters is subdivided into fricative-plosive and plosive-fricative sequences) is organized in a very similar way, starting with a single figure giving an overview of the temporal structure of the given consonant sequence, continuing with detailed consideration of each measurement parameter relevant to that particular sequence, and finishing with a brief summary. Thus, a quick way of following the main trends in the results, and relating them to the issues summarized at the end of Chapter 2, is to examine in each subsection the initial overview figure (with its accompanying comments) together with the summary at the end of the subsection.

The work closes in Chapter 7 with an overall summary and discussion, leading to perspectives for future work.

The appendix on Icelandic, already mentioned in the general preface to this monograph, has a preface of its own indicating the context in which it should be read.

2 Laryngeal-oral coordination in single voiceless consonants and consonant sequences: results and implications of previous investigations

2.1 Introduction[2]

We will preface the discussion of the two main topics in this section, i.e (1) organisation of the devoicing gesture in clusters containing voiceless consonant(s) plus sonorant, and (2) laryngeal kinematics in purely voiceless clusters, with consideration of some of the basic kinematic properties of laryngeal articulation in single voiceless plosives and fricatives. This will form the background for the central discussion of longer consonantal sequences. As we will see, some interesting questions already emerge here, that can then be picked up again with respect to these longer sequences.

2.2 Properties of single voiceless consonants

We will organize this section around a series of comparisons: in the first part with respect to manner of articulation (i.e plosives vs. fricatives) and in the second part with respect to place of articulation. Within each subsection we will compare various aspects of the amplitude and timing of the devoicing gesture. To set the scene, typical traces of the transillumination and audio signal for a voiceless aspirated plosive and for a voiceless fricative are given in the following figure[3].

2.2.1 Manner of articulation

There is a fairly widespread finding in the literature that the amplitude of the devoicing gesture is larger for fricatives than plosives. A straight comparison of single plosives and fricatives with this result is found for example in McGarr & Löfqvist (1988), Löfqvist & McGarr (1987), Munhall & Ostry (1985) (based on the ultrasound measurements in the latter, the difference in

---

[2]A preliminary version of this chapter, not covering the more recent literature, appeared in Hoole, 1999a.

[3]Detailed discussion of the methodology of transillumination (photoelectroglottography) is given in the next two chapters.

the amplitude of vocal fold abduction for plosives and fricatives is a mere 0.25mm, though still significant). A group of papers in which clusters rather than single voiceless sounds were examined points clearly in the same direction (Löfqvist & Yoshioka, 1980a,b; Yoshioka, Löfqvist & Hirose, 1980, 1981). Summarizing these latter studies, Yoshioka et al. (1980, p.306) go as far as to say regarding the more vigorous abduction in fricatives that

> *"this finding for fricatives is also consistent with our recent studies using American English, Icelandic and Swedish although the phonologies differ, among other things, in the significance of stop aspiration. Therefore, we are inclined to conclude that at least the difference in the peak value between a voiceless fricative and a voiceless stop is universal."*

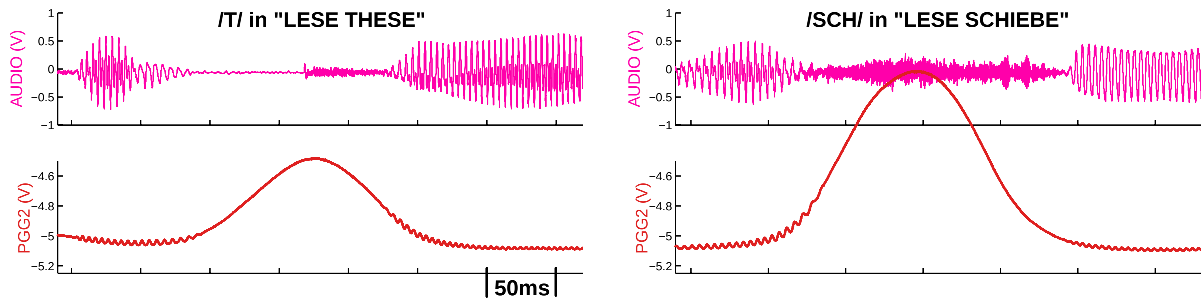


Fig. 2.1: Example of transillumination signal for voiceless aspirated plosive (left) and voiceless fricative (right)

In fact, this may slightly overstate the situation: the amount of aspiration required for stops in specific languages may occasionally override this tendency. In an extensive study of Danish (with the unusually large number of 5 subjects in the kinematic part of her study) Hutters (1984) found slightly but significantly larger peak glottal opening in aspirated stops than in fricatives[4]. She notes that aspiration is more extensive in Danish than e.g Swedish. She also notes the possibility, in view of the subtlety of the differences, that differences in larynx height for the different sounds compared may interfere with the interpretation of the amplitude of the transillumination signal.

With regard to the timing of the devoicing gesture, one robust difference between fricatives and (aspirated) plosives that emerges clearly from the literature is that the onset of glottal abduction is earlier for fricatives, relative to the formation of the oral closure (e.g Hutters, 1984; Hoole, Pompino-Marschall & Dames, 1984; Löfqvist & McGarr, 1987; Butcher, 1977; Jessen, 1998, 1999; for further comparative information on glottal timing in fricatives and aspirated stops see Löfqvist & Yoshioka, 1984). The reason is probably to be found in the aerodynamic requirements of fricative production. Löfqvist & McGarr (1987) discuss reasons for the larger glottal gesture in fricatives, but their remarks could equally well apply to the early onset of abduction in fricatives (p. 399):

> *"the larger gesture for a voiceless fricative is most likely due to the aerodynamics of fricative production, in that a large glottal opening not only prevents voicing but also reduces laryngeal resistance to air flow and assists in the build-up of oral pressure necessary for driving the noise source."*

---

[4]There is also a report by Butcher (1977) for one (probably) English speaker showing greater peak glottal opening on plosives than fricatives; but very few experimental details are given, so the significance of this result is difficult to assess.

The aerodynamically crucial phase of a fricative is probably its onset, whereas for an aspirated plosive the crucial phase is the offset (in addition, Löfqvist & McGarr suggest that early onset of glottal abduction is avoided in English stops as inappropriate preaspiration might otherwise occur). Related to this is a tendency for fricatives to show higher velocities and tighter timing control in the abduction phase compared with the adduction phase on the one hand, and compared with plosives on the other hand. However, the picture to be found in the literature is not completely consistent (cf. Löfqvist & McGarr, 1987).

A visual impression of the typical differences between fricatives and plosives is given in the following figure adapted from Yoshioka et al. (1981, Fig. 12); refer back also to the two individual examples at the start of this section.

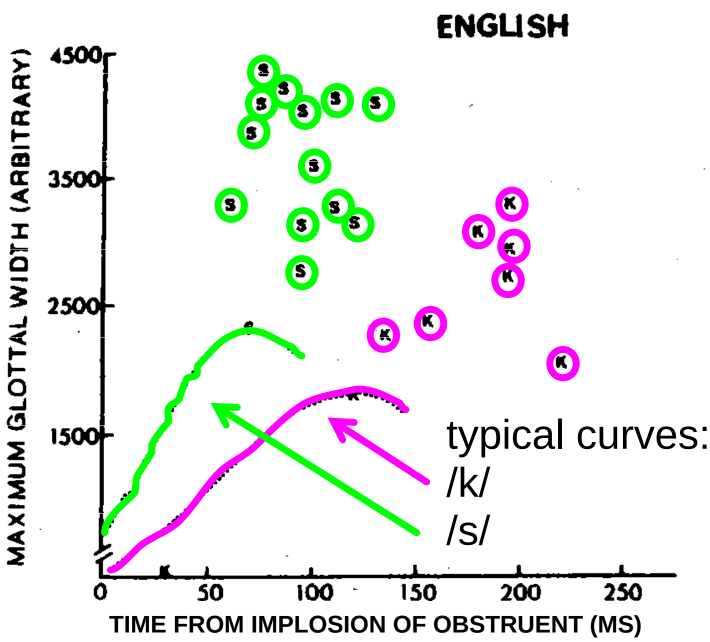


Fig. 2.2: Comparison of devoicing gesture for fricatives and plosives. Adapted from Yoshioka et al., 1981, Fig. 12

Another way of looking at the early onset of glottal abduction in fricatives is with respect to the onset of the preceeding vowel. It is well-known that vowels tend to be longer before fricatives. Hoole et al. (1984) suggested (on the basis of material that was not ideally suited since it compared a plosive with an /**st**/ combination rather than a singleton fricative) that the timing of glottal abduction could be identical for plosives and fricatives when viewed from the onset of the previous vowel. This is illustrated in the following figure.

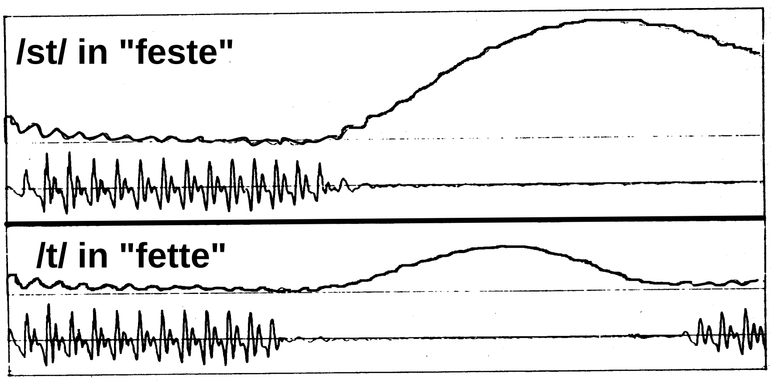


Fig. 2.3: Fig. 2 from Hoole et al. (1984). Glottis signal and audio signal for /**st**/ from “feste” (top panel) and /**t**/ from “fette” (bottom panel) lined-up at the onset of the preceding vowel (/**ɛ**/)

However, the more balanced material of Hutters (1984) failed to confirm this, since although the expected differences in vowel length were found, they were not large enough to completely compensate for the difference in time of glottal abduction relative to fricative constriction and stop closure; significant timing differences between stops and fricatives remained. Nonetheless, the theme of the relative amount of reorganisation of laryngeal and oral articulations is one that we will be returning to.

2.2.2 Place of Articulation

There are surprisingly few studies that compare the laryngeal devoicing gesture with respect to place of articulation. Regarding the amplitude of the gesture Hutters (1984) found in Danish that peak glottal opening was greater for /**s**/ than for /**f**/, and for /**t**/ than for /**p**/ (although the latter comparison did not reach statistical significance, perhaps being complicated by the fact that /**t**/ is affricated in Danish). Cooper (1991) compared the stops /**p, t, k**/ in two speakers of American English and found a significant place of articulation effect for peak glottal opening, but the pattern of results was not straightforward since the different stops were differently affected by the experimental variation of stress and position of the stop in the word[5].

Probably the more interesting issue is whether the ***timing*** of the devoicing gesture is influenced by place of articulation, particularly in aspirated stops. Refer to the following figure for a schematic illustration of the relation between the time-course of a typical devoicing gesture and the oral occlusion and aspiration phases in such sounds (i.e a schematic counterpart to the actual utterance shown above in the first figure in this chapter). This schematic figure will be used below as a framework for explaining the possible control hypotheses.

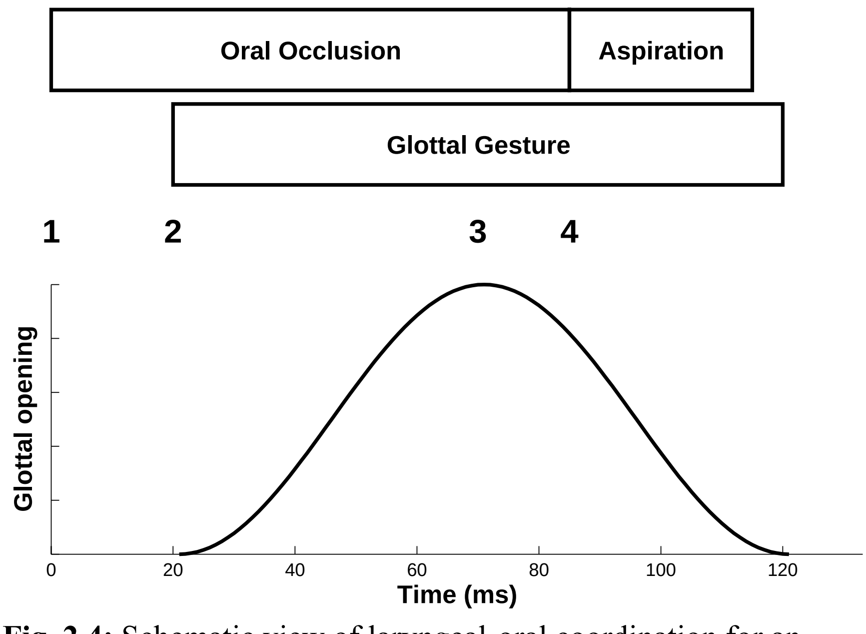


Fig. 2.4: Schematic view of laryngeal-oral coordination for an aspirated plosive. The four time-points marked with digits are: (1) Onset of oral occlusion; (2) Onset of glottal gesture; (3) Peak glottal opening (PGO); (4) Release of oral occlusion

---

[5]Regarding amplitude of the glottal gesture see also discussion of uvular fricatives on p. 14 below.

The interest derives from the widespread observation that place of articulation has a significant effect on VOT. The most robust finding is that /**p**/ has shorter VOT than /**t**/ or /**k**/. Whether there is a general relationship of the form p<t<k (i.e longer VOT for more retracted consonants) is more open to debate (see e.g Docherty, 1992, for discussion). Disregarding possible additional aerodynamic effects for the moment, this suggests that peak glottal opening is timed earlier with respect to release for /**p**/ than for the other plosives. On the other hand, /**p**/ also generally has a longer occlusion duration than the other stops. Taken together this raises the possibility that the devoicing gesture has essentially the same duration for all stops, and that the differences in VOT are a simple passive effect of different oral occlusion durations superimposed on a constant laryngeal gesture. A suggestion along these lines has been put forward by Weismer (1980) and by Suomi (1980; cited in Docherty, 1992, p.137) on the basis of durational analysis of acoustic data. Hutters (1985) also presents some evidence for a similar effect operating across languages rather than across place of articulation: i.e languages with short occlusion phases have long aspiration phases, and vice versa. Docherty notes that Suomi's conclusion was based on consideration of mean duration (occlusion, VOT, total devoicing) for each stop category and himself applies what he regards as a more stringent test of the hypothesis: in addition to examining mean duration values (which confirmed the existence of a reciprocity between occlusion and VOT duration) he also tested for a negative correlation between the two variables, since under the hypothesis of an invariant gesture a strong negative correlation should occur. The evidence for this was, however, rather weak. In comparison with the rather weak negative correlations for occlusion vs. VOT, Docherty found fairly strong positive correlations between total abduction duration and VOT, which can be seen as a test that there *are* laryngeal differences, and these are responsible for VOT[6].

Of the few relevant transillumination studies, Hoole et al. (1984) found over the German stops /**p**/ and /**t**/ a reciprocal relationship between occlusion duration and the duration of the interval from peak glottal opening to release, but did not test the constancy of the devoicing gesture directly.

Jessen (1999) looked closely at this question on the basis of his extensive material for one German speaker. Assuming that aspiration is closely related to the interval from PGO to Release, the question to be answered is what is primarily responsible for differences in this interval between consonants. Jessen (p.993) very usefully formulates this in terms of three hypotheses (refer to the labels for laryngeal and oral events in the above figure)[7]:

A. The ***late glottal gesture hypothesis***: with "*glottal gesture duration*" and "*oral occlusion duration*" remaining constant, "*PGO to release*" is shortened (i.e becomes less positive, or even negative) by shifting the glottal gesture rightwards. That this happens can be inferred from increased values of "*onset of oral occlusion to onset of glottal gesture*".

---

[6]It should be noted that, however, that the interpretation of these latter correlations is somewhat problematic, as they are part-whole correlations (cf. Benoit, 1986).

[7]We have replaced Jessen's own abbreviations for the various intervals and events to be consistent with those used in Fig. 2.4 on p. 6 above. In particular, we define the interval PGO to release to be positive when PGO comes *before* release of the oral occlusion, whereas Jessen uses a definition that gives the opposite sign for this interval.

B. The ***long glottal gesture hypothesis***: with "*onset of oral occlusion to onset of glottal gesture*" and "*oral occlusion duration*" remaining constant,"*PGO to release*" is shortened by lengthening "*glottal gesture duration*". As an increase in "*glottal gesture duration*" is often accompanied by an increase in "*PGO amplitude*", the latter can also be taken into account here.

C. The ***short stop closure hypothesis***: with "*glottal gesture duration*" and "*onset of oral occlusion to onset of glottal gesture*" remaining constant, "*PGO to release*" is shortened by shortening "*oral occlusion duration*"

Overall, his results favoured Hypothesis C: There was quite a close reciprocal relationship between oral occlusion duration and PGO-to-Release - the correlation between these two variables was about as strong (with opposite sign) as the correlation between PGO-to-Release and aspiration duration. Unfortunately for a clear picture, the duration of the glottal gesture was not completely constant over places of articulation; however, its variation was not related in any systematic way to PGO-to-Release. (This investigation will be returned to below with regard to sonorant clusters.)

Hutters' (1984) Danish data (leaving /**t**/ out of consideration in view of its affrication) showed that occlusion release comes earlier relative to peak glottal opening for /**k**/ than for /**p**/, but there were no differences in either occlusion duration or vowel plus occlusion duration for these stops; the interval from vowel onset to peak glottal opening did in fact turn out to be shorter for /**p**/ than for /**k**/, so there do appear to be some active laryngeal differences between the two stops.

A further direct test of this question is to be found in Cooper's already mentioned 1991 study, where /**p**/, /**t**/ and /**k**/ were compared for two speakers of American English. He found the expected reciprocal relationship between duration of oral occlusion and VOT, /**p**/ contrasting with /**t, k**/ (VOT was shorter in /**p**/), but neither his acoustic data nor the associated transillumination data allowed a strict interpretation in terms of an invariant laryngeal gesture over place of articulation. The duration of the devoicing gesture was longer for /**t**/ than for /**k**/. But it is not clear what the motivation for this difference could be since it was not, for example, related to duration of VOT. VOT *was* directly related to the timing of peak glottal opening relative to release, and this probably reflects an active process of interarticulator timing, rather than emerging passively from variation of occlusion duration. But it is still not clear *why* this form of organisation should occur.

The idea of an invariant glottal gesture for all stops thus does not appear completely justified by the data. Weismer (1980) even went so far as to suggest an invariant gesture for stops and fricatives - which as we have seen is also probably not justified. Nevertheless it is interesting at this juncture to pick up Weismer's conjectures as to why voiceless fricatives have a constriction duration that is clearly longer than the occlusion duration of voiceless plosives. Assuming that it is inappropriate for fricatives to be aspirated (at least for English) then it may be easier *"to 'fit' the supraglottal constriction to the time course of the devoicing gesture"* (Weismer, p. 436) than vice-versa. This concept may still have some merit (cf. the similar discussion of clusters below) even if the invariance of the devoicing gesture is not correct in a hard and fast sense (see also Shipp's, 1982, suggestion that the highly preprogrammed nature of the abductory-adductory cycle may make the larynx "*one of the basic metronomes of the speech production process*", p.111)

2.3 Devoicing organisation in consonant sequences with sonorants

2.3.1 Introduction

We will be arguing in this section that clusters consisting of a voiceless element or elements, followed by a sonorant such as /**l**/ or /**n**/ can be potentially very useful for highlighting the gaps in our knowledge about the principles underlying interarticulatory coordination. This is for the simple reason that the most basic hypothesis is probably that sonorants are quite simply irrelevant: if the devoicing gesture is organized with respect to the underlyingly voiceless segments then the timing of landmarks in the glottal gesture relative to those in the oral gesture(s) might be expected to stay the same regardless of whether a sonorant is present or not. If the situation turns out to be not so simple, then currently one would have to admit that we are essentially ignorant about how the coordination relations are to be formulated. As we will see below, even the very radical possibility that addition of an underlyingly voiced sonorant to an otherwise voiceless syllable onset can actually lead to an *increase* in the magnitude of the glottal gesture is by no means out of the question.

We will look first at the basic background for stop-sonorant and fricative-sonorant sequences separately, and then move on to consider both kinds of sequences in the light of some more recent intstrumental investigations.

2.3.2 Coarticulatory devoicing in stop-sonorant and fricative-sonorant sequences: Background (Docherty, 1992)

Stop-sonorant sequences

One of the most accessible sources of systematic data is Docherty's (1992) acoustic investigation, and this will accordingly form the basis for much of the discussion.

Two simple regularities can at once be stated for sequences of stop or fricative plus sonorant[8]:

1) VOT (i.e the period of voicelessness following release of the stop or fricative) is longer in these sequences than in simple CV sequences;
2) It is well documented that stops and fricatives generally have a shorter occlusion duration when they occur in clusters (e.g Klatt, 1973, 1975; Haggard, 1973; Hawkins, 1979), though in fact we will also be encountering some cases where this effect is quite weak.

Docherty notes (p. 146) that there have been virtually no attempts to explain the longer VOT's in stop-sonorant clusters. One exception discussed further by him is a speculative suggestion by Hoole (1987)[9] that the above two findings can be simply related in a manner entirely analogous

[8]We will restrict consideration here to word-initial clusters. See e.g Docherty (1992) and Dent (1984) for investigations of coarticulatory devoicing in such clusters across word-boundaries.

[9]We call the suggestion speculative since it was not based on any data for English, but resulted from the attempt to place some findings on devoicing in Icelandic in a coarticulatory perspective. Specifically, Icelandic shows essentially a mirror-image of the phenomenon being discussed here: for example, sequences of voiceless nasal or lateral plus stop. These can

to the attempt (discussed above) to explain place of articulation differences in VOT for simple plosives in terms of the superimposition of different occlusion durations on an invariant devoicing gesture. In other words, pairs such as English "keen" and "clean" may have the same glottal gesture, but a shortened occlusion duration of /**k**/ in "clean", resulting in a largely voiceless /**l**/. In terms of the schematic illustration given in Fig. 2.4 on p. 6 one can think of the devoiced sonorant replacing the phase labelled "aspiration", this phase being proportionally longer and the preceding phase labelled "oral occlusion" proportionally shorter in the consonant clusters under discussion here than in the simple aspirated plosives.

As with the place of articulation data above, Docherty's acoustic data did not, however, provide much support for this hypothesis:

Occlusion durations clearly reduced in /**t**/-sonorant clusters, and VOT increased. In clusters with /**k**/ and /**p**/, however, the reduction in occlusion duration was rather slight (especially for /**p**/) but VOT nonetheless increased reliably. Accordingly, overall in the stop-sonorant-vowel case the total duration of devoicing was *longer* than in the simple stop-vowel case; in other words there was a greater increase in VOT than could be accounted for by the reduction in stop occlusion duration alone. We find this result most intriguing, perhaps more so than Docherty himself seems to do, since it is difficult to think of a speech production model that could predict this finding. In rather overstated terms, it appears that the effect of adjoining a voiced consonant to a voiceless aspirated plosive is to *increase* the magnitude of the devoicing gesture, which is most definitely not how coarticulatory effects are generally considered to work. Before indulging in further speculation we must hasten to point out that there may well be one simple passive explanation for the unexpected fact that the stop-continuant cluster has a longer period of devoicing than the simple stop, namely that the aerodynamic conditions in the continuant are not conducive to initiation of phonation (due to the fact that the oral tract is still partially occluded; see Hanson and Stevens, 2002, especially p. 1175 ff). Thus the acoustically measured period of voicelessness may not be an accurate reflection of the duration of the laryngeal gesture itself. Further articulatory data could thus still save the invariant laryngeal gesture hypothesis, although at the time Docherty was writing he seemed to be of the opinion that the magnitude of the effects made this rather unlikely. In the meantime some relevant experimental data has emerged; this is outlined below. Given that we have information on the pattern of laryngeal activity we will as part of our own results also be in a better position to estimate how crucial aerodynamic effects may be in determining the precise duration of voicelessness in clusters terminating with a sonorant.

Even if it remains an open issue whether devoicing duration is genuinely longer in stop-sonorant clusters, it would seem at least very unlikely that devoicing duration is *shorter*. This is in itself a significant finding since given the shorter occlusion duration a shorter devoicing could well be expected under the plausible assumption that the component gestures of an aspirated plosive become modified in parallel. For example, working within the framework of the Task-Dynamics model, Saltzman & Munhall (1989) point to evidence from perturbation experiments that the

---

be regarded as (voiced) sonorants coarticulated with a following preaspirated stop. The temporal relationships were such that a very similar devoicing gesture occurred for simple preaspirated stops and for the continuant-stop sequences, while the occlusion duration of the stop shortened in the latter case. An expanded version of this paper is included as an appendix.

laryngeal gesture is modified when the bilabial closure for /**p**/ is interfered with experimentally. They cite this as evidence for a level of intergestural cohesion that undoubtedly must exist (cf. VOT). These workers further introduce a concept of gestural "dominance" (op. cit. p.349): in other words, different segments have a different degree of dominance over the timing of the glottal peak. This concept is used to explain the ways in which glottal gestures merge in voiceless clusters (see below). The problem in the present context is that in /**kl**/ clusters, for example, no other segment should be competing with /**k**/ for dominance of the larynx, yet it may be necessary to assume that the position of peak glottal opening relative to /**k**/ release is shifted from the non-cluster case. Kingston's (1990) concept of binding (of laryngeal to oral articulations) would seem to run into similar problems[10].

One way around this problem, which would certainly be in the spirit of the task-dynamics approach, is that in clusters the acoustic manifestation of occlusion duration in plosives or constriction duration in fricatives is no longer very directly related to the underlying gestural activation. For example, in /**sm**/ it is conceivable that the acoustic manifestation of /**s**/ is partly 'hidden' and thus shortened by an overlapping bilabial gesture (cf. Borden & Gay, 1979).

Is it possible to come up with an explanation as to why the devoicing gesture conceivably lengthens? In an analysis of voiceless clusters (to which we return below) Browman & Goldstein (1986) come to the conclusion that it can be stated as a regularity of English that a word (syllable) can only begin with one devoicing gesture (and we assume initially, because of the similar status of aspiration, that German is likely to pattern in a similar way).

The relevant laryngeal-oral coordination patterns are captured in two rules (op. cit. p.228):

(1) If a fricative gesture is present, coordinate the peak glottal opening with the midpoint of the fricative.

(2) Otherwise, coordinate the peak glottal opening with the release of the stop gesture.

This single-gesture model could be extended, certainly with a good deal of violence to the authors' original intentions, to suggest that in some sense the devoicing gesture is a property of the whole syllable onset even in the present cases, too, where the onset is purely voiceless. The devoicing gesture may then lengthen as the syllable onset becomes longer. These rules thus form a major point of orientation throughout this work. A common theme will be an assessment of the extent to which they may, in fact, require modification: Just as one of the motivations for closer examination of fricative-plosive and plosive-fricative clusters in the next section was the desire to test whether rules of the above kind really capture the principles of coordination underlying speakers behaviour, it is equally the case with sonorant clusters that they could induce changes in coordination patterns (versus single consonants) that are not captured by them (for example timing of PGO relative to release).

An alternative, more output oriented style of explanation for lengthening of the glottal gesture might be that it is perceptually important to have a substantial amount of devoicing on the second element in a cluster when the first element is underlyingly voiceless (e.g to separate "played", "blade", "prayed", “braid”). A further alternative is that given the aerodynamic conditions in the vocal tract, early adduction might not lead to reliable re-initiation of voicing anyway, so speakers

---

[10]It is also unclear whether current formulations of the dominance and the binding concepts can handle the possibly different timing of peak glottal opening relative to release for different places of articulation in aspirated plosives (cf. discussion above).

find it easier to use a somewhat longer gesture. (Note: according to Docherty, p.147, the VOT of English phonologically voiced stops is also slightly longer in stop-sonorant sequences, such as /**bl**/, than in the singleton case.)

Fricative-sonorant sequences

To set the scene, a basic superficial difference between plosive-sonorant and fricative-sonorant sequences should be noted: sonorant devoicing after fricatives can be expected to be less extensive than after plosives, since - regardless of what the details of the coordination patterns turn out to show - the glottal gesture is more likely to be approaching completion at the end of a fricative occlusion than at the burst of a plosive (this is assumed to be a natural consequence of the basic single consonant patterns shown at the outset of this chapter, and in the schematic figure 2.4). In addition to Docherty (especially pp. 147-150), older acoustic analysis can be found e.g in Klatt (1973, 1975)

Docherty's results for fricative-sonorant sequences are essentially comparable to those for stop-sonorant sequences. For /**s**/ plus nasal sequences the constriction duration for /**s**/ was reduced in comparison with single /**s**/ by about 20ms, but total devoicing duration increased (by about 15-20ms), so again it seems that the amount of nasal devoicing does not simply result from the reduction in /**s**/-duration. The other fricative-sonorant combinations mostly indicated the same pattern (actually, for his measurements based on the time-wave he did not try to measure shortening of the fricative constriction - but assumed on the basis of the literature that this was the case[11]. The clear point was that total devoicing duration increased). One interesting exception was that /**f**/-sonorant clusters did not show a significant increase in total devoicing duration, leading Docherty to speculate that this may be related to the potential for coproduction of the oral components of the cluster (which is presumably higher in the labiodental fricative case; in fact the labial stop in Docherty's data also shows a relatively weak increase in devoicing duration in clusters). Thus, in the /**sl**/ case, with little coproduction possible he suggests that "one might hypothesize the existence of a temporal constraint delaying voicing onset until the lateral gesture is complete" (op. cit. p. 154). This seems to be close to the suggestion made above that the devoicing gesture may be influenced by the length of the whole syllable onset - independently to some extent of the intrinsic voicing characteristics of the segments making up that onset. If rules of this kind should prove necessary they would have interesting implications for the patterns of inter-gestural coordination that a production model would have to account for.

2.3.3 Clusters with sonorants: More recent investigations

A few more recent investigations have looked directly at laryngeal behaviour in clusters with sonorants.

---

[11]For our own experiments we used EPG in parallel with the transillumination recording as a source of supplementary information on oral occlusion durations in cases where acoustic analysis was difficult.

Jessen (1998, 1999)

The study of Jessen (1999), which we have already discussed above with regard to place of articulation effects in single stops, also considers stop-sonorant combinations. Thus he considers explicitly the question of whether duration of voicelessness in the sonorant is directly related to shortening of the stop occlusion. His material covered /**p, t, k**/ for the plosives combined (to the extent possible in German) with /**l, r, n**/.

Regarding combinations with /**l**/ (/**pl**/, /**kl**/) it turned out that occlusion duration for the bilabial shortened only neglibly in the cluster (a tendency already observed on p. 10 above). Nonetheless, VOT clearly increased. However, this could not be unambiguously attributed to increased duration of the glottal gesture. This did increase in one session, but actually decreased (non-significantly) in the other one. (There did not appear to be significant differences in glottal opening amplitude.) For /**k**/ vs. /**kl**/ the expected shortening of k-occlusion in the cluster was found. Glottal gesture duration did not change, so an increase in VOT would be expected. This in fact occurred, but in one session only rather weakly. Overall, unlike for /**p**/, the results would be consistent with Jessen's hypothesis C, the "short stop closure hypothesis" (see above p. 8).

An interesting combination in his corpus was /**kn**/, since this cannot be investigated in English. This was one of the sound sequences we decided not to investigate because of the possible influence of velum movement on endoscope position. This consonant sequence showed clear reduction in /**k**/-occlusion in the cluster, and a clear increase in VOT. Duration of the glottal gesture did not change in one session, and was shorter in the other session. Although at first sight less likely in nasal than lateral clusters, it would seem that there could be an aerodynamic component partly determining the duration of voicelessness: In session 2 /**k**/ and /**kn**/ had the same duration of voicelessness, but glottal gesture duration was shorter for /**kn**/. In session 1, /**kn**/ had longer voicelessness but glottal duration was the same.

This partial dissociation of devoicing gesture duration from voicelessness duration may explain the rather variable patterns. In view of the aerodynamic influence on voice timing the speaker may not need to control all aspects of the glottal gesture so precisely[12].

A rather striking finding in this investigation involved stop-/**r**/ clusters. All the combinations (/**pr, tr, kr**/) showed a substantial increase in duration of the glottal gesture (and also in amplitude of the gesture - as far as this can be estimated reliably from the transillumination signal: see methodological discussion in the next two chapters). The changes in stop closure duration were not very consistent (only clearly shorter for /**tr**/) and the VOT duration was not longer than for stop-/**l**/ (no comparison possible for /**tr**/ of course). Also, total voiceless duration was longer than for the single stop, but not consistently longer than for stop-/**l**/. What could explain the extensive glottal gesture then? Possibly it is an active mechanism to ensure realization as a uvular fricative (whereas /**l**/ in stop-/**l**/ hypothetically shows only passive coarticulatory devoicing). A uvular fricative could indeed require a large glottal opening: Hanson & Stevens (2002; in turn quoting Stevens, 1998) estimate that for maximum frication noise generation the supraglottal constriction area should be about half the glottal area. Given that a typical uvular

[12]In the present case there may have been some changes in the shape of the glottal gesture since PGO-to-Release does not always change as much as would be expected from change in occlusion duration when glottal gesture duration stays the same.

constriction most likely has a larger cross-sectional area than e.g an alveolar fricative constriction, then a correspondingly larger glottal area would not be unexpected[13].

Jessen does not discuss fricative-sonorant sequences explicitly (although they formed part of his corpus); however, in Jessen (1998) in a discussion of the principles determining the number of observable glottal gestures in consonant sequences (one of the topics of the next section) he shows a few examples of /**ʃpr**/ vs. /**ʃpl**/ sequences. The /**ʃpr**/ sequences were remarkable for being apparently the only ones in the whole corpus to consistently show double opening peaks. Admittedly, the speech rate appears to have been fairly slow, but this does confirm the link between fricative realization of /**r**/ and large glottal opening.

There is reason for caution with the results for /**r**/, as Jessen himself acknowledges (e.g. 1999, p.996) - the reason in fact that we did not include them in our corpus. Uvular consonants are certainly not optimal for transillumination since movement of the tongue root or dorsum could disturb the position of the endoscope, or -depending on endoscope position - partly obscure the passage of light through the glottis (in the meantime we have successfully recorded uvular stops and fricatives in Berber and Moroccan Arabic, so this might have been too cautious). If the latter were the case the most likely artefact would be a sudden reduction in light during articulation of the /**r**/. Inspection of the traces in Jessen (1998; and other plots he kindly made available) do not give any obvious indication of this. He also considers the possibility that the magnitude of the glottal opening could be a passive effect of a sudden increase in intraoral pressure (particularly sudden because of the retracted constriction location). It is true that Stevens, in particular, has emphasized that increased intraoral pressure leads to increased abduction of the vocal folds (see e.g Hanson & Stevens, 2002, p. 1164, the effect being modelled as part of the HLSyn synthesis system). Personally, I think this is unlikely. I am not sure that there would be such a marked drop and rise in air-pressure at the transition from plosive to /**r**/.

Tsuchida et al., 2000

These authors examined clusters with /**l**/ (as well as the non-sonorant clusters to be discussed in the next section) for one speaker of American English. Apparently /**r**/-clusters were also recorded but unfortunately, in the light of the preceeding discussion, were not considered further as they "*.... showed some complications, due to durational properties that are beyond the scope of the present discussion*" , p. 170.

For plosives, they found the expected slight shortening of the oral occlusion (for cluster vs. singleton), but the glottal gesture was actually somewhat longer (especially in the adductory (closing) phase) in the latter case.

For fricatives the pattern was slightly different. Both the oral occlusion for fricatives reduced, as well as the length of the glottal gesture. This reduction in the length of the glottal gestures was mainly attributed to the adduction phases (the abduction phase remained remarkably constant over all experimental conditions). Although different places of articulation were recorded it is not reported whether this was a relevant effect (in fact, no statistics are presented). Even if the results are not completely conclusive, this study very usefully indicates that for these clusters we

[13]A recent transillumination recording we have made of Moroccan Arabic confirms this supposition: the corpus included fricatives at several places of articulation, including uvular. The latter gave consistently the largest glottal opening.

simply do not know whether to expect a glottal gesture that is longer, shorter, or the same - and whether this depends on further factors such as manner of articulation. This thus encapsulates the basic motivation for our own recordings that for many speech sound sequences the principles of gestural coordination are essentially unknown,

The observed differences between plosives and fricatives in this paper are interesting, but - given that the robustness of the results is difficult to assess - they are used to argue for, I would say, a prematurely radical suggestion for the phonological representation. English is suggested to have two privative laryngeal features: <voice> and <spread glottis>. Only voiceless stops, it is claimed, are specified for <spread glottis>, while voiceless fricatives are specified neither for <voice> nor <spread glottis>. Glottal abduction for fricatives is assumed to be implemented in the phonetic component due to the aerodynamic requirements of fricatives. They assume that the longer glottal gesture for /**pl**/ results from two segments being linked to <spread glottis> whereas in /**fl**/ there is no <spread glottis> in the phonological representation.

Larson, 2003

Finally, Larson (2003) addressed specifically the question of whether increased VOTs in plosive-sonorant vs. plosive-vowel are best put down to aerodynamic conditions or whether differences in the glottal gestures could be involved. The study investigated combinations of /**t**/ with /**w**/ and /**r**/ for two speakers of American English (of at least equal concern - from the same point of view - was the source of differences in VOT associated with differences in height of the following vowel). For both speakers, /**r**/ belonged to the sounds with longest VOT and shortest occlusion duration. Over the whole material, however, there was no relationship between VOT and the interval from PGO-to-Release. This could be interpreted as indicating that other factors such as aerodynamics (and not laryngeal-oral timing) determine the details of VOT. However, not enough details are given in this paper to be able to confidently interpret the results in this way: the aerodynamic supposition might lead one to expect particularly long total devoicing duration in combinations with sonorants. It is not clear from the figures that this is the case. But no information is given on the total length of the laryngeal gesture, thus it is not possible to ascertain - in line with the aerodynamic hypothesis - whether total devoicing duration is particularly long relative to gesture duration for sonorants (and perhaps also high vowel contexts). As already mentioned, the question of the status of aerodynamic effects in determining the precise duration of devoicing is a point that will be picked up in the discussion of our own results.

2.4 Devoicing patterns in voiceless clusters

2.4.1 Introducton

Clusters of voiceless consonants provide one of the most suitable fields for examining processes of coarticulation or coproduction at the laryngeal level by studying how the simple, ballistic-looking pattern of ab- and adduction found in single consonants is modified when sequences of voiceless consonants occur. The most convenient source of basic information on this topic is a series of articles published in the 1980s by Löfqvist and colleagues, in which sequences of voiceless sounds in American English, Swedish, Icelandic, Dutch and Japanese were studied (Löfqvist & Yoshioka, 1980a,b; Yoshioka, Löfqvist & Hirose, 1980, 1981; Yoshioka, Löfqvist & Collier, 1982). These papers have the advantage of sharing a common methodology, namely transillumination/fiberoptics together with EMG (the latter not for Icelandic). The corpora are also quite comparable, consisting for the four Germanic languages mainly of combinations of /**s**/ and a stop to left and right of a word boundary - giving sequences of up to 5 voiceless consonants. For Japanese, which does not have clusters of this kind, long voiceless sequences were obtained by exploiting the phenomenon of vowel devoicing, preceeded and followed by voiceless stop or fricative.

One emphasis in these papers is in arriving at a qualitative understanding of the time course of laryngeal ab- and adduction as a function of the structure of the consonant sequence, i.e in predicting where 1, 2 or more peaks in the transillumination signal will occur (in addition these articles also provided the consistent result of larger, faster abduction in fricatives vs. stops, as dicussed above).

In a later paper (Munhall & Löfqvist, 1992) the question of the relationship between the number of peaks in the transillumination signal and the number of underlying laryngeal gestures is examined - specifically whether a single peak in the surface behaviour can plausibly be regarded (in appropriate contexts) as a blending of two (or more) underlying gestures. In Saltzman & Munhall (1989) some of the additional assumptions likely to be required to predict the details of the blending process are discussed.

Each of these developments will be discussed in turn.

2.4.2 Basic kinematic patterns

With regard, then, to the observable kinematics of laryngeal behaviour in voiceless consonant sequences the results have been summarized by Löfqvist (1990, p.296) that

> *"sounds requiring a high rate of airflow, such as fricatives and aspirated stops, are produced with a separate gesture".*

Perhaps the clearest example of this behaviour is to be found in fricative-plosive clusters. For the three Germanic languages English, Swedish and Icelandic, when these clusters occur word-initially or finally (e.g /#**sp**/ or /**sp**#/) the plosive is unaspirated, and only one abduction peak occurs. When the cluster spans a word boundary the stop is aspirated in all languages, and two peaks are found. In apparent contrast, more recent work by Jessen (1998) presents some evidence that in German initial /#**sp**/ clusters two glottal opening peaks can occur. He notes, however, that in such cases the /**p**/ actually has appreciable aspiration. He sees this then as actually further evidence for the power of Löfqvist's generalization, and as speaking against the generalization of Browman & Goldstein, to which we will be continually referring (cf. p. 11 above and p. 20 below) that words in English and other Germanic languages begin with at most one glottal gesture[14]. One of the main aims of the experiments below is in fact to provide further material to test the power of this generalizaton.

As the number of voiceless segments in the cluster increases, then more peaks can occur, e.g /**sks**#**k**/ (or equivalent thereof) showed three peaks in all three languages. On the other hand, there are a number of cases when fewer peaks are observed than the above summary might lead one to expect. For example, the long voiceless sequence /**ks**#**sp**/ showed only one peak in all three languages. This may well be related to the homorganicity of the fricatives: simple /**s**#**s**/ sequences also showed only one peak in English, Icelandic and Dutch (the corresponding Swedish data was not shown). /**k**#**k**/ in English showed only one peak, whereas the non-homorganic sequence /**k**#**p**/ in Swedish had two[15]. Compared with the Germanic languages Japanese appears to show in general a weaker tendency to multiple peaks. A sequence such as stop-devoiced vowel-geminate stop shows only one; even the very long voiceless sequence fricative-devoiced vowel-geminate fricative showed only comparatively weak evidence of more than one peak. Possibly this situation is related to the fact that aspiration is not a prominent feature of Japanese stops, so the air-flow requirements in sequences involving stops may not be particularly stringent.

In recent work on Tashlhiyt Berber (Ridouane et al., 2006), we obtained findings broadly compatible with those of the Löfqvist group, with one exception. Berber is an interesting language in this connection, because very long sequences of voiceless sounds can occur (even voiceless words). Compatible results were apparent., for example, in sequences of 4 voiceless consonants. The sequences **sk**#**sk**, **ks**#**ks** and **k**#**sks** all showed two opening peaks whereas **k**#**kss** and **kk**#**ss** showed only one. Thus where fricatives are separated by an intervening unaspirated plosive then separate glottal peaks occur, whereas a geminate fricative preceded by unaspirated (geminate) plosive shows only one. The departure from previous results involved in particular initial #**sk** sequences. The plosive showed consistent aspiration but the cluster showed only one glottal opening (this is thus the contrary phenomenon to that observed by Jessen for /**sp**/ clusters just discussed above). It was speculated that this was related to the widespread occurrence in Berber of aspirated geminates (also perhaps somewhat unusual), which also have a single (large) glottal opening. One relevant figure from this paper will be shown below as a postscript to the results of our own experiments for fricative-plosive clusters (see p. 97).

Some typical figures from these papers of Löfqvist and co-authors are shown below, illustrating 2 and 4-segment clusters of /**s**/ and /**k**/ with various word boundaries for American English (Figs. 1 and 9 from Yoshioka et al., 1981).

---

[14]cf. also Iverson & Salmon's (1995) generalization that at most one glottal gesture can occur morpheme-internally in English (further discussion in Jessen, 1998).

[15]This may also be related to a greater tendency of Swedish to aspirate and a lesser tendency to glottalize word-final plosives than English.

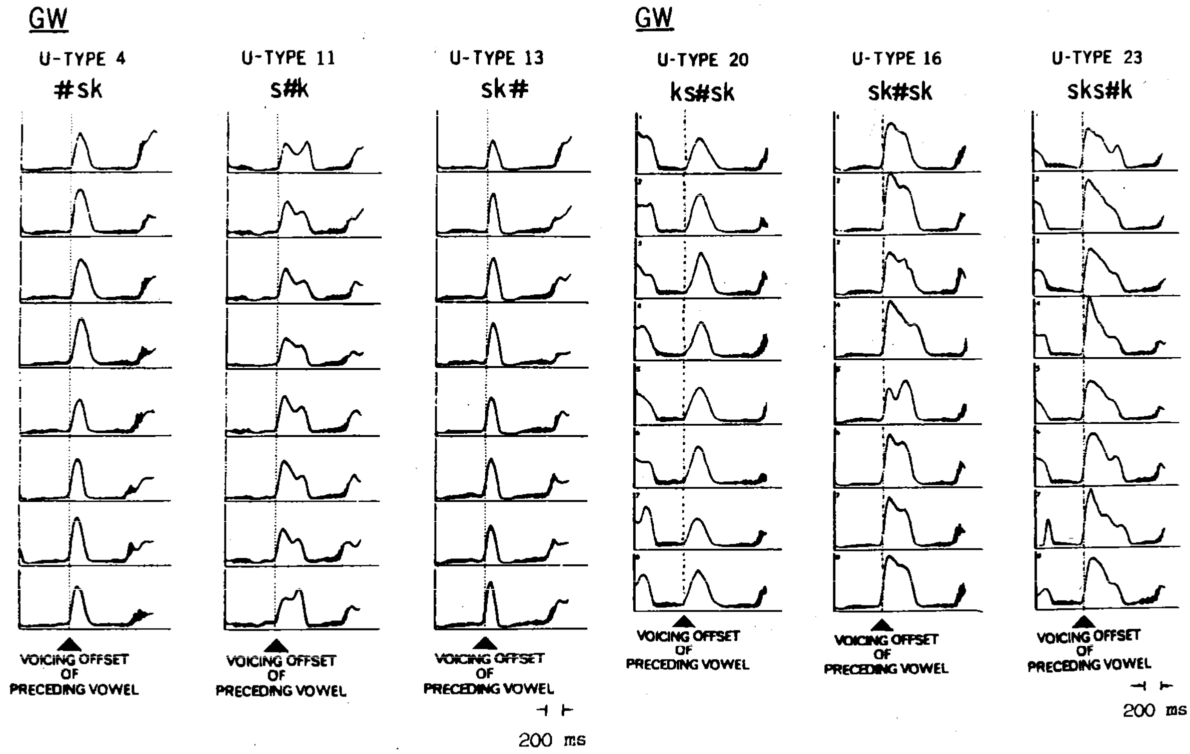


Fig. 2.5: Fig. 1 (left three columns) and Fig. 9 (right three columns) from Yoshioka et al. (1981). Original legends: "*Glottographic patterns for eight productions of three utterance types containing the /sk/ sequence in various contexts*" (left columns), and "*Glottographic patterns for eight productions of three utterance types containing clusters of four voiceless phones*" (right columns).

2.4.3 Surface observations vs. underlying organization

Following these qualitative remarks, we immediately reach the stage, of course, at which it becomes important to distinguish between the observable kinematic behaviour and the putative underlying gestural input. Clearly a homorganic cluster could be realized with a particularly large degree of overlap of discrete underlying oral and laryngeal gestures. However, here we reach the limits of the interpretability of the papers discussed in the immediately preceding section, since no figures are given allowing, for example, fricative constriction duration to be compared in the singleton vs. the homorganic cluster case. Nonetheless, the authors did note in the Icelandic paper that where different repetitions of a given cluster were spoken with widely varying durations then the number of observable peaks might be less at the shorter duration; e.g for /**t**#**k**/ two peaks clearly corresponding to each stop at the long duration, only one peak at short durations. It is then tempting to assume that underlyingly two peaks are present at the shorter duration, too; they have simply become merged together. This is illustrated schematically in the next figure.

Munhall & Löfqvist (1992) then examined the plausibility of this assumption more systematically by running an experiment in which only one cluster was examined (/**s**#**t**/ from "kiss Ted") but where a wide range of speech rates was elicited (and stress was also varied) in order to obtain something approaching a continuum of cluster durations.

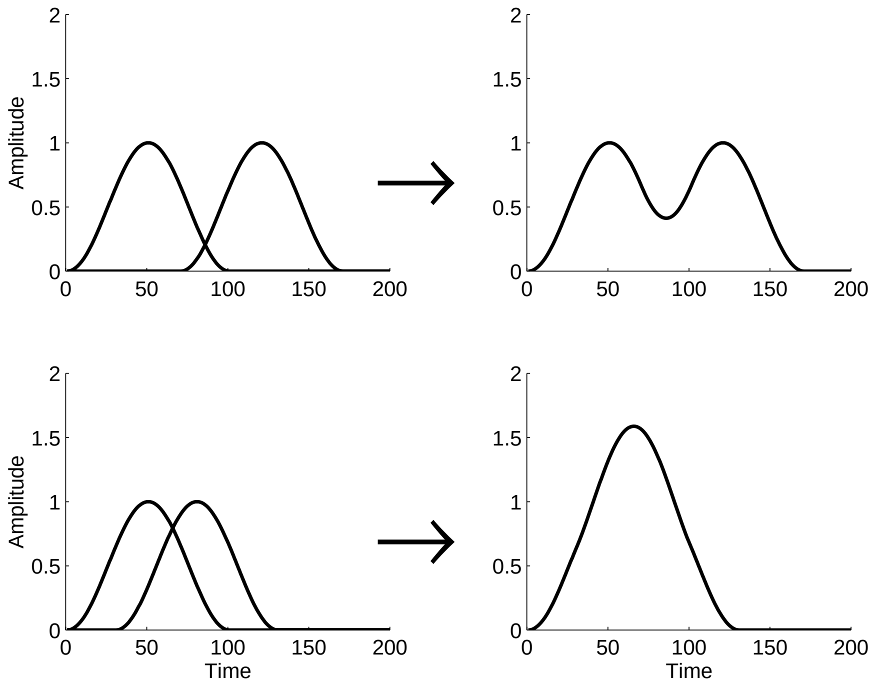


Fig. 2.6: Simplified illustration of relationship between hypothesized underlying gestural input (left panels) and observable behaviour (right panels) at different degrees of gestural overlap (top vs. bottom). Amplitude of the basic laryngeal gesture has been arbitrarily set to 1, and duration to 100. See Munhall & Löfqvist (1992, Fig. 5) for a wider range of more realistic simulations.

The result showed by and large a gradual merging from two separated gestures at the slowest rates via a more complexly shaped movement at intermediate rates to a simple single-peaked movement at the fastest rates. Single-peaked patterns for this kind of cluster may thus be seen as simply one end of a continuum, rather than a completely different mode of organisation compared with the multi-peaked tokens. For the cross-word clusters examined here, and for example the /**s**#**s**/ homorganic clusters mentioned above, the approach is undoubtedly rather persuasive. Whether word-initial clusters (e.g /#**sp**/) can by the same line of reasoning (cf. Saltzman & Munhall, 1989; Löfqvist, 1990; also Pétursson, 1977) be regarded as underlyingly two gestures is more contentious (see below); leaving aside the exceptions presented by Jessen (1998) to be discussed again with our results, they never, as far as we know, show two gestures on the surface. Munhall & Löfqvist are also quick to admit that alternative explanations are not completely ruled out:

> *"One problem in the area of coarticulation and in the present study is that it is difficult, in practice, to distinguish between alternative explanations. At the fastest speaking rates in the present data, a single movement is observed. By examining the kinematics of these movements in isolation it is impossible to determine the nature of the underlying control signal. For two reasons, we have favored the overlap account for the present data. While any individual movement could be accounted for by many approaches, it is more parsimonious to attribute all the data to a single pattern of serial ordering. It would appear, particularly*

from the intermediate rate observations, that two separate gestures are blended. This style of coordination can produce the full range of observed data and thus seems a likely candidate even for the fastest speaking rates. A second factor that supports this approach is evidence from other motor activities......" (p.122).

One of the main motivations for the present work was to test whether this remains the most parsimonious explanation, by examining a wide range of consonantal structures.

The Munhall & Löfqvist study remains a significant experiment for speech production studies as a whole (one might even say that it was long overdue, following the pioneering studies just discussed under the heading of basic kinematic patterns): the great simplicity of the devoicing gesture (in spatial terms) in comparison, for example, to tongue movements makes it probably the speech sub-system where the existence of blending processes can be most convincingly demonstrated.

Some suggestions for principles underlying the details of the blending process are to be found in Saltzman & Munhall (1989). As mentioned above, they make use of the concept of dominance:

"The dominance for a voiceless consonant's oral constriction over its glottal timing appears to be influenced by (at least) two factors. The first is the manner class of the segment: Frication intervals (at least for /s/) dominate glottal behavior more strongly than stop closure intervals.....The second factor is the presence of a word-initial boundary: Word-initial consonants dominate glottal behavior more strongly than the same nonword-initial consonants." (p. 369)

Motivation for the idea of fricative dominance is developed especially in Goldstein (1990). The observation that peak glottal opening in fricative-stop clusters is firmly located within the fricative determines, in particular, the order of the two rules given in the earlier paper of Browman & Goldstein (1986) and already quoted on p. 11 above[16].

Saltzman & Munhall illustrate the process first with some unpublished data on word-final clusters. English /**s**#/, /**ks**#/, /**sk**#/ all have only one glottal peak, which for single /**s**/ is smaller than in the other two cases (observable in Yoshioka et al., 1981), suggesting that in the cluster case blending of two gestures is involved. The specific location of the peak glottal opening in the clusters could be interpreted as indicating that /**s**/ is the 'dominant' partner, but with the location of the peak being perturbed slightly away from midfrication by the adjacent stop (midfrication being the normal location of peak glottal opening in isolated fricatives). It will be recalled that one motivation for this kind of approach is that a more parsimonious analysis results if the single glottal peak can be assumed to be the result of two underlying gestures. The only problem in the above example is that word-final voiceless plosives in English are often glottalized (see e.g Yoshioka et al., 1981) so the blending approach is here not necessarily more parsimonious since these plosives are clearly not glottalized, and thus some additional rule is in any case required to state when the laryngeal gesture for a word-final voiceless plosive can be reorganized from devoicing (abduction) to glottalization (adduction) (on the problem of glottalization see Browman & Goldstein, 1992, and Kingston & Cohen's (1992) comment).

---

[16]We repeat them here for convenience: (1) "*If a fricative gesture is present, coordinate the peak glottal opening with the midpoint of the fricative*." (2) "*Otherwise, coordinate the peak glottal opening with the release of the stop gesture*."

In a further example Saltzman & Munhall compare such word-final clusters with corresponding word-initial clusters. We have already noted that in e.g /#**st**/ only a single peak occurs. We have also mentioned that for Munhall & Löfqvist the "kiss Ted" results make it reasonable to assume that these single-peaked word-initial clusters consist underlyingly of two blended gestures. On the other hand, we have further noted that for Browman & Goldstein (1986) it is a significant generalization of the (articulatory) phonology of English that a word can begin with no more than one glottal gesture[17]. There is thus an interesting divergence of views even among quite closely related approaches (cf. Saltzman & Munhall, op. cit. p. 365).

Saltzman & Munhall state that for these word-initial clusters in English peak glottal opening occurs at mid-frication in both single /**s**/ and in /**st**/ and thus that, in contrast to the word-final case, location of peak glottal opening has not been perturbed by the adjacent plosive. In terms of the dominance concept, this would be due to the intrinsically high dominance of /**s**/, reinforced by its word-initial position. In fact, however, as far as we can tell, the relevant literature does not state that peak glottal opening in /**st**/ is at mid-frication, only that it is during the frication phase. Pétursson (1977; not cited by Saltzman & Munhall) in fact notes that in Icelandic it occurs in the first half of the frication phase. Goldstein (1990), on the other hand, notes that it may be delayed somewhat, i.e. later than mid-frication. This reflects a paucity in the literature of precise information on constriction and occlusion duration in those clusters for which we have information on the laryngeal kinematics. Of the small amount of more recent literature with relevant data, the paper by Tsuchida et al. (2000), already discussed in connection with the sonorant clusters, also does not indicate that a simple scheme like mid-frication will be adequate.

There is also some ambivalence in the literature as to what constitutes a clearly more extensive devoicing gesture. The blending hypothesis would lead us to expect a larger gesture on /**st**/ than on /**s**/. Almost the only accessible source of numeric data showing this to be the case is for one American speaker in McGarr & Löfqvist (1988). For Swedish, Löfqvist & Yoshioka (1980b) say /#**sp**/ is similar to /**s**/, as do Yoshioka, Löfqvist & Collier (1982) for Dutch. Goldstein (1990, p.447), following on from the articulatory phonology analysis, also seems to view the gestures as about the same size.

Finally, in order to link up with the discussion of mixed-voicing clusters above, it should be noted that even if the laryngeal gesture for /**st**/ does indeed turn out to be reliably larger than for /**s**/ then this may not be sufficient grounds for suspecting the presence of two underlying gestures if, in turn, it emerges that such sequences as /**pl**/ and /**sl**/ also have a larger devoicing gesture than the singleton case. More generally, if, for example, both /**st**/ and /**sl**/ can result in shifts in laryngeal-oral coordination relative to the singleton case then one will have to be cautious about interpreting the pattern for /**st**/ as blending that has resulted from competition between two underlying gestures.

Let us return briefly to the second factor suggested by Saltzman & Munhall to determine dominance strength, namely position of the consonant in the word. This is a very reasonable

---

[17]Pétursson (1977) points out that this may be a generalization that is specific to the Germanic languages. Some Indian languages contrast unaspirated stop with aspirated stop following /**s**/ (within the same word). In the latter case it is tempting to assume that two peaks in the glottal abduction will be observed. It is true that we have already pointed to the case of Berber (Ridouane et al., 2006), where somewhat aspirated /**sp**/ occurs but with only one peak of glottal abduction, however Berber does not contrast aspirated and unaspirated /**sp**/.

principle since it is quite clear that the devoicing gesture for a word-final fricative is smaller than for a word-initial one (see Yoshioka et al., 1981), while for stops Cooper (1991) has also shown clear effects of stress and position in the word (see also discussion of reduction of devoicing gestural magnitude in Browman & Goldstein, 1992). However, some aspects of cross-word boundary clusters do not seem to quite accord with expectations. American English, Swedish and Icelandic all have data for sequences with a structure like /**st**#**st**/ (Am. English has /**sk**#**sk**/), i.e the same kind of cluster before and after the word boundary. In all these cases two peaks are observable, but the first one (presumably corresponding to the word-*final* position) generally appears to be higher. Similarly for American English, and Dutch in /**s**#**s**/ only one peak occurs, but it is skewed to the left (not, however in Icelandic, where /**s**#**s**/ also occurs), suggesting more vigorous devoicing early in the sequence. On the other hand American English /**k**#**k**/ peaks late in the sequence (skewed to the right). These examples suggest that the amplitude of the devoicing gesture may also be modulated on-line depending on the aerodynamic conditions in the vocal tract: as already mentioned above, the critical laryngeal phase of a fricative is the onset, since voicing must be terminated and air-pressure built up to drive the frication source. However, once these demands have been met the requirements for the following devoicing gesture (i.e for the second /**sk**/ in /**sk**#**sk**/, or the second /**s**/ in /**s**#**s**/) are probably not so stringent, and the amplitude may then be smaller. For plosives the reverse applies: the more stringent demands are at offset rather than onset. In short, the procedures by which dominance is determined in any particular case may have to make more explicit reference to the air-flow demands of the sequence of sounds being produced.

Rounding of this section it should be said the great advantage of the rather specific proposals for blending and dominance put forward in Saltzman & Munhall resides in the fact that they provide a very efficient framework for pinpointing the current state of our knowledge.

2.5 Conclusion: Issues to resolve

We will now summarize this review of previous investigations with respect to the issues that it appears worthwhile to try and resolve in our own set of experiments.

Turning first to the mixed-voicing clusters, it is striking to observe that there are (probably) three main logical possibilities for the form that the coordination relations can take, and that at the current state of our knowledge it is simply not possible to choose between them. Expressed in terms of straightforward acoustic measurements, it is surprisingly unclear what combination of articulatory adjustments leads to the well-documented longer voice onset time in voiceless+sonorant clusters (VOT used here in the sense of the duration of the interval from release of the last underlyingly voiceless segment, whether plosive or fricative, to onset of voicing). Let us run through the possibilities again in turn:

(1) The initial voiceless segment(s) reduce(s) in duration in the /l/-cluster, and the duration of the glottal gesture reduces in parallel.

This scheme sounds very reasonable, since it is tempting to assume that laryngeal-oral coordination relations are indeed expressed along the lines sketched out in the two rules of Browman & Goldstein (peak glottal opening roughly at midpoint of a voiceless fricative, and at release of an aspirated plosive) and that these relations are maintained when the duration of the underlyingly voiceless segments changes. Curiously, this is the scheme for which probably least evidence is available in the literature (note that it is not necessarily incompatible with increased VOT in the sonorant clusters, if aerodynamic conditions in the sonorant are disadvantageous for voicing).

(2) The laryngeal gesture remains the same, but timing of oral articulations changes.

This possibility is a logical extension of the conception that differences of VOT related to place of articulation in plosives emerge from different oral occlusion durations overlaid on a constant glottal gesture. Thus, this scheme would account immediately for increased VOT in sonorant clusters (i.e without ‘help’ from the aerodynamic conditions). Sporadic evidence has been found for this pattern (and we argued for it indirectly on the basis of a small amount of Icelandic data relating preaspiration and voiceless sonorants), but it is certainly not yet clear whether it should count as one of the major candidates.

Although the first two scenarios give a different viewpoint on how complexer sound sequences are assembled from simpler elements, both would be consistent with a view of the devoicing of sonorants as a simple coarticulatory phenomenon, not requiring any particular account in phonological representations. This is part of one of Browman & Goldstein’s original examples motivating a gestural representation (1986, p.228). As long as the representation of a plosive (in English, German ...) includes a representation of laryngeal-oral timing relations (rather than some essentially atemporal feature like [+spread glottis], then no rule is required to ‘explain’ the voicelessness of the sonorant, and also no explanation is required for the different amounts of devoicing following plosives and fricatives. In their original formulation, Browman & Goldstein appear to have been thinking more in terms of scenario 1 above. The big difference between scenario 2 and 1 is that in scenario 2 we move much closer to a conceptualization of the glottal gesture as a property of the whole syllable onset (which will be picked up again below), this being in fact consistent with Browman & Goldstein's basic generalization of one glottal gesture in syllable onsets (see also Kehrein & Golston, 2004). Even if scenario 2 leads to a perturbation of the timing relations away from the 'ideal' values expressed in the two rules, the categorical timing difference between a fricative and an aspirated plosive should remain present.

(3) Glottal gestures are amplified in voiceless+sonorant clusters

Common to the first two scenarios is the fact that there are no active changes in the glottal gesture that can be attributed to the presence of the sonorant (and hence the characterization of sonorant devoicing as a passive coarticulatory process). This is where the third scenario is radically different.

In a simple segment-by-segment view of speech production one could even say that this possibility would be sensational if it turned out to be robustly present, since it is not easy to understand how addition of an underlyingly voiced segment can result in longer voicelessness. More soberly, one could see this scenario as pushing even more strongly than scenario 2 towards a conceptualization of the glottal gesture as a property of the whole syllable onset. If the onset lengthens then the glottal gesture may also lengthen. The radical scenario does seem to be a realistic possibility, but on the other hand it also does not emerge everywhere where it conceivably could. The original hint in this direction came from those acoustic studies that showed longer total duration of voicelessness (i.e not just VOT alone, but voiceless occlusion duration + VOT) in occlusive plus sonorant clusters, with the suspicion that this could not be accounted for purely by aerodynamic effects. A particularly striking case was the realization of German /**r**/ as a voiceless uvular fricative in Jessen's study. Tsuchida et al. also found some lengthening of the glottal gesture in plosive-lateral clusters and modelled this by associating both plosive and lateral to the [spread glottis] specification. We suspect that such an approach may be unduly categorical. Based on the rather scanty information from previous investigations there does seem to be a real possibility that there is not a single predominant scenario. Obviously speakers must have a representation of the laryngeal-oral coordination relations (otherwise they would be simply unable to speak). The challenge to be faced is to find a formulation of the coordination relations that is consistent with both constancy and variability in the articulatory patterns.

Turning to sequences of purely voiceless consonants, the basic motivation for looking at these more closely comes from the observation that even investigators working within a broadly similar tradition can arrive at a strikingly different perspective. Regarding the case of fricative-plosive clusters, Saltzman & Munhall (1989, p. 364-365) characterize the analysis based on two underlying gestures (directly followed up in the experiment of Munhall & Löfqvist, 1992) as linguistically conservative and empirically radical, contrasting this explicitly with Browman & Goldstein's model (one gesture per onset as a regularity of English) as linguistically radical and empirically conservative. Linguistically conservative in this context means that they "*assume that gestures cohere in bundles corresponding, roughly, to traditional segmental descriptions, and that these segmental units maintain their integrity in fluent speech*" (p. 365), while empirically radical refers to the apparent discrepancy between the assumption of two underlying gestures, while almost invariably only one is visible on the surface.[18]

The perspective we will try and find evidence for below in fact deviates from both these approaches. Briefly, the argument will be that the observable movement patterns are most simply understood as fulfilling the aerodynamic constraints of the syllable onset as a whole. This may sound almost like a truism. Putting it more specifically, the suggestion on the one hand is that at least for the frequently occurring sound sequences that make up syllable onsets speakers are well able to plan a complete movement pattern that does not need to be resolved into multiple underlying gestures that are blended according to rather abstract principles of dominance. On the other side of the coin, it will have been noticed that the Browman & Goldstein approach formulates the coordination relations in terms of a single glottal landmark (peak glottal opening) relative to a single oral landmark. As soon as we move from single segments to more complex sequences this kind of formulation may lose much of its attractiveness. While Goldstein (1990) envisages the possibility of peak glottal opening being perturbed away from mid-frication in fricative-plosive sequences (but with the fricative still 'winning' the competition with the plosive) it appears necessary to ask whether this perturbation can be so large that peak glottal opening is no longer the appropriate landmark for formulating the coordination relation. This may, for example, be the case in plosive-fricative sequences for which currently very little data is available (but see Yohsioka et al., 1982, and comments in Saltzman & Munhall, 1989, p. 370).

Since we would like to question the appropriateness of formulating the coordination relations in terms of a mapping between single points in the laryngeal and oral gestures we leave aside here details of the 1990 discussion between Goldstein and Kingston regarding such issues as whether the glottal gesture binds (in Kingston's terms) more tightly to oral onsets or offsets, and whether plosives bind the glottal gesture more tightly than fricatives. The basic guideline to use while reading the results in the next chapter is that each onset, whether single segment or complex sequence, defines its own specific set of aerodynamic constraints, which speakers then devise appropriate but at times perhaps quite variable ways of fulfilling.

The final point to make in recapitulating the work presented above also links the analysis of voiceless+sonorant clusters and purely voiceless clusters: If amplification of the glottal gesture (in duration or magnitude) can occur for both kinds of cluster (and also if shifts in timing occur relative to the basic voiceless segment) then interpreting modifications of this kind as evidence for gestural blending in the purely voiceless clusters becomes hazardous.

---

[18]Linguistically conservative should not probably be taken to mean a consensus view among linguists, because there is a long phonological tradition of regarding sequences like /sp/ as complex segments, intermediate between simple segments and true clusters.

3 The development of fiberoptic and photoelectroglottographic techniques for the investigation of laryngeal articulation

3.1 Introduction[19]

It can be assumed that the immediate aim of any investigation of devoicing is to obtain measurements of the amplitude, form and timing of the abductory-adductory cycle, either directly in terms of the time-course of the separation of the vocal processes of the arytenoid cartilages, or indirectly through the resulting transillumination signal, or through the underlying electromyographic activity. As discussed further in the previous chapter the wider aim of such analyses is to gain further insight into, firstly, the nature of laryngeal-oral coordination in different categories of consonants and secondly into the blending processes occurring in clusters of voiceless sounds. We will discuss how the data for such analyses can acquired most effectively using a combination of fiberoptics and transillumination, and will also look more briefly at EMG and pulse-echo ultrasound. The emphasis on the former techniques is motivated by the fact that they are probably the ones most readily available in a normal phonetic laboratory environment. The reader is referred to Löfqvist (1990) for very convenient illustrations of some typical voiceless sequences since transillumination and EMG signals are shown in parallel.

3.2 Fiberoptic endoscopy, transillumination, and their combination

3.2.1 Use of fiberoptics for filming the larynx in running speech

In 1968 Sawashima and Hirose presented an endofiberscope that for the first time made it possible to routinely investigate laryngeal activity in running speech. For a description of fiberscope construction see Sawashima and Ushijima (1971). For a representative early study examining the voiced and voiceless consonants of English see Sawashima, Abramson, Cooper & Lisker (1970).

One of the main problems in analyzing laryngeal films (ceteris paribus, this also applies to transillumination) is caused by the fact that the distance between objective lens (i.e distal end of

[19] An earlier version of this chapter appeared in Hoole, 1999b.

the fiberscope) and the glottis is not constant and not known. Differences in this distance can be caused, of course, by vertical movements of the larynx but also by the influence of velar movement on endoscope position. These influences must be minimized by careful choice of the speech material since methods that have been proposed to control for the varying distance between endoscope and glottis, such as radiographic monitoring (Kiritani, 1971) or stereoscopic procedures (Sawashima & Miyazaki, 1974, Fujimura, Baer & Niimi, 1979, Yoshioka, 1984), are unlikely to prove suitable for routine use. Thus most investigations must currently content themselves with relative rather than absolute measurements of glottal opening. Assuming suitable phonetic material is employed the stability actually achieved in a particular recording must essentially be judged by the experienced investigator.

One reason why a combination of fiberscopy and transillumination is attractive is that at standard video or cine frame rates the temporal resolution of an endoscopic film is rather low, particularly if it is desired to locate the offset and onset of voicing in the devoicing gesture (cf. Hirose and Niimi, 1987).

One of the main attempts to circumvent the above limitation involves the use of specialized digital video equipment (Honda, Kiritani, Imagawa & Hirose, 1987; Kiritani, Imagawa & Hirose, 1992). CCD cameras allow trade-offs between spatial and temporal resolution, so an adequate frame-rate is certainly achievable. Use of image-processing to extract parameters such as glottal area or distance between the vocal processes should be feasible. However, in view of the fact that such algorithms tend to be computationally expensive and that some compromises with regard to spatial resolution have to be made it seems that at the present time this represents essentially a rather roundabout way of arriving at something that is not very different from a straightforward transillumination signal, at least as far as the gross characteristics of the devoicing gesture are concerned. Nonetheless, for the investigation of the way in which patterns of vocal fold vibration are modified at transitions between vowels and consonants the technique already appears to offer much promise (it should also be ideally suited to investigating irregular or asymmetric phonatory phenomena such as creak or diplophonia). With further technological advances more routine application of this technique should become possible. Certainly, with rigid rather than flexible endoscopes use of high-speed imaging is clearly increasing. Examples of recent work can be found in Murano et al. (2003) and Dollinger et al. (2003).

3.2.2 Initial developments in transillumination

The transillumination technique initially developed independently of fiberoptic endoscopy. It essentially involves a light source and a phototransducer located on opposite sides of the glottis; the amount of light passing through the glottis, and accordingly the output voltage of the phototransducer amplifier, is modulated by the changes in the size of the glottal aperture occurring during speech and respiration. Over the years various different arrangements of these two basic components have been tried out (see e.g Sonesson, 1960; Malécot & Peebles, 1965; Lisker, Abramson, Cooper & Schvey, 1969; Ohala, 1966; Frøkjær-Jensen (1967), mainly regarding whether the light source is applied externally to the neck below the glottis, and with the phototransducer in the pharynx, or vice-versa. See Hutters (1976) for a valuable overview of methodological issues[20].

Once fiberoptic endoscopy started to become widespread in phonetic research, it was natural to employ an arrangement with the fiberscope functioning as light source in the pharynx and with the transducer applied externally to the neck[21] (Löfqvist & Yoshioka, 1980a,b). The great advantage of this approach is, of course, that the endoscopic view allows the stability of the positioning of the light-source in the pharynx to be monitored, at least qualitatively.

There are essentially two positions in which the transducer can be applied to the neck, either between the cricoid and thyroid cartilage, or below the cricoid cartilage. Following Frøkjær-Jensen (1971) we can assume that devoicing mainly involves modulation of the width of the posterior glottis, while in phonation the main modulation of the width occurs at more anterior locations. It appears (and is anatomically plausible) that the lower transducer position weights the posterior devoicing activity more strongly, while the upper position weights phonatory activity more strongly[22]. Accordingly, the sub-cricoid position is preferable for studies of devoicing; in particular, a more stable (albeit sometimes rather weak) signal is obtained since in the upper position the signal can be strongly influenced by changes in laryngeal height and orientation, related, for example, to the intonation pattern of the utterance. Thus, Löfqvist & Yoshioka found a good correlation between the amplitude of the transillumination signal and the glottal width as measured from a fiberoptic cine film, but only when the lower transducer postion was used[23].

3.2.3 Transillumination with video-fiberscopic filming

In our own implementation (Hoole, Schröter-Morasch & Ziegler, 1997), we became convinced that it is essential to record the whole experimental session on videotape (cf. Hirose, 1986) since if a permanent record of the endoscopic view is not available it is very easy for the human observer to overlook brief retractions of the tongue-root or epiglottis that have a massive influence on the transillumination signal. A specific example is given in Chapter 4 (Fig. 4.2).

---

[20]One aspect of Hutter's work that deserves to be more widely followed is a procedure to determine the relationship between change in output voltage of the PGG amplifier for defined changes in light-intensity.

[21]In contrast to custom-designed equipment such as the Frøkjær-Jensen photoelectroglottograph, the stability of standard endoscopic light-sources should not be taken for granted. Appreciable ripple at the first few harmonics of the mains frequency can be troublesome. Use of a DC power source is advisable. See further discussion of experimental procedures in Chapter 4.

[22]cf. Baer, Löfqvist & McGarr (1983) for this conception of the transillumination signal as the weighted sum of the glottal width at a series of points along its length.

[23]It should not be overlooked, however, that the correlation is not based on two completely independent measurement methods; it will be probably be the case, for example, that changes in the distance between larynx and fiberscope will have a similar effect on the analysis of both the cine film and the transillumination signal (e.g smaller image, lower signal amplitude).

Explicit synchronization of the video and transillumination signals is also very important[24]. In order to further facilitate the recognition of such artefacts, it seemed to us that one obvious approach would be to record from **two** phototransistors simultaneously, one between cricoid and thyroid, and one below the cricoid cartilage. Based on the above discussion of transducer positioning one can expect that the reduction in light intensity caused by shadowing will in general either be greater at the upper transducer position or will at least start earlier there. In other words, when the two signals diverge, a departure from ideal recording conditions may have occurred[25].

In the next figure typical traces obtained with the twin transducer system are displayed for the German sentence "Lies 'die Schiffe' bitte" ("Read 'the ships' please"). The sentence contains the voiceless sounds /**s**/, /**ʃ**/, /**f**/ and /**t**/. Each of these sounds is associated with a peak in the amplitude of the transillumination signals. Main stress in this sentence fell on the first syllable of "Schiffe", and it will be observed that much the largest amplitude of the devoicing gesture in this utterance is associated with the fricative /**ʃ**/ at the onset of this syllable, peak glottal opening occurring at about the temporal midpoint of the fricative. Vocal fold vibration is visible in the transillumination signal as a high-frequency modulation overlaid on the gross abductory and adductory movements. As is to be expected from the above discussion, this phonatory activity is a more salient feature of the signal labelled PGG1, which corresponds to the transducer position between thyroid and cricoid cartilages. A consistent feature of voiceless fricatives that can be easily followed here (especially for the stressed /**ʃ**/) is the hysteresis effect by which vocal fold vibration dies out rather gradually during glottal abduction for the fricative, but does not recommence for the following vowel until the glottis is almost completely adducted again.

Normally, as here, the relative amplitude of glottal abduction for the different voiceless sounds in the sentence is the same at both transducer positions, the two traces proceeding essentially in parallel. If, however, this is not the case, (for example, if abduction for /**ʃ**/ were to appear greater than for /**f**/ in the PGG 2 trace but smaller in the PGG 1 trace) then this is an indication that recourse should be had to the video film to check for shadowing of the glottis etc.

---

[24]We have found the following solution very convenient: a commercially available video-timer (FOR-A VTG33) was modified according to a technique developed by N.R. Petersen at the phonetics lab of Copenhagen University. In addition to its normal function, which is to insert date and time into the video signal with a resolution of 1 cs, pulses for the one-second intervals were led out in a form which allowed them to be recorded on one track of an instrumentation recorder that was also used to record the photoglottographic and audio signals. Alternatively, if the measurement signals are recorded online by computer, a straightforward solution would be to remote-control the video-timer using a digital I/O port, with the computer resetting and starting the timer at the start of each measurement sequence. A procedure of this kind has been implemented with a FOR-A VTG55 video timer for parallel video filming and electromagnetic articulography. Further synchronization procedures are discussed in the next chapter (p. 37).

[25]This double transduction may also increase the utility of transillumination for studies of phonation, since vibratory parameters may differ along the front-back dimension of the glottis (Baer et al., 1983).

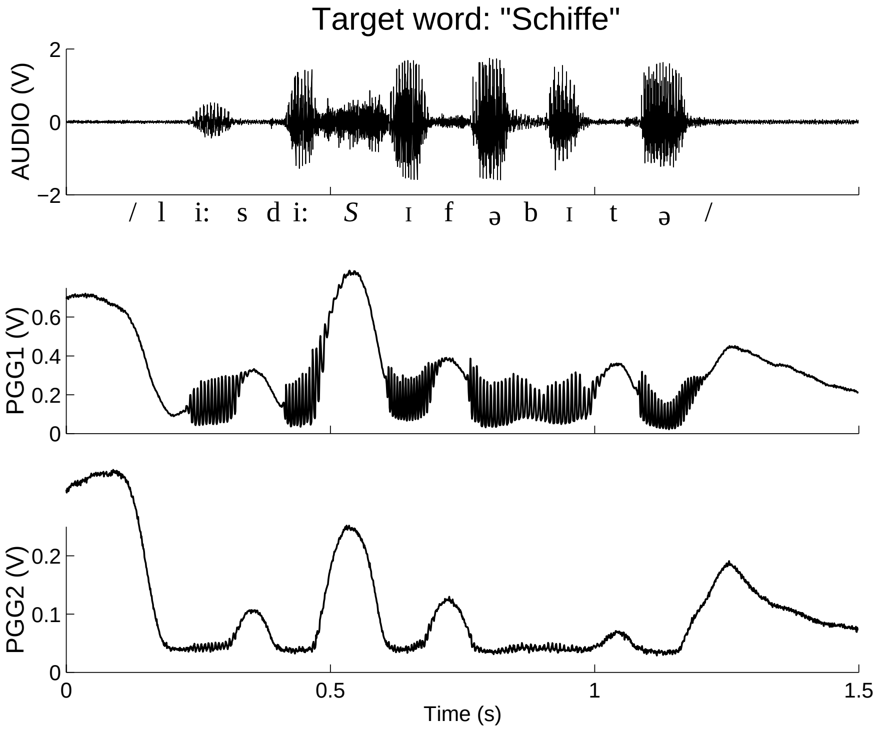


Fig. 3.1: Example of two-channel transillumination signal for a German sentence ("Lies die Schiffe bitte") containing several voiceless sounds. The trace labelled PGG 1 (middle panel) is derived from a phototransistor applied externally to the neck between thryroid and cricoid cartilages. For PGG 2 (bottom panel) the phototransistor was located below the cricoid cartilage.

3.3 Electromyography of the larynx for the study of devoicing

Laryngeal EMG has clearly contributed much to our understanding of articulatory functions of the larynx. See especially the work of Hirose (1975) in clarifying the status of the PCA (posterior cricoarytenoid) as a speech muscle, and in establishing the typical reciprocal pattern of activity of PCA and INT (interarytenoid) in the devoicing gesture (Sawashima & Hirose, 1983). The specific role of the cricothyroid (CT) in the control of voicing was the main topic of Part I of this work, where general background to the use of EMG in speech research on the larynx can be found. Perhaps the only point worth drawing attention to here is that - unlike CT - recording from the muscles most closely involved in the abductory-adductory movement of the arytenoids for devoicing, namely PCA and INT, requires a perioral rather than a percutaneous approach (see e.g Hirose, 1971). The perioral approach is probably rather more demanding for the subject; accordingly EMG was not seriously considered in the context of the present experiments.

3.4 Pulse-echo ultrasound

The ultrasound technique has provided much useful kinematic data on several articulatory systems (e.g. Keller & Ostry, 1983; Munhall, Ostry & Parush, 1985). Munhall & Ostry (1985) report on its application to laryngeal kinematics. Briefly, the procedure is designed to measure the distance between an ultrasound transducer applied laterally to the thyroid cartilage and the moving vocal fold, sample rates of 1 kHz or more being possible. A potential advantage of the ultrasound technique is that, in contrast to transillumination, absolute values for the excursion of the vocal fold can be achieved. However, this may only be an apparent advantage because it is never possible to indicate precisely which point (or rather small area) on the vocal fold forms the basis of the measurement at any given moment or in any given session. Perhaps the main drawback currently with this interesting technique is that there are relatively few published investigations, so that little is known about possible sources of error; there are no studies of which we are aware which compare the technique with parallel recordings made with an alternative procedure such as fiberoptics (but see Hamlet, 1981, for extensive discussion of the utility of ultrasound in the analysis of phonation).

4 Experimental Procedures

This chapter is divided into the following four sections:

In the first section the choice of speech material used in the experiments is presented.

In the second section the recording procedures used for synchronized acquisition of audio, transillumination together with video filming of the larynx, and electropalatography are outlined. The use of the video images to check for artefacts in the transillumination signal is illustrated.

The third section presents the pre-processing procedures applied to the transillumination data, including discussion of procedures for making the estimate of the amplitude of glottal movement as robust as possible.

Finally, in the fourth and longest section, the procedures for identifying temporal landmarks in the glottal and oral signals are outlined and illustrated with typical acoustic, glottographic and palatographic signals.

4.1 Speech Material

The speech material on which the analyses were based was acquired as part of a larger corpus designed to provide material for different purposes. Basically, the corpus included voiced and voiceless consonants in word-initial and word-medial position, using real German words as far as possible within the constraints of the transillumination setup, i.e avoiding sounds with tongue retraction, and also nasal consonants, in the vicinity of the target sounds.

Each sentence of the corpus included two target words, spoken in the frame

"Lese WORD1 wie WORD2 bitte" ("Read WORD1 like WORD2 please").

The corpus consisted of two groups of sentences:

Group 1: Spoken in three loudness conditions: Loud, Normal, and Whisper. The Whisper condition will not be considered further here[26]. The Loud condition will be used to form a source of supplementary material on selected contrasts.

Group 2: Spoken only at normal volume. Mainly intended to cover consonant clusters:

The first of the two tables immediately below lists the two target words in each sentence. The consonant material used for the present analyses is underlined and in bold-face. In the present investigation medial consonants were not analyzed; in addition analysis was restricted to either one-syllable words, or to two-syllable words with schwa in the second syllable:

[26] But it is potentially intriguing: Do devoicing gestures occur in voiceless utterances?

Group 1:

biete_**T**ype
büßte_**T**iefe
fühle_Lüde
Pfister_Olive
Pfütze_Hüfe
piepe_**z**ipfe
Püterich_hüpfe
schiebe_**T**ische
Schübe_Diele
teste_Thesaurus
These_testiere
Tschüss_düste
Tüte_Hütsche
viele_Lüwe
Wiese_**T**ide
wüsche_**Sp**itze
Ziele_Hitsche
Zypern_**C**hile

Group 2

Flüsse_**Sp**itze
Pflicht_**St**ühle
Pflüge_**St**iele
Plitsche_**p**iepse
Plüsch_**Schl**itze
Psi_**Z**yste
Psyche_**P**iste
Schwüpse_**fl**itze
Spiele_**Spl**itter
Spüle_Zwiebel

Table 4.1: List of the two target words in each utterance (separated by underscore). Spoken in the carrier phrase "Lese WORD1 wie WORD2 bitte". Words in Group 1 were spoken at all three volume levels (normal, whisper, loud), words in Group 2 at normal volume only. Analyzed consonants are bold and underlined.

Single consonants:

p	**piepe**, piepse, Piste
t	**These**, **teste**, **Tüte**, **Tide**, **Tiefe**, **Tische**, **Type**
f	**viele**, **fühle**,
ʃ	**schiebe**, **Schübe**

Fricative plus plosive:

ʃt	Stiele, Stühle
ʃp	Spiele, Spüle, Spitze, **Spitze**

Plosive plus fricative:

pf	**Pfister**, **Pfütze**
ps	Psi, Psyche
ts	**Ziele**, **Zypern**, **zipfe**, Zyste
tʃ	**Tschüss**, **Chile**

Clusters with / l /:

pl	Plitsche, Plüsch
fl	Flüsse, flitze
ʃl	Schlitze
ʃpl	Splitter
pfl	Pflicht, Pflüge

Table 4.2: List of target words actually analyzed, broken down by consonant sequence. Words in bold face are from Group 1

Five randomizations of each of these two groups of sentences was prepared. The material was then spoken in the following order:

3 repetitions of Group 1
3 repetitions of Group 2
2 repetitions of Group1
2 repetitions of Group 2

For the Loud/Normal/Whisper condition of Group 1, the three loudness conditions for a given sentence were spoken consecutively, but also randomized so that there was not a fixed order of the loudness conditions over the different sentences.

There was thus a different amount of material available for the different target consonant sequences. This is broken down in the second table on the preceding page. Words in bold face are from Group 1, and thus available in the different loudness conditions:

4.2 Subjects

Three subjects took part in the experiments. An important criterion for participation was that they were all experienced in talking with the EPG artificial palate in place

SF: Female, born and raised in eastern Germany.

CG: Male, born and raised in southern Germany.

RW: Male, born and raised in eastern Germany.

All subjects were in their thirties at the time of the experiments and spoke standard German with minor regional colouring. SF and CG had already participated in a previous fiberscopic experiment and also took part in the EMG experiments presented in Part I of this monograph.

4.3 Recording Procedures

4.3.1 Experimental setup

An overview of the experimental setup is shown in the following figure. We are not aware of many previous investigations that have attempted simultaneous measurement of both laryngeal and oral articulation; one notable exception is to be found in Fujino et al. (2000) who combined videofiberscopy, transillumination and electromagnetic articulography (and intraoral pressure measurement!). Details of the individual elements of the setup are as follows:

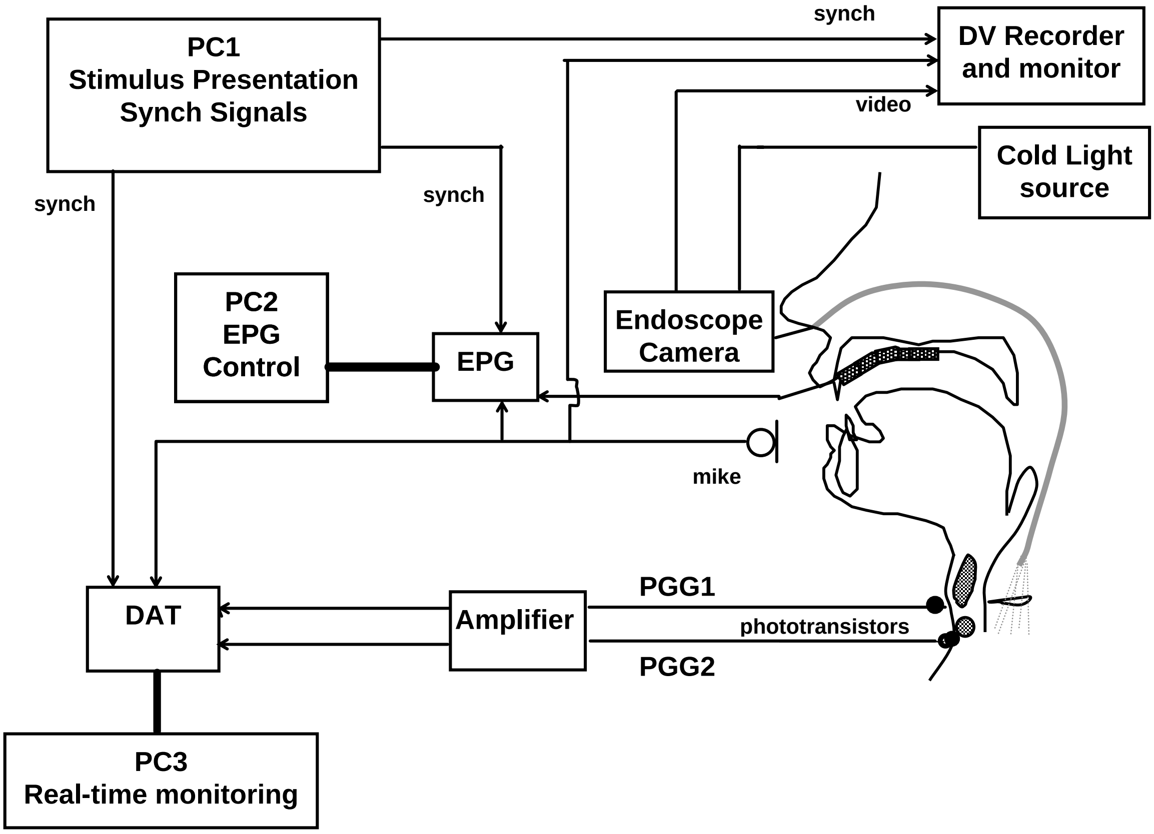


Fig. 4.1: Experimental setup for synchronized acquisition of audio, transillumination, electropalatographic and video signals.

An Olympus fiberscope (ENF P3) was inserted pernasally into the pharynx to provide illumination of the larynx for the transilluminaton technique. The light-source was a Timcke KQ−14/S endoscopic light source specially modified to run on a 12V battery power supply. The light passing through the glottis was picked up by two phototransistors (BPX 81) attached externally to the neck, one between thyroid and cricoid (PGG1), one just below the cricoid (PGG2). These signals were amplified by a Frøkjær-Jensen photoelectroglottograph (LG900; the light-source built into this equipment was not used), modified to provide amplification for two channels, and recorded on DAT tape on a multichannel instrumentation recorder (Sony PC208x) running in 4-channel mode at a sample rate of 24kHz. The endoscope was connected to a black-and-white CCD camera (Hitachi KP-M1/K) and the video signal was digitally recorded on MiniDV cassettes using a Sony GV-D1000E. The LCD display of the recorder was used by the otolaryngologist Dr. Klaus Dahlmeier to monitor the position of the endoscope.

Tongue-palate contact patterns were registered by means of a Reading EPG3 system.

The audio signal was picked up by a Sennheiser MKH20 microphone and recorded on the DAT tape along with the transillumination signals. The microphone signal was also fed to one audio input of the DV recorder, as well as to one of the analog inputs of the EPG3 device.

The final basic element in the setup was a PC (PC1 in the figure) that displayed the speech stimuli to the subject (on an additional VT200 alphanumeric terminal attached to the serial port) and also generated a synchronization signal marking the start and end of each trial that was output via the PC's parallel port and recorded on the DAT instrumentation recorder, the DV recorder, as well as on the second analog channel of the EPG3 device.

The DAT instrumentation recorder was connected via a digital interface to a notebook computer (PC3 in the figure) that provided a realtime oscilloscopic display of the transillumination and audio signals during the experiment, and was used after the experiment to digitally transfer the measured signals from DAT tape to hard-disc.

The basic procedure during the experiment for each trial was as follows:

- Stimulus for upcoming trial was displayed to the subject on the alphanumeric terminal (PC1)
- EPG data acquisition was initiated by investigator monitoring EPG on PC2
- Start signal given to subject (beep) and synchronization signals generated (PC1)
- Subject speaks
- After preset time (e.g 3s) synchronization signal reset (PC1)
- EPG data acquisition terminated (PC2)

The DAT and DV recorders ran continuously throughout the experimental session.

4.3.2 Alignment of the different data streams

After the experiment the following procedures were necessary to extract and align the data:

The complete DAT recording of the experimental session was transferred to computer disk.

Then the portions marked by the synch signal were extracted to individual files for each trial.

A similar procedure was necessary for EPG data. This data was already stored in individual files for each trial during acquisition, but it was then necessary to trim the data to the portion marked by the synch signal[27].

The audio signal recorded by the EPG hardware was used as a cross-check with respect to the audio recorded on the DAT tape that trials had not been missed or datafiles corrupted.

Video processing

Synchronization of the video signals with the other signals followed procedures that are conceptually similar to those used for the signals on DAT tape, but were not actually carried out until about two years after the original experiments because of the very high demands made on computer storage (selected trials were processed earlier)[28]:

---

[27]This was based on procedures we originally developed to allow synchronized acquisition of EPG and articulographic signals (Hoole, 1996, pp. 91ff), and modified by C. Mooshammer for use with a newer version of the EPG system.

[28]A variation on this procedure has also been used for recordings where only the speech audio but not the synch signal was recorded on the DV tape:

- The complete audio track is extracted from the film.
- One well-defined time-point (e.g a plosive burst) is identified both in the audio track from the video film and the complete audio track from the DAT tape.
- The synchronization information on the DAT tape can then be adjusted to give a first

- The video film of the experimental session was digitally transferred at full resolution via IEEE1394 interface to computer.
- From this film the audio track with the synchronization signal was extracted and processed to extract the time-points of the synch marks.
- All frames of the complete video track were exported to individual TIFF files
- All TIFF files corresponding to the same trial were merged into a single MATLAB file.

At this stage an additional processing step was applied to the video data to simplify optimal use of the restricted temporal resolution of the video signal:

For historical reasons video is normally generated in an interlaced format, i.e each frame (25/s in Europe) consists of two interlaced fields, alternating between odd and even video lines in the full frame (the field rate accordingly is 50/s). When one considers that many speech movements (e.g in our case a glottal abduction-adduction gesture) may not last more than 150ms (i.e about four frames at 25/s), it is obviously advantageous to arrange the final video data field-wise rather than frame-wise. We thus deinterlaced the frames, i.e separated each frame into its odd and even field, then interpolated the “missing” lines in each field, so that adjacent frames in the final data are spatially aligned and have the aspect ratio of the original frames.

4.3.3 Using video to identify unreliable transillumination data

An example of the aligned and deinterlaced video is shown in the next figure. The main purpose of this figure is to illustrate why it is important to carefully peruse the video information in conjunction with the transillumination signals.

The figure shows an extract from the target word “Teste”. Here the target sound (initial /**t**/) is followed by the short lax vowel /**ɛ**/. Although this is notionally a front vowel (i.e might be considered appropriate for transillumination) the configuration in the pharynx is actually quite constricted. This has the effect that the view of the glottis is increasingly obscured by the tongue-root during the CV movement (this becomes apparent in the film roughly from the release of /**t**/). It will be observed in the transillumination signal that the baseline in /**ɛ**/ is lower than in the vowel preceding /**t**/, and the shape of the signal appears unusual compared to other examples of plosives in the corpus (or indeed in general in the literature). Since this pattern was consistent over all three subjects it was decided to exclude this item from further analysis, to avoid the risk of performing measurements on distorted signals.

---

approximation for synch markers with respect to the DV audio (typically the alignment of these two audio streams only drifts apart very slowly).

- Cross-correlate the DAT and DV audio for those signal portions corresponding to each trial. The lag corresponding to the maximum in the cross-correlation function can be used to fine-adjust the position of the synch markers for the DV audio stream.

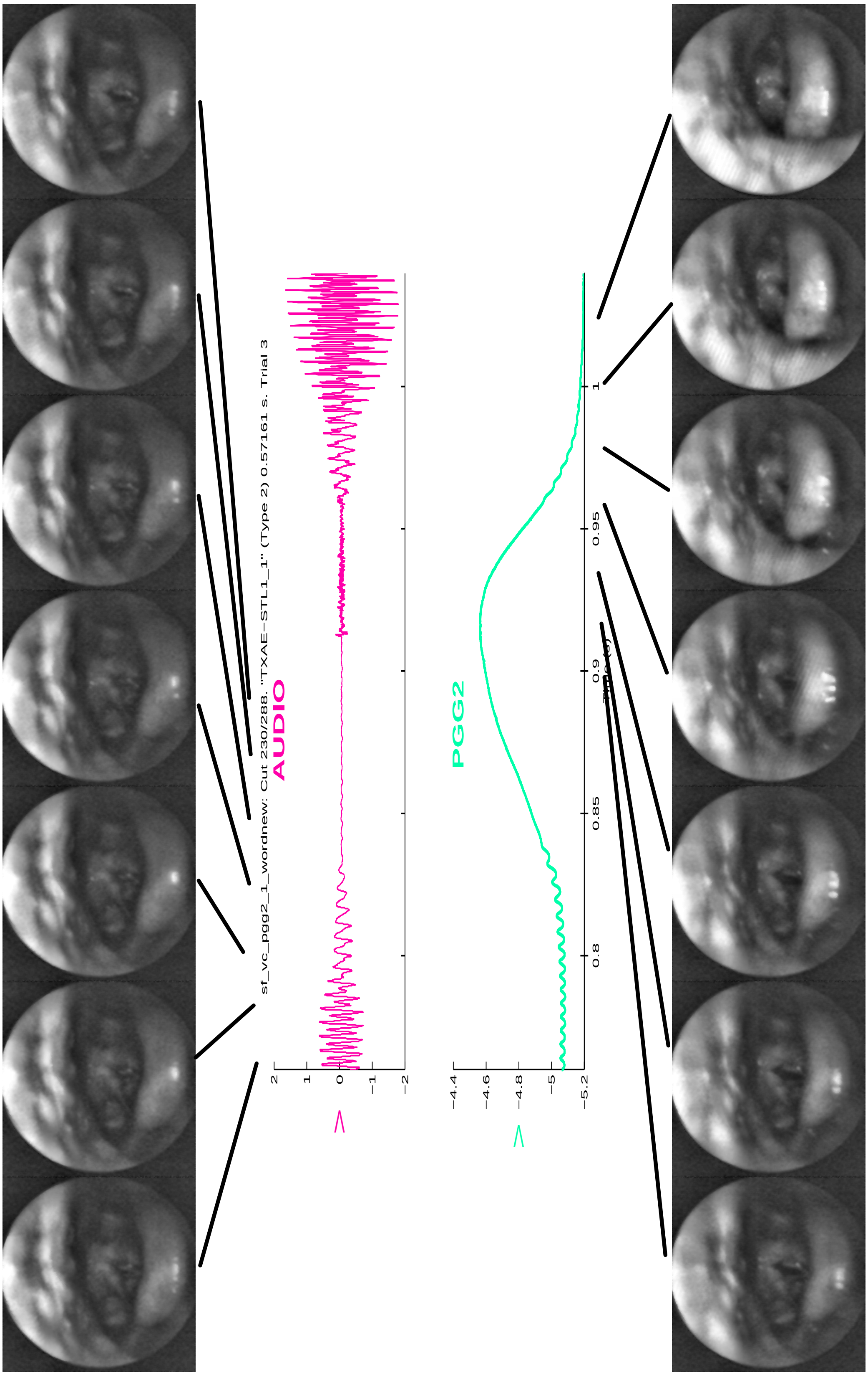


Fig. 4.2: Transillumination signal and synchronized video frames for initial /t/ of "Teste". Sequence starts just before end of schwa of "lese" in the carrier phrase and continues into the first part of the stressed vowel /ɛ/ of "Teste". Note increasing retraction of epiglottis (upwards in the images) towards the end of the sequence. Interval between video frames is 20ms.

4.4 Preprocessing of the transillumination signal

4.4.1 Introduction

The glottal signals were first downsampled with appropriate anti-aliasing filtering from the DAT sampling rate of 24kHz to 1kHz. They were then additionally smoothed with an FIR filter (Kaiser window design, -6dB at 27.5 Hz, -70dB at 50Hz and above), and first-differenced to obtain a velocity signal.

4.4.2 Normalization procedures for glottal amplitude

In this section we will discuss procedures for removing some extraneous sources of variation from the amplitude of the transillumination signal. Partly due to the nature of the corpus this was a somewhat complicated task. It should, however, be borne in mind when reading the following section that a major part of the information we hope to derive from the transillumination data, namely information on laryngeal *timing*, is essentially unaffected by the precise stability of the signal amplitude.

As discussed in the previous chapter it is scarcely possible to calibrate the transillumination signal in, for example, mm^2 of glottal opening area. Moreover, the measured signal amplitude can be affected by changes in the position of the fiberscope during the course of the experiment, and probably even by changes in the height of the larynx during the course of the utterance. Thus interpretation of apparent glottal opening amplitude differences for different sounds or linguistic contexts must be carried out with great caution. Since, however, it is of some interest to compare the amplitude of, for example, single fricatives vs. fricative-plosive clusters we will outline in this section the procedures followed to improve the robustness of the estimates of glottal opening that can be made from the transillumination signal.

As will be seen, there are a number of potential approaches for compensating for extraneous sources of influence on the signal amplitude. One useful procedure in cases where the target items are spoken in a constant carrier phrase would be to include in the carrier phrase a sound with a clearly defined devoicing gesture such as a single voiceless fricative. The amplitude of the target items could then be normalized relative to the amplitude of the reference sound. This was not feasible in the current experiment since a large part of the material was spoken at different loudness levels, which obviously would also affect a reference sound.

As an alternative means of obtaining a reference signal we first extracted the minimum signal amplitude in the lowpass-filtered transillumination signal over the 100ms preceding glottal abduction for the target consonant sequence (i.e essentially within the preceding vowel). This baseline signal was then used in two ways:

First, it was subtracted from the raw glottal amplitudes, so that a value of zero would effectively correspond to glottal closure[29].

---

[29]To maximize the dynamic range of the recording equipment the transillumination amplifier was equipped with an offset adjustment so that the signal could be biased to give negative voltages at low light levels. Thus 0V in the raw signal has an arbitrary relationship to the state of the glottis.

Second, the recordings were scanned for the minimum signal level, i.e a situation where it was assumed that essentially no light passes through the glottis. Note that this is not generally the case during glottal closure for phonation; rather, a case is required where the tongue root and epiglottis retract sufficiently to completely obscure the view into the larynx. Potential cases were cross-checked in the video signal, and the corresponding transillumination voltage was noted.

A gain adjustment for each utterance was then calculated that was proportional to the voltage difference between the vowel baseline value and this "dark reference".

This normalization procedure was not equally successful for all subjects. The effect of the normalization was estimated by computing the average of the standard deviations of peak glottal opening amplitude of each target item in the corpus (usually 5 repetitions) before and after normalization. For RW the procedure was particularly successful, giving a reduction in average standard deviation of over 25%. This was not entirely unexpected since RW was the subject where changes in the strength of the illumination of the glottis over the course of the experiment definitely occurred: During the first part of his session the voltage of the battery power supply for the light source started to decrease markedly, and for the second part we reverted to use of the mains power supply (on reasons for normally preferring battery power supply see previous chapter).

For CG a slight variation on this procedure was used: rather than using the difference between baseline and dark reference to compute a gain adjustment, we simply computed a regression function to predict glottal opening amplitude from the baseline value, then adjusting the amplitude by the regression prediction. This reduced average standard deviation by just under 25%, which was about 6% better than the dark reference method.

For speaker SF a baseline-oriented normalization procedure proved ineffective. While for both RW and CG the correlation between baseline voltage and peak glottal opening voltage was about 0.5, for SF it was only about 0.14. Accordingly, for this subject normalization was carried out by computing that gain adjustment for each block of repetitions which would give each block the same average peak glottal opening amplitude. This is also a very standard kind of normalization procedure; its disadvantage in the present case is that it does not operate on an utterance by utterance basis, and it is not necessarily to be expected that, for example, shifts in endoscope position directly follow each block of repetitions. However, it can certainly compensate for gross changes in recording conditions over the course of the experiment, and inspection of the glottal images extracted from the video at peak glottal opening did indicate some fairly consistent differences between repetitions. The other disadvantage with a block-wise normalization procedure in the specific context of the present experiment relates to the fact that the corpus was divided into two sections (the loudness variation and cluster sections; see details on corpus and organization of the experimental sessions at the beginning of this chapter) so it was necessary to normalize the loudness and cluster sections separately. Nevertheless, the final result of the normalization for speaker SF was a reduction in the average standard deviation of just over 30%, i.e in fact somewhat more than was achieved for the other two speakers[30].

The second area in which adjustment of the measured amplitudes was performed related to a rather obvious effect of position of the word in the carrier phrase for speakers SF and RW: Consonants from Word 2 typically showed lower glottal opening amplitude than those from

[30]Blockwise normalization was also examined for RW and CG but it was somewhat less effective for these speakers than the baseline-oriented procedures.

Word 1. Such an effect is not entirely unexpected since - as a kind of declination effect - there may be a shift in the vertical position of the larynx over the course of the utterance, and/or the relative position of cricoid and thryroid cartilage may change. Both effects could conceivably alter the amount of light reaching the phototransistor on the external surface of the neck[31].

Since the differences did not appear attributable to a simply more reduced articulation in Word 2 position (there were no salient differences in duration - and indeed the carrier phrase was specifically designed to elicit words with comparable stress: "lese WORD 1 wie WORD 2 bitte") it was decided to apply a correction factor to the amplitude of material in Word 2 position. The corpus actually contained a fair number of cases where the same consonant or cluster appeared in both positions (e.g /**sp**/ from the cluster part of the corpus, and single /**t**/ from the loudness part). First of all the amplitudes were compared for this directly comparable material. The average ratios of Word 1 to Word 2 amplitudes worked out at 1.64 and 1.55 for RW and SF respectively. As a cross-check the ratios obtained by pooling all material for the two word positions (broken down into Loud, Normal,and Cluster corpus items) were computed. These averaged out at 1.53 and 1.67 for RW and SF respectively. For each subject the more conservative ratio was chosen as final adjustment factor to apply to Word 2 amplitudes, i.e 1.53 for RW and 1.55 for SF (these ratios were also computed for CG, but as they were close to 1 no adjustments were made (approx. 1.09)).

Examples of normalized and unnormalized data

The first example shows results for the /**f**/ in "viele" for subject SF. Here the normalization gave a particularly strong reduction in variability: the standard deviation of peak glottal opening decreased about 70%.

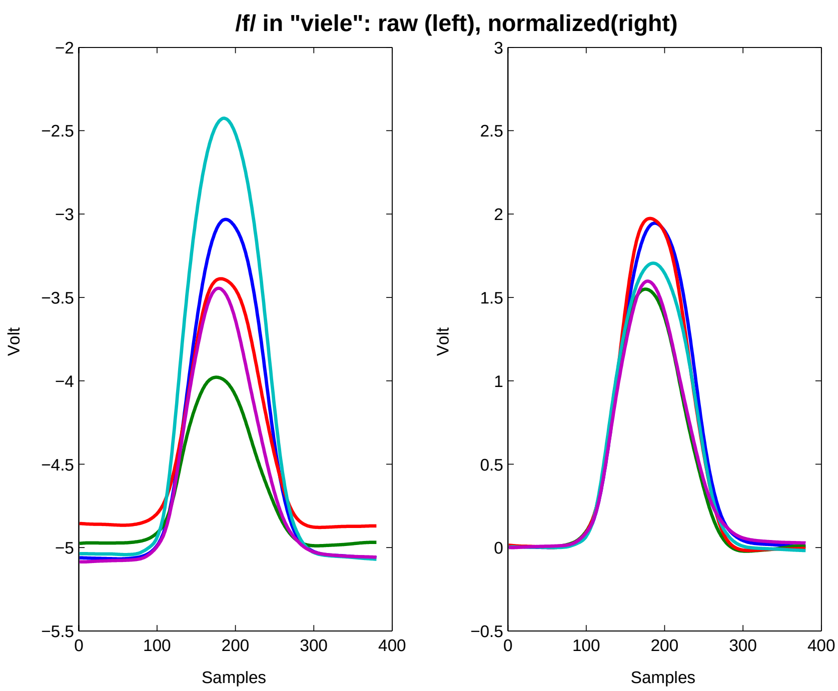


Fig. 4.3: Example of results of normalization procedure: 5 repetitions of glottal movement for /f/ of "viele" (speaker SF) before (left) and after (right) normalization.

[31]And in the case of RW, baseline-normalization was not able to remove this effect.

The second example includes the effect of adjusting amplitudes for Word 2 material. "Spiele" (Word 1 position) shows quite a considerable reduction in variability (about 40%). For "Spitze" (Word 2) the reduction in variability is negligible (partly counteracted by the increase in amplitude), but the increased amplitude results in greater similarity to "Spiele"

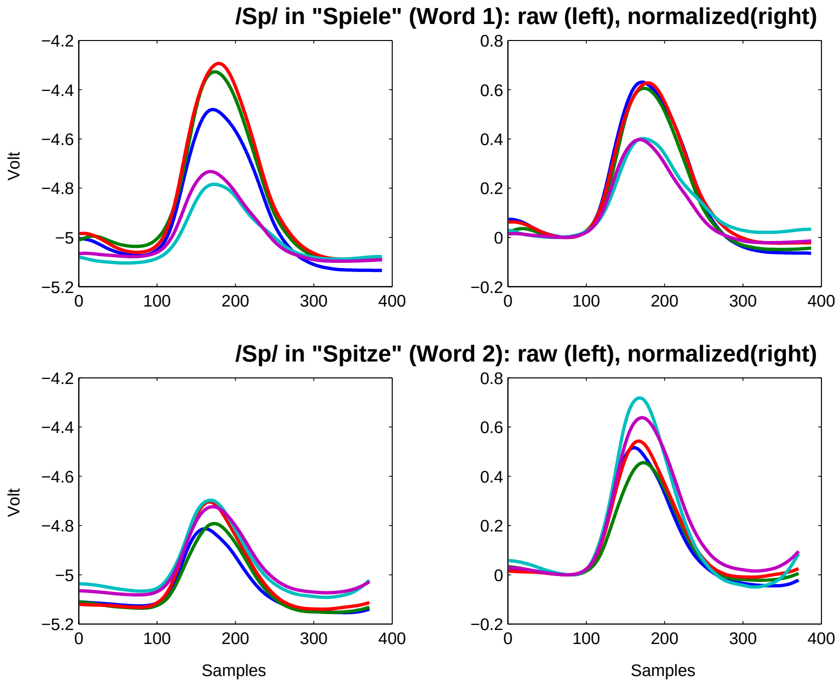


Fig. 4.4: Example of results of normalization procedure relative to position of word in phrase for cluster /Sp/. Top panels: First position in phrase ("Spiele"). Bottom panels: Second position in phrase ("Spitze"). Left panels: Before normalization. Right panels: After normalization.

A further methodological point also due to the rather convoluted nature of the corpus is the fact that for analysis it is convenient to combine the material from the Cluster part of the corpus (Group 2, spoken only at normal volume) with the Normal material from the Loudness part of the corpus (Group 1). Since the word 'Spitze' occurred in both parts of the corpus this was used to check whether this is likely to lead to any distortion. With respect to the duration of the glottal gesture the differences between the two conditions amounted to -3, +8 and -17ms for speakers SF, CG, and RW respectively. The differences were clearly not significant for SF and CG, and just significant at $p<0.05$ for RW in a nonparametric Kruskal-Wallis test. For glottal opening amplitude the differences for SF and CG were of the order of 5% and again clearly not significant. For RW the difference was larger and again just reached $p<0.05$, however it was the condition with the longer gesture duration (cluster corpus) that had the smaller opening amplitude. In short, while we admittedly have only a very small amount of material for a direct comparison (5 repetitions of each word per condition), there is essentially no evidence of consistently more vigorous glottal activity (i.e a longer, larger gesture) in one of the two conditions.

4.5 Segmentation

In this section we will present the choice of characteristic time-instants in the various signals that will be used to analyze the patterns of laryngeal-oral coordination. The first, much briefer, sub-section will present the time parameters determined in the glottal signal, and then in a substantially longer section the time-points related to oral articulation are introduced. This section is inevitably longer because of the variety of consonant sequences involved, but it is also used to present a series of typical examples of both the oral and glottal signals.

4.5.1 Glottal activity

The following time-points were determined in the glottal opening-closing movement for each target consonant or cluster (this opening-closing cycle will be referred to as the glottal gesture):

(1) Onset and offset of the glottal gesture

This was determined from the velocity signal using a threshold criterion of 20% of peak velocity in each consonant analyzed. This is a fairly high value to use for such a velocity threshold, but experience has shown that it gives stable results: there can be quite a lot of fluctuation in the baseline of the signal in the vowels before and after the voiceless segment so fairly high velocities can occur even in portions of the signal where it can be assumed that no glottal gesture is active[32].

(2) The time of peak glottal opening

(3) The time-points of maximum opening and closing velocity

In addition the following time point was determined from the acoustic signal:

(4) Voicing onset

All these time points are illustrated in Fig. 4.5 below.

It is worth emphasizing that in the present material no cases of multiple glottal opening peaks were encountered, so it was possible to apply this scheme straightforwardly to all utterances.

4.5.2 Oral articulation

In order to relate the time-course of glottal articulation to oral articulation the time-points of the onsets and offsets of the various articulatory constrictions were determined, based partly on the audio signal and partly on electropalatographic information. They will be outlined separately for plosives, fricatives, and clusters with laterals and illustrated, in conjunction with the corresponding transillumination signal, with displays of some typical utterances.

---

[32] In a different context, namely oral kinematics monitored by EMA, a 20% velocity criterion has also given good results in difficult situations. See Kroos et al. (1997) for more background on choice of velocity thresholds (Kroos, C., Hoole, P., Kühnert, B. and Tillmann, H. (1997). "Phonetic evidence for the phonological status of the tense-lax distinction in German," Forschungsberichte des Instituts für Phonetik und Sprachliche Kommunikation der Universität München 35, pp. 17-25).

Plosives

For /**t**/ there were many cases in the EPG signal of a very clear transition between adjacent frames (i.e within 10ms) from clearly no closure (i.e at least 4 contacts free in the front rows) to complete closure. These cases were used as a guideline to determine a consistent point in the sonagram for closure onset. In practice this corresponded to the traditional point of a sharp decrease in amplitude in the higher formants. Since there is no EPG information for /**p**/ it was clearly necessary to have a sonagram-based procedure. There were also a few cases for /**t**/ where, for example, one contact remained off in the row with most contact during part of the closure phase. The basic procedure followed was to locate the boundary at the end of the last pitch period with clear excitation of the higher formants for the two male speakers, and at the end of the following pitch period for the female speaker, i.e within roughly 10ms or 1 EPG frame from the start of the relevant pitch period. Because the recordings were not made in a sound-treated room (i.e with some background noise and reverberation), in cases where the acoustic recordings were unclear it was always ensured (for /**t**/) that the start of closure was not located *later* than complete EPG closure.

An example with sonagram and the sum of free contacts in EPG rows 1 and 2 is shown in Fig. 4.5. The close temporal relation between a sharp change in the sonagraphic pattern and a sharp decline in the number of free contacts in the front rows is clearly apparent (the figure also illustrates the time-points labelled in the transillumination signal).

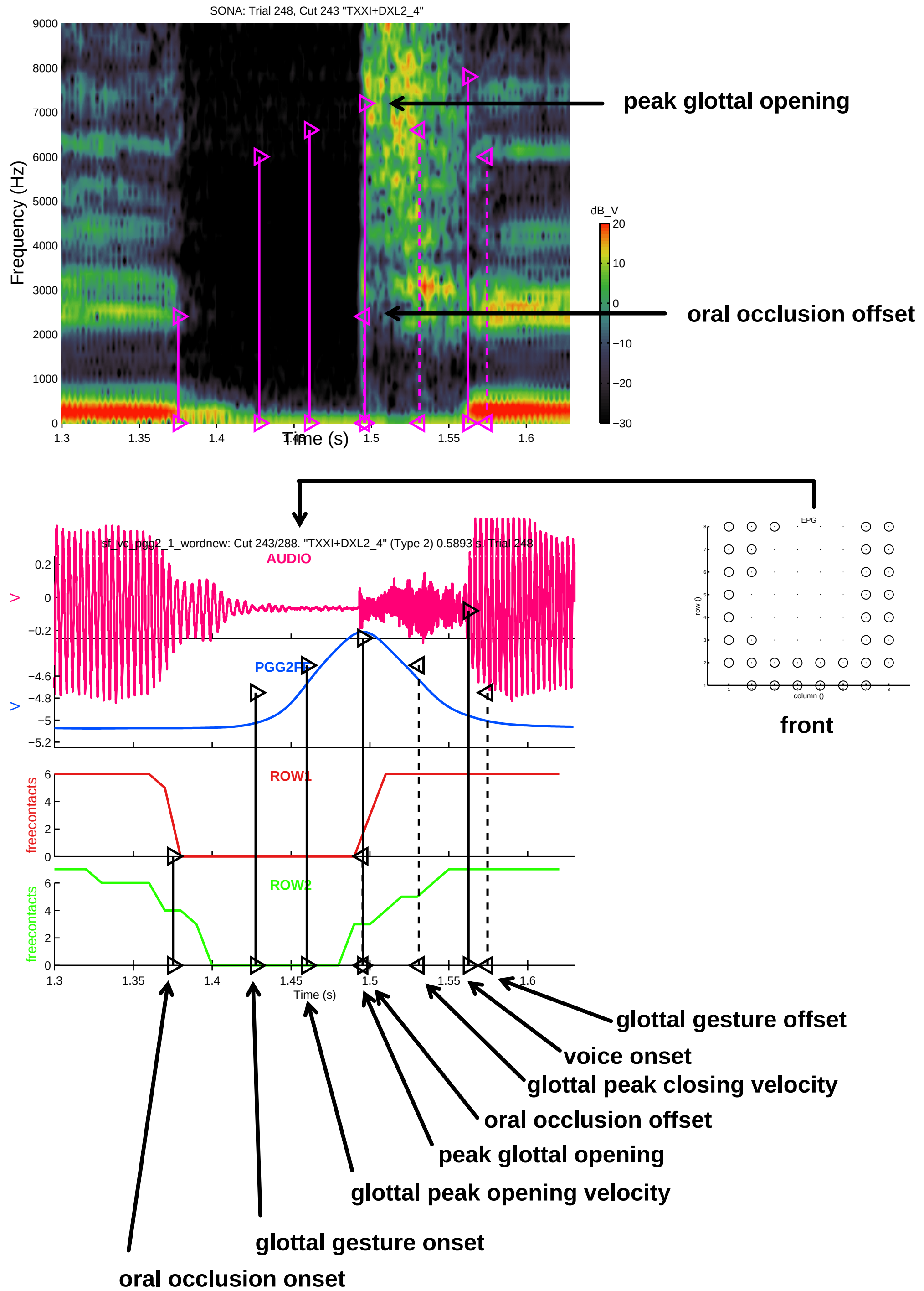


Fig. 4.5: Acoustic, glottographic and palatographic signals for the consonant /**t**/ (from target word "Tiede", speaker SF), exemplifying labelled time-points. Onset and offset of the glottal gesture, and maximum opening and closing velocity are shown here with respect to the smoothed transillumination waveform; determination of these points actually used the first-difference of this waveform (not shown here). Since Peak Glottal Opening and offset of the oral occlusion occur almost simultaneously they have been labelled additionally in the sonagram panel for clarity. The time-functions based on the palatographic signal show the number of free contacts separately for Row 1 (frontmost) and Row 2 on the artificial palate. Zero corresponds to complete closure. Complete lack of contact corresponds to 6 free contacts for Row 1 (8 for all other rows). A full frame of EPG data taken from the middle of the oral occlusion phase is shown in the small panel on the right (the two outermost electrode positions in the 8x8 matrix are not present for Row 1)

Fricatives

Onset and offset of single fricatives was located at the onset and offset of clear frication in the higher frequencies. This usually locates the onset of the fricative a short time after the amplitude of the higher formants has begun to weaken. However, since glottal abduction (as already mentioned, and as will be emphasized below) starts relatively early for fricatives compared to plosives, it was reasoned that the weakening of the higher formants could well be attributable to the weakening of the glottal closure phase. Since our segmentation point for plosives is very close to complete closure (usually within 10ms) it seems preferable to choose for fricatives a point where the oral constriction is clearly narrow enough for the generation of high-frequency noise. It should be borne in mind, however, that an unambiguous segmentation point for fricatives is inherently more difficult than for plosives using acoustic information: because of the later glottal abduction for plosives, the acoustics at the approach to closure phase are influenced essentially by the changing oral articulation; for fricatives the oral and glottal constriction are - to a much greater degree - changing simultaneously. The strength and nature of the generated frication noise is crucially dependent on the relative size of these constriction. The acoustic consequences of the complex and fast-changing aerodynamic conditions in VC and CV transitions have been extensively modelled by Stevens (e.g 1998)[33].

It was not possible to use the EPG information to provide an articulatory definition of the onset and offset of the fricative constriction for /**ʃ**/ (even for /**s**/ there were problems that will be mentioned below). In the high front vowel contexts necessary for transillumination the tongue movement from vowel to fricative can be quite subtle, and no patterns were found that were sufficiently unambiguous to be useful.

It was expected that EPG might provide a clearer pattern for /**s**/ than /**ʃ**/since the resolution of EPG is particularly good for the frontmost constriction locations. Using the clusters /**ps**/ and /**ts**/ it was hoped that it might be possible to find articulatory confirmation - at least for fricative offset - of the plausibility and interpretability of the acoustic criterion used. However, this search proved fruitless, for different reasons.

For speaker SF, for example, a plausible criterion for onset of the fricative in /**ps**/ seemed to be when the number of free contacts in row 1 or 2 reduced to 4 or less. However, at offset, this criterion often remained fulfilled until well into the vowel. Using a stricter criterion for constriction, e.g 3 or less free contacts, would have meant that onset did not occur until well after clear acoustic onset of frication (this even happened with a criterion of 4 occasionally)

A very typical pattern for SF, which made problems with offset determination unavoidable, was that in /**ts**/ sequences the number of contacts in the front two rows increased very gradually over the course of the fricative. In fact, the criterion of 4 free contacts would probably have worked out on average as matching the acoustic offset criterion quite well, but, because of the very gradual change in the EPG pattern, using it directly would have given a segmentation point with excessive variability relative to the glottal timing pattern.

The group of four figures on the following double page illustrates these points for /**ps**/ and /**ts**/ for subject SF, followed by /**ps**/ and /**ts**/ for CG.

---

[33]With respect to plosives this outline is something of a simplification since it is in fact possible to find subtle changes in voice-quality during the approach to articulatory closure; see e.g Ní Chasaide & Gobl, 1993.

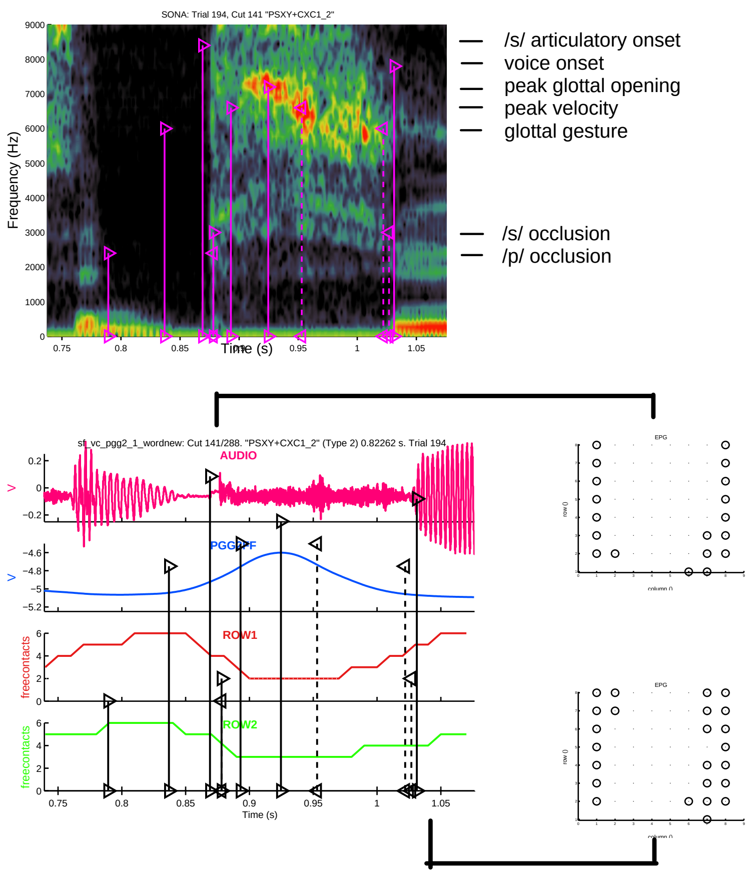


Fig. 4.6: Segmentation example for /**ps**/ in "Psyche" (subject SF). Top EPG frame located at frication onset (burst of /**p**/). Bottom EPG frame located at offset of articulatory criterion for fricative, Labels to the right of the sonagram panel indicate the identity of the markers with the corresponding height. Refer back to Fig. 4.5 for more detailed labels of the same segment types.

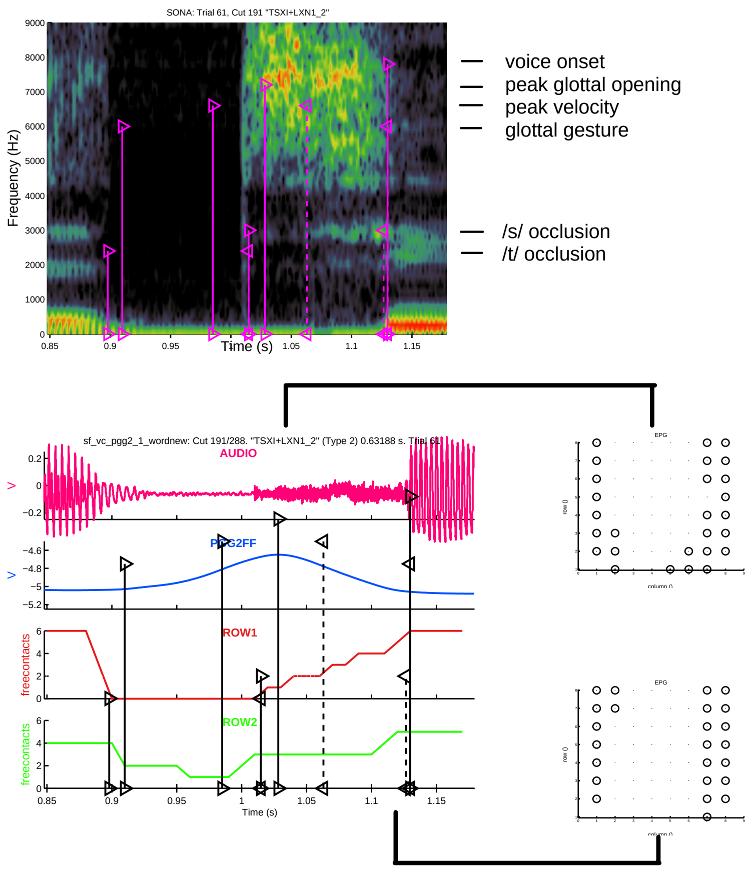


Fig. 4.7: Segmentation example for /**ts**/ in "Ziele" (subject SF). Top EPG frame located during fully developed frication. Bottom EPG frame located at offset of articulatory criterion for fricative

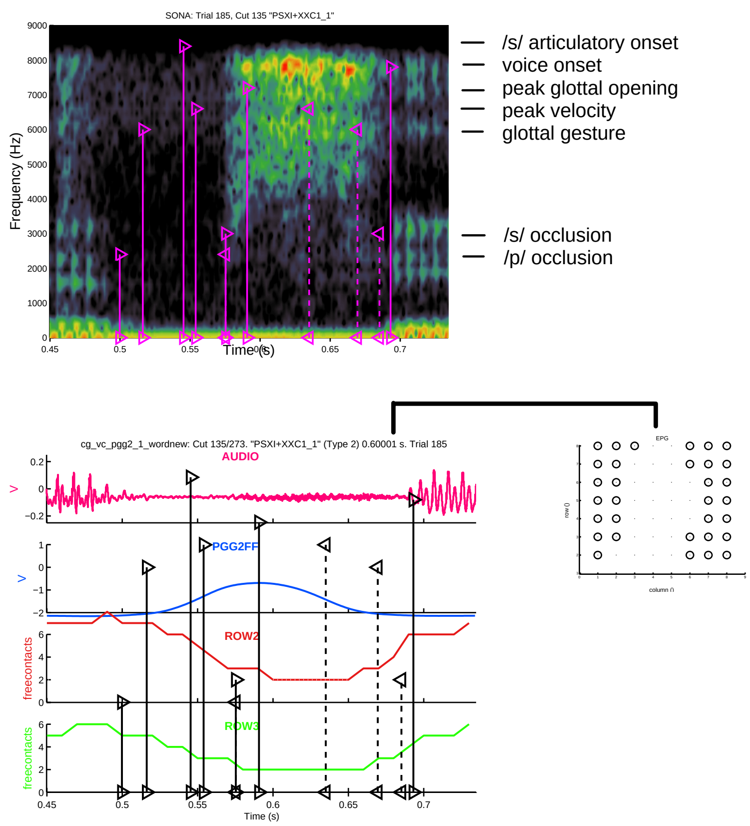


Fig. 4.8: Segmentation example for /**ps**/ in "Psi" (subject CG). EPG at last frame where palatographic criterion for fricative is met

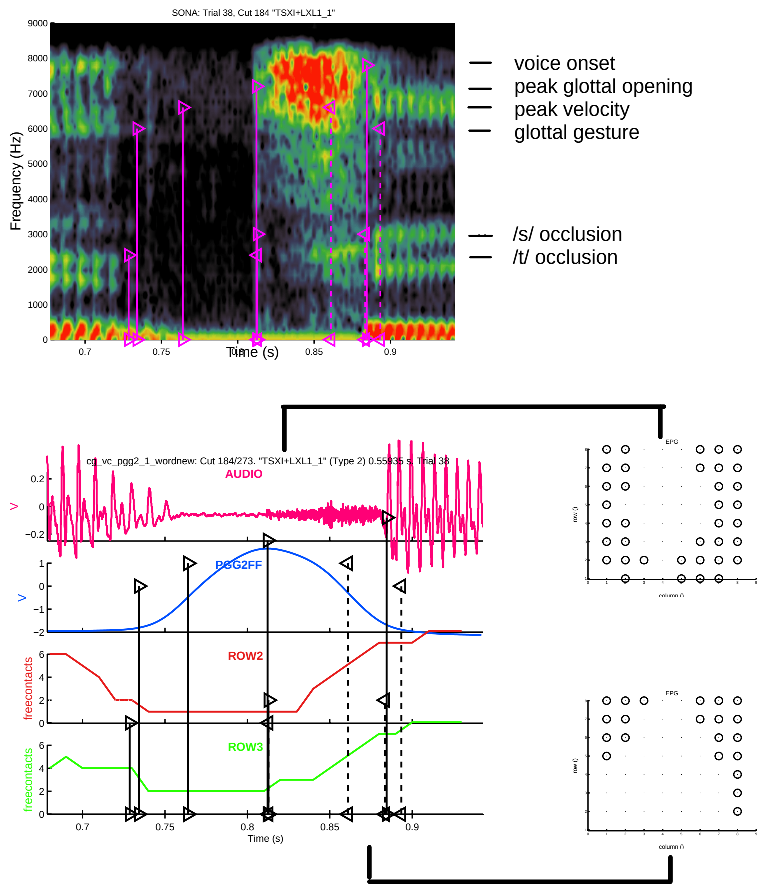


Fig. 4.9: Segmentation example for /**ts**/ in "Ziele" (subject CG). Top EPG frame located just after acoustic onset of fricative. Bottom EPG frame located just before acoustic offset of fricative

For speaker CG a resonable criterion appeared be free contacts <=3 in any of rows 1 to 3. For most of the /**ps**/ sequencies a clear pattern of decrease and increase in the number of free contacts in the front rows is observed, with the increase not being so gradual as to make determination of fricative offset by this criterion too variable. However, applying this criterion to fricative offset in the /**ts**/ sequences would have led in many cases to highly aberrant results.

The offset criterion often coincided with an increase in frication energy at lower frequencies - especially in the region of F3 of the following vowel /**i**/. Thus there is certainly a very systematic relationship between EPG pattern and the acoustics. However, given that the aim of the segmentation is to provide a basis for capturing patterns of laryngeal-oral coordination it hardly seems warranted to regard the offset of the /**s**/ gesture as occurring often roughly midway through the frication phase. Since EPG only reacts to contact it is possible for a constriction to be quite narrow and yet leave no palatographic trace. In this case probably the tongue blade is just sufficiently far away from the alveolar ridge for no contact to register (and at the same time the palatal constriction for the following /**i**/ has probably narrowed somewhat).

For speaker RW, too, it proved unfeasible to define a consistent criterion that could be applied to /**s**/ onset and offset. Free contacts <=4 appeared reasonable as criterion for /**s**/ onset in /**ps**/ sequences, but once again no consistent change could be identified at the CV transition. And again a narrower criterion would often have located onset well after frication onset and the offset well before frication offset (RW showed features already observed in the other speakers, such as very gradual change in contact pattern in /**ts**/ sequences, and a shift to probably a more palatal configuration in the latter part of the fricative).

In summary, EPG did not provide useful information for articulatory *offset* of the fricative. For the *onset* in /**ps**/ an articulatory criterion was interesting for a more substantive reason than simply supporting the appropriateness of an acoustically-based segmentation: One problem that could not be avoided in this study is that acoustically defined segments are an imperfect basis for assessing gestural onset and offset (cf. Fowler & Smith, 1986, on the conventional basis of traditional acoustic segementations). Thus if, for example, there are grounds for believing that the glottal gesture in /**ps**/ is closely coordinated with the /**s**/, then it may not be appropriate to use the acoustic manifestation of /**s**/ as the basis for determining whether peak glottal opening, for example, is coordinated with fricative midpoint, i.e the /**s**/ gesture may start while the lips are still closed for /**p**/. Although we have expressed reservations about EPG criteria for delimiting the fricatives, the articulatory onset in /**ps**/ sequences was considerably more tractable than the offsets in /**ps**/ and /**ts**/. Thus we will use this information to at least roughly test whether a substantial difference in interpretation of laryngeal-oral timing patterns would ensue, depending on whether an articulatory or acoustic criterion for fricative onset is chosen[34].

[34]A more output oriented approach might argue that it only makes sense to time laryngeal behaviour relative to the acoustic manifestation of the fricative, since after all the laryngeal constriction is critically involved in shaping the arerodynamic conditions necessary for fricative production.

Combinations with laterals

Unlike for the fricatives, for the lateral /**l**/ it is very easy to use EPG as a basis for segmentation, and there are definite advantages in doing so. The combination of often more apical articulation than for the fricatives together with complete rather than partial closure across the midline results in very clear transitions in the EPG patterns. After inspection of the patterns of the three speakers we used free contacts <=2 as criterion for onset and offset of /**l**/-closure. Because of the typically very fast-changing EPG patterns the choice of criterion is also less critical than for the fricatives; a slightly different criterion would very often result in choice of exactly the same EPG frame.

The criterion for /**l**/-offset often gave a point very slightly before (approx. 1 pitch period) the one that would probably be chosen in a sonagram display (apparent, for example, in the next figure), but since the effect was small and quite systematic the articulatory criterion was maintained. This moreover had the advantage of being easy to apply to the few cases where /**l**/-offset occurred before voicing onset, this being difficult to identify in the sonagram. The offset of /**l**/ is in fact one of the less important segment boundaries. It is necessary in order to determine what proportion of /**l**/ is voiceless, but is not directly required for assessment of laryngeal-oral coordination.

Of more significance is /**l**/-onset, since this also delimits voiceless segments that presumably are directly involved in determining coordination patterns. The different cases will be considered in turn.

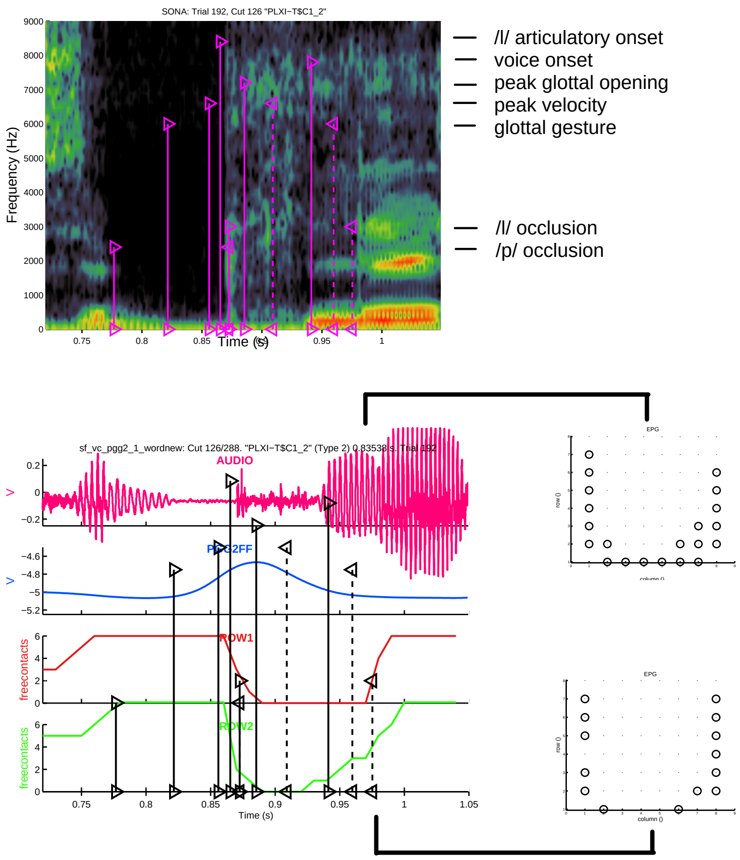


Fig. 4.10: Segmentation example for /**pl**/ in “Plitsche” (subject SF). EPG frames are at articulatorily defined /**l**/−offset (bottom) and one frame before this (top), thus illustrating very fast change in contact pattern

/pl/ and /ʃpl/: These are the simplest cases. An acoustically determined segment boundary was located at the burst of /p/, but in addition the articulatory onset of /**l**/ (typically earlier than the /**p**/-burst) was located in the EPG signal. An illustration of /**pl**/ for speaker SF is shown in the previous figure. It exemplifies the very fast change in EPG pattern at onset and offset of /**l**/ (compare to the fricatives above), and also the articulatorily determined onset of /**l**/ in addition to the segment boundary at the acoustic burst of the initial /p/. Here the articulatorily defined /**l**/ onset is just very slightly before the burst for /**p**/, but in other cases the articulatorily defined boundary could lead the burst by a good deal more; this was especially the case in the /**ʃpl**/ sequences.

/ʃl/: Here the boundary between the /**ʃ**/ and /**l**/ was based entirely on the EPG signal. In the acoustic signal this boundary was not surprisingly generally marked by a considerable reduction in frication amplitude, but the precise location would have been, for example, much less easy to determine for CG than SF. Given the clearness of the EPG signal for all speakers, it was considered to allow a more consistent location to be chosen.

This consonant combination is also unproblematic in the sense that substantial gestural overlap probably does not need to be taken into account when considering the coordination of the fricative with the glottal gesture. Since both /**ʃ**/ and /**l**/ are coronal and competing for the same articulator it is unlikely that the criteria used to delimit /**ʃ**/ result in a badly distorted estimate of its gestural onset and offset.

A representative example for each speaker is shown in the next group figures. Note the widely differing nature and clarity of the sonagraphic pattern, but the clear EPG patterns (note also that articulatory offset for /**l**/ precedes voice onset for RW)

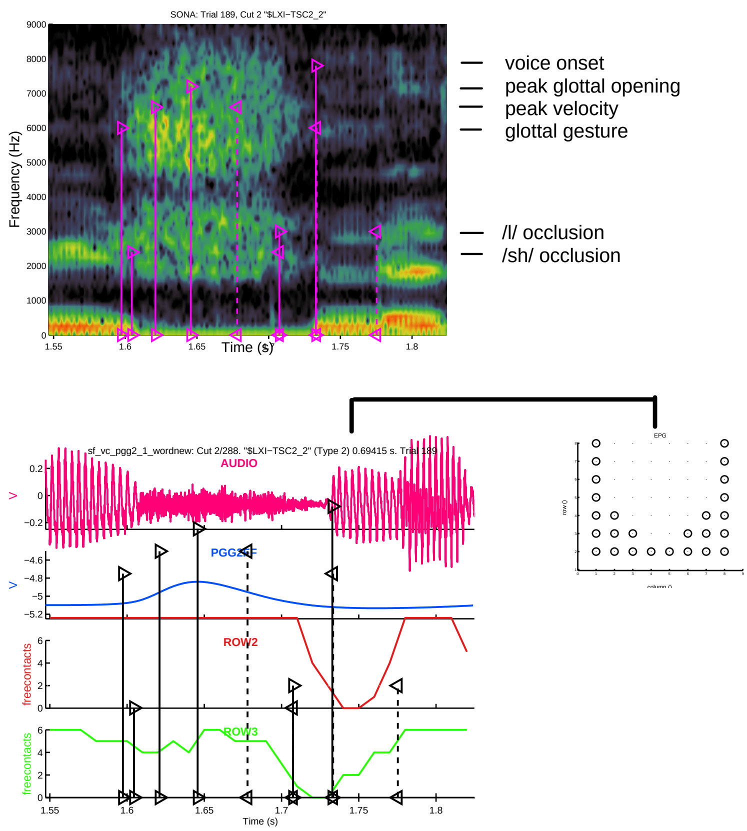


Fig. 4.11: Segmentation example for /**ʃl**/ in "Schlitze" (subject SF). EPG frame is located at about midpoint of /**l**/-occlusion, just after voice onset

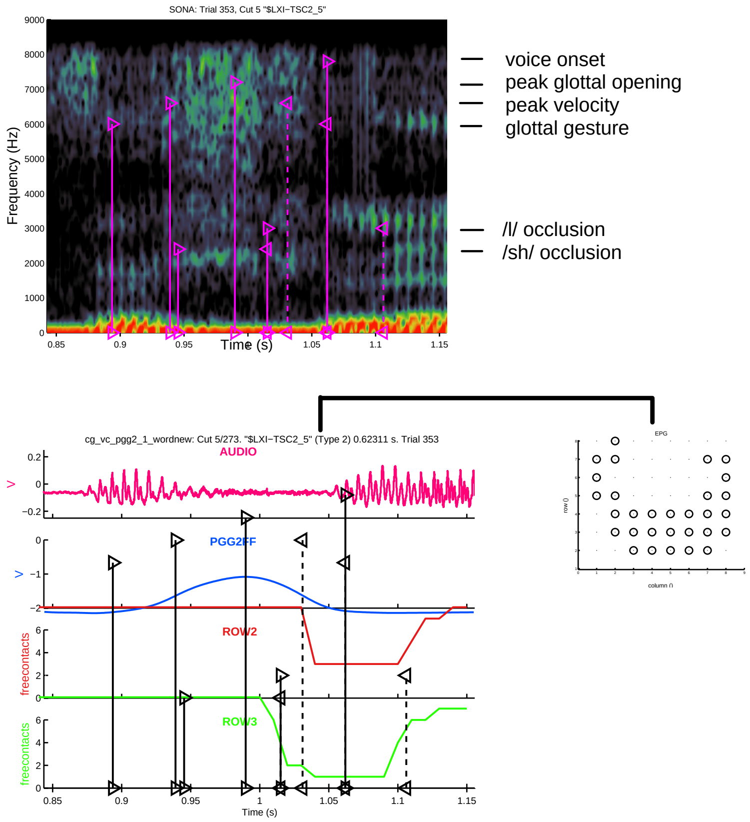


Fig. 4.12: Segmentation example for /ʃl/ in “Schlitze” (subject CG). EPG frame is located in /l/-occlusion, at start of region of minimum free contact in Row 2 and Row 3.

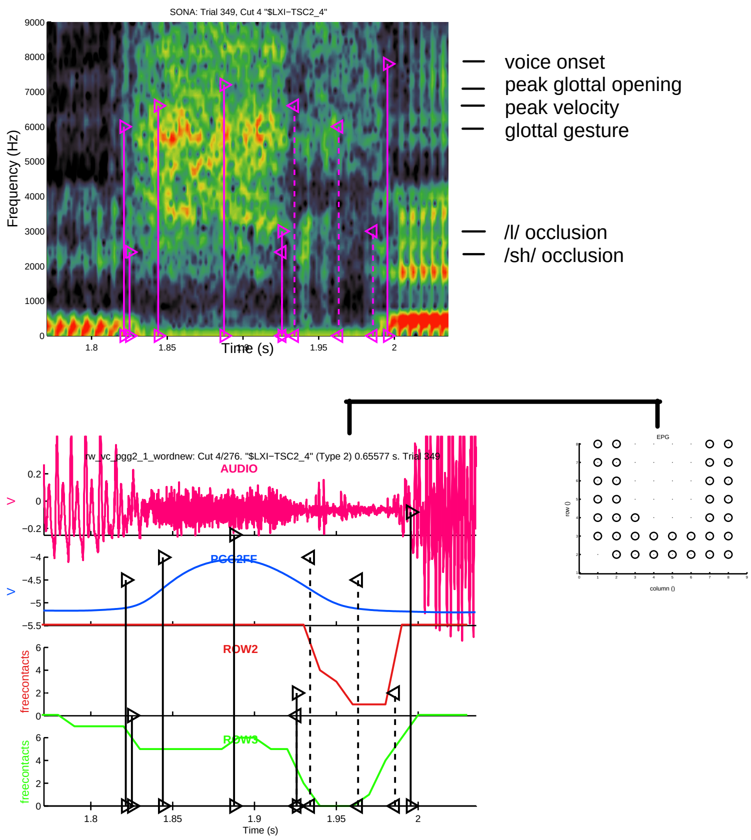


Fig. 4.13: Segmentation example for /ʃl/ in “Schlitze” (subject RW). EPG frame is located at about midpoint of /l/-occlusion, where Row 2 and Row 3 are simultaneously at their respective minimum free contact. Note that voicing onset is slightly later than articulatory offset of the /l/.

/fl/ and /pfl/: These cases require somewhat more discussion. Here, too, the boundary between the fricative and /**l**/ was determined purely on the basis of the EPG criterion. As might be expected this once again often coincided with a marked drop in frication amplitude, though with differences in the acoustic pattern over subjects and tokens. For example, for SF an acoustic boundary was clear in “Flitze” but much less so in “Fluesse” (perhaps an effect of lip rounding/constriction in the latter case). For CG, due to overall poorer quality of the audio recording, determination of any acoustically based boundaries would have been highly problematic. RW showed less a reduction in frication amplitude, but rather an increase in saliency of lower-frequency noise with formation of the /**l**/-constriction (and would also not have allowed very precise temporal determination of boundaries in the acoustics). Thus the EPG signal clearly offered a criterion that could be more consistently applied to all speakers and tokens. However, it is important to note that this is a further case (like /**ps**/ above) where potential patterns of gestural overlap cannot be pinned down exactly. Thus the temporal relationship of the chosen segmentation point to the relaxation of the labial constriction cannot be known precisely. Certainly it did not appear feasible from the acoustics to estimate whether the labial constriction is relaxed before or after the formation of the lingual constriction. This is thus a slightly different situation from the case of labial plosives followed by /**l**/. In those cases the release burst of /**p**/ represents a well-defined point in the life of the labial gesture that can be determined independently of the lingual gesture for /**l**/. We speculate however that relaxation of the constriction for /**f**/ may generally be timed close to formation of constriction for /**l**/: Any lingual constriction will immediately have a major impact on the possibility for generation of frication at the more anterior labiodental location, whereas overlap of the lingual and labial constriction during the plosive will have essentially no effect on the acoustics in the occlusion phase of the plosive. This is a case of where simple articulatory data is still very much needed (we are not aware of relevant information in the literature).

The range of patterns involved is shown in the next group of figures (one page per subject, two examples of /**fl**/ sequences per page).

This also concludes this chapter on experimental methodology. In the following chapter on the results it is worth pointing out that no specific movement patterns will be shown, so it may be useful to refer back to the large number of examples in this chapter on occasion.

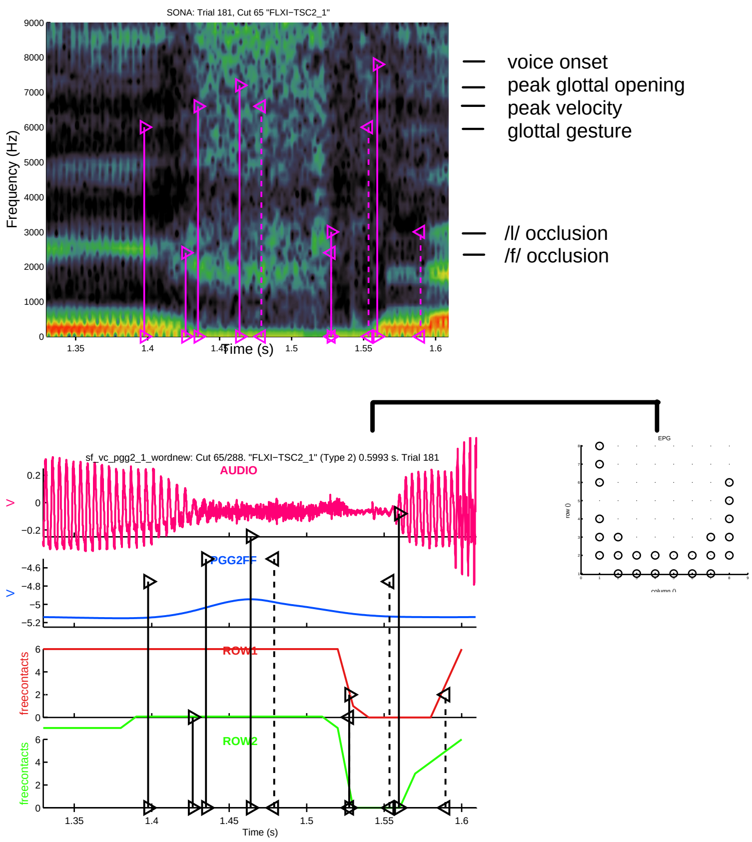


Fig. 4.14: Segmentation example for /**fl**/ in "Flitze" (subject SF). EPG frame is located in /**l**/-occlusion, just after start of region of minimum free contact in Row 1 and Row 2

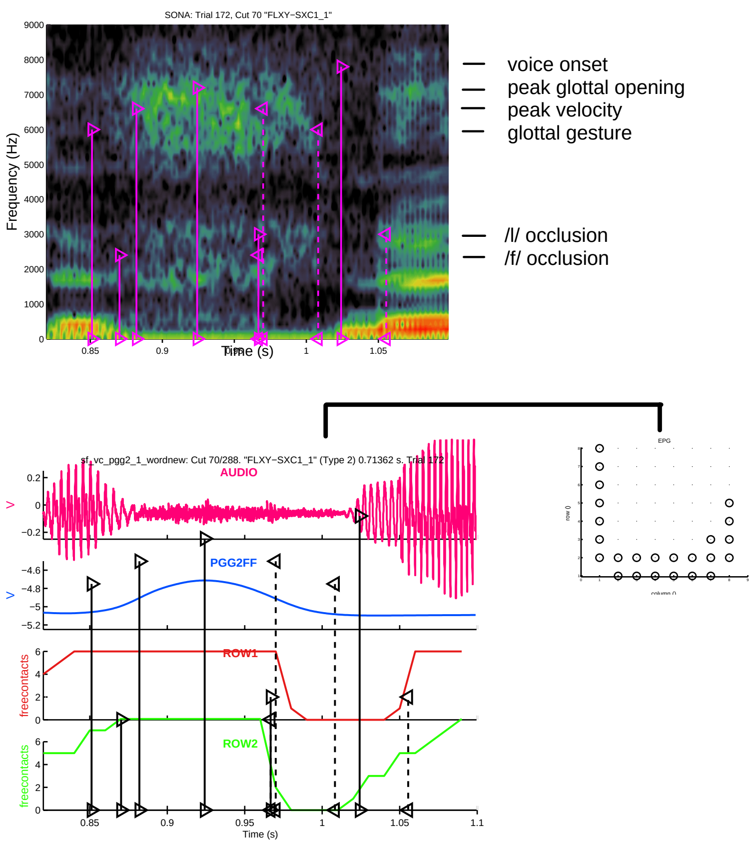


Fig. 4.15: Segmentation example for /**fl**/ in "Fluesse" (subject SF). EPG frame is located in /**l**/-occlusion,in region of minimum free contact in Row 1 and Row 2

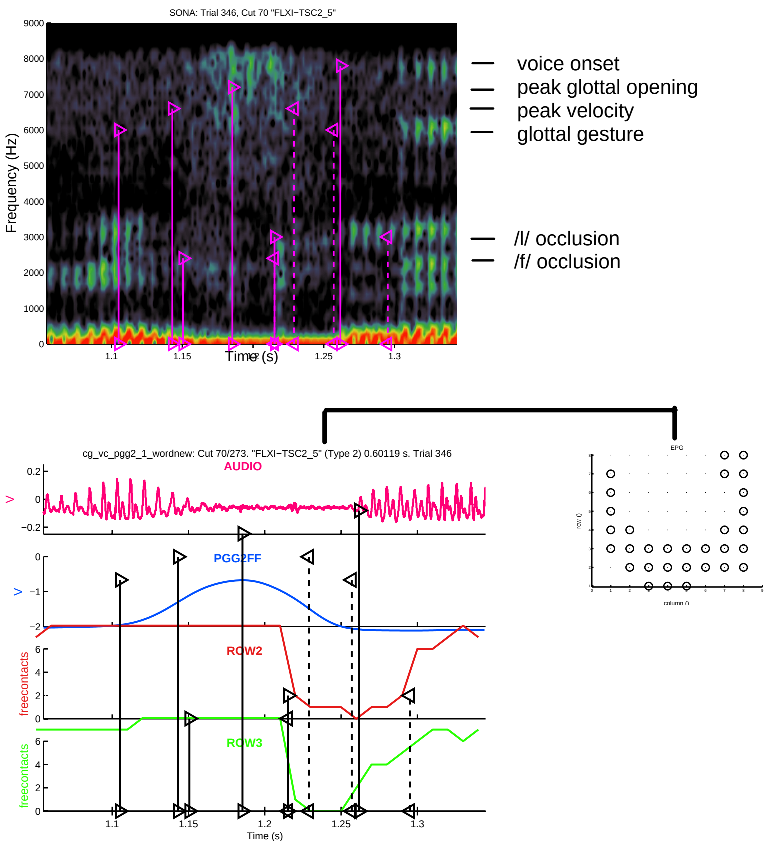


Fig. 4.16: Segmentation example for /**fl**/ in "Flitze" (subject CG). EPG frame is located in /**l**/-occlusion, at about midpoint of minimum free contact in Row 3

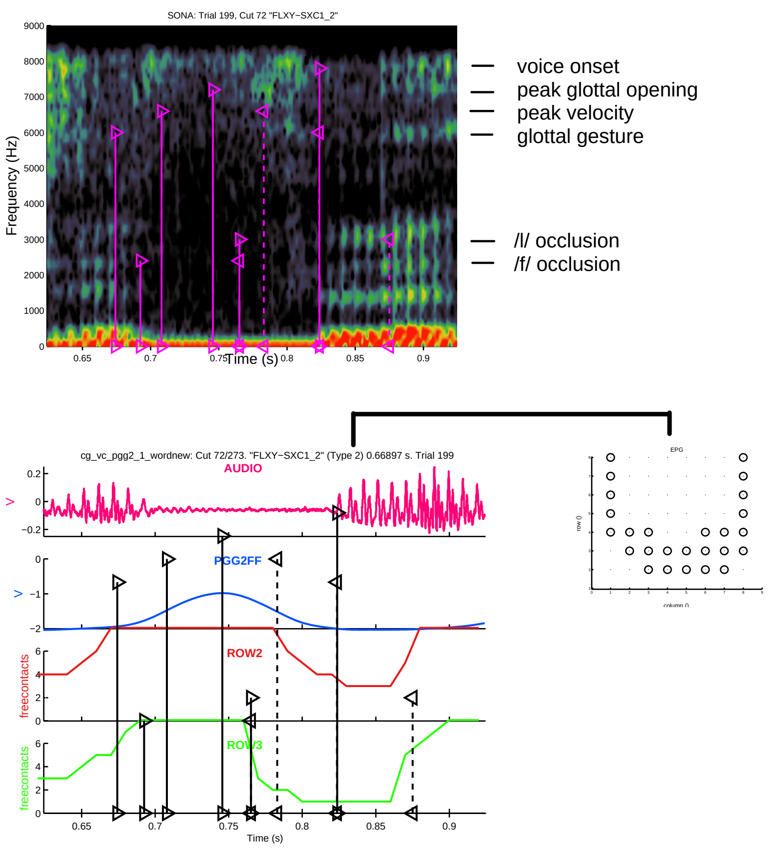


Fig. 4.17: Segmentation example for /**fl**/ in "Fluesse" (subject CG). EPG frame is located in /**l**/-occlusion, at about midpoint ofminimum free contact in Row 3

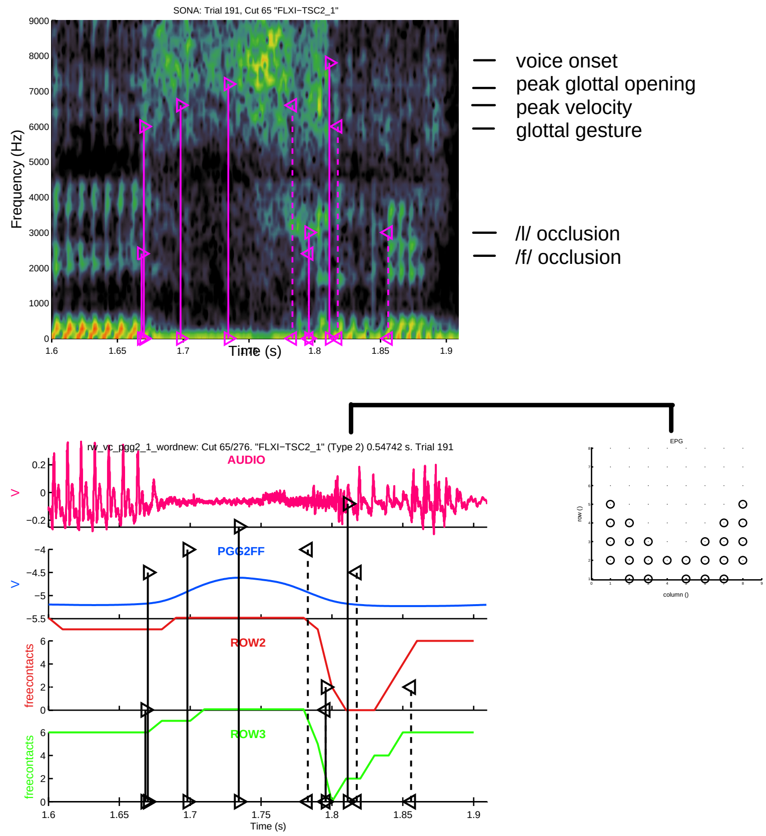


Fig. 4.18: Segmentation example for /**fl**/ in "Flitze" (subject RW). EPG frame is located in /**l**/-occlusion, at first frame with zero free contact in Row 2

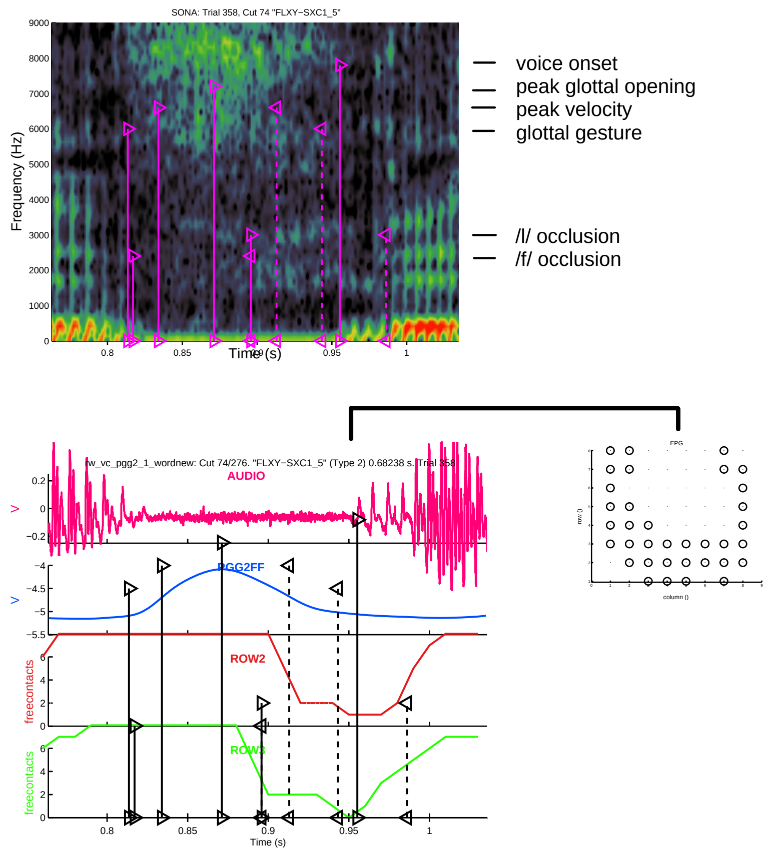


Fig. 4.19: Segmentation example for /**fl**/ in "Fluesse" (subject RW). EPG frame is located in /**l**/-occlusion at frame with simultaneous minimum free contact in Row 2 and Row 3

5 Results: Background and Statistical Overview

In this chapter, some general background to the results will be given, and, in particular, a complete tabulation of the statistical results will be given. These tables have been collated here to make it easier, while progressing through the detailed results of the next chapter, to quickly refer to the basic results of any other part of the investigation.

In the next chapter, the results will be presented in three main sections, essentially built up around a series of comparisons:

1. Single fricatives and single plosives. In addition to comparing fricatives and plosives, the influence of place of articulation in plosives will be examined
2. Combinations of fricative and plosive. Fricative-plosive clusters will be compared with single fricatives. Plosive-fricative combinations will be compared explicitly with single plosives, but will also be used to give a preliminary overview of much of the material up to that point by comparing plosive-fricative, fricative-plosive, single fricative and single plosive
3. Clusters containing /**l**/, i.e /**pl**/, fricative-/**l**/, /**pfl**/, and /**ʃpl**/ These will be compared with the corresponding consonant or consonant group without /**l**/

Before turning to the detailed results in the next chapter, the present chapter will give, firstly, one preliminary result in overview, namely the effect of loud vs. normal speech volume. Secondly, the background to the statistical procedures used to analyze the detailed results will be given, together with an overview of what kind of information is contained in the tables. This is followed, thirdly, by the statistical tables themselves.

5.1 Comparison of glottal activity for loud vs. normal speech

As already mentioned in the presentation of the experimental procedures, loud vs. normal speech is treated within the framework of the present analyses not as an issue in its own right, but rather as a means of increasing the amount of material in the available corpus that can be analyzed with respect to the central issues of laryngeal-oral coordination (and of providing further insight into the robustness of any effects observed). Since speech volume thus crops up as an independent variable in many of the analyses below without itself being of central concern it seems worthwhile giving an overview of its effect here. This will make it possible to avoid some redundancy in discussion of the detailed results in the next chapter in cases where the effect of speech volume simply follows the basic trends visible in the next figure.

The figure shows a scatter-plot of peak glottal opening amplitude versus glottal gesture duration. Each data point is the average over the (usually) five repetitions of each word form and is coded with respect to speech volume. The pattern is very consistent over the 3 speakers: peak glottal opening amplitude is clearly larger for loud speech (for all speakers about double the amplitude for loud vs. normal speech), whereas the ***duration*** of the glottal gesture is essentially the same for both volume conditions. This is a convenient result for the investigation of timing patterns below, since it means that identification of timing patterns is not complicated by substantial differences in speech rate between volume conditions. A further baseline result than can be noted here is that glottal gesture durations are obviously very similar for subjects SF and CG, while RW has a tendency to slightly shorter durations.

With regard to glottal amplitude it will be recalled from the methods section above that measurements of glottal opening amplitude made by transillumination have to be interpreted with caution. The difference between loud and normal speech is obviously an example of a very robust effect that transcends any of these methodological problems and thus provides a useful background against which to view the potentially more subtle sound-specific effects in the main results.

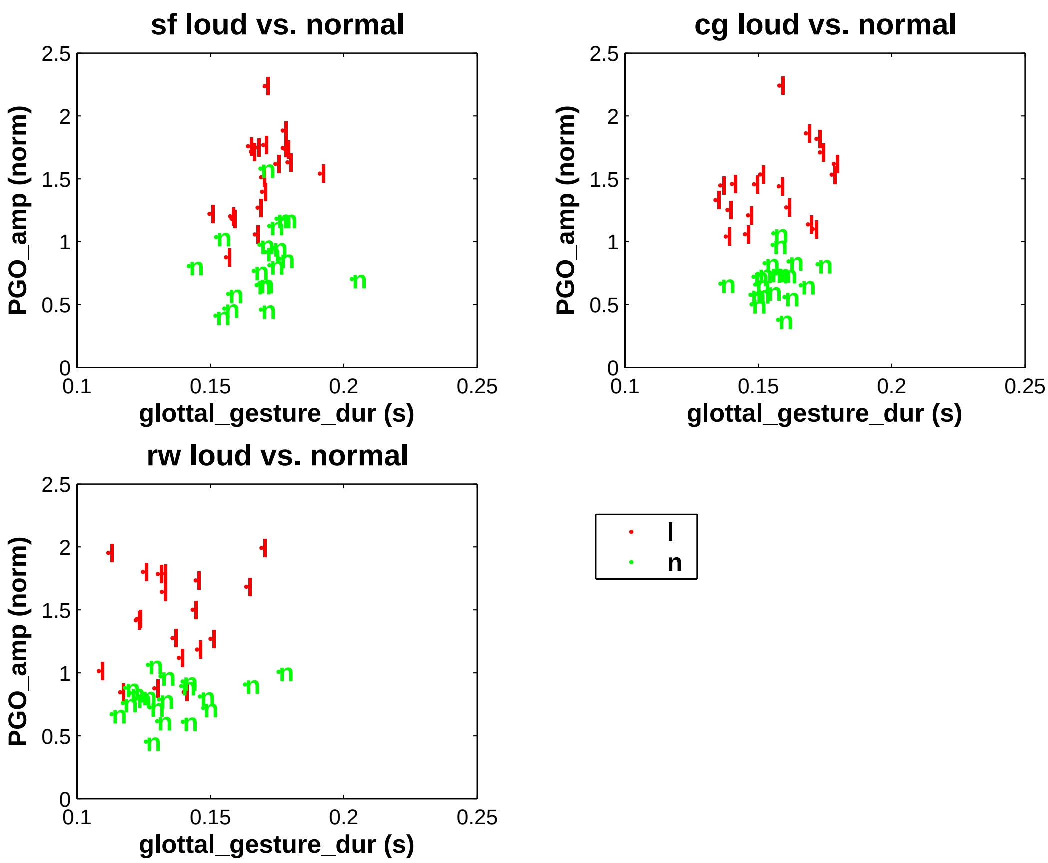


Fig. 5.1: Scatter plot of peak glottal opening amplitude and glottal gesture duration separately for each subject. Plot symbol and colour indicates speech volume condition (loud vs. normal)

5.2 Background to statistical procedures

The main technique for comparing timing patterns based on durational measurements across the various linguistic categories was analysis of variance[35]. However, for glottal opening amplitude there is much to be said for using a non-parametric technique (risk of large outliers due to shifts in endoscope position, possible nonlinear relationship between glottal opening and measured light). For durational measurements, too, there are a number of potential situations where a non-parametric procedure might be advisable: some comparisons may only be based on one or two word-forms each, resulting in low values for n (i.e maximum n=5 if only one word-form is involved); some comparisons (e.g all material with single initial /**p**/ vs. /**t**/) involve different numbers of word-forms per condition (three for /**p**/, seven for /**t**/) that are not perfectly matched for example with respect to following vowel. Thus departures from the distributional assumptions of analysis of variance could occur. In order to define an approach that could be applied consistently across all analyses we thus decided to back up ANOVA in all cases with the Kruskal-Wallis nonparametric analysis of variance procedure (essentially an ANOVA is performed after replacing the measurement values by their rank position), using the Kruskal-Wallis results as the basis for any pair-wise comparisons made[36].

5.2.1 Typographical conventions in statistical tables

The tables collated in the next section use the following typographical conventions to indicate the level of significance of the results. The basic scheme is:

lowercase, UPPERCASE, **UPPERCASE BOLD** to indicate significance at $p < 0.05$, 0.01 and 0.001, respectively.

Thus the symbols 'x', 'X' and '**X**' indicate correspondingly significant results in the expected direction. Very occasionally the small lowercase symbol 'x' is used to indicate results that just miss the $p < 0.05$ level.

The symbols 'o', 'O' and '**O**' indicate significant results contrary to the expected direction.

In two-factorial ANOVAs the italic symbols '*i*', '*I*' and '***I***' are used to indicate interactions. When an interaction occurs the corresponding italic symbol is placed in the table cell corresponding to both factors involved.

Each table of ANOVA results includes a column indicating the expected direction of the effect. Sometimes this must be taken with a pinch of salt. Essentially, this column shows how significant results indicated with the notation above are to be interpreted. For some of the parameters none

---

[35] ANOVA was performed using the anovan function of the statistics toolbox version 4.0 available in MATLAB version 6.5. This ANOVA function implements the general linear model (GLM) procedure, allowing the use of unbalanced designs. The Type 3 sums of squares model was used.

[36] This is available as an option in the multcompare function in the statistics toolbox. The documentation recommends use of the Tukey-Kramer procedure for determining critical values for multiple comparisons when using Kruskal-Wallis. A criterion significance level of p=0.01 was used for reporting siginificant pair-wise comparisons.

of the hypotheses discussed in earlier chapters provide a specific expectation. In this case, the 'expected' direction is simply chosen to follow the predominant direction actually found, and the expected direction is placed in brackets (in cases where there was no specific expectation and no significant results were found then the expected direction is left blank). Note that in two-factorial ANOVAs an expected direction has only been given for the factor of predominant interest. For example, in the comparison of single plosives and fricatives the ANOVA includes volume level as a second factor, but no expected direction is given for it in the table. For details of effects related to volume level it will be necessary to refer to the figures given in the next chapter.

In all of the sub-sections below in which a two-factor ANOVA is employed then in addition to the main ANOVA table there is also a table giving the results of paired comparisons for the main pairs of interest (e.g plosive vs. fricative, matched for volume level). These paired comparisons are based on the Kruskal-Wallis nonparametric analysis of variance, and use a criterion level of p=0.01 for inclusion in the table (there is also one case of a one-factor ANOVA with more than two levels; here, too an additional table of pairwise comparisons is given).

In the case of the one-factor ANOVAs, the results of the nonparametric ANOVA are shown in addition to the normal ANOVA in the main table (for the sake of succinctness only explicitly shown when the two procedures lead to different results).

5.2.2 Overview of Tables

In the list below the numbered headings indicate exactly the sections and sub-sections used for presentation of the detailed results in the next chapter. Each bullet in the list corresponds to one table in the next section of the present chapter. The right-justified cross-references in the list below are repeated in the top left cell of each ANOVA table and indicate where in the following chapter the results are presented and discussed in detail.

1. Single plosives and fricatives
   1.1 Plosives vs. fricatives — Chapter 6, pp. 75 - 80
   - 2-factor ANOVA: plosive vs. fricative and normal vs. loud
   - Nonparametric pairwise comparisons of plosive vs. fricative, matched for volume level

   1.2 Place of articulation in plosives — Chapter 6, pp. 81 - 87
   - 2-factor ANOVA: /**p**/ vs. /**t**/, and normal vs. loud
   - Nonparametric pairwise comparisons of /**p**/ vs. /**t**/ matched for volume level
2. Combinations of fricatives and plosives
   2.1 Fricative-plosive clusters compared to single fricatives — Chapter 6, pp. 89 - 97
   - 1-factor ANOVA (parametric and nonparametric): /**ʃ**/ vs. /**ʃt**/ vs. /**ʃp**/
   - Nonparametric pairwise comparisons of /**ʃ**/ vs. /**ʃt**/ and /**ʃ**/ vs. /**ʃp**/

   2.2 Plosive-fricative sequences compared to single plosives — Chapter 6, pp. 101 - 109
   - 2-factor ANOVA: plosive-fricative vs. single plosive, and normal vs. loud. Results tabulated for both complete and reduced selection of material
   - Nonparametric pairwise comparisons of plosive-fricative vs. single plosive, matched for volume level. Results tabulated for both complete and reduced selection of material
3. Consonant combinations with /**l**/
   3.1 /**pl**/ clusters compared to single /**p**/ — Chapter 6, pp. 112 - 117
   - 1-factor ANOVA (parametric and nonparametric): /**pl**/ vs. /**p**/

   3.2 Fricative-/**l**/ clusters compared to single fricative — Chapter 6, pp.121 -126
   - 2-factor ANOVA: /**ʃ**/ vs. /**f**/, and singleton vs. /**l**/-cluster
   - Nonparametric pairwise comparisons of fricative-lateral vs. single fricative, matched for place of articulation

   3.3 /**pfl**/ clusters compared to /**pf**/ — Chapter 6, pp.129 -133
   - 1-factor ANOVA (parametric and nonparametric): /**pfl**/ vs. /**pf**/

   3.4 /**ʃpl**/ clusters compared to /**ʃp**/ — Chapter 6, pp.135 -139
   - 1-factor ANOVA (parametric and nonparametric): /**ʃpl**/ vs. /**ʃp**/. Results tabulated for both complete and reduced selection of material

5.3 Tabulation of statistical results

The order of the parameters in the tables follows the order of presentation in the corresponding section of the next chapter.

5.3.1 Single plosives and fricatives

Plosives vs. Fricatives

Chapter 6, pp. 75 - 80	Expected trend C1CAT	C1CAT			VOLCAT		
		SF	CG	RW	SF	CG	RW
1 ges_dur	F > P	**X**	**X *I***	**X**		***I***	
2 PGO_amp	F > P	**X**		O	**X**	**X**	**X**
3 occ_to_abd	P > F	**X**	**X**	**X**	X		
4 abd_vs_add	P > F	**X**	**O**			**X**	x
5 occ_to_PGO	P > F	**X**	**X**	**X**	X	**X**	x
6 mean_abd_vel	F > P	**X**		**O** *i*	**X**	**X**	**X** *i*

Table. 5.1: Summary of ANOVA results for factors C1CAT (plosive vs. fricative) and VOLCAT (normal vs. loud). Abbreviations of parameter names in first column: **1**. Glottal gesture duration; **2**. Amplitude of peak glottal opening; **3**. Duration of interval from start of oral occlusion to start of glottal abduction; **4**. Ratio of glottal abduction to adduction duration; **5**. Duration of interval from start of oral occlusion to peak glottal opening; **6**. Average glottal abduction velocity

	SF	CG	RW
1 ges_dur	(FN > PN)	FL > PL	FL > PL; FN > PN
2 PGO_amp	FN > PN; (FL > PL)		
3 occ_to_abd	FL > PL; FN > PN	FL > PL; FN > PN	FL > PL; FN > PN
4 abd_vs_add	PL > FL		
5 occ_to_PGO	PL > FL; PN > FN	PL > FL; PN > FN	PL > FL; PN > FN
6 mean_abd_vel	FN > PN; (FL > PL)		

Table. 5.2: Summary of nonparametric pairwise comparisons reaching $p<0.01$ for fricative vs. plosive, matched for volume level (small font in brackets just misses criterion level). ‘F’=Fricative, ‘P’=Plosive, ‘L’=Loud, ‘N’=Normal.

Place of articulation in plosives

Chapter 6, pp. 81 - 87	Expected trend C1	C1			VOLCAT		
		SF	CG	RW	SF	CG	RW
1. ges_dur	(t > p)	x				X	
2. PGO_amp	(p > t)	X			**X**	**X**	**X**
3. asp_dur	t > p	**X**	X		x	X	**X**
4. occ_dur	p > t				**X**		x
5. PGO_to_rel	p > t	**X** *i*	**X**		*i*		x
6. occ_to_abd	(t > p)	x	X	**X**	X		

Table. 5.3: Summary of ANOVA results for factors C1 (/p/ vs. /t/) and VOLCAT (normal vs. loud). Abbreviations of parameter names in first column: **1**. Glottal gesture duration; **2**. Amplitude of peak glottal opening; **3**. Duration of aspiration; **4**. Duration of oral occlusion; **5**. Duration of interval from peak glottal opening to release of the oral occlusion; **6**. Duration of interval from start of oral occlusion to start of glottal abduction

	SF	CG	RW
1. ges_dur			
2. PGO_amp			
3. asp_dur	pN > tN		
4. occ_dur			(tN > pN)
5. PGO_to_rel	pN > tN		
6. occ_to_abd			tN > pN

Table. 5.4: Summary of nonparametric pairwise comparisons reaching p<0.01 for /**p**/ vs. /**t**/, matched for volume level (pN = /**p**/, Normal volume. tN = /**t**/, Normal volume)

5.3.2 Combinations of fricatives and plosives

Fricative-plosive clusters compared to single fricatives

Chapter 6, pp. 89 - 97	Expected trend	SF	CG	RW
1 ges_dur	ʃP > ʃ			**X**
2 PGO_amp	ʃP > ʃ	**O** (O)	X (x)	
3 occ1_dur	ʃ > ʃP	**X**	X	**X**
4 PGO_relpos	ʃ > ʃP	**X**	**X**	**X**
5 abd_vs_add	ʃ > ʃP	**X**		**X**
6 abd_dur	(ʃ > ʃP)	**X**		ns (x)
7 add_dur	(ʃP > ʃ)	X		**X**
8 rel_to_glott_end	ʃ > ʃP	**X**	**X**	x (X)

Table. 5.5: Overview of one-factor ANOVA results using /**ʃ**/ vs. /**ʃt**/ vs. /**ʃp**/ as levels. Results in brackets are from Kruskal-Wallis analysis of variance where different from parametric results. Expected trend in second column refers to single fricative vs. fricative-plosive (‘P’ stands for ‘/p, t/’). Abbreviations of parameter names in first column: **1**. Glottal gesture duration; **2**. Amplitude of peak glottal opening; **3**. Oral occlusion duration of fricative segment (/S/); **4**. Relative position of peak glottal opening in the fricative segment; **5**. Ratio of glottal abduction to adduction duration; **6**. Duration of glottal abduction; **7**. Duration of glottal adduction; **8**. Duration of interval from release of last occlusion to end of glottal adduction

	SF	CG	RW
1 ges_dur			**ʃt** > **ʃ**
2 PGO_amp	(**ʃp**, **ʃt**) < **ʃ**		
3 occ1_dur	**ʃp** < **ʃ**	**ʃp** < **ʃ**	(**ʃp**, **ʃt**) < **ʃ**
4 PGO_relpos	**ʃp** < **ʃ**	**ʃp** < **ʃ**	(**ʃp**, **ʃt**) < **ʃ**
5 abd_vs_add	(**ʃp**, **ʃt**) < **ʃ**		(**ʃp**, **ʃt**) < **ʃ**
6 abd_dur	**ʃp** < **ʃ**		
7 add_dur			(**ʃp**, **ʃt**) > **ʃ**
8 rel_to_glott_end	**ʃp** < **ʃ**	(**ʃp**, **ʃt**) < **ʃ**	**ʃp** < **ʃ**

Table. 5.6: Overview of cases where /**ʃp**/ and/or /**ʃt**/ differ from /**ʃ**/ in Kruskal-Wallis pairwise comparisons at p<0.01

Plosive-fricative sequences compared to single plosives

Chapter 6, pp. 101 - 109	Expected trend C2CAT	C2CAT			VOLCAT		
		SF	CG	RW	SF	CG	RW
1 ges_dur	PF > P	**X** (X)	**X** *i*	**X**		*i*	
2 PGO_amp	PF > P	**X** *I (i)*	**X** (X)		**X** *I (i)*	**X**	**X**
3 abd_vs_add	PF > P		**X**	**X**		**X**	x (ns)
4 occ1_dur	P > PF	**X**		**X** (x)	**X**		x (ns)
5 PGO_to_rel	P > PF	**X**	X (ns)	**X**	x (ns)	x (ns)	
6 occ_to_abd		*i (ns)*	*I*		x *i (ns)*	*I*	

Table. 5.7: Significance levels in two-way ANOVAs for plosive-fricative vs single plosive (C2CAT) and normal vs. loud volume (VOLCAT). Brackets indicate results for reduced material where different from complete material (see Chapter 6, p. 99, for details). Abbreviations of parameter names in the first column are : **1**.Glottal gesture duration; **2**. Amplitude of peak glottal opening; **3**. Ratio of duration of glottal abduction to adduction; **4**. Duration of occlusion of plosive segment; **5**. Duration of interval from peak glottal opening to release (of plosive segment); **6**. Duration of interval from start of plosive occlusion to start of glottal abduction

	Pairwise comparisons		
	SF	CG	RW
1 ges_dur	PF_N > P_N (none)	PF_L > P_L (none)	PF_L > P_L; PF_N > P_N (almost PF_L > P_L)
2 PGO_amp			
3 abd_vs_add			PF_N > P_N; (none)
4 occ1_dur	PF_L < P_L		PF_L < P_L; (none)
5 PGO_to_rel	PF_L < P_L PF_N < P_N	PF_N < P_N (none)	PF_L < P_L PF_N < P_N
6 occ_to_abd			

Table. 5.8: Results of pairwise comparisons of plosive-fricative vs single plosive, matched for volume level. PF_N = Plosive-fricative (normal); P_N = Single plosive (normal); PF_L = Plosive-fricative (Loud); P_L = Single plosive (loud). Brackets indicate results for reduced material where different from complete material. See previous table for full names of parameters.

5.3.3 Consonant combinations with /l/

*/**pl**/ clusters compared to single /**p**/*

Chapter 6, pp. 112 - 117	Expected trend	SF	CG	RW
1 ges_dur	pl > p		ns (x)	x
2 PGO_amp	pl > p	o		x (ns)
3 occ1_dur	p > pl		X (x)	
4 PGO_to_rel	p > pl	ns (x)	**X**	x
5 occ_to_abd	(pl > p)	X		X
6 VOT	pl > p	X (x)	**X** (X)	**X**
7 add_to_vo	pl > p			
8 voiceless_dur	pl > p	X	ns (x)	**X**

Table. 5.9: Results of ANOVA for /p/ vs. /pl/. Results of nonparametric test in brackets where different from ANOVA. Abbreviations of parameter names in first column: **1**. Glottal gesture duration; **2**. Amplitude of peak glottal opening; **3**. Oral occlusion duration of /p/ segment; **4**. Duration of interval from peak glottal opening to release of plosive; **5**. Duration of interval from start of plosive to start of glottal abduction; **6**. Duration of interval from release of /p/ occlusion to voice onset; **7**. Duration of interval from end of glottal adduction to voice onset; **8**. Total duration of voicelessness

*Fricative-/**l**/ clusters compared to single fricative*

Chapter 6, pp.121 - 126	Expected trend C2	C1			C2		
		SF	CG	RW	SF	CG	RW
1 fric_dur	F > FL	x		X	**X**	**X**	**X**
2 ges_dur	FL > F			X *i*	**O**		*i*
3 PGO_amp	FL > F			**X**	**O**		
4 voiceless_dur	FL > F		x	**X**	**O**		X
5 glott_end_to_VO	FL > F		**X** *i* [ʃ < f]	**X** *I* [ʃ > f]		*i*	x *I*
6 occ_to_abd		**X**					
7 PGO_rel pos	F > FL					**X**	**X**
8 occ_end_relpos	F > FL				**X**	**X**	**X**
9 abd_vs_add	F > FL				x		x

Table. 5.10: Results of two-way ANOVA with factors C1 (/**ʃ**/ vs. /**f**/) and C2 (singleton vs. **l**-cluster). ‘F’ in second column refers to the fricative category, not to phoneme /**f**/. Abbreviations of parameter names in first column: **1**. Oral occlusion duration of fricative segment; **2**. Glottal gesture duration; **3**. Amplitude of peak glottal opening; **4**. Total duration of voicelessness; **5**. Duration of interval from end of glottal adduction to voice onset; **6**. Duration of interval from start of fricative to start of glottal abduction; **7**. Relative position of peak glottal opening in the fricative segment; **8**. Relative position of end of fricative segment within the glottal gesture; **9**. Ratio of glottal abduction to adduction duration

Pairwise comparisons	SF	CG	RW
1 fric_dur	fl < f	fl < f	
2 ges_dur			
3 PGO_amp	fl < f		
4 total_voiceless			
5 glott_end_to_VO			
6 occ_to_abd			
7 PGO_relpos		fl < f	fl < f
8 occ_end_relpos	fl < f	fl < f	fl < f
9 abd_vs_add			

Table. 5.11: Results of pairwise comparisons of fricative-lateral vs single fricative (matched for place of articulation). Same abbreviations of parameter names as previous figure.

*/**pfl**/ clusters compared to /**pf**/*

Chapter 6, pp.129 - 133	Expected trend	SF	CG	RW
1 p_dur	pf > pfl	X	X	
2 f_dur	pf > pfl	x	X	
3 ges_dur	pfl > pf		ns (x)	ns (x)
4 PGO_amp	pfl > pf			
5 voiceless_dur	pfl > pf	x		x
6 glott_end_to_VO	pfl > pf			**X** (X)
7 PGO_relpos	pf > pfl	**X**	ns (x)	
8 occ2_end_relpos	pf > pfl	**X**	**X**	X

Table. 5.12: Statistical results of /pf/ vs. /pfl (results of nonparametric test in brackets; only given where different from parametric test). Abbreviations of parameter names in first column: **1**. Oral occlusion duration of plosive segment /p/; **2**. Oral occlusion duration of fricative segment /f/; **3**. Glottal gesture duration; **4**. Amplitude of peak glottal opening; **5**. Total duration of voicelessness; **6**. Duration of interval from end of glottal adduction to voice onset; **7**. Relative position of peak glottal opening in the fricative segment; **8**. Relative position of end of fricative segment within the glottal gesture

*/**ʃpl**/ clusters compared to /**ʃp**/*

Chapter 6, pp.135 -139	Expected trend	SF		CG		RW	
		all	best pair	all	best pair	all	best pair
1 S_dur	Sp > Spl	ns (o)					
2 p_dur	Sp > Spl	X (x)					
3 ges_dur	Spl > Sp	**X** (X)	x	**X**	X		
4 PGO_amp	Spl > Sp			x (ns)			x
5 voiceless_dur	Spl > Sp						
6 PGO_relpos	Sp > Spl		X	x	X	**X** (X)	ns (x)
7 occ2_end_relpos	Sp > Spl	**X** (X)	**X** (X)	**X** (X)	X		
8 abd vs add	Sp > Spl					o	

Table. 5.13: Statistical results of /**ʃp**/ vs. /**ʃpl**/ (results of nonparametric test in brackets; only given where different from parametric test). Results in column marked "best pair" based on /Sp/ word-form matched for corpus part and position in carrier phrase with /Spl/ word-form. Abbreviations of parameter names in first column: **1**. Oral occlusion duration of fricative segment /S/; **2**. Oral occlusion duration of plosive segment /p/; **3**. Glottal gesture duration; **4**. Amplitude of peak glottal opening; **5**. Total duration of voicelessness; **6**. Relative position of peak glottal opening in the fricative segment; **7**. Relative position of release of second occlusion (plosive /p/) within the glottal gesture; **8**. Ratio of glottal abduction to adduction duration

6 Results in detail

6.1 Results 1: Single plosives and fricatives

In the first part of this section glottal behaviour for single plosives and fricatives will be compared; in the second part, single plosives will be examined for place of articulation effects.

6.1.1 Plosives vs. Fricatives

This comparison is of interest in its own right, but moreover forms necessary background for consideration of consonant sequences in later sections. The main expectation one would have from the literature reviewed above is that the glottal gesture is more vigorous for fricatives, in particular in the abduction phase.

Overview of temporal structure

We set the scene with a graphical overview of the temporal structure of these sounds. Figures of this kind will be used to provide basic orientation throughout the results section. One of their main purposes is to make the detailed discussion of specific parameters easier to follow, so a general description of the temporal features that can be gleaned from such figures will be given here.

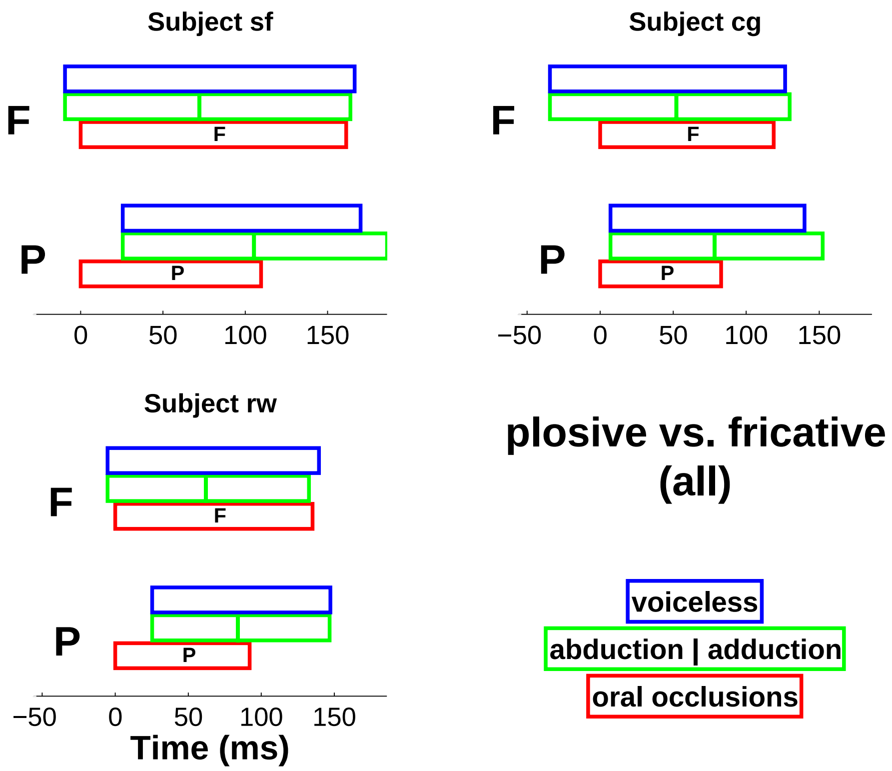


Fig. 6.1: Comparison of temporal structure of plosives and fricatives. For details see text

The bottom red bars indicate the durations of oral occlusions and are labelled either with a specific phoneme label (if lower case), or, as here with upper case letters indicating a sound category (**P**losive vs. **F**ricative). In the present figure, the start of the occlusion phase has been used as the line-up point, i.e this corresponds to zero on the time axis. The middle green bar indicates the time-course of the glottal gesture, divided into abduction phase and adduction phase. Accordingly the green vertical division usually close to the middle of the bar indicates the time instant of peak glottal opening. The main purpose of the blue bar at the top is to indicate the time of voicing onset at the end of the consonant, as determined from the acoustic signal; this occurs at the right end of the bar. These bars are labelled “voiceless”, i.e they give an idea of the period over which no vocal fold vibration occurs. In fact, this is not strictly accurate: the start of the voiceless phase has simply been set identical with the start of glottal abduction, i.e no attempt was made in the signals to determine the offset of vocal fold vibration. Because of the acoustic environment of the recordings this would have been difficult and inaccurate, and none of the questions in which we were interested depended crucially on having this information. As examples of the information on temporal structure that can be derived from this figure one could note the following (details of the most revealing parameters, with statistical tests, discussion and interpretation are given below):

- Shorter oral occlusion duration for plosives than fricatives
- Start of glottal abduction earlier relative to start of oral occlusion for fricatives
- Peak glottal opening roughly at midpoint of oral occlusion for fricatives, and close to burst for plosives

As an example of a more subtle point:

- Voicing onset is somewhat earlier relative to termination of glottal adduction for plosives than fricatives.

Detailed results for individual parameters

Discussion in this section will focus on the following parameters:

- Glottal gesture duration
- Amplitude of peak glottal opening
- Duration of interval from start of oral occlusion to start of glottal abduction
- Ratio of glottal abduction to adduction duration
- Duration of interval from start of oral occlusion to peak glottal opening
- Average glottal abduction velocity

The first two figures in this section that detail specific parameters provide basic information on gesture duration and peak opening amplitude.

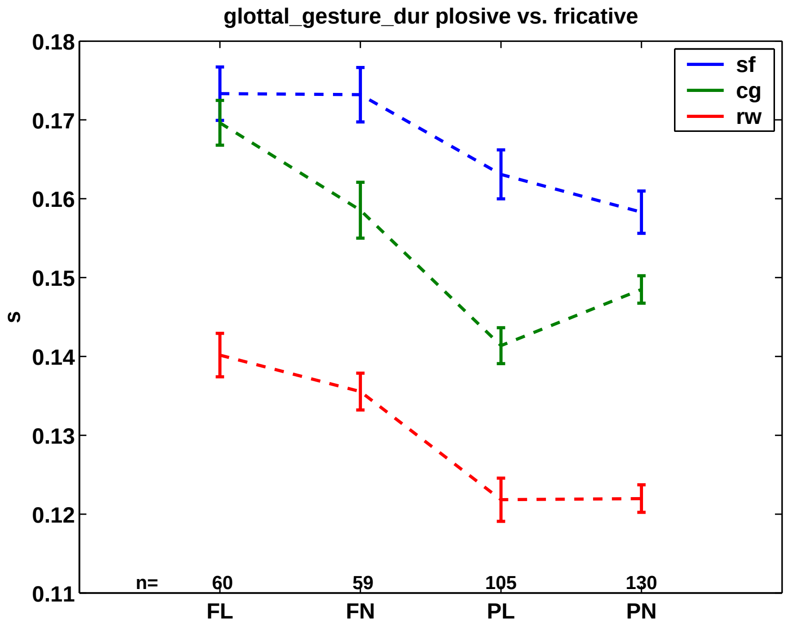


Fig. 6.2: Mean of glottal gesture duration, fricatives vs. plosives. FL= Fricatives, Loud; FN = Fricatives, Normal; PL= Plosives, Loud; PN = Plosives, Normal. Error bars show standard error of mean. Indicated n per condition is total over all 3 subjects.

The bulk of the results will be presented by means of figures of this kind, so a word is in order about some of the conventions used in these displays. The labels on the abscissa should always be interpreted with respect to the explicit information in the figure legend. Generally, capital letters refer to sound categories rather than individual phonemes (in lower case). Thus ‘F’ here refers to the category ‘Fricative’, not the phoneme /**f**/ (and ‘P’ to ‘Plosive’, not to /**p**/). These labels will also be used to identify categories in a posteriori pair-wise comparisons (see background to statistical procedures in the previous chapter). Since n per condition can vary quite widely it was decided to include this information in the figures to make it easy to find. Note that every number always represents the total over the 3 subjects. Since differences between subjects are essentially negligible this was simply considered typographically more convenient than giving average over subjects (which could require a fractional part) or giving separate n for each subject.

Turning to the results themselves, it will be observed that glottal gesture duration is longer for fricatives than plosives, which conforms to expectations. The effect is not enormous - about 15ms difference - but highly consistent and significant at $p<0.001$ for all subjects in a two-way ANOVA with C1-Category and Speech Volume as factors (Subject CG had a highly significant interaction between the two factors, but Speech Volume itself was non-significant for all subjects). A summary of all ANOVA and nonparametric pairwise comparison results is given in the two tables on p. 64 of the previous chapter.

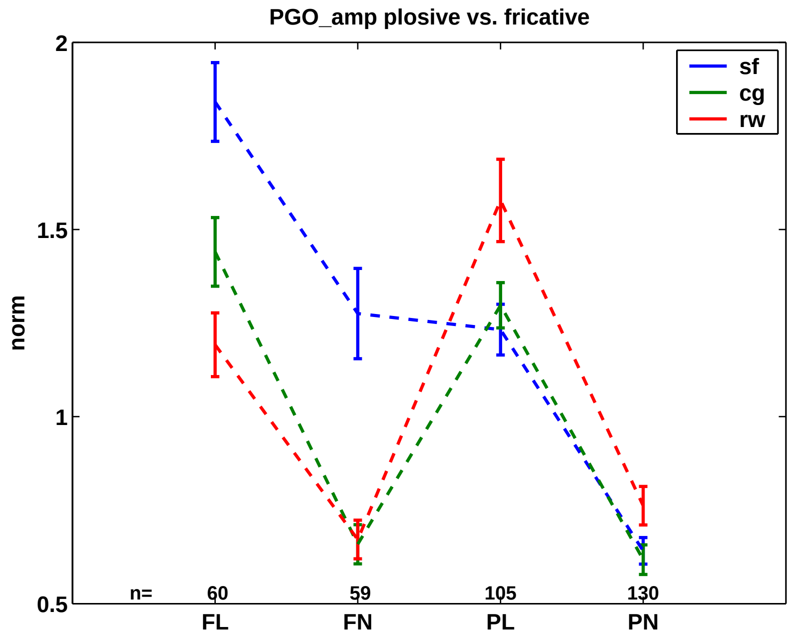


Fig. 6.3: Peak Glottal Opening Amplitude. Other details as in previous figure

In contrast to the result for gesture duration the results for gesture amplitude are much less clear. Of the three subjects only SF shows the expected pattern of greater glottal opening for fricatives. This was significant at $p<0.001$ in the normal ANOVA and was basically confirmed by the non-parametric Kruskal-Wallis test: In pairwise comparisons (i.e comparing fricatives and plosives matched for volume level) using 0.01 as the criterion significance level, we found FN>PN at this level, while FL>PL just missed reaching this level.

RW actually has a significant effect in the other direction in the ANOVA ($p<0.01$), though neither the FN vs. PN nor FL vs. PL pairs reached 0.01 in the non-parametric test.

No significant effects at all were found for CG.

Thus there is no consistent evidence in this data for the expected larger glottal opening in fricatives than plosives.

The next features of glottal behaviour to be looked at relate to what is probably the other most frequent observation regarding differences between fricatives and plosives, namely that abduction starts earlier relative to formation of the oral occlusion for fricatives (and may be faster).

The next figure shows the relevant temporal information, i.e the duration of the interval from formation of the oral occlusion to start of the glottal opening movement.

As can be seen, the results are absolutely consistent and clear-cut for all three subjects with glottal abduction for the fricatives starting some 30ms earlier than for the plosives (the only difference between the subjects is that CG favours overall an earlier abduction onset). Not surprisingly, the ANOVA results for fricative vs. plosive are significant at $p<0.001$ for all subjects. The two important pairwise comparisons FL vs. PL and FN vs. PN both reach the 0.01 level for all subjects.

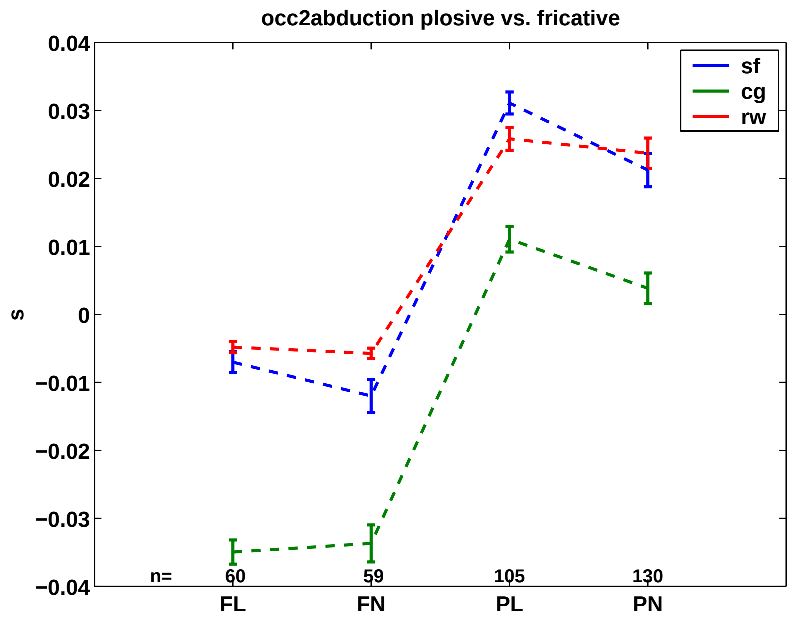


Fig. 6.4: Duration of interval from start of oral occlusion to start of glottal abduction (negative values indicate abduction starts before formation of the occlusion). Other details as in Fig. 6.2

Of course, this result is by no means unexpected. The important point is to note just how clear-cut it is: This particular distinction between fricatives and plosives will turn out to remain quite salient throughout the more complex consonant sequences as well. It is precisely by noting what distinctions are completely stable and what distinctions tend to become smeared by speaker-specific effects or the sometimes rather heterogeneous nature of our speech material that we can build up a clearer perspective on what are the key features of laryngeal control.

The next parameter to be considered is very much a case in point. There are plausible aerodynamic reasons that opening the glottis early or rapidly is particularly important for fricatives compared to plosives. We have just seen that early opening of the glottis is a robust effect. A possible counterpart in the shape of the glottal gesture that would capture rapid abduction is the ratio of abduction duration to adduction duration (this is also a convenient parameter for capturing the shape of the glottal cycle in the more complex consonant sequences). Accordingly, one would expect lower values of this parameter for the fricatives than for plosives.

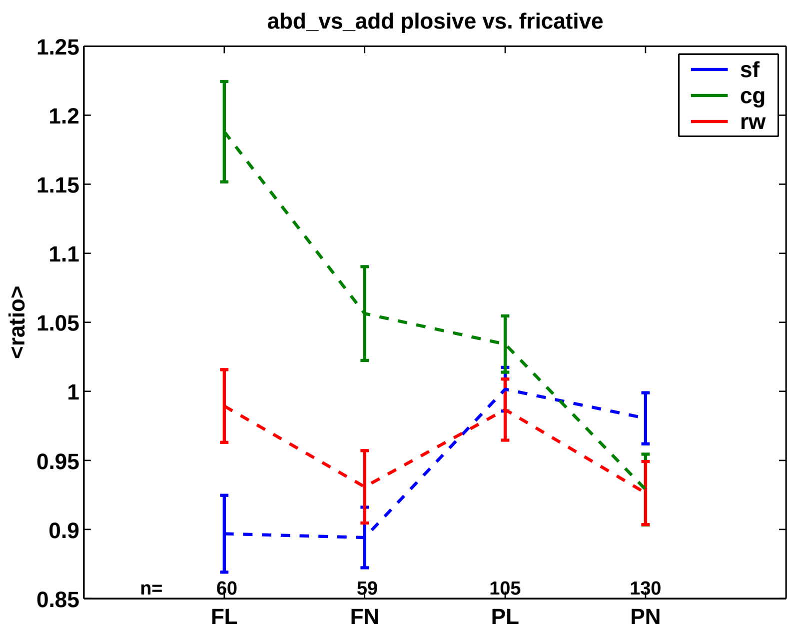


Fig. 6.5: Ratio of abduction duration to adduction duration. Other details as in Fig. 6.2

Even a quick glance at the figure above is enough to reveal that the situation is much less clearcut than in the previous figure showing the time interval from occlusion to abduction.

The only subject who fits in with the expectation is SF: For the plosives the ratio is about 1, indicating similar duration for abduction and adduction while for the fricatives the ratio is consistently below 1 indicating shorter abduction than adduction duration ($p<0.001$ in the ANOVA; of the relevant pairwise comparisons in the non-parametric test FL vs. PL reaches the 0.01 level, while FN vs PN does not quite reach it).

For RW there is clearly no fricative-plosive difference (merely a slight difference between loud and normal volume).

The most unexpected result - at least at first sight - is for CG, who has higher values of the ratio for the fricatives; for the loud fricatives the value is actually well over 1. The ANOVA shows a highly significant effect, but in the opposite direction to SF (though neither the FL vs. PL nor FN vs. PN pairs quite reaches 0.01 in the nonparametric test)[37].

In fact, the previous figure may offer a clue as to the reason for this unexpected pattern: CG is the subject who shows particularly early glottal abduction[38]; thus, fast abduction may simply not be so crucial when it is initiated in good time. Seen in this light, speaker-specific differences are not simply a disturbing feature impeding interpretation, but can actually help to make clearer what are the crucial elements in laryngeal control in terms of fulfilling the aerodynamic requirements of the sounds to be uttered, and what on the other hand are elements that retain freedom to vary. This is a theme that will be encountered on several occasions in the presentation of the results.

Finally in this section, we will look at two further parameters, namely interval from onset of occlusion to peak glottal opening, and average abduction speed. This is mainly for the sake of completeness since parameters of this kind have frequently been discussed in the literature. Referring back to the figure from Yoshioka et al. (1981) on p. 5, which was used to summarize previous findings on fricatives versus plosives from the literature, the expectation is clearly that peak glottal opening will occur closer to onset of oral occlusion in fricatives than plosives, and that the slope of the opening movement will be steeper in fricatives, i.e higher abduction velocity. In fact, the patterns found in our data are essentially predictable from parameters already shown, and thus add little new information.

The interval from start of oral occlusion to peak glottal opening patterns very similarly to the interval from start of oral occlusion to start of glottal abduction shown in Fig. 6.4 above, and is similarly clear-cut (i.e same significance levels both in ANOVA and nonparametric pairwise comparisons). In other words, peak glottal opening is reliably located closer to start of oral occlusion in fricatives than plosives (this was to be expected precisely because of the early onset of abduction in fricatives, coupled with the fact that overall gesture duration is not a great deal longer, and supported in the case of SF by the relatively short abduction phase). This thus parallels findings from the literature quite closely.

---

[37]We are not sure as to the interpretation of the higher values of the ratio for the loud volume condition, but as this is marginal to the main concerns of the investigation we will not pursue it further here.

[38]See Part I of this monograph for a discussion of this finding from the point of view of the EMG activity.

There is actually a good deal more to discuss about the relative timing of peak glottal opening within the occlusion phase of fricatives, in particular the strength of the generalization that it occurs at fricative midpoint, however this will be postponed until later sections where single fricatives are compared with consonantal sequences containing fricatives (comparison of fricatives and plosives in the context of the present section is not particularly meaningful as they are quite simply categorically different).

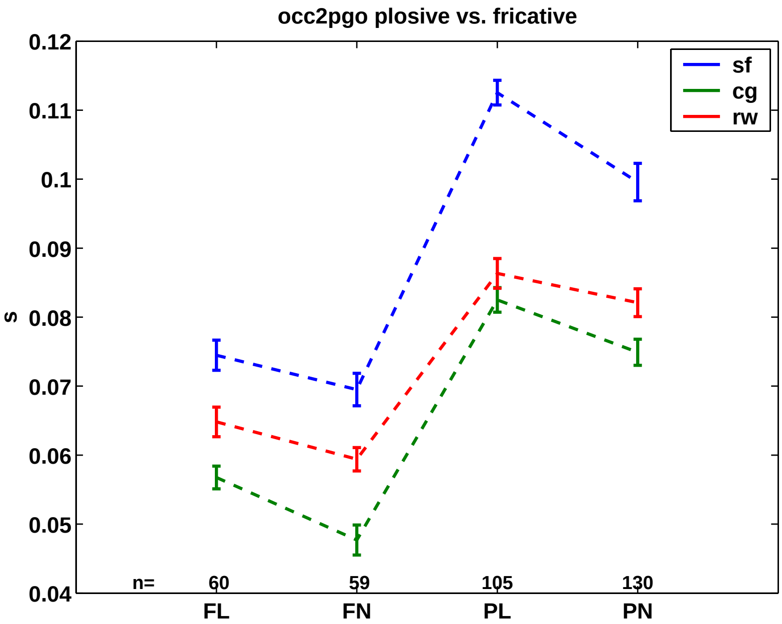


Fig. 6.6: Interval from start of oral occlusion to peak glottal opening. Other details as in Fig. 6.2

The final parameter, average glottal abduction velocity, shows a pattern very similar to the amplitude of peak glottal opening, and thus does not generally confirm the expectation from the literature of higher velocity in fricatives.

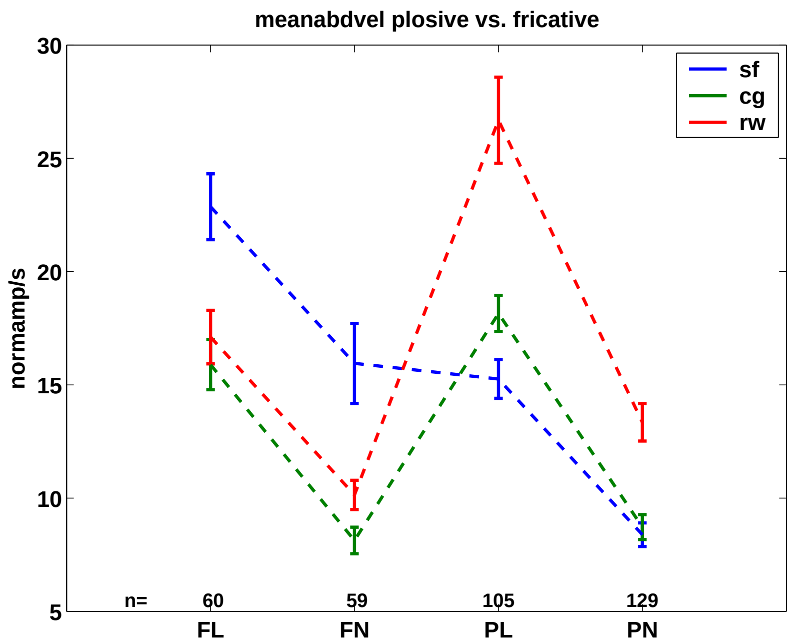


Fig. 6.7: Average glottal abduction velocity. Other details as in Fig. 6.2

In other words, as seen above with peak glottal opening amplitude, the expectation is confirmed only for SF while CG shows no effect, and RW a tendency contrary to the expectation. This pattern is essentially predictable taking into account the specific combinations of peak glottal opening, gesture duration and abduction-to-adduction ratio seen for the individual speakers above. (For SF significance levels are identical to those for peak glottal opening, i.e $p<0.001$ in the ANOVA, with FN vs. PN reaching 0.01 in the pairwise comparison, and FL vs. PL just missing this level. For RW we find $p<0.001$ in the ANOVA. This strongly significant result (peak glottal opening only reached $p<0.01$) was mainly due to a particularly high value for PL, the Consonant*Volume interaction being weakly significant. But once again neither the FL vs. PL or FN vs. PN pairs reach 0.01 in the pairwise comparisons.)

Preliminary summary

It has been seen that the results of the comparison of fricatives and plosives only partly conformed to expectations. The time at which abduction starts relative to the formation of the oral occlusion indeed gave as expected a very robust differentiation, so it will be interesting to observe whether it continues to prove so in more complex fricative vs. plosive syllable onsets. The overall duration of the glottal gesture was also fairly clearly longer for fricatives. On the other hand, the shape of the gesture, the speed of abduction, and the overall amplitude showed no clear pattern.

Since this result was not really expected it is worth pointing out that in this work we have only analyzed the sounds in word-initial position, at the start of the syllable bearing word-stress. This is the location where aspiration of voiceless plosives is typically strongest. It was very noticeable in this corpus that the many plosives in medial position (refer back to corpus list in Chapter 4, p. 34) might be only weakly aspirated and show very restricted glottal opening (discussed in more detail in Fuchs, 2005; an example with transillumination data was also given in the concluding discussion of Part I of this monograph). Fricatives, however, seem to reduce their glottal opening much less at prosodically weaker locations in the word, so in such cases they would certainly be found to have larger glottal opening than plosives.

6.1.2 Place of articulation in plosives

Introduction

As was seen in the discussion of previous work, the possible relevance of place of articulation in plosives is a useful issue for bringing into clearer focus the extent of our understanding of laryngeal-oral coordination[39].

The immediate reason for this is that VOT - as a readily measurable acoustic parameter - is known to vary fairly systematically with place of articulation (Cho & Ladefoged, 1999; Docherty, 1992), and thus it is interesting to determine the underlying reasons for this. Since aspirated plosives (which are the only ones of concern here) must involve appreciable glottal opening at oral release, the most succinct characterization of laryngeal-oral coordination is the duration of the time interval from peak glottal opening to oral release. Should this vary systematically over plosives then the interesting question in turn for speech motor control is whether this represents an active adjustment of the glottal gesture or whether it emerges as a passive effect of changes in the duration of the oral occlusion.

Detailed results for individual parameters

The parameters presented below are the following (with a particular emphasis on the relationship between pairs of durational parameters):

- Glottal gesture duration
- Amplitude of peak glottal opening
- Relationship between duration of aspiration and duration of oral occlusion
- Relationship between duration of aspiration and duration of interval from peak glottal opening to release of the oral occlusion
- Relationship between duration of interval from peak glottal opening to release of the oral occlusion and duration of oral occlusion
- Duration of interval from start of oral occlusion to start of glottal abduction

The tabulation of the statistical results for each individual parameter is to be found on p. 65.

To provide a background against which to discuss what is actually being controlled, the most convenient point of departure (as at the beginning of this section for plosives vs. fricatives) is to examine the duration and amplitude of the glottal gesture.

Regarding duration (see next figure) there is a very weak trend for longer values for /**t**/ for SF ($p<0.05$), with a nonsignificant tendency in the same direction for CG, but no pairwise comparisons reach $p=0.01$.

---

[39]Place of articulation in fricatives is not expected to be as revealing, and will not be examined here. Indirectly, some information on this is provided by later sections, where some comparisons involve fricatives at specific places of articulation.

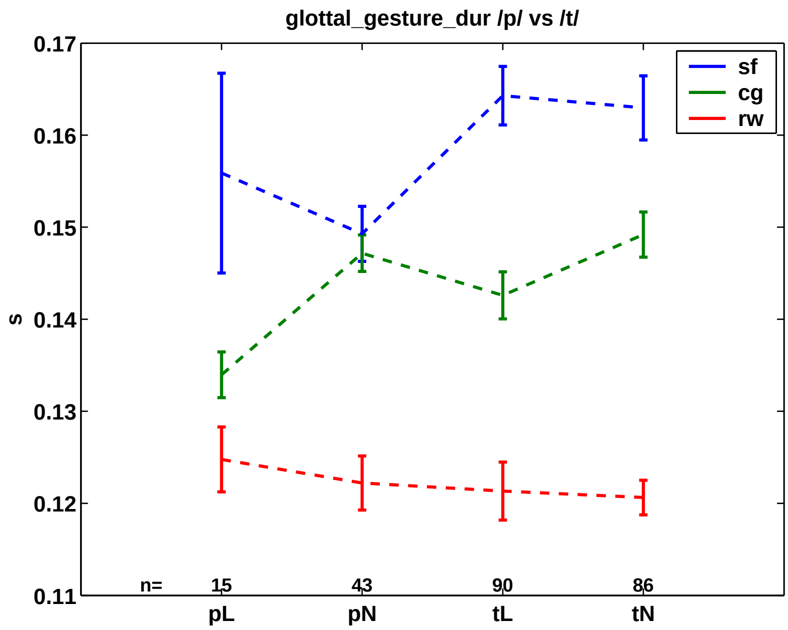


Fig. 6.8: Glottal gesture duration for /**p**/ vs. /**t**/. Abscissa labels: pL= /**p**/, Loud; pN = /**p**/, Normal; tL= /**t**/, Loud; tN = /**t**/, Normal. Other details as in Fig. 6.2, p. 75

No evidence for differences in glottal opening amplitude were found (next figure). A significant effect for /**p**/ vs. /**t**/ was found in the 2-way ANOVA for speaker SF, but the robustness of this is doubtful: in the non-parametric analysis neither direct /**p**/ vs /**t**/ pairwise comparison came near reaching the p=0.01 criterion (it should be noted throughout these analyses that the /**p**/-Loud condition has very small n (only one word form), and thus may not give very stable results).

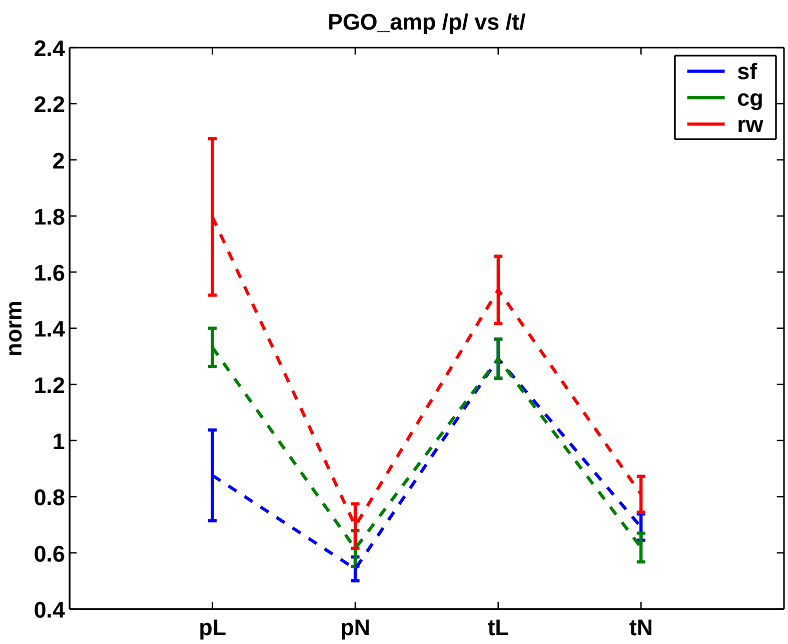


Fig. 6.9: Peak Glottal Opening Amplitude for /**p**/ vs. /**t**/. Other details as in previous figure, and Fig. 6.2

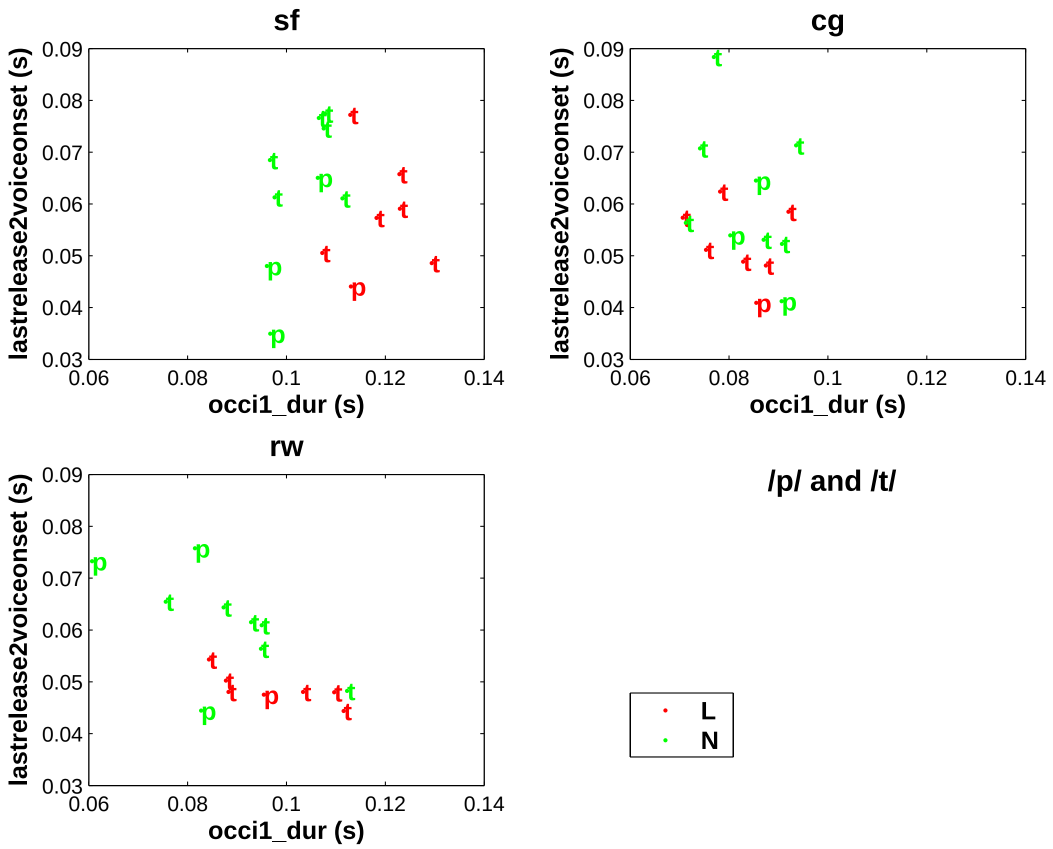
Thus as a background for consideration of further parameters it seems safest to assume essentially the same gestural duration and magnitude for both places of articulation.

The other data required as background before considering coordination parameters are occlusion duration and aspiration duration for /**p**/ and /**t**/. In other words, do the present data confirm the expected pattern of longer occlusion and shorter aspiration for /**p**/, and thus reciprocal relationship between these two parameters, viewed over both consonants?

It turned out that occlusion duration clearly did not conform to the expected pattern. The main effect of consonant was non-significant in the ANOVA for all speakers. At the normal volume level speaker RW actually had an effect contrary to the expected direction that almost reached p=0.01 in the non-parametric pairwise comparison for pN vs. tN.

For aspiration duration there was a significant main effect in the expected direction of longer aspiration for /**t**/ for two speakers (SF: p<0.001; CG: p< 0.01), but no effect for RW. The only pairwise comparison with matched volume levels to reach p=0.01 in the nonparametric test was pN vs. tN for SF.

Given the above pattern of results it is hardly to be expected that a simple reciprocity of occlusion vs. aspiration duration over /**p**/ and /**t**/ will be found. This is illustrated in the following figure. This shows a scatter plot of aspiration duration vs. occlusion duration where each data point represents the average value for one word-form from the corpus (thus as already mentioned, the conditions pL, pN, tL, tN are not very evenly represented).

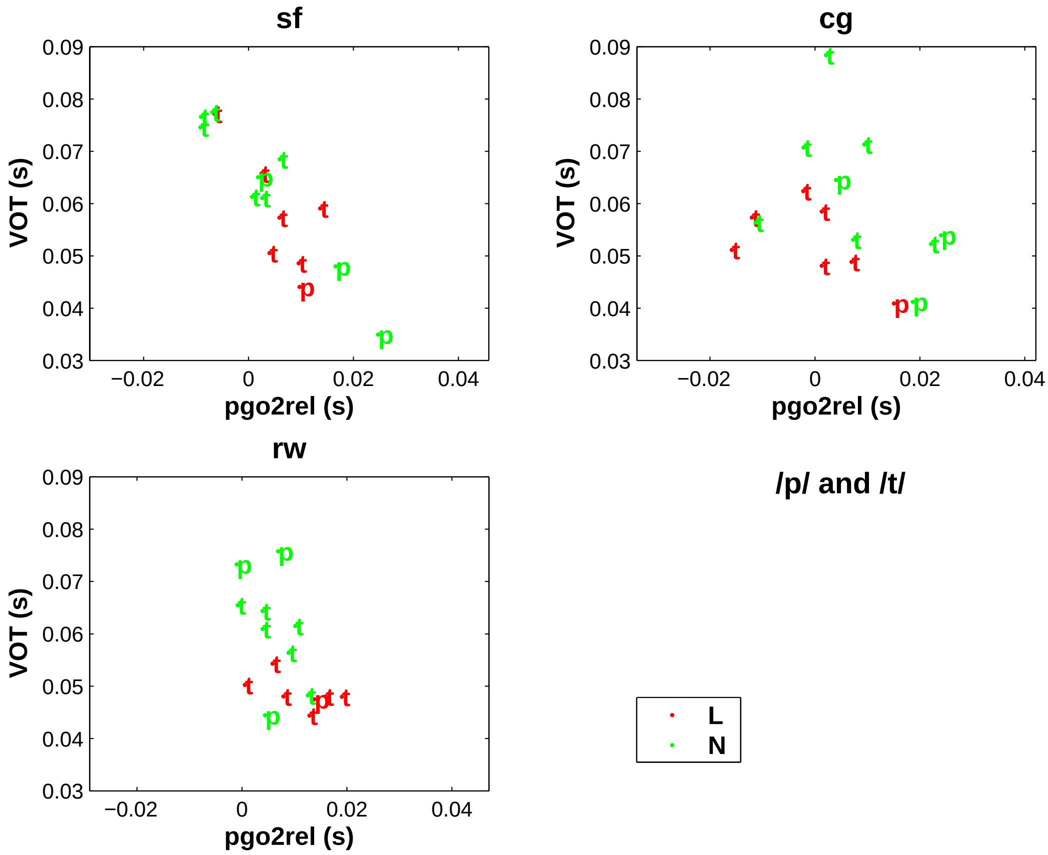


Overall			Normal			Loud		
SF	CG	RW	SF	CG	RW	SF	CG	RW
0.04	-0.25	-0.67	0.61	-0.35	-0.62	0	-0.28	-0.76

Fig. 6.10: Aspiration duration (y axis) vs. occlusion duration (x axis) for /**p**/ and /**t**/. Volume condition indicated by colour: Red = Loud, Green = Normal. Table below figure shows correlation coefficients (Pearson's r).

Interestingly, the only speaker to show a clear negative relationship is RW. However, in his case, this has nothing to do with a place of articulation effect. SF even has a positive correlation in the normal volume condition (see tabulation of correlation coefficients below the scatter plots).

Clearly the simple scenario taken as point of departure is not reproduced in this data. Nevertheless it is still worth enquiring into what laryngeal-oral coordination pattern leads to the longer aspiration in /**t**/ for SF and CG. The key timing parameter is, of course, the duration of the interval from peak glottal opening to release. In the ANOVA the main effect of C1 was significant at p<0.001 for SF and CG (and not significant for RW). In the nonparametric test pN vs. tN reached p=0.01 for SF. For CG neither of the straight pairs pN vs. tN nor pL vs. tL quite reached this level of significance, but one of the less direct pairs (pN vs. tL) did. Thus, as might have been expected, there is some evidence that this interval is longer for /**p**/ than /**t**/. Typically, for /**p**/ the interval is almost always positive, i.e peak glottal opening occurs before release, whereas for /**t**/ the interval is more weakly positive and sometimes even slightly negative. Rather than showing a figure with the means and standard errors per condition it is probably more useful to show the results for this interval in a scatterplot, relating it to aspiration duration:



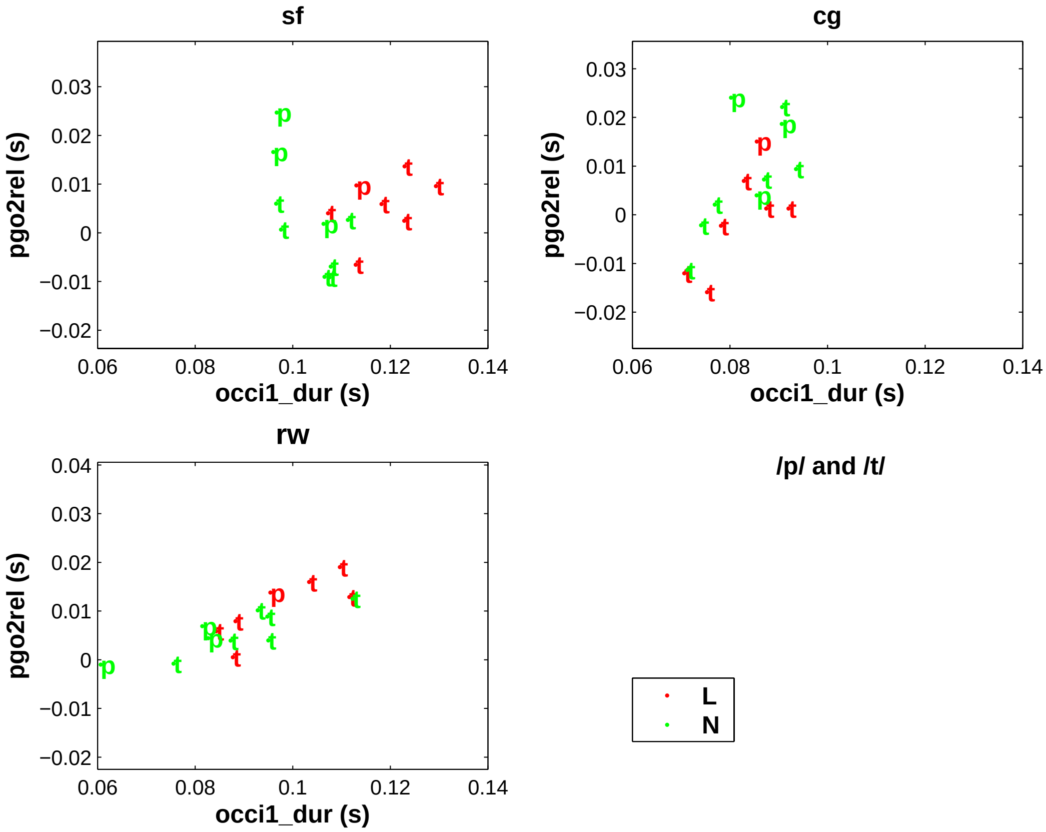
Overall			Normal			Loud		
SF	CG	RW	SF	CG	RW	SF	CG	RW
-0.89	-0.26	-0.56	-0.96	-0.47	-0.46	-0.75	-0.55	-0.54

Fig. 6.11: Aspiration duration ('VOT') vs. PGO to release. Other details as in previous figure

This figure thus makes it possible to examine how closely aspiration duration is related to the interval from peak glottal opening to release. The expectation would be for a negative correlation. Once again each data point in the plot is the mean of one word-form in the corpus.

Clearly, negative correlations are found for all speakers (including RW) albeit with different strengths. For SF the relationship is particularly strong, and the slope of the relationship is about -1, indicating that changes in PGO to release result in a corresponding increment or decrement in aspiration duration. For speaker RW on the other hand the differences in PGO to release are quite restricted compared to those in aspiration duration.

For speaker SF and, to a lesser extent, speaker CG (whose pattern is simply less clear) we thus find a different timing of peak glottal opening for /**p**/ vs. /**t**/, which can be related in turn to the differences in aspiration duration. In view of what has been seen already it seems unlikely that the speaker derives this timing difference as a passive consequence of modifying occlusion duration. However, to make this point explicit we will next look at a corresponding scatter plot.



Overall			Normal			Loud		
SF	CG	RW	SF	CG	RW	SF	CG	RW
-0.1	0.66	0.83	-0.7	0.67	0.85	0.43	0.69	0.81

Fig. 6.12: PGO to release vs. Occlusion duration for /**p**/ and /**t**/. Other details as in previous two figures

Let us consider speaker SF first, since she was the speaker who showed the clearest tendency to shorter aspiration and more positive PGO-to-release intervals for /**p**/. If variation in PGO-to-release simply falls out from differences in occlusion duration we would expect a positive correlation between these two parameters. For SF this is quite clearly not the case: the overall pattern is very unclear, and in fact there is a negative correlation in the normal volume condition. For speaker CG there is a fairly clear positive relationship with a slope of roughly +1. The problem for a simple interpretation is that /**p**/ and /**t**/ are not in fact reliably distinguished in occlusion duration. The positive relationship between the two parameters here is as much due to variation within the categories (particularly evident for /**t**/, for which the largest number of word-forms was available), as to differences between the categories. Finally, RW has already been mentioned as showing different behaviour from the other two subjects: here he actually has the strongest positive correlation of all three subjects but - as for his relationship between occlusion and aspiration duration - it is completely unrelated to the consonant categories. It should also be noted that the slope of the relationship is quite flat, with occlusion duration varying over a much larger range than PGO-to-release, so despite the high correlation there is not a straight one-to-one trade-off between duration of the occlusion and PGO-to-release.

Preliminary summary

Taking stock in terms of Jessen's three hypotheses, the hypothesis that we originally found most intriguing was Hypothesis 3, the "short stop closure hypothesis". As just seen from the previous paragraph, the speaker who comes closest to this is CG, but not very strikingly. Speaker SF clearly does not fit in with this hypothesis; based on the measurements reviewed to date she would come closest to Hypothesis B, the "long glottal gesture hypothesis", but again this was not a very strong effect. Of the logical possibilities forming the basis for Jensen's hypotheses we have not yet considered Hypothesis A, the "late glottal gesture hypothesis", since we considered it at first sight the least plausible one. It turns out that in fact for all speakers there is a significant difference in the direction of a longer interval from onset of occlusion to onset of glottal abduction for /**t**/. This is illustrated in the following figure.

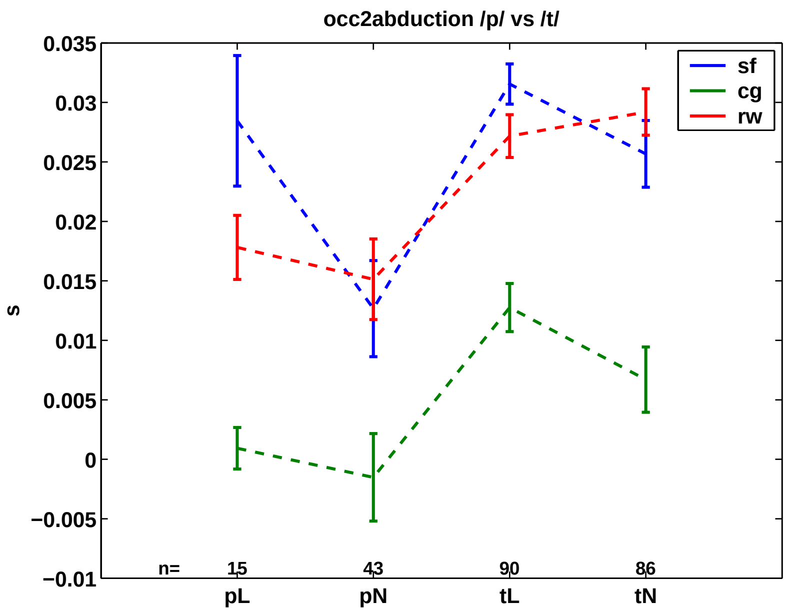


Fig. 6.13: Duration of interval from onset of oral occlusion to onset of glottal abduction for /**p**/ vs. /**t**/ Other details as in Fig. 6.8

The level of significance in the ANOVA increased from SF ($p<0.05$) via CG to RW ($p<0.001$). RW was the only speaker for whom a pair-wise comparison matched for volume level reached $p=0.01$ in the non-parametric test. In any case, though not expected, it does not seem to be possible to discount the possibility that the timing of the onset of glottal abduction may have some role to play.

The overall impression one receives from this analysis of place of articulation effects in consonants is that one is confronted with a variety of rather weak effects. This illustrates once again what is both a strength and a difficulty with the current corpus: Since the phonetic make-up of the words is not completely rigorously controlled (e.g vowel and medial consonants can affect the overall length of the word) effects that emerge clearly can indeed be considered robust, but some weaker effects may remain submerged. For example, the point of departure in this section was the possibility of a trade-off between duration of oral occlusion and duration of VOT viewed over the consonant categories, i.e a negative correlation. However, there is a simple situation in which the expectation would be for a ***positive*** correlation between these two intervals, namely if one word is spoken more slowly or with greater emphasis than another. Thus the actual effect of interest here may only emerge with very restricted, carefully controlled material, not fulfilled by the present material (for example, even within one consonant category a fair range of occlusion durations occurred).

Returning to the implications of the different hypotheses, the "short stop closure" hypothesis seemed particularly interesting because it could account for shifts in laryngeal-oral coordination without requiring reorganization of the laryngeal gesture itself (and also seemed potentially relevant for the plosive-sonorant combinations to be looked at shortly). Some evidence was found that this could be a relevant mechanism for CG and RW, but for RW it was not linked to the place of articulation distinction. The other two hypotheses, the long glottal gesture, and the late glottal gesture, would both imply active changes in the glottal gesture itself. Active mechanisms are probably particularly relevant for SF, since she represented the clearest case of longer aspiration for /**t**/, probably related to the longer gesture for /**t**/ supported by the later abduction onset. Both RW and CG showed later abduction onset for /**t**/, but the functional relevance of this is unclear, particularly for RW, since he showed no place-of-articulation related difference in aspiration.

Nevertheless, as we move now from single consonants to sequences of consonants, it will be seen that the timing of the start of glottal abduction is an aspect that will continue to be important.

6.2 Results 2: Combinations of fricatives and plosives

Having reviewed the properties of single consonants in the previous section, the focus in this second main section of the results, as well as subsequent sections, will be on consonants in sequence. This second part of the results is in turn in two parts:

First of all, fricative-plosive combinations will be dealt with; here the main comparison will be with single fricatives. Following that, plosive-fricative combinations will be discussed. While the main comparison here will be with single plosives, many of the figures will juxtapose the four main consonantal structures encountered hitherto, i.e single plosives and fricatives, together with fricative-plosive and plosive-fricative combinations.

6.2.1 Fricative-plosive combinations

Introduction and notes on material

In this section we move to the first groups of consonants in the material. The immediate aim of this section will be to review the magnitude, shape and timing of the glottal gesture in fricative-plosive combinations. This will be done by comparing them with single fricatives. This is a natural choice for several reasons: we expect peak glottal opening to be located during the fricative rather than the plosive. It would certainly not be expected to be anywhere near the release of the plosive since the plosives are unaspirated. Based on the discussion of the literature the main issues to be elucidated are whether there is evidence that the glottal gesture in fricative-plosive combinations can be captured as the overlapping of activity related to single fricatives and plosives (cf. Munhall & Löfqvist, 1992), whether the timing of peak glottal opening remains associated with the midpoint of the fricative as it is (roughly speaking) in single fricatives, and - if it does not remain associated in this way - whether it is possible to identify the underlying timing principle.

The nature of the material is somewhat different from that in the previous section: for the fricative-plosive material only one word-token out of a total of 6 (2 x /**ʃt**/, 4 x /**ʃp**/) was spoken at loud volume, so most of the discussion in this section will be based on material spoken at normal volume only. Only a few supplementary analyses will include the volume contrast explicitly as a factor. Since all the fricative-plosive combinations have /**ʃ**/ as initial fricative, we will use single /**ʃ**/ as the fricative for comparison (i.e material with /**f**/ will not be included).[40]

Overview of temporal structure

As in the previous section we present here right at the outset an overview of the temporal structure of the relevant sequences so that this can be referred back to throughout the section.

---

[40]In the following section (on plosive-fricative combinations) most of the figures actually show all of the four categories fricative-plosive, fricative, plosive-fricative and plosive. In those figures /**f**/ is included in the fricative data, and the figures also show both normal and loud speech volume for all categories, i.e including fricative-plosive (see Fig. 6.25 and subsequent figures on p. 102ff).

Among the most salient points in this figure are:

- Total duration of voicelessness clearly longer in the clusters
- Differences in length of the glottal gesture are, however, much less obvious
- Duration of fricative occlusion shorter in the clusters

We will however postpone any comment or interpretation until the results have been presented in detail and statistically analyzed.

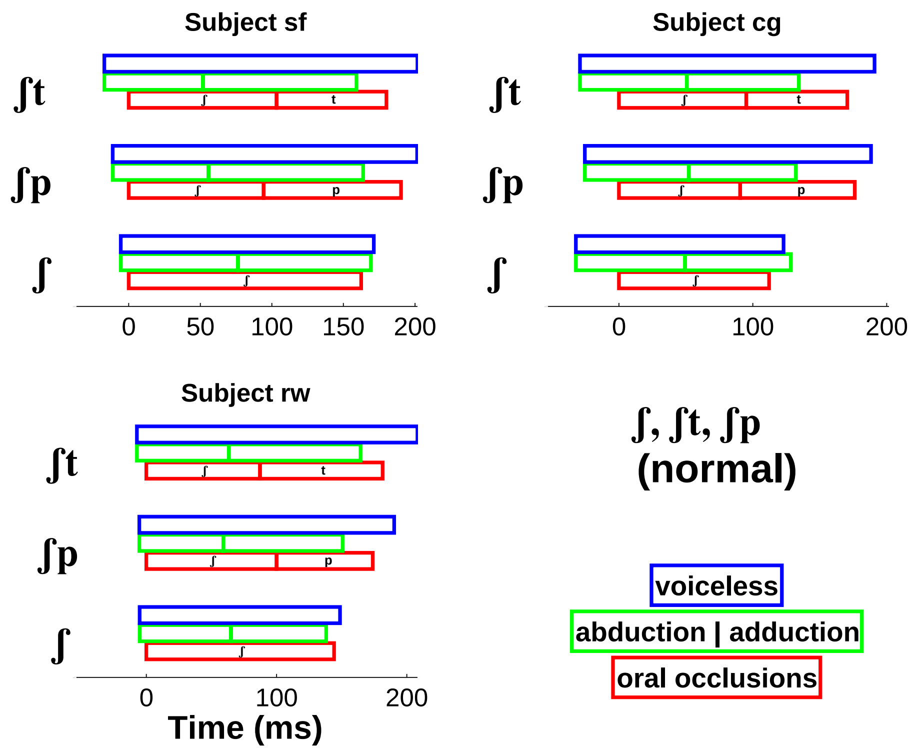


Fig. 6.14: Comparison of temporal structure of /**ʃp**/ vs /**ʃt**/ vs /**ʃ**/ (normal volume). Zero on time axis corresponds to onset of oral occlusion for initial /**ʃ**/. For details see text (and refer back to explanations of Fig. 6.1 on p. 73)

Detailed results for individual parameters

Results of the following parameters will be covered here:

- Glottal gesture duration
- Amplitude of peak glottal opening
- Oral occlusion duration of fricative segment (/**ʃ**/)
- Relative position of peak glottal opening in the fricative segment
- Relationship between interval from peak glottal opening to release of fricative and duration of fricative occlusion
- Ratio of glottal abduction to adduction duration
- Duration of interval from release of last occlusion to end of glottal adduction
- Duration of interval from release of last occlusion to voice onset (“VOT”)

Tables of the statistical results for the individual parameters are to be found on p. 66

The first two figures in this section that detail specific parameters provide basic information on gesture duration and peak opening amplitude. Longer and/or larger glottal gestures for the clusters might be expected if they consist underlyingly of a blending of activity for two consonants.

Regarding gesture duration, it is very clear from the figure below (backed up by the ANOVA results) that two subjects show no difference between clusters and singletons. Only for RW is there a significant effect in the expected direction of shorter durations for the singletons (mainly due to a particularly clear difference between /**ʃt**/ and /**ʃ**/.

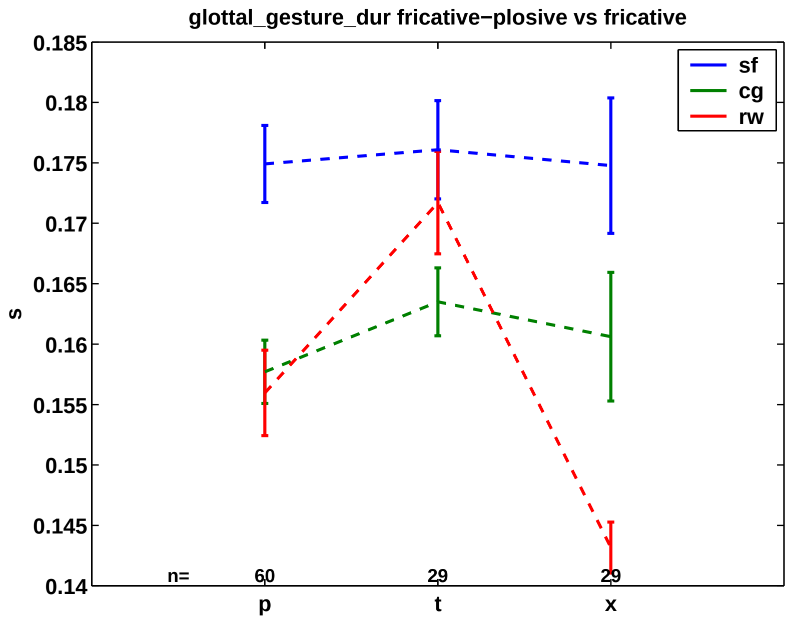


Fig. 6.15: Glottal gesture duration, comparing single fricatives with fricative-plosive sequences. Abscissa labels: p = /**ʃp**/, t = /**ʃt**/, x=/**ʃ**/, i.e the label indicates C2 in the cluster, with ‘x’ indicating absence of C2.

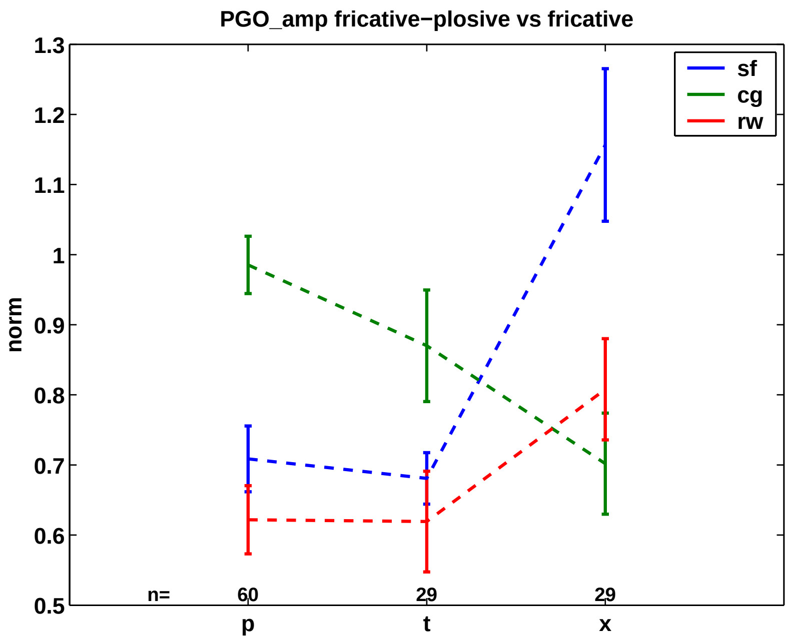


Fig. 6.16: Peak glottal opening amplitude, comparing singles fricatives with fricative-plosive sequences. Other details as in previous figure

Regarding gesture amplitude (see figure immediately above), there is also clearly no consistent evidence for a larger glottal opening for the clusters: CG shows a weakly significant result in the expected direction, RW has no difference, and SF actually has a strong effect in the opposite direction[41].

With respect to measurements of glottal opening, it will be recalled from the section on pre-processing of the transillumination data in Chapter 4 (p. 40ff) that normalization of glottal amplitude was not entirely straightforward. For example, items in the cluster section of the corpus had to be normalized separately from the loudness variation section. It also appeared that position of the word in the carrier phrase introduced some complications. In an attempt to cross-check whether different results might have emerged if the structure of the corpus had been simpler, we carried out two supplementary analyses using only a subset of the material:

(1) Material only from the loudness variation part of the corpus (Group 1 of the material)

This involved two word tokens with single initial /**ʃ**/ and one word-token with initial /**ʃp**/.
In this special case it was possible to use both loud and normal speech volumes. No evidence was found that using such a subset would indicate a shift in the results towards larger glottal opening for the clusters: for CG and RW no significant difference between clusters and singletons was found, while for SF - as for the complete material above - significantly larger glottal opening for the singleton was found.
A drawback of this subset was that both words with singleton fricatives were in the first position in the carrier phrase, while the cluster was in second position. Possibly the normalization procedures might not have precisely compensated for the tendency for smaller opening amplitudes in Word 2 position. This leads to the second variant:

(2) Tokens only from first position in the carrier phrase, but now mixing across the loudness and cluster parts of the corpus.

This was based on two word-tokens with singleton /**ʃ**/ from Group 1 of the corpus and two word-tokens with /**ʃp**/ from Group 2 . The results were very similar to those from the complete material: Once again SF showed significantly larger opening for the singleton ($p<0.01$ in Kruskal-Wallis) and CG had weakly significantly larger openings in the cluster. For RW larger openings for the singleton were found, that just reached significance (in the complete material he had a non-significant trend in the same direction). RW's results for overall gestural magnitude are in any case mixed, since here, too, he had significantly *shorter* gestural duration for the singletons.

In any case, the analysis of the two additional subsets only serves to reinforce the evidence from the complete material that our data shows absolutely no indication of greater glottal opening magnitudes in clusters compared to singletons.

Before we turn to the coordination of glottal and oral articulations, we first need to consider the duration of the oral occlusion for the fricative. In the overview of the temporal structure shown as the first figure in this section it was fairly obvious - and of course hardly surprising - that the

---

[41]Since in this section the analyses of variance generally involve only one independent variable it is possible to compare the normal analysis of variance directly with the Kruskal-Wallis non-parametric analysis of variance. Thus for this section the significance levels from the non-parametric procedure have been explicitly included in the tabulation of the ANOVA results where they differ from the parametric result. As discussed in the previous chapter, we believe that consideration of the non-parametric results is particularly advisable in the analysis of glottal opening amplitude.

duration of the fricative in the clusters was shorter than the singleton fricative. The next figure shows this in detail.

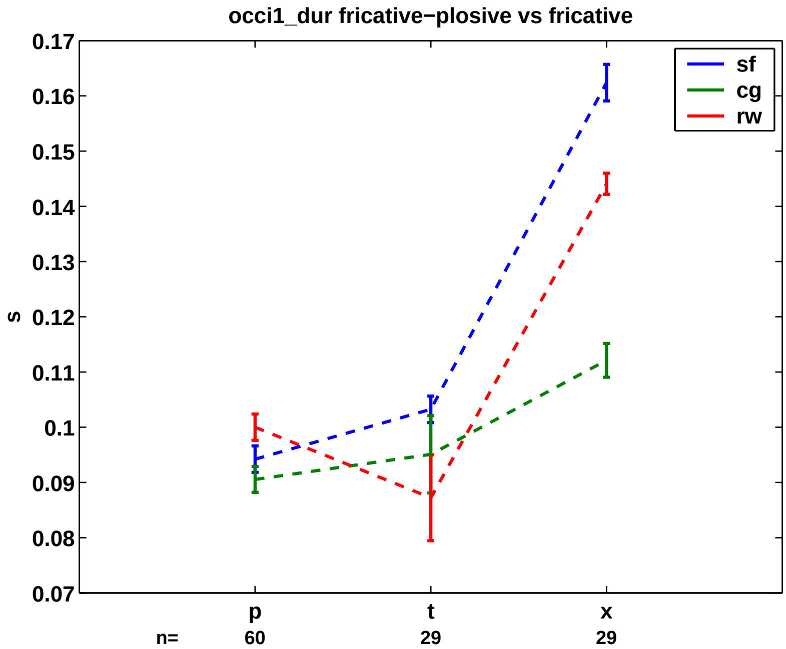


Fig. 6.17: Oral occlusion duration of fricative, comparing single fricatives with fricative-plosive sequences. Other details as in Fig. 6.15

Clearly this is a very consistent effect; the only point to note is that the magnitude of the difference is smaller for speaker CG. In the analysis of variance the identity of C2 (/**p**/ vs. /**t**/ vs. none) was significant at at least $p<0.01$ for all speakers. In the non-parametric pairwise comparison all speakers reached $p=0.01$ for /**p**/ vs. none (and RW also for /**t**/ vs none). A supplementary analysis was run by grouping /**ʃp**/ and /**ʃt**/ clusters together (inspection of the figure indicates this to be justifiable) and comparing clusters and singletons directly. This gave a significant result at $p<0.001$ for all subjects.

It might seem that here the significant changes in fricative duration from singleton to cluster are rather trivial, but we will be encountering several other cases where the reduction in occlusion durations going from singleton to cluster is by no means as clear.

Given this robust difference with regard to the length of the fricative, the first issue with regard to laryngeal-oral coordination is the location in the fricative of peak glottal opening. This is illustrated in the next figure. The relative position of peak glottal opening was calculated by dividing the duration from peak glottal opening to release by the duration of the fricative. Values of 0.5 thus indicate that PGO occurs at fricative mid-point, values below 0.5 indicate a location after the midpoint.

Once again. the pattern over the speakers is quite consistent (leaving aside the strikingly low value for /**ʃt**/ for speaker RW) in the sense that values are higher for single fricatives than for fricative-plosive clusters (i.e PGO is located relatively late in the oral occlusion in the latter case).

The influence of C2 is highly significant for all speakers, with the /**p**/ vs. none comparison reaching $p=0.01$ in every case (and, once again, /**t**/ vs. none for RW). In a further supplementary analysis merging /**ʃp**/ and /**ʃt**/ the cluster vs. singleton difference was significant at at least $p<0.01$ for all speakers. It thus appears that timing of PGO does not remain at a constant location in the fricative.

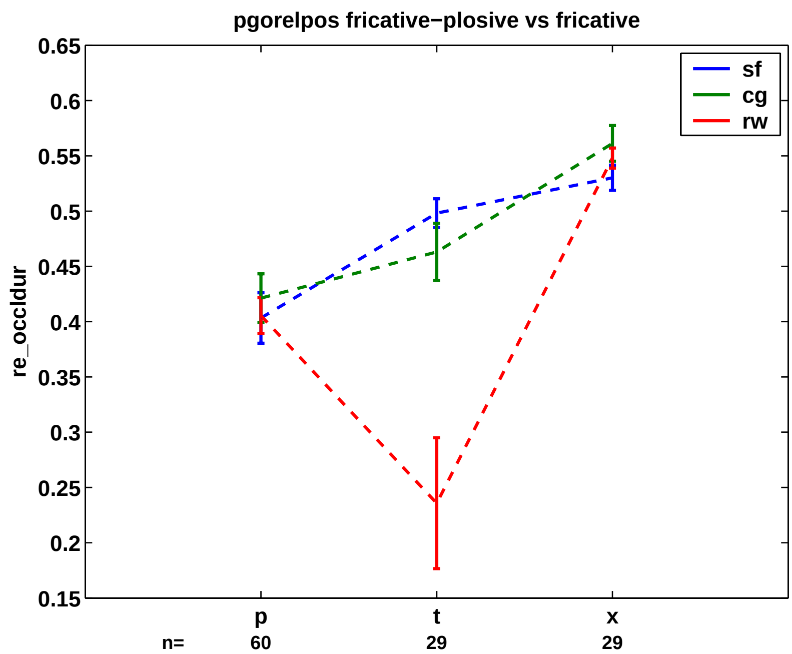


Fig. 6.18: Relative position of peak glottal opening in fricative, comparing single fricatives and fricative-plosive sequences. A value of 0.5 indicates PGO occurs at midpoint of fricative. Lower values indicate later point in the fricative (0 = end of fricative occlusion). Other details as Fig. 6.15

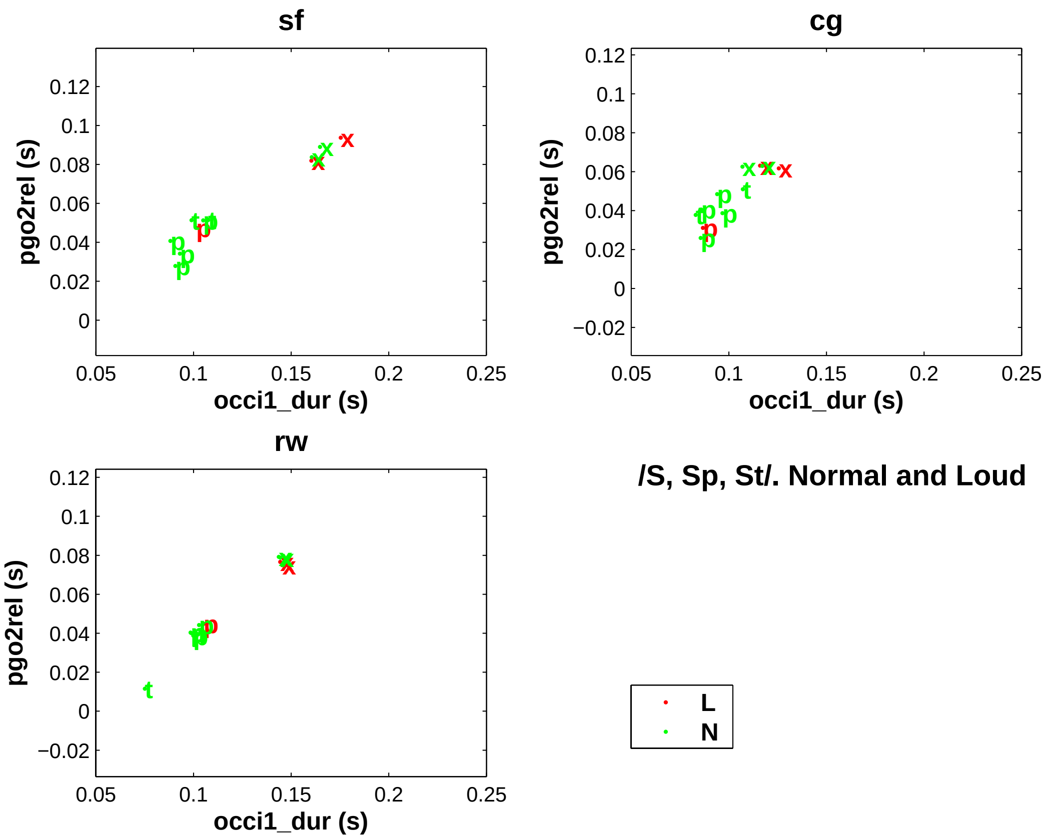


Fig. 6.19: Relationship between the interval from PGO to release and the duration of the oral occlusion for the fricative. Data points are labelled with the identity of C2 ('x' = no C2), i.e same conventions as for abscissa labels in preceding figures. Volume levels are colour-coded (L vs. N). Since there are very few word-tokens at the loud volume level in this part of the corpus, correlations were calculated only over the complete material (rather than also subdividing by loudness level as was done for the corresponding analysis of place of articulation in plosives).

The correlations were: SF: 0.98; CG: 0.89; RW: 0.99

It might be tempting to interpret this result as showing that in clusters the location of PGO is perturbed away from mid-frication under the influence of the following consonant and thus perhaps that glottal behaviour reflects a merging of influences from fricative and plosive. Such an interpretation could be misleading, however. To support this contention, the figure immediately above shows the relationship between the duration of the fricative occlusion and the interval from PGO to release using the means of each individual word-token (i.e a similar procedure to that followed when examining place-of-articulation effects in plosives).

Clearly the relationship is an extremely close one (in the analysis of place of articulation in plosives the situation was by no means so clear), and the slopes for all three speakers are reasonably close to unity. In other words, a large part of the shifts in the relative location of peak glottal opening can be explained as a passive effect of the change in occlusion duration. This indicates on the one hand that in clusters it may not be necessary to assume the influence of the plosive on the organisation of the glottal gesture itself (i.e contra Munhall & Löfqvist), but also, on the other hand, that expressing laryngeal-oral timing in terms of the timing of the peak glottal opening may not be the most appropriate route to take, i.e this may not capture the control principle on which speakers actually base their behaviour (contra Browman & Goldstein).

The remaining analyses in this section will test the well-foundedness of these preliminary interpretations.

We have seen that there are no gross differences in duration or amplitude of the glottal gesture for clusters vs. singletons. However, one potentially relevant feature that has not yet been considered is the ***shape*** of the gesture, which can be most easily captured in the ratio of abduction to adduction duration. This is shown in the next figure.

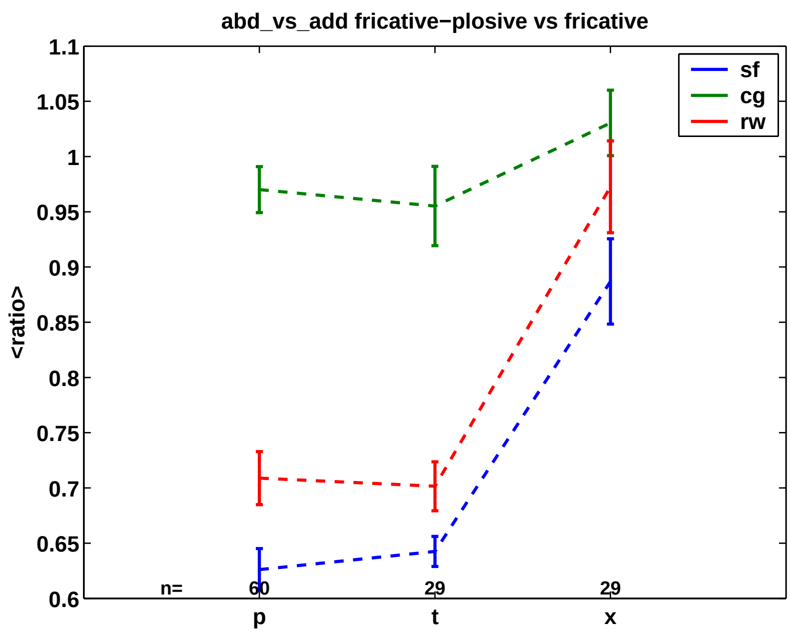


Fig. 6.20: Ratio of glottal abduction to adduction duration, comparing single fricatives with fricative-plosive sequences. Values below 1 indicate that abduction is shorter than adduction. Other details as in Fig. 6.15

Even though our data do not indicate that peak glottal opening is closely tied to fricative midpoint there remains the constraint in the clusters that appreciable glottal abduction must occur early in the fricative, whereas the time available for adduction is potentially longer since it only needs to be completed around the time of the plosive release. Accordingly, one might expect lower values for the abduction/adduction ratio in the clusters compared to the singletons. The figure makes it clear that a consistent effect in this direction occurs. It is highly significant for SF and RW in the ANOVA, with both clusters contrasting with the singleton at $p<0.01$ in the pairwise nonparametric test. For CG the tendency goes in the same direction but is too weak to reach significance. For the latter speaker this finding is reminiscent of the discussion of the abduction-adduction ratio in the comparison of fricatives and plosives. There, too, it was noted that he did not show the expected tendency to lower ratios for fricatives and it was suggested that this could be related to the fact that he is the speaker who initiates glottal abduction earliest. Accordingly his abduction movement is under less of a constraint to be completed quickly for fricatives, and, in the present case too, this may tend to smear out shifts in the abduction-adduction ratio between clusters and singletons.

Expressing the shift in the abduction-adduction ratio in terms of actual time, the duration of glottal abduction is about 6-7ms shorter in the clusters (averaging over the three speakers) and the duration of glottal adduction about 13ms longer (for reference, the statistical tests of these durations are given in the tables on p. 66).

At first sight, one could argue that the shift in the abduction-adduction ratio is consistent with a pattern of gestural organisation that consists of a dominant fricative-related gesture in the initial part of the cluster, with the latter part of this gesture then being overlapped by a weaker plosive-related gesture, this being manifested in a more gradual adduction phase. Based on consideration of the next timing parameter, namely the relation of the end of the glottal gesture to the offset of the oral occlusion, we will try to argue that this may not be the most persuasive interpretation.

This is shown in the next figure. The time-point of the end of the articulatory closure (of the plosives in the clusters, and of the fricative in the singletons) corresponds to zero, with negative values of the parameter corresponding to end of the glottal gesture *before* this time-point.

Clearly a very robust effect is present, distinguishing the clusters from the fricatives, with glottal adduction being completed well before oral release for the clusters (in the ANOVA $p<0.001$ for SF and CG, $p<0.05$ for RW; for all speakers the non-parametric pair-wise comparison p vs. x reaches $p=0.01$). To help put these timing relationships into perspective the figure juxtaposed immediately below shows the values for the interval from release of the last oral constriction to voice onset (“Voice onset time”). The range of the ordinate is the same for both figures. (It may be helpful here to refer back to the overview of the temporal structure displayed at the beginning of this section.) VOT is fairly similar for the clusters and singletons. It is actually somewhat longer for the clusters, which serves to reinforce the observation that in the clusters the time-point of voice onset has become dissociated from the time-point of glottal adduction whereas for the single fricatives glottal adduction and voice-onset would coincide at a value of about +5ms if one were to average over speakers. For the clusters we believe that this observation has interesting implications. Clearly, the duration of VOT in the clusters must simply reflect the time required for aerodynamic conditions to become appropriate for voicing following the burst; the appropriate glottal configuration for voicing can be assumed to have been reached well before the burst. This in turn means that it actually does not matter precisely when speakers complete glottal adduction. As long as it occurs somewhere between onset of the plosive occlusion and about 10ms after the plosive release it will have a negligeable effect on the acoustic output.

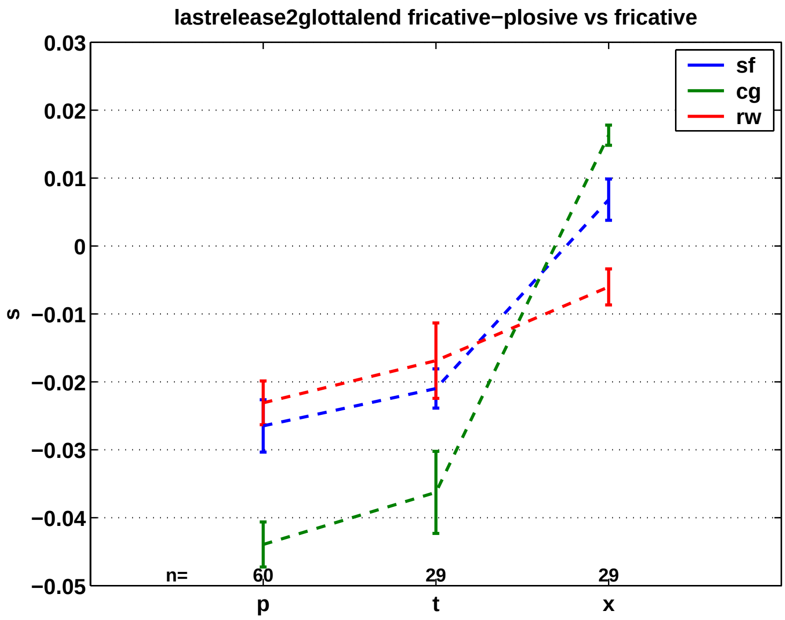


Fig. 6.21: Duration of interval from release of last occlusion to end of glottal gesture, comparing single fricatives and fricative-plosive sequences. Negative values indicate glottal adduction completed before oral release. Other details as in Fig. 6.15

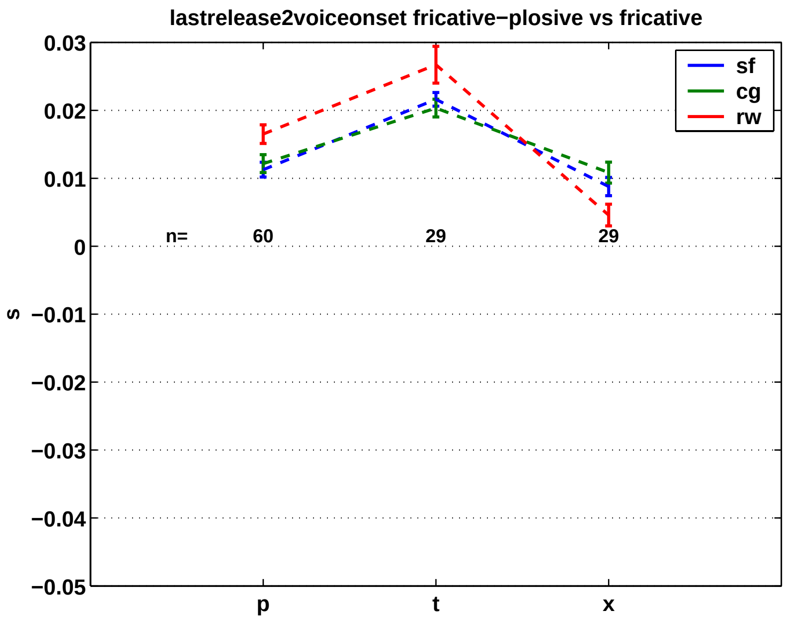


Fig. 6.22: Duration of interval from release of last occlusion to voice onset ("VOT"). Although only positive values occur for this parameter the scaling of the y-axis has been deliberately chosen to match that in the previous figure.

We believe that the strikingly early completion of glottal adduction speaks against modelling the glottal movement pattern as the overlap of fricative and plosive gestural activation. If the basic glottal opening pattern for plosives in German is assumed to involve glottal opening at release then even if it is assumed that the plosive here occurs in a structurally much weaker position in the syllable onset than the fricative then completion of glottal adduction well before release (as much as 40ms in the case of speaker CG) seems an implausible scenario. We believe that a more parsimonious scenario simply involves assuming that speakers are very well able to learn the relations between movement and its aerodynamic consequences of the kind outlined in the previous paragraph. Given that plosives in such clusters are essentially unaspirated in German the crucial aerodynamic task for the speaker is to abduct early enough for the fricative. Once the fricative is up and running the precise glottal movement pattern has few constraints. Speakers could, for example, take advantage of this freedom by lengthening the adduction phase relative to the abduction phase. This could explain the above finding for the abduction-adduction ratio, which - as just pointed - might at first sight seem consistent with a gestural overlap interpretation. A further prediction that could be made from the line of reasoning being put forward here, is that movement patterns may well vary over speakers. Specifically, one could predict that speakers may latch on to different habitually preferred timepoints for the completion of glottal adduction in such clusters. From this would also follow that variation in the abduction-adduction ratio could be expected. Obviously, we cannot conclusively document this prediction with only three speakers, but at least the behaviour found in our speakers is not inconsistent with it: thus we found one speaker, CG, with noticeably earlier completion of glottal adduction than the other two speakers, and for him also a different preferred region for the abduction-adduction ratio[42].

Preliminary conclusions

This section has shown that fricative-plosive sequences do not differ robustly in terms of overall glottal gesture duration or amplitude from single fricatives. The only finding that gave a prima facie indication of overlap of two underlying gestures related to segments differing in their dominance of glottal articulation was the shift in the abduction-adduction ratio. However, it was proposed that the sometimes very early completion of glottal adduction argued against an underlyingly separable fricative- and plosive-related component. Similarly, it was found that the relative position of PGO in the fricative shifted quite reliably between singletons and clusters. However, rather than indicating a blending process of two underlying gestures this appeared to be to a large extent a passive result of the change in fricative duration. This also raised the question as to whether the timing of PGO may be less central than is often assumed for the expression of coordination relations, an issue that will recur in subsequenct sections.

A postscript on fricative-plosive clusters in Berber

As a postscript to this section it is perhaps interesting to point out that the finding of early completion of glottal adduction in fricative-plosive clusters is the timing pattern that German appears to make use of, but that in superficially similar clusters in other languages the coordination pattern may well be different. This is illustrated here for two speakers from recent work on Berber (Ridouane et al., 2006). In the initial fricative-plosive clusters we investigated in this language it was reliably found that glottal adduction is not completed until somewhat after release of the plosive, with noticeable aspiration of the plosive. The general shape of the glottal movement pattern is nevertheless remarkably similar to that found in our German data with the same tendency to a slightly longer adduction than abduction phase (and note also that peak glottal opening occurs well after the midpoint of the fricative).

---

[42]The report of Jessen, 1998, of aspirated plosives and dual glottal opening peaks in German /**ʃp**/ clusters (cf. Chapter 2, p. 16, above) could be seen as simply representing the extreme end of this potential for variability, and speakers' ability to adapt flexibly to the communicative demands, but not as indicating that two underlying gestures are *in general* present. As already mentioned in Chapter 4, p. 44, in the section on segmentation, no cases of double-peaked glottal movements were found for any of the consonant sequences in the present material, not even at the loud speaking volume, which presumably represents a clear style of speech. This is consistent with the vast majority of data for comparable languages in the literature.

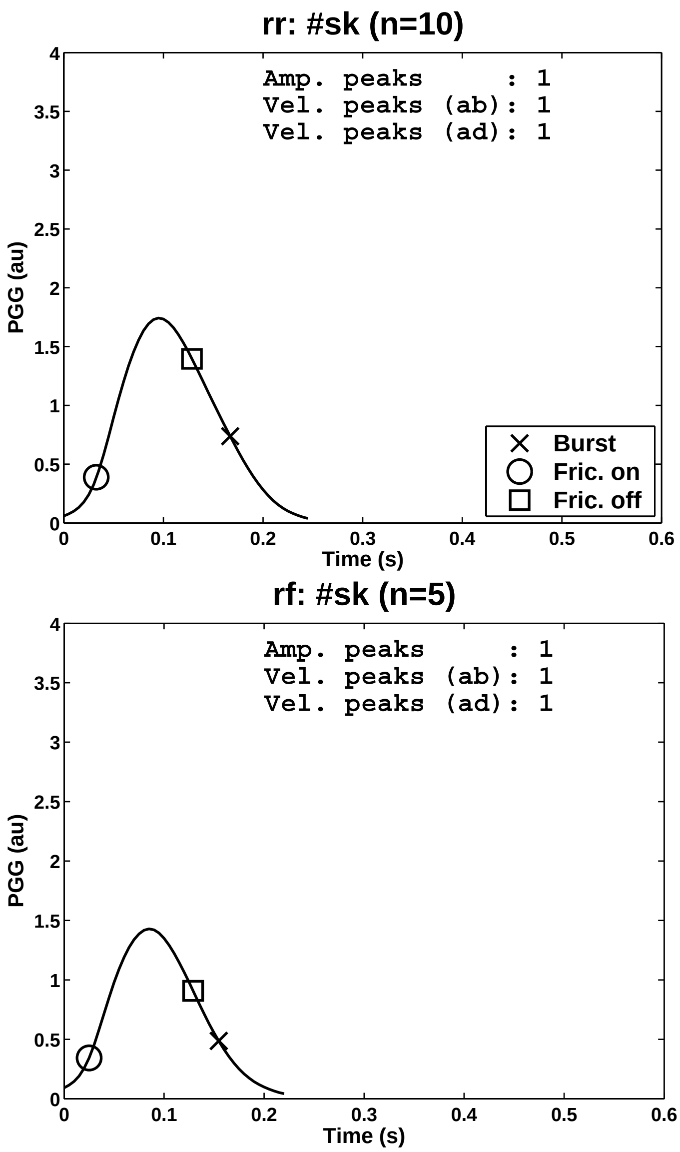


Fig. 6.23: Ensemble-averaged glottal activity for word-initial /**sk**/ sequences of two speakers of Berber (from Ridouane et al., 2006, Fig. 16.2). Release of /k/ is marked by 'X'.
(The 'peak counts' indicate how well the averaged patterns reflect the raw data: they represent the average over individual tokens. Since all peak counts are precisely 1 there were simply no exceptions from a kinematic pattern consisting of 1 amplitude peak, 1 velocity peak for abduction and 1 velocity peak for adduction.)

6.2.2 Plosive-fricative sequences

Introduction and notes on material

We will discuss the structure of plosive-fricative sequences principally via comparison with the single plosives. This appears a priori the most apposite comparison simply because the initial element in a cluster can be considered the dominant one. These two groups also match well in the more incidental sense that they are both particularly well represented at both loudness levels. This gives a clear and manageable factorial design for the bulk of the statistical tests that we well present (2 consonant categories x 2 volume levels). However, in order to relate the plosive-fricative results to the other categories, and in order to provide a first overview of all the material that has been presented so far in this investigation, most of the figures of individual parameters will include not only the plosive-fricative and single plosive categories, but also the single fricatives and fricative-plosive categories (usually in parallel for normal and loud volume level).

There is an important exception to the basic procedure of comparing plosive-fricatives with plosives. As part of the key issue of elucidating what time-points in the glottal gesture are the crucial ones for formulating coordination relations with the oral gestures we will look not only at timing of PGO relative to the initial plosive but also with respect to the fricative, in view of the important role that has been assigned to the fricative in this respect. The single plosive material is obviously irrelevant, so the comparison will be with the other consonantal onsets that include fricative segments.

Two remarks on the material are necessary at the start of this section.

The first concerns the plosive-fricative combinations themselves. For these consonant sequences a large amount of material was recorded, but it is also rather heterogeneous compared to the other categories. First of all there are two monosyllabic words (“Psi” and “Tschüss”), and secondly there are several words that are somewhat peripheral to native German vocabulary (“Psi”, “Psyche”, “Zyste”, possibly also “Zypern” and “Chile”; for “Chile”, single fricative as initial consonant also occurs and appeared to be the preferred form for speaker RW). These are clearly factors that could have some influence on the temporal and spatial characteristics of the initial consonants. Accordingly, the statistical tests were also carried out for a reduced version of the material, eliminating the items just listed (giving two /**pf**/ and two /**ts**/ word-forms); the corresponding results are given in the tables below where they differ from those based on the complete material.

The number of word-forms accordingly is:

Complete material:	10 at normal volume	7 at loud volume
Reduced material:	4 at normal volume	4 at loud volume

The second remark on the material concerns the fricative-plosives and fricatives compared in the previous section and included in the figures in this section: While the data values for fricative-plosives are identical in the previous and present section the values for fricatives are slightly different since the present section uses all fricative material (/**ʃ**/ and /**f**/) whereas the previous section used /**ʃ**/ only. In addition, the present section shows data for loud and normal volume utterances in parallel, but the data for fricative-plosive at loud volume should be viewed with caution as the number of tokens is very limited (for this reason it was not discussed explicitly in the previous section). In fact, the slightly different material in the present section does not suggest any noteworthy qualification of the trends identified in the previous section, thus supporting their robustness.

Overview of temporal structure

As in previous sections, we will preface this section with figures showing the overall temporal structure of the utterances involved. Since preliminary inspection of the data indicated that volume level had a fairly minor influence on the temporal structure, for the purpose of these illustrations the data were averaged over both levels (detailed results of individual parameters will consider volume levels separately, however). In the following figure, the large panel shows results for the complete material, based on just the two categories plosive-fricative and plosive, while the two smaller panels are based on the more restricted material and divided according to place of articulation, giving /p/ vs./pf/ on the left and /t/ vs. /ts/ on the right.

Features that can be pointed to, as a framework for the detailed results below, include:

- Longer duration of voicelessness (blue bars) for clusters than singletons.

  This is hardly a surprising result, but we will need to consider below whether it is really due to a longer glottal gesture duration (green bars). These overview figures suggest that a relevant factor is also where voicing starts relative to the end of the glottal gesture (end of blue bar relative to end of green bar). Generally the position appears to be relatively later in the clusters[43].
- Time-point of peak glottal opening later relative to the end of the plosive occlusion in the clusters compared to the singleton plosive

  In the singletons PGO is usually just before the end of the plosive occlusion, while in clusters it is usually located in the early part of the fricative. As with the fricative-plosive clusters one of the main points to consider is whether peak glottal opening can be considered as closely linked to a specific point in the fricative, or whether, for example, its location simply varies passively with changes in the duration of the plosive occlusion. One would probably expect plosive occlusion duration to be shorter in the clusters than

[43]There is actually an important methodological point here: Since the temporal relationship between voicing onset and completion of the glottal gesture can vary over sound categories purely acoustic measures of the duration of voicelessness must be interpreted with caution as evidence for the duration of the underlying glottal gesture.

A particularly drastic dissociation of voicing onset from the end of the glottal gesture has already been seen in the previous section in the discussion of fricative-plosive sequences: completion of the glottal gesture well before release of the plosive and consequently even longer before the onset of voicing.

We will not go into any details of the duration of the interval from end of the glottal gesture to onset of voicing with respect to plosive-fricatives vs. plosives. Significant differences were found for all 3 speakers, though differing somewhat in strength over speakers and different subsets of the material (as can be seen to a certain extent even in the figures shown here). The important point is simply that quite clearly differences between sound categories can potentially occur. As a footnote to this footnote: This is, in turn, certainly not surprising given the expected difference in oral constriction and intraoral air-pressure towards the end of the glottal gesture in aspirated plosives compared to fricatives.

in the singletons. This will be considered in detail below; from the overviews shown here it is not completely clear whether this is really the case.

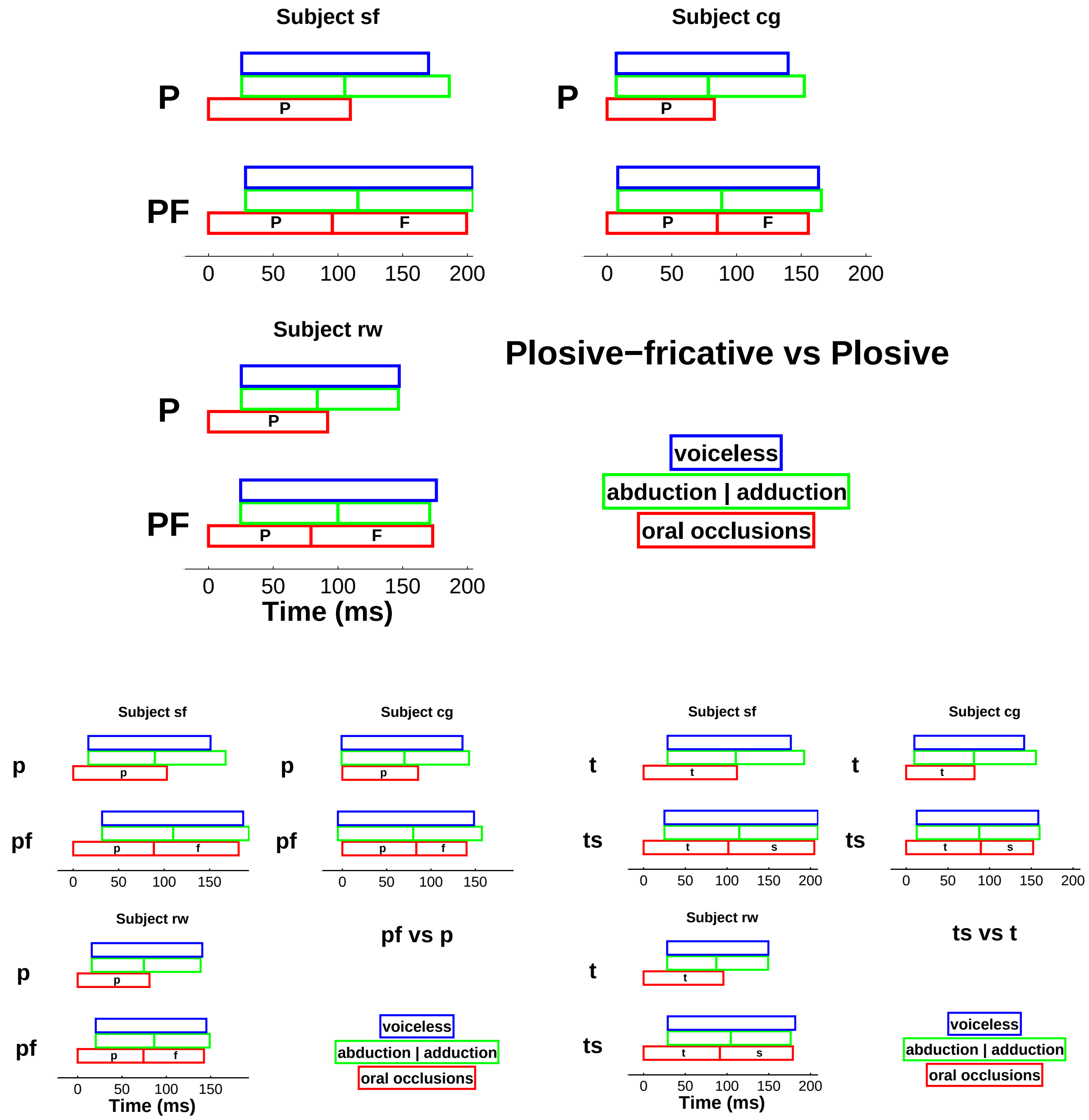


Fig. 6.24: Overview of temporal structure of plosive-fricative sequences, and single plosives. Top: complete material (upper case labels ‘P’ and ‘PF’ indicate ‘Plosive’ and ‘Plosive-Fricative’, respectively). Bottom: reduced material, subdivided by place of articulation and labelled phonemically (labial left, alveolar right) . Data in all panels represents average over both volume levels

Detailed results for individual parameters

The following parameters will be presented in this section:

- Glottal gesture duration
- Amplitude of peak glottal opening
- Ratio of duration of glottal abduction to adduction
- Duration of interval from start of plosive to start of glottal abduction

- Relationship between interval from peak glottal opening to release of plosive and duration of plosive occlusion
- Relative position of peak glottal opening in the fricative segment (this compares the plosive-fricative sequences with fricative articulations from previous sections)

The tables summarizing for all parameters firstly the ANOVA results and secondly the pairwise comparisons for plosive-fricative versus plosive matched for loudness condition are to be found on p. 67

Again, as in previous sections we will look first at the duration of the glottal gesture, as one of the most basic characterizations of the laryngeal activity. The figure makes it apparent that the plosive-fricative combinations share very similar durations to the single fricatives, and to the fricative-plosive combinations, but appear to have somewhat longer duration than the single plosives. It will be recalled from previous sections that fricative-plosives and fricatives did not differ consistently while single fricatives and plosives did. So the main question here is how robust the longer duration for plosive-fricatives compared to plosives is. In the ANOVAs clearly significant results were found for all subjects for both the complete and the restricted material, even though the magnitude of the differences is not enormous: usually of the order of 10-20ms. In the non-parametric direct pairwise comparisons of plosive-fricative and plosive all subjects had at least one comparison reaching p=0.01 for the full material, but no comparisons quite reached this level for the restricted material.

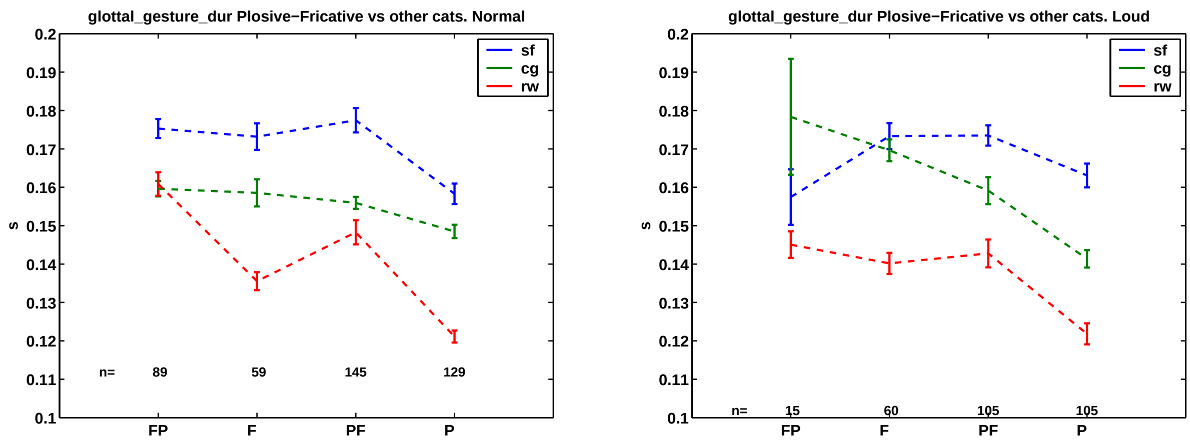


Fig. 6.25: Glottal gesture duration for four consonant categories. Labels on abscissa: FP = Fricative-Plosive; F = Single Fricative; PF = Plosive-Fricative; P = Single Plosive. Normal volume level in left panel, loud volume in right panel.

Turning to peak glottal opening amplitude and picking up again on results from previous sections it will be recalled that evidence for a consistent difference between fricative-plosives and fricatives, or between fricatives and plosives, was not found. Inspection of the current figures indicates a tendency to greater glottal opening for the plosive-fricative than plosive at normal volume for all three subjects, and at loud volume for two. The ANOVA did give a significant result for SF and CG (albeit with a significant interaction with volume level for SF), however the non-parametric test (which, as repeatedly emphasized, we find more apposite for analysis of glottal opening) gave no pairs reaching p=0.01. Given that all results for RW were non-significant we feel that on balance there is little evidence for a robust difference in glottal opening, thus continuing the general picture found for this parameter with the other consonant categories.

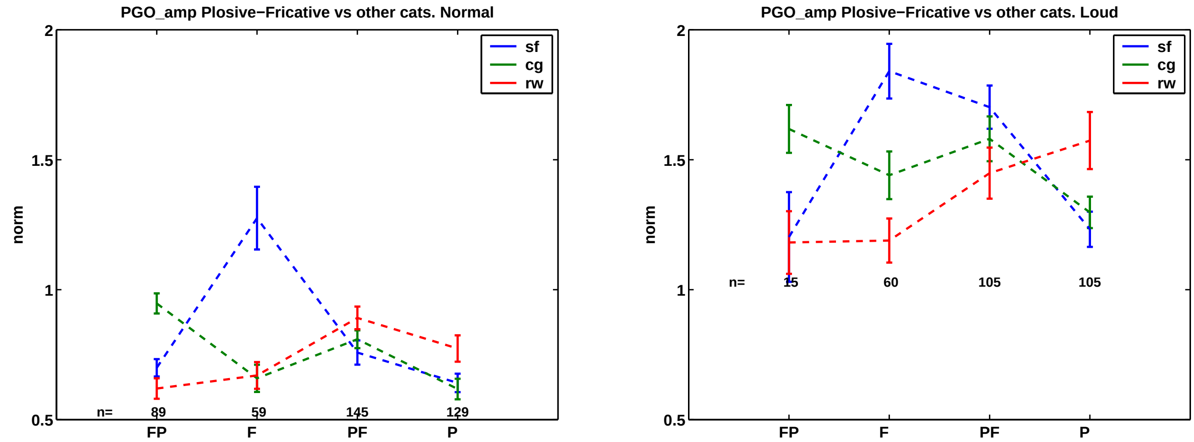


Fig. 6.26: Peak glottal opening amplitude for four consonant categories. See Fig. 6.25 for details.

The next parameter to consider is the ratio of glottal abduction to adduction duration; at the same time it is the last one in which the glottal gesture is considered in isolation, i.e without being put in relation to aspects of the supraglottal activity. As seen in earlier sections the shape of the glottal gesture can provide an indication of hypothetical underlying gestural input, or at least that speakers adapt the movement pattern to changing aerodynamic demands. Since plosive-fricative combinations are the mirror-image of fricative-plosive combinations, and since fricative-plosives showed a fairly clear tendency to have the lowest values of the abduction-adduction ratio, then a possible expectation is that plosive-fricatives will show the most extreme skewing in the opposite direction. In terms of the aerodynamic demands there are very few constraints on the abduction gesture: there is simply the necessity for sufficient abduction at the start of the fricative to generate frication; the peak opening need not have been reached by then. Completion of abduction does have to be quite well coordinated with the end of the oral occlusion for the fricative, however, and cannot show the kind of anticipation of the end of the consonant as we found for the plosive in fricative-plosive combinations.

As for the two previous parameters in this section we compared the plosive-fricatives with the plosives in the statistical tests. Inspection of the figure below indicates that the effects are not very substantial, and also not consistent across the subjects. CG and RW in fact show an effect in the hypothesized direction that is significant at $p<0.001$ in the ANOVA (the criterion level in the paired comparisons is however only reached by RW for PF_N > P_N[44], and then only in the full, not in the restricted material). For SF, on the other hand, there is no significant difference, with not even an indication of a trend in the expected direction.

In addition, it will be recalled from the previous section comparing plosives and fricatives that one of the main reasons why the expected pattern of lower ratios for fricatives was not found, was that speaker CG showed unusually high values for the fricatives. Thus, the functional significance

[44] '_N' and '__L' will be appended to the sound categories to indicate the loudness condition

of his significantly higher ratios for plosive-fricative than plosive is unclear given that over all four categories shown in the figure below he still has highest ratios on the single fricatives.

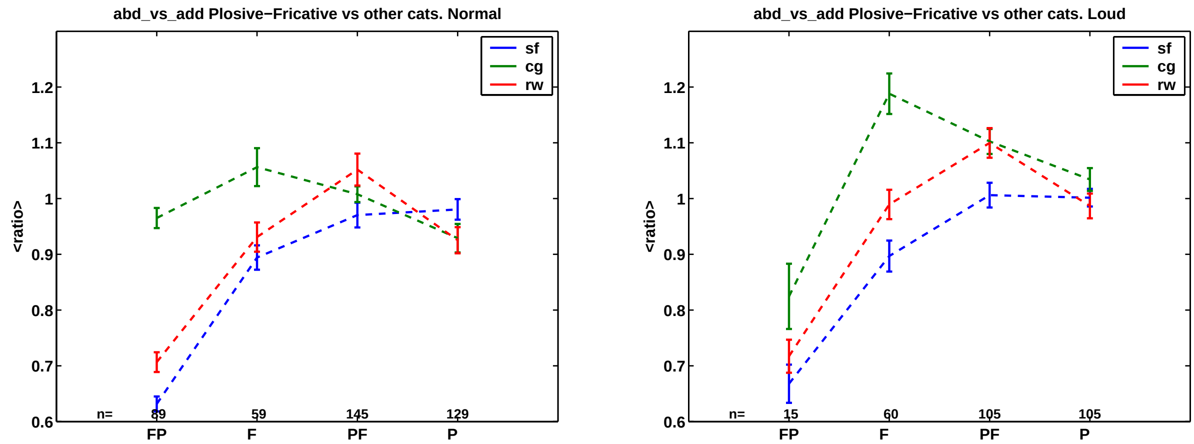


Fig. 6.27: Ratio of duration of glottal abduction to adduction for four consonant categories. See Fig. 6.25 for details.

Summarizing the findings for all these four categories then, it appears that the most robust finding is for low values of the ratio on fricative-plosive combinations, whereas within the other three categories (fricative, plosive-fricative, plosive) only relatively slight and/or inconsistent differences occur.

The first parameter in which we consider the timing of glottal events relative to oral ones is the duration of the interval from onset of oral occlusion to onset of glottal abduction (see next figure).

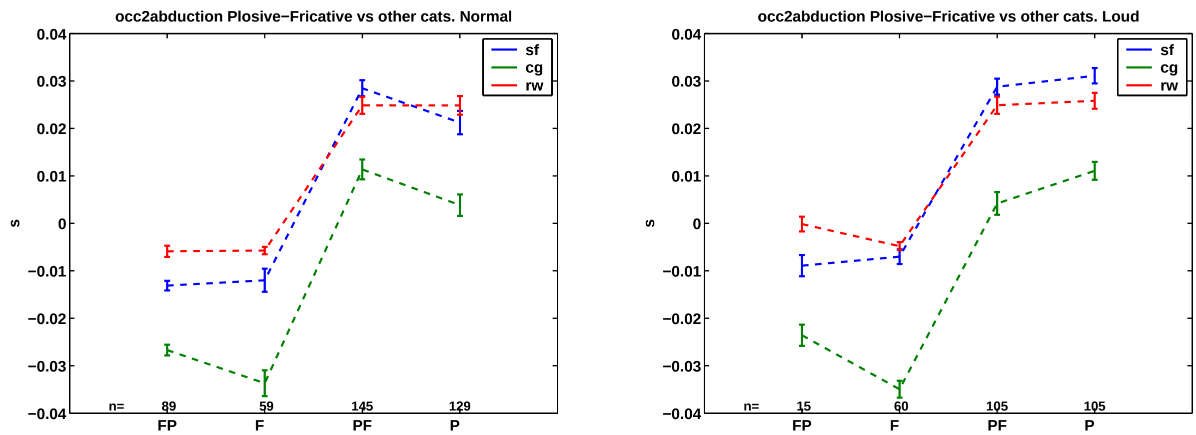


Fig. 6.28: Duration of interval from start of first oral occlusion to start of glottal abduction for four consonant categories. See Fig. 6.25 for details.

The interest of the parameter at this juncture is not with regard to any specific features for plosive-fricative combinations, but rather with regard to the pattern over all four consonantal categories. It was emphasized in the first section, in which fricatives and plosives were compared, just how clear-cut the distinction was there, making this parameter a potentially important one for understanding key control features. The figure above reveals that the distinction remains just as clear-cut when all four categories are considered, i.e fricative-onset vs. plosive-onset categories are perfectly separated. We had no particular hypothesis as to whether finer distinctions should occur within the plosive-onset group: in fact absolutely no significant effects occur[45].

Similarly, in the previous section on fricative-plosive vs. fricative this parameter was not explicitly analyzed, simply because we had no specific hypothesis to test concerning these two categories. The present figure confirms that indeed there are no consistent differences between them.

Following glottal abduction, the next aspect of coordination to discuss is the timing of peak glottal opening. This requires more extensive consideration. We will proceed from two main points of view, firstly comparing plosive-fricatives and plosive, and, secondly reviewing the relative position of peak glottal opening in the fricative for the categories single fricative, fricative-plosive and plosive-fricative.

Timing of peak glottal opening (1): plosive-fricative vs. plosive

Regarding plosive-fricative vs. plosive the feature apparent from the illustrations of the temporal structure displayed at the start of this section is that peak glottal opening shifts from typically just before the end of the plosive in single plosives to somewhat after the end of the plosive in plosive-fricatives (i.e in the early part of the fricative). Here we will consider how robust this difference is. Of more interest for understanding of control mechanisms, however, is the extent to which any shift in the location of peak glottal opening can be directly attributed to changes in the duration of the initial consonant (which one could expect to be shorter in the cluster). This is thus a counterpart to the approach followed with regard to place of articulation effects in single consonants.

The basic statistical results for the duration of peak glottal opening to release and the duration of the plosive occlusion are as follows:

For peak glottal opening to release the difference between plosive and plosive-fricative was significant at $p<0.001$ for SF and RW. For CG the difference was significant at $p<0.01$ for the complete material, but not significant for the restricted material. These differences between the subjects were reflected in the results of the pairwise comparisons: for SF and RW both the PF_N vs P_N and the PF_L vs. P_L pairs reached $p=0.01$ (regardless of the selection of the material), confirming the robustness of the difference for these two speakers, whereas for CG only one comparison (PF_N vs P_N) reached this level for the complete material, and none for the restricted material. The differences in the position of peak glottal opening relative to the release were of the order of 25ms for SF and RW, and under 10ms for CG.

The differences in plosive duration between single plosives and plosive-fricatives were smaller than the differences in peak glottal opening to release: of the order of 10-15ms for SF and RW, and negligible for CG. This is by and large reflected in the significance levels for the statistical tests: only for SF was the difference significant at $p<0.001$ for both the complete and restricted material. For CG no significant differences were found. Over the three subjects the number of significant pairwise comparisons also reduced compared to the PGO-to-release results (see tables on p. 67 for details). Since we are particularly interested in the relationship between these

[45]Just conceivably, plosive-fricatives might be expected to attract even longer intervals, i.e delayed glottal abduction, because of the weak constraints on the time-course of the abduction phase alluded to above.

parameters we will proceed directly to a corresponding scatterplot based on each relevant word category, and refrain from showing further details of these parameters in isolation.

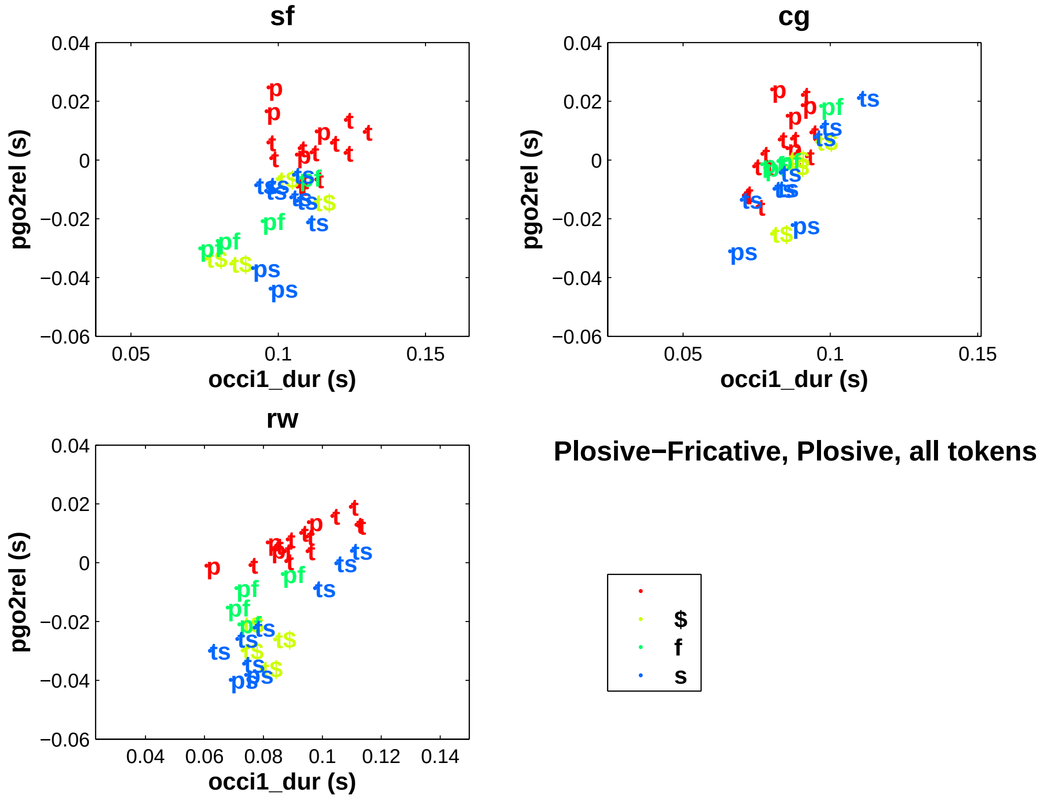


Fig. 6.29: Relationship between interval from peak glottal opening to release and duration of plosive occlusion in single plosives and plosive-fricative combinations. Identity of C2 is colour-coded. Red indicates single consonant, i.e no C2. ‘$’ = /ʃ/. Both volume levels.

For all subjects there is clearly quite a strong positive relationship (as in earlier sections, the scatterplots are shown with equal scaling for x and y axes, to make it easier to gain a visual impression as to whether there is a one-to-one tradeoff between occlusion duration and PGO-to-Release). The correlation coefficients over the complete material (taking singletons and clusters together) amount to 0.59 (SF), 0.67 (CG) and 0.71 (RW).

More interesting than the overall correlations, however, is the question whether data for the singletons (red data points in the figure) lie on a different regression line than data for the clusters (all colours except red). If this were the case, it would indicate that any shifts in timing of PGO going from singletons to clusters are not just a spin-off from the changing occlusion durations, but could reflect more active gestural changes on the part of the speaker. This approach looks potentially most revealing for CG and RW since they have a wide range of occlusion durations for C1 in the clusters (probably reflecting the heterogeneous nature of the corpus, and thus leading to the at first sight unexpectedly small differences in occlusion duration between singletons and clusters). Visual inspection of the plots however indicates that the regression line for the clusters could be shifted down on the y-axis relative to the singletons, i.e at a given occlusion duration the cluster will attract a lower value of PGO-to-Release than the singleton. This can be tested with analysis of covariance[46].

---

[46]Here we encounter again the flip side of the coin with respect to the heterogeneity of the corpus: it does not just muddy the waters. The diversity of the material should actually

For all speakers statistically significant differences were indeed found for the relationship between plosive duration and PGO-to-release for singletons vs clusters: the intercepts differed at $p<0.01$ for SF, $p<0.05$ for CG and $p<0.001$ for RW. The slopes also differed for SF and RW ($p<0.01$ and $p<0.05$, respectively) but not for CG.

These results are of least import for SF, since she shows the smallest amount of overlap in occlusion durations between singletons and clusters, and - at least for singletons - she has only a very weak relationship between occlusion duration and PGO-to-release in the first place. However, for CG and RW, where occlusion durations overlap substantially, it is useful to be able to show that there is *not* a common function relating occlusion duration and PGO timing over the complete plosive-initial material. In other words, other more active gestural changes must influence this timing parameter, and in fact we already know which they are: gestural duration had proved relevant for all speakers, and the abduction-to-adduction ratio may have some role to play for CG and RW.

Timing of peak glottal opening (2): all syllable onsets containing fricatives

The second approach to assessing the timing of peak glottal opening is at the same time a summary of all the material involving fricatives considered up to now, i.e how does the relative position of PGO in the fricative segment vary over the categories fricative, fricative-plosive, and plosive-fricative. (As an exception to the other analyses in this section single plosives are now, of course, irrelevant, but we will retain the format of the previous figures in this section showing all four categories.).

The results are shown in the next figure (as in the comparable figures in this section separately for normal and loud material).

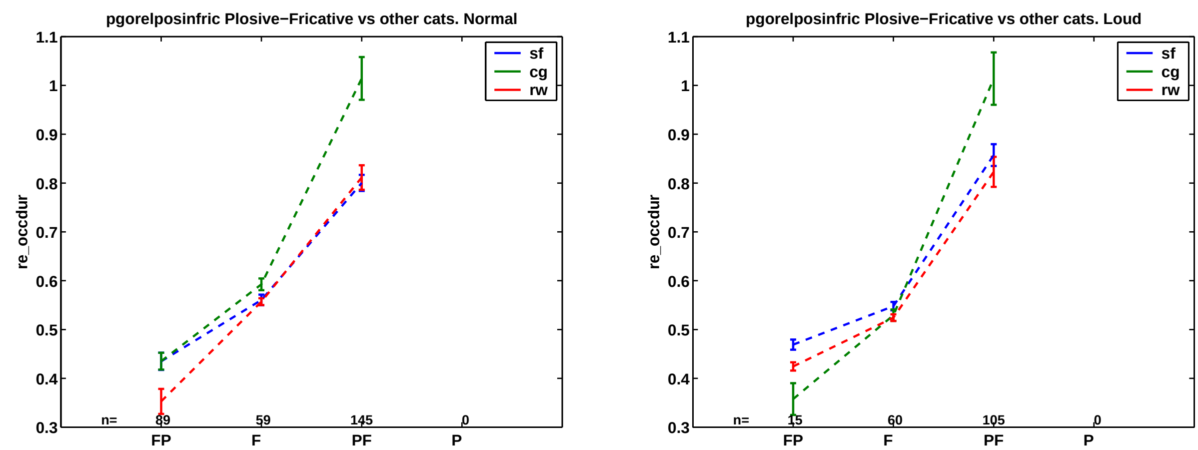


Fig. 6.30: Relative position of peak glottal opening in fricative segment. Higher values indicate earlier locations (1 = fricative onset; 0 = fricative offset). Since the same arrangement of the abscissa has been retained as for comparable figures in this section there is an empty position corresponding to material with single plosives. See Fig. 6.25 for other details.

The figure makes clear - as even a cursory inspection of the overview of the temporal structure also shows - that peak glottal opening is timed very early in plosive-fricative sequences; for CG

lead to better estimates of the regression lines for the cluster vs. non-cluster groups.

some cases were even encountered where PGO occurs in the preceeding plosive. The differences between the three categories are obviously quite clear-cut. This agrees with the small amount of data available in the literature for such sequences. Yoshioka et al. (1982) included the sequence /**ps**/ (also in our corpus) in their investigation of Dutch. Although precise figures are not given, it is stated that timing of peak glottal opening "*approximately coincides with the timing of the oral release for the stop /p/*" (p. 31). In our recent study of Berber just referred to above, the sequence /**ks**/ was investigated; peak glottal opening was typically just after the onset of frication for the /**s**/ (Ridouane et al., 2006, Fig. 16.3).

At this juncture it is necessary to take up a methodological point made in Chapter 4 (p. 50), namely that estimating the timing of peak glottal opening relative to the *acoustic* manifestation of the fricative may not be entirely appropriate. For a sequence like /**ps**/ the constriction for the /s/ may be achieved before the release of /p/, and thus not audible; a different picture of the coordination relations could emerge if the articulatorily rather than acoustically defined period of constriction for the fricative is used. Specifically for the sequence /**ps**/ it was pointed out in the earlier chapter that the EPG data could be used to check this idea. In fact, it turned out for all speakers that the articulatorily defined onset of the /**s**/ constriction was within 5ms of the burst of the preceding /**p**/, so the effect of taking this into account for quantifiying the coordination relations would be absolutely negligible[47].

A summary of the pairwise comparisons matched for loudness level among the three sound categories containing a fricative segment in the syllable onset is given in the following table.

	Normal			Loud		
	SF	CG	RW	SF	CG	RW
FP vs F	x(see note)	x(see note)	x			
FP vs PF	x	x	x	x	x	x
F vs PF	x	x	x	x	x	x

Table. 6.1: Breakdown of pairwise comparisons of relative position of peak glottal opening in fricative for the three sound categories containing a fricative segment. Cells containing 'x' indicate pairwise comparisons reaching p=0.01 in nonparametric test.
Note: By a quirk of the non-parametric procedure in the two indicated cases the criterion was only reached when the data selection used the restricted material for plosive-fricative. With the unrestricted material the criterion was just missed, even though the data for the actual pair compared is exactly the same in both cases. For all other cells in the table the choice of restricted or unrestricted plosive-fricative material had no impact on the result.

[47]Since closer examination of this point does not appear necessary we have not included a figure explicitly showing the timing relations for both articulatorily and acoustically defined onsets of the fricative segment. If desired, reference can be made forward to the illustration of the temporal structure of /**pl**/ sequences (Fig. 6.31 on p. 112), which shows the time-point of the onset of the /**l**/-constriction as determined from EPG. For speakers SF and CG the /**l**/ shows slightly more overlap with the preceding /**p**/ than does the /**s**/ in /**ps**/. Speaker RW shows very slightly more overlap for /**s**/.

Plosive-fricative sequences are clearly robustly different from both fricative-plosive sequences and single fricatives. The differences between the latter two categories are clearly smaller in magnitude. Statistical results for them have already been presented in the previous section on fricative-plosive vs. fricative, where the differences were found to be in fact quite consistent and reliable. For normal volume speech, the pairwise differences hover around the criterion of p=0.01, so whether the criterion is met or not depends somewhat on the precise selection of material examined (see further discussion in the previous section (p. 92f), and in the note in the caption of the table below). The fact that the criterion is not met at loud volume is not particularly important given the very small number of tokens for fricative-plosive in this condition; the figure makes clear that the loud data show the same general tendency as the normal data.

Final notes on plosive-fricative vs. plosive

Returning to the immediate subject of this section, namely the comparison of plosive-fricative and plosive (PF vs. P), this can conveniently be rounded off by summarizing some further points of comparison with the pair just discussed: fricative-plosive vs. fricative (FP vs. F):

- For PF vs. P the difference in C1 duration between the members of the pair was smaller than between the members of the FP vs. F pair (p. 105f and p. 91ff)
- On the other hand, there was quite a consistent difference in glottal gesture duration for PF vs. P but not for FP vs. F (Fig. 6.25 on p. 102 , also Fig. 6.15 on p. 90).
- Related to first bullet point: The relationship between occlusion (C1) duration and duration of PGO-to-release was of course examined for FP vs F, but the ANCOVA analysis used for PF vs P was not employed because of the much smaller amount of material for FP vs F, and because the groups were clearly separated by occlusion duration rather than overlapping (Fig. 6.29 on p. 106 and Fig. 6.19 on p. 93).

Before moving on to the last section of results in which we include clusters with /l/ in the discussion we will give a preliminary review of the results up to now in the light of the models of laryngeal-oral coordination that have been at the centre of discussion.

Taking stock of the results to this point

The model of Browman and Goldstein attributes a major role to the fricative if one is present in the syllable onset. Their basic rule assumed coordination of PGO with fricative midpoint. In fact, they (Goldstein, 1990) also envisaged the possibility that PGO could be perturbed away from the midpoint depending on the phonetic make-up of the syllable onset. However, when one considers the massive variations in relative position of PGO in the fricative shown above, then it seems to be a plausible conclusion that the coordination is simply not formulated in terms of the timing of PGO, i.e the variation goes beyond perturbation of a basic underlying coordination relation (recall that some cases were encountered where PGO is not located in the fricative at all)[48].

---

[48]Regarding the timing of peak glottal opening for e.g /**ps**/ vs. /**sp**/, Saltzman & Munhall (1989) give the following account in terms of dominance relations: ".. *when the inherently stronger /s/ is augmented by word-initial status in /#sp/, the glottal peak cannot be perturbed away from midfrication by the following /p/. However, when the intrinsically weaker /p/ is word-initial, the glottal peak is pulled by the /p/ from midfrication to the*

At this juncture, we would prefer a formulation more along the following lines:

Speakers obviously time the start of glottal abduction precisely, clearly distinguishing between fricatives and plosives in initial position.

For plosive-fricatives the completion of glottal adduction must be timed to match the end of the fricative occlusion (unlike for fricative-plosives it would not be aerodynamically acceptable for adduction to substantially anticipate the end of the occlusion). Accordingly, speakers plan the duration of the glottal gesture to take into account the expected overall length of the oral occlusions. Within this framework, PGO is not planned directly but emerges from the basic constraints on onset and offset of the gesture, with the precise position relative to the fricative depending purely incidentally on the varying duration of C1 for the different clusters.

Although differing from Browman & Goldstein's model in terms of how the coordination is formulated, this account for plosive-fricatives (and also for fricative-plosives) is consistent with their general point that syllable onsets in English (and here presumably then for German) are characterized by one glottal gesture. As repeatedly emphasized this is the main point of difference from an account based on overlap of gestures related to each element in the onset. In the present specific case of plosive-fricatives there is perhaps actually slightly more evidence for gestural overlap than for fricative-plosives, since a fairly consistent lengthening of the gesture and (to a lesser extent) a shift in its shape compared to plosives was found[49]. Thus the results certainly do not come down unequivocally in favour of our preferred account, namely that planning is based around the aerodynamic requirements of the cluster as a whole. One of the main aims of the next section is to determine whether consideration of the additional sonorant cluster material will tip the balance in its favour.

---

closure-frication boundary" (p. 370).

[49]Perhaps this formulation is unduly cautious: The differences between plosive-fricatives and single fricatives are smaller than the differences from single plosives, so taking fricatives as the basis for comparison gives less suggestion of gestural overlap.

6.3 Results 3: Consonant combinations with /l/

Introduction

This section will consider in turn each consonant combination with /**l**/ by comparing then with the corresponding consonant or cluster without /**l**/. Since all the combinations with /**l**/ occur in the cluster part of the corpus there are no matching loud volume utterances available. In fact the material in this part of the study is fairly restricted in size. The four subsections (shown with an indication of the number of word-forms for the combinations with /**l**/) are as follows:

1. /**pl**/ vs. /**p**/ — 2 word-forms for /**pl**/
2. fricative-/**l**/ vs. fricative — 1 word-form for /**ʃl**/, 2 word-forms for /**fl**/
3. /**pfl**/ vs. /**pf**/ — 2 word-forms for /**pfl**/
4. /**ʃpl**/ vs. /**ʃp**/ — 1 word-form for /**ʃpl**/

At the end of Chapter 2 (p. 23ff) we reviewed the laryngeal-oral coordination patterns that appeared logically conceivable for consonant combinations with sonorants, pointing out that despite major differences between them it would be very difficult on the basis of previous investigations to identify the most likely candidate. Recapitulating briefly, the three basic possibilities envisaged were:

1. If the segmental durations of the voiceless segment(s) in the cluster reduce, then the duration of the glottal gesture would reduce in parallel, keeping the same coordination relations with respect to the voiceless segments.
2. The duration of the glottal gesture stays basically the same, while the durations of the individual oral occlusions change when a sonorant is added to the syllable onset
3. Regarded as the most radical possibility: The duration (and possibly the amplitude) of the glottal gesture actually increase in clusters with /**l**/.

One very simple reason why these clusters are intriguing is that the total duration of voicelessness may increase in the clusters with sonorant, and one would hope to be able to decide whether this is due - quite mundanely - to the aerodynamic conditions during the sonorant segment, or whether more active glottal mechanisms are involved. Since we have not performed direct aerodynamic measurements we will be concerned throughout this section to at least roughly estimate the role of aerodynamic conditions by observing where voicing re-starts relative to the time-course of the glottal gesture. Should it turn out to be necessary to assume the presence of active glottal adjustments, then the more far-reaching question that then arises is what the functional significance could be.

6.3.1 /pl/ vs. /p/

Overview of temporal structure

As in previous sections we first show an overview of the temporal structure (with start of the /**p**/ occlusion as the line-up point). Note that unlike previous figures of this type there is an additional light blue bar labelled “articulatory onset” for the /**pl**/ sequences. The left boundary of this bar is actually what gives it its name and indicates the onset of the /**l**/ tongue-tip occlusion as determined in the EPG data (refer back to the section on segmentation in Chapter 4, p. 51ff for further illustration; the right edge of the bar is redundant, being simply located for convenience at the end of /**l**/ segment as shown in the red oral occlusion bar). This additional information on timing of the oral gestures is principally of interest for comparing /**pl**/ and /**ʃpl**/ (Fig. 6.55, p. 134) with respect to the amount of overlap between /**l**/ and the preceding /**p**/. (For reasons given in Chapter 4 this approach cannot be used for /**fl**/ sequences. With regard to assessment of /**s**/ constriction onset in /**ps**/ sequences see p. 108 above.)

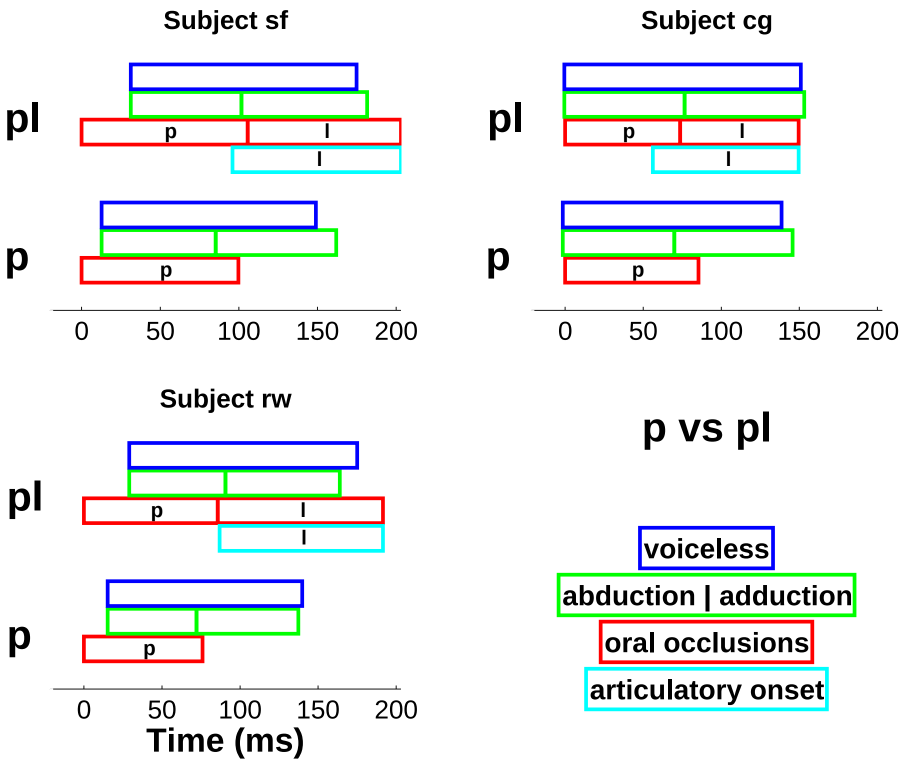


Fig. 6.31: Overview of temporal structure of /**p**/ and /**pl**/, line-up at start of occlusion for /**p**/. See text for further details.

Features of this display that can serve as orientation for the following discussion include:

- /**l**/ is substantially devoiced (this was expected)
- Peak glottal opening is later relative to /**p**/ release for /**pl**/ than for single /**p**/ (though whether this is actually statistically significant will not be seen til later)
- Duration of /**p**/ occlusion and glottal gesture duration do not appear to be massively different between /**p**/ and /**pl**/.
- Duration of interval from start of /**p**/ occlusion (left edge of red bar) to voice onset (right border of dark blue bar) appears longer for /**pl**/

Detailed results of individual parameters

We will now consider the individual parameters in detail. In several cases there will be an impression of inconsistent patterns over the three speakers but we hope to show that at the end of the analysis an interestingly consistent picture emerges that takes this variability into account and provides a fresh perspective on the original hypotheses.

The parameters that will be examined in this section are:

- Glottal gesture duration
- Amplitude of peak glottal opening
- Oral occlusion duration of /**p**/ segment
- Duration of interval from peak glottal opening to release of plosive
- Duration of interval from start of plosive to start of glottal abduction
- Duration of interval from release of /**p**/ occlusion to voice onset ("VOT")
- Duration of interval from end of glottal adduction to voice onset
- Total duration of voicelessness (stats and discussion but no figure)

The table summarizing the statistical results for each parameter is to be found on p. 68.

Looking first at the basic properties of the glottal gesture, the first figure immediately below indicates a somewhat longer glottal gesture for /**pl**/ for two subjects (CG and RW), but this was only significant for RW (at $p<0.05$).

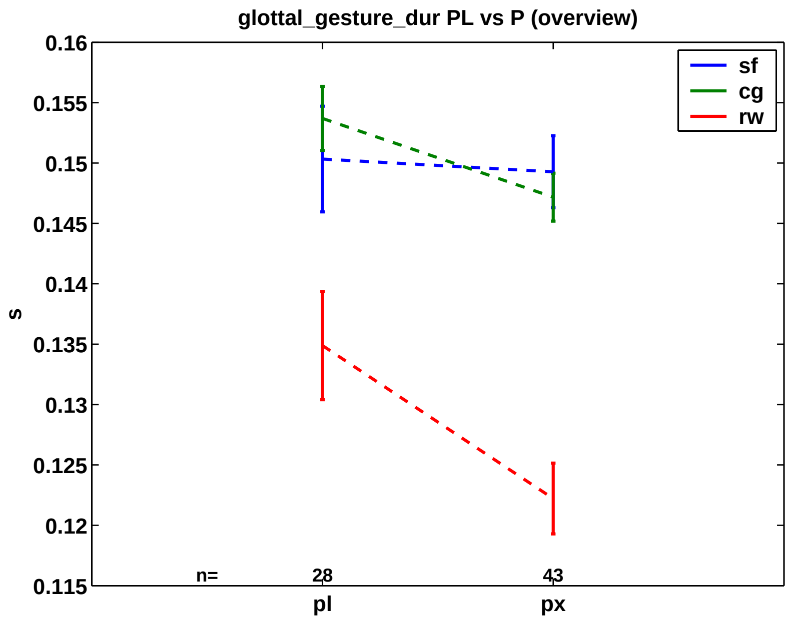


Fig. 6.32: Duration of glottal gesture for /**pl**/ vs. /**p**/. In this and the following figures in this section singleton /**p**/ is labelled 'px' on the abscissa, i.e as in many earlier figures 'x' stands for empty C2 position in syllable onset.

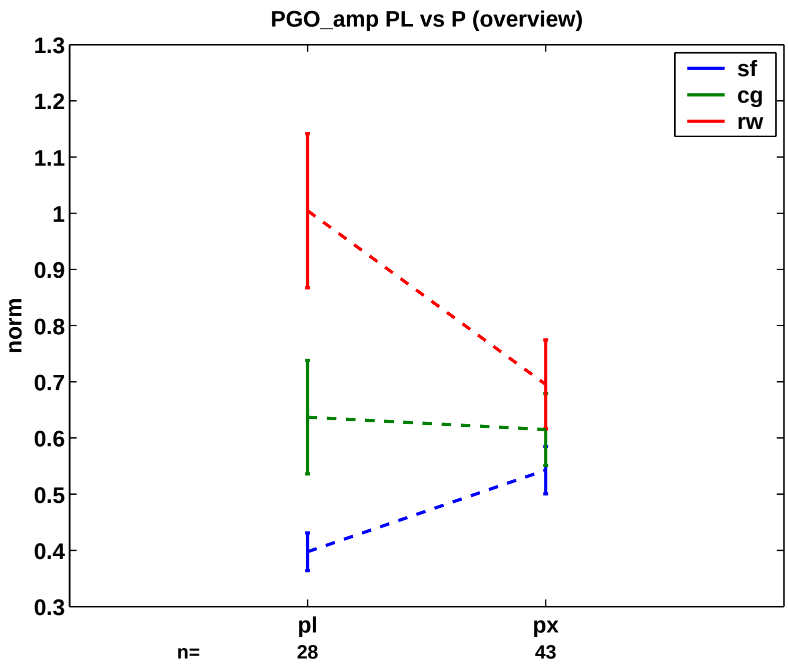


Fig. 6.33: Amplitude of peak glottal opening for /**pl**/ vs. /**p**/. Other details as in Fig. 6.32.

RW also had a tendency to a larger glottal opening amplitude for /**pl**/, but this was not significant in the crucial non-parametric test (and absolutely marginal in the parametric one: p=0.0500). Overall, results for glottal amplitude were mixed since SF had a weakly significant effect in the opposite direction (/**p**/ > /**pl**/), while absolutely no tendency was observable for CG. Thus while RW shows a tendency in the direction of the radical hypothesis of a more extensive devoicing gesture for /**pl**/ this is clearly in no way a general tendency, and even for him statistically not very robust. And in fact, consideration of the next parameter - duration of the occlusion of the initial /**p**/ - may provide a more prosaic interpretation for this tendency (see next figure).

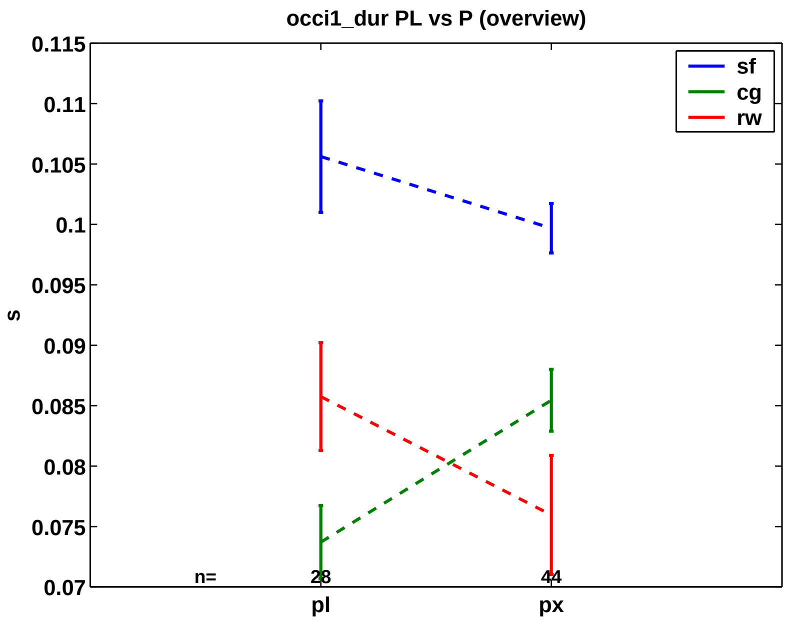


Fig. 6.34: Occlusion duration of /**p**/segment, comparing /**pl**/ vs. /**p**/. Other details as in Fig. 6.32.

The figure again shows a very mixed picture over the subjects, but only the shorter duration of /**p**/ in /**pl**/ for CG reached signficance (p<0.01, albeit only p<0.05 in the nonparametric test), while the apparently *longer* durations of /**p**/ in /**pl**/ for SF and RW in neither case reached sigificance. CG is thus the only speaker who would fit in the coordination model for clusters that is based on the idea of a basically constant glottal gesture superimposed on shifting oral articulations. In fact, it would seen that there is only a weak tendency for /**p**/ to shorten when occurring in a cluster like /**pl**/. Although some shortening was our original expectation, some of the literature reviewed in Chapter 2 (see e.g p. 10 and p. 13) did in fact also show only rather minor changes in /**p**/ occlusion duration[50]. Regarding the slightly (nonsignificantly) longer duration for /**p**/ in /**pl**/ for RW this may indicate simply a slightly more deliberate mode of speech for the /pl/ words (one of them ('Plüsch') also being monosyllabic) in turn also underlying the tendency to a longer glottal gesture.

If we now turn to the first coordination parameter, the duration of the interval from peak glottal opening to release (next figure), the patterns over the speakers suddenly look more similar: for all speakers the cluster has lower values, being generally slightly negative, i.e peak glottal opening is reached just after the release of the /**p**/ occlusion.

---

[50]This probably reflects the considerable scope for coproduction of the labial and apical gestures

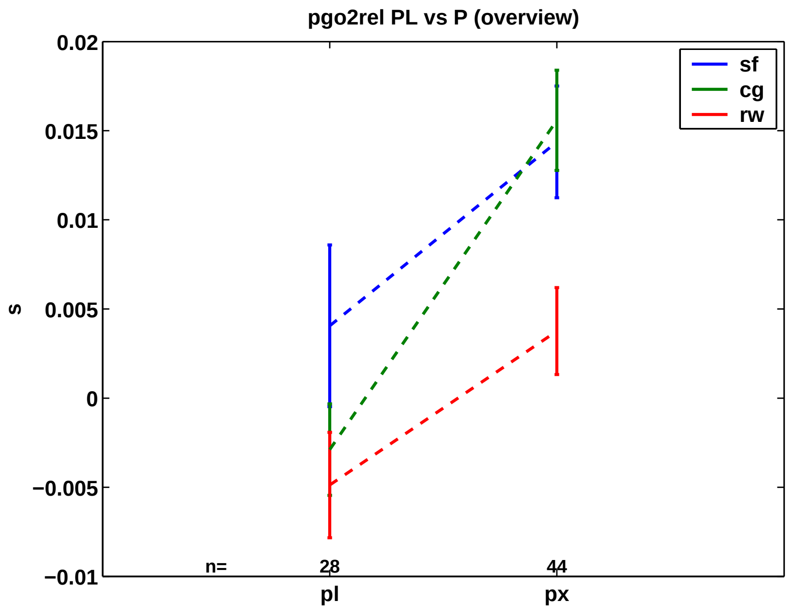


Fig. 6.35: Duration of interval from peak glottal opening to release of /**p**/ occlusion, comparing /**pl**/ vs. /**p**/. Other details as in Fig. 6.32.

The difference is not enormous - not much more than 10ms - but significant at $p<0.001$ for CG and at $p<0.05$ for RW. SF also reaches $p<0.05$ in the nonparametric test, though just missing significance in the parametric one. On balance, then, there does seem to be some evidence for a shift in the location of peak glottal opening between singleton and cluster. Up to now, it is only for CG that it is easy to explain why this shift occurred: for him it is a simple consequence of the shortened oral occlusion in the clusters. For SF and RW the slight lengthening of overall glottal gesture duration could be a partial explanation, but we found a clearer explanation in a location where we would not have expected it, namely in the duration of the interval from oral occlusion onset to glottal abduction (next figure).

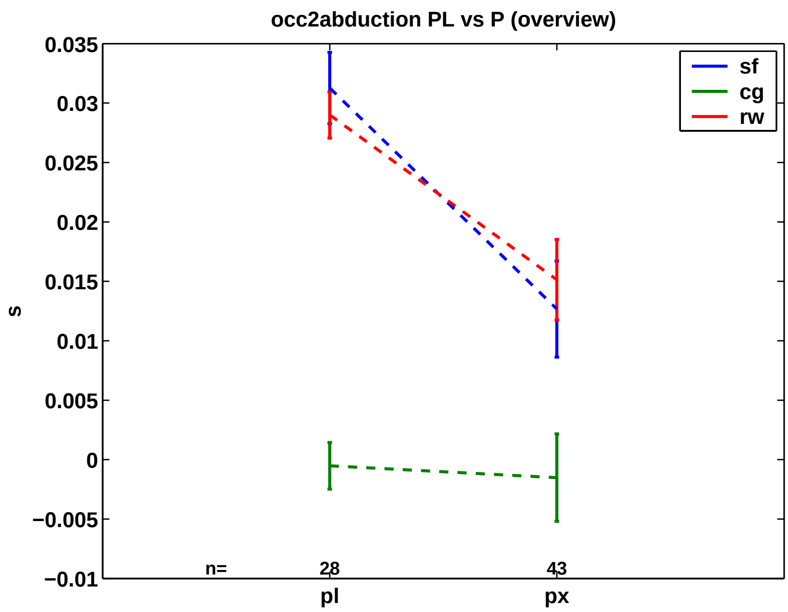


Fig. 6.36: Duration of interval from onset of oral occlusion for /**p**/ to onset of glottal abduction, comparing /**pl**/ vs. /**p**/. Other details as in Fig. 6.32.

This shows very similar behaviour for SF and RW, with glottal abduction starting later for /**pl**/ than for /**p**/. In other words, the whole glottal gesture is shifted towards later times. For these two speakers the difference was significant at $p<0.01$ (and comparable in magnitude to the shift in peak glottal opening location). For speaker CG, there is absolutely no difference with respect to this parameter. The net result of the variety of timing adjustments over the speakers is the very similar pattern with respect to peak glottal opening to release. What we have not yet considered is the functional significance of the change in the timing of peak glottal opening.

The most direct way to approach this is to now consider a parameter that would be available even in a purely acoustic investigation of these sounds, namely VOT, i.e the duration of the interval from release of the /**p**/-occlusion to voice onset.

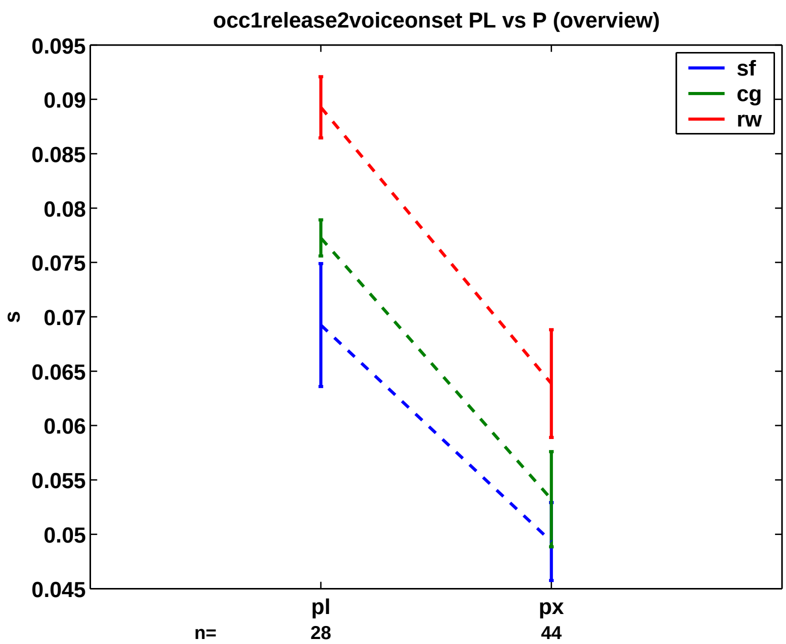


Fig. 6.37: Duration of interval from release of /**p**/ occlusion to voice onset, comparing /**pl**/ vs. /**p**/. Other details as in Fig. 6.32.

At the level of this key aspect of the acoustic output we now find remarkably similar behaviour over all three speakers, i.e VOT about 20ms longer for /**pl**/ than for /**p**/, and the statistical significance of this result appears to be robust: the lowest level reached is $p<0.05$ for SF in the nonparametric test. Otherwise at least $p<0.01$ is reached in both the parametric and nonparametric tests (see table on p. 68 for details). A result of this kind has, of course, already been found in many acoustic investigations, but we are now in a better position to explain how it comes about. One possibility that has not yet been explicitly considered is that longer VOT in /**pl**/ is a purely aerodynamic effect, due to slower reduction in intraoral pressure when /**l**/ rather than a vowel follows /**p**/ (this was raised as a possible explanation for the finding in the literature of longer total voicelessness; we will turn to this latter parameter in a moment). Evidence for the aerodynamic mechanism would be present if it turned out that the time-instant of voicing onset is delayed relative to the end of the glottal gesture for /**pl**/ vs /**p**/. The corresponding data is shown in the next figure.

Clearly there is a tendency in the expected direction for all speakers, but it is very weak and in no case does it reach statistical significance. This is not to deny that aerodynamic effects can be present: within the group of /p/+Vowel words there was a tendency for lower values of this parameter in ‘Piste’, with a relatively unconstricted lax vowel, than in ‘Piepe’ and ‘Piepse’ (with tense vowel). However, at least when - as in this corpus - the comparison involves high vowels versus /**l**/ the aerodynamic effects seem to be too weak to explain the longer VOT in clusters with /**l**/. In other words this finding can be assumed to represent planned behaviour on the part of the speaker. Nevertheless, in the discussion we will pick up the idea that aerodynamic effects may provide a bias favouring the development of particular articulatory patterns, even if they are not the *immediate* cause of longer VOT in this specific case.

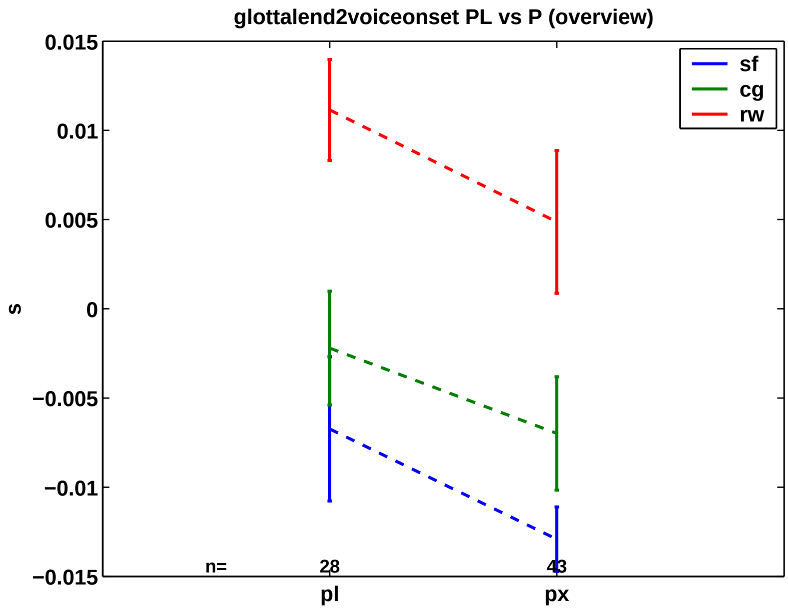


Fig. 6.38: Duration of interval from completion of glottal adduction to voice onset for /**pl**/ vs. /**p**/ (negative values indicate that voicing starts before the end of the glottal gesture). Other details as in Fig. 6.32.

Preliminary conclusions

Let us now draw the threads together in this section by reviewing the results in the light of the original hypotheses:

(1) One hypothesis received no support at all: This was the hypothesis that the relative pattern of coordination of the glottal gesture with the oral gesture for /**p**/ (as the only underlyingly voiceless segment involved) would stay the same over any changes in /**p**/ (for example in occlusion duration) when /**p**/ forms part of a cluster. This was the hypothesis for which actually least evidence was available from previous investigations, but it nevertheless seemed plausible enough to be worth considering explicitly, given the popularity of expressing coordination relations for plosives in terms of the timing of peak glottal opening. The negative evidence for this hypothesis opens the way for a theme that has already been introduced in previous sections, namely the glottal gesture as reflecting the demands of the syllable onset as a whole.

(2) The second hypothesis was that a constant glottal gesture would be overlaid with changing oral articulations, specifically a shorter occlusion for /**p**/ in the cluster. We have seen the idea of a trade-off between oral occlusion duration and peak glottal opening to release in various guises both in the literature and in previous sections of the results. Even though it usually did not work out in a hard and fast sense this mechanism does seem to have enough of a potential contribution to timing patterns to be worth taking seriously. In fact, in the present case it only worked well as an explanation for speaker CG, perhaps because the tendency overall to reduce the occlusion duration of /**p**/ in clusters is rather weak. It would be probably be worth investigating another place of articulation e.g /**kl**/, but there was no room for this in the present corpus.

(3) The third hypothesis was originally regarded as the most radical one since it assumed a possible increase in glottal gestural magnitude in the cluster in order to account for the

reports in the literature of longer total duration of voicelessness in the cluster than in the singleton case. Only very tenuous evidence for an increase in gestural magnitude was found. We have not, however, reported on the duration of total voicelessness directly. If this is computed as the time interval from onset of the oral occlusion for /**p**/ to onset of voicing then in fact a longer duration in the clusters is found for SF and RW, but not for CG. Based on this definition, a fair part of the contribution to the longer voiceless interval in SF and RW comes from the later onset of glottal abduction. This helps to drive home the unifying feature of these results, which is clearly not increased gestural magnitude and increased voiceless interval since CG is an obvious exception, but rather increased duration of VOT. If it is accepted that this is the acoustic feature that speakers directly aim to control, then the varied articulatory strategies we have seen are immediately understandable: both reducing /**p**/ occlusion duration (CG) as well as delaying abduction onset (and perhaps lengthening the gesture somewhat) are ways of achieving this aim. Thus although the original radical hypothesis is not confirmed there is evidence that speakers actively aim to produce a substantial amount of devoicing on /**l**/; this suggests that the original account that we were inclined to favour, namely that devoicing of /**l**/ is essentially a passive consequence of inevitable coarticulatory effects, falls short of the mark.

In order to summarize the findings we show another overview of the temporal structure, but this time lined-up at the release of the /**p**/ occlusion, to show up the VOT differences more clearly, i.e note the later location of the right edge of the blue "voiceless" bars for the /**pl**/ sequences:

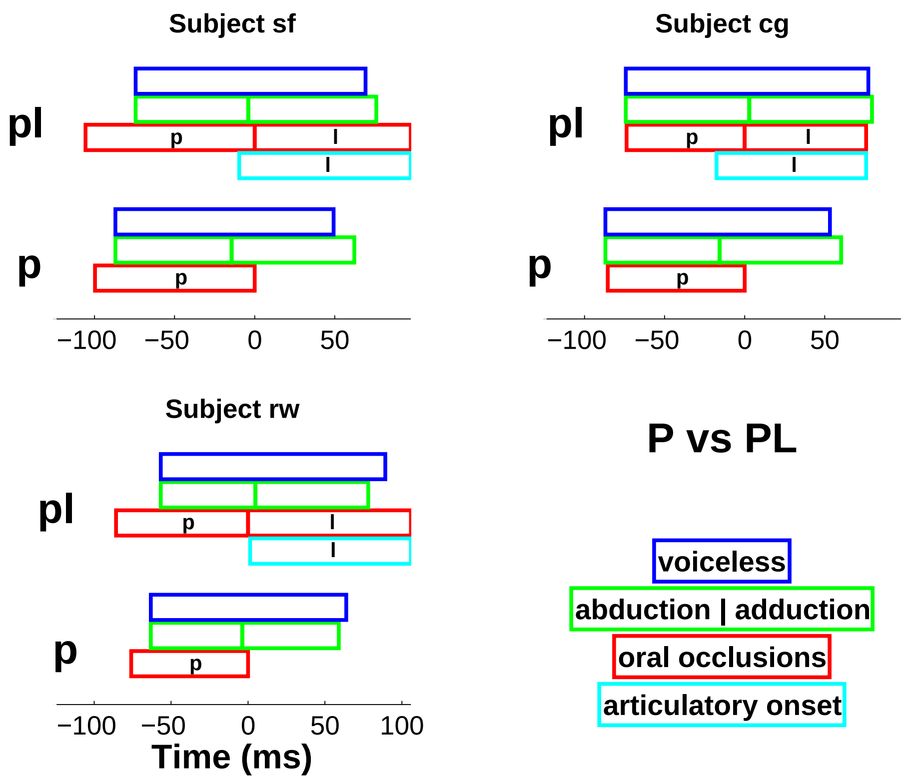


Fig. 6.39: Overview of temporal structure of /p/ and /pl/, line-up at end of occlusion for /p/

Finally, these results make it necessary to address the question of *why* voicelessness should be enhanced.

One possibility is that the long VOT duration in /**pl**/ helps to ensure sufficient contrast from /**bl**/. As we saw in Chapter 2 (p. 12), Docherty had noted a tendency for longer VOT in /**bl**/ vs. /**b**/ (presumably here aerodynamic effects are at work) so this could in turn make longer VOT in /**pl**/ advantageous. This on its own might not be a very strong reason, particularly in German where the functional load of /**pl**/ vs. /**bl**/ is probably less than in English, owing to the German development of /**pfl**/ and /**pf**/ sequences. Perhaps an additional factor is that - once sonorants have become largely voiceless - then it helps for the voicelessness to be long, in order to make contrasts like /**pl**/ and /**pr**/ more salient (we made a similar argument concerning the mirror-image sequences of Icelandic: voiceless sonorants *before* plosives).

We will re-assess these issues after considering if, and where, further evidence for active enhancement of voicelessness on the sonorants emerges in the remaining sections of this chapter.

6.3.2 Fricative-/l/ versus fricative

Introduction

The most basic issue of interest is whether the total duration of voicelessness increases in fricative-lateral sequences compared to the single fricatives, and, if so, whether this can be attributed to an actual increase in glottal gesture magnitude, or, more prosaically, to differing aerodynamic conditions. The evidence hitherto in the literature is, as we have seen in Chapter 2 (p. 12), somewhat mixed. Docherty's (1992) acoustic evidence makes an increase in glottal gesture appear conceivable, but the direct glottal analysis of one speaker in Tsuchida et al. (2000) did not confirm this. Even if total voicelessness does not increase there remains the question of how oral and laryngeal activity is coordinated, given that a reduction in the duration of the fricative occlusion is to be expected. Will the glottal gesture retain the same phasing relative to the oral occlusion as the latter changes, or will there be shifts in the coordination even though /l/ should be irrelevant to the glottal articulation?

Unlike the previous section, for the sound categories in this section the comparisons can be based on two places of articulation, i.e /**f**/ vs. /**fl**/ and /**ʃ**/ vs /**ʃl**/. Accordingly the statistical results are based on two-factorial analyses, factor C1 referring to place of articulation and factor C2 referring to the /**l**/-cluster versus singleton distinction (see tables on on p. 69). Since we have no specific hypotheses regarding place of articulation in fricatives the results for C1 have been tabulated, but otherwise will be referred to only in passing.

The complete breakdown of the material used in this section is as follows:

/**f**/:	two word forms;	both from Group 1;	both in first word position
/**fl**/:	two word forms;	both from Group 2;	one in first, one in second word position
/**ʃ**/:	two word forms;	both from Group 1;	both in first word position
/**ʃl**/:	one word form;	from Group 2;	in second word position

Overview of temporal structure

As before we start with an overview of the temporal structure, pointing out some of the most salient features, before proceeding to detailed statistics below:

- Very clear that the duration of the fricative occlusion is reduced in fricative-lateral sequences.
- On the other hand, no obvious indication that total voicelessness or glottal gesture duration *increase*; in fact for SF both appear to decrease, somewhat in parallel to fricative occlusion duration.
- Not easy to estimate whether the relative location of peak glottal opening in the fricative may have shifted from fricative to fricative-lateral.
- In contrast, very clear that the *end* of the glottal gesture does not stay in the same location relative to the end of the fricative occlusion. Thus, for CG and RW in particular, the glottal gesture overlaps the /l/ occlusion to quite an extent (though, not, of course, as much as in plosive-lateral combinations).

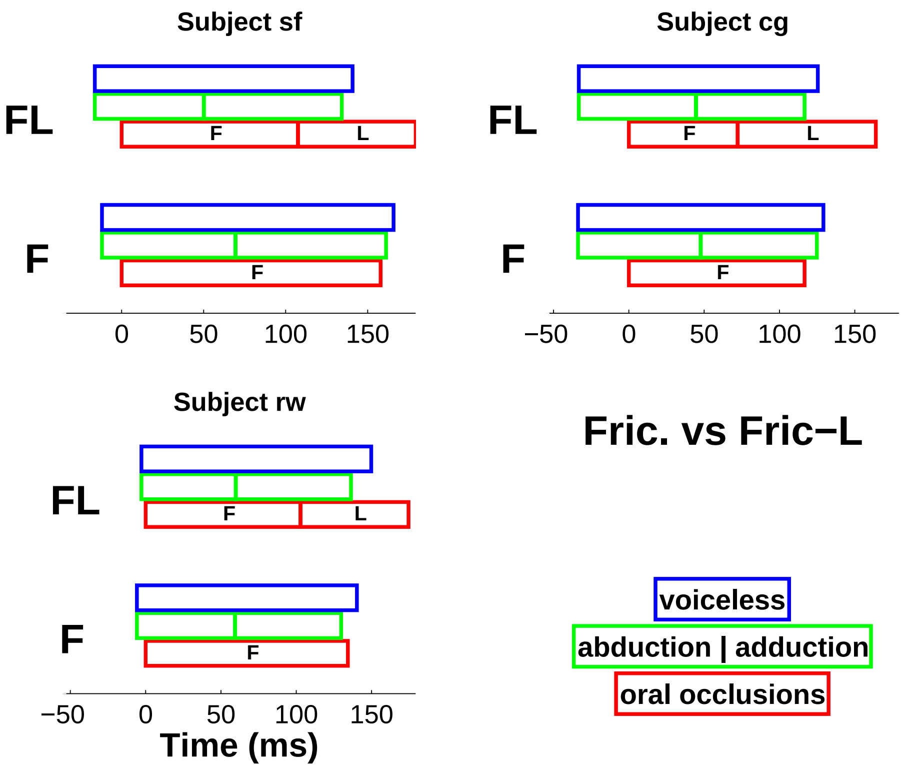


Fig. 6.40: Overview of temporal structure of single fricatives (lablelled ‘F’) and fricative-/l/ sequences (labelled ‘FL’); line-up at start of fricative occlusion.

Detailed results of individual parameters

The parameters that will be examined in this section are:

- Oral occlusion duration of fricative segment
- Glottal gesture duration
- Amplitude of peak glottal opening (no figure)
- Total duration of voicelessness
- Duration of interval from end of glottal adduction to voice onset
- Duration of interval from start of fricative to start of glottal abduction (no figure)
- Relative position of peak glottal opening in the fricative segment
- Relative position of end of fricative segment within the glottal gesture
- Ratio of glottal abduction to adduction duration

First we will examine the duration of the fricative occlusion: there was a very consistent result over all three subjects, so this thus provides a useful framework for consideration of more variable effects.

As the figure shows, the pattern of shorter occlusion duration for the cluster than for the singleton is consistent not only over subjects, but also over the two places of articulation. In the ANOVA the main effect of C2 (cluster vs. singleton) reaches $p<0.001$ for all speakers. A table of the ANOVA results is given on p. 69 above. It is followed by a further table giving those pairwise comparisons of cluster vs. singleton (matched for place of articulation) that reached the criterion of p=0.01 in the non-parametric test. In the present case, the criterion was only reached for /**fl**/ vs. /**f**/ (for 2 out of 3 speakers), the /**ʃl**/ vs. /**ʃ**/ comparison usually just missing this level of significance. Note that the change in occlusion duration is apparently more robust for the fricative in the fricative-lateral combinations here than for the plosive in the plosive-lateral combination /**pl**/ examined in the previous section (Fig. 6.34 on p. 114).

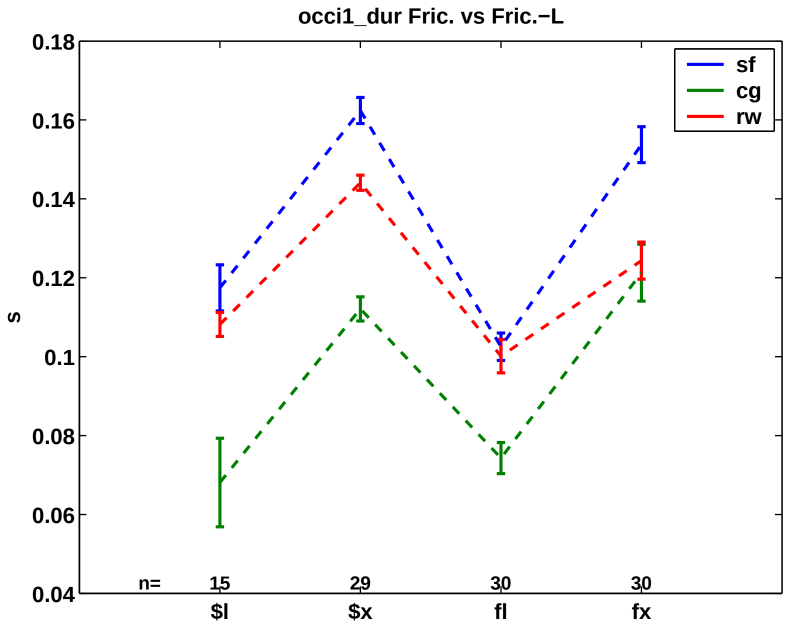


Fig. 6.41: Duration of fricative occlusion for single fricative vs. fricative-/**l**/ clusters. Interpretation of the abscissa labels for this and the following figures: $ = /**ʃ**/; $x and fx indicate the singleton consonants /**ʃ**/ and /**f**/, respectively (i.e as in previous sections 'x' indicates absence of C2).

We next consider the total duration of the glottal gesture. As already intimated above, we now no longer have a consistent pattern over all three speakers (see next figure).

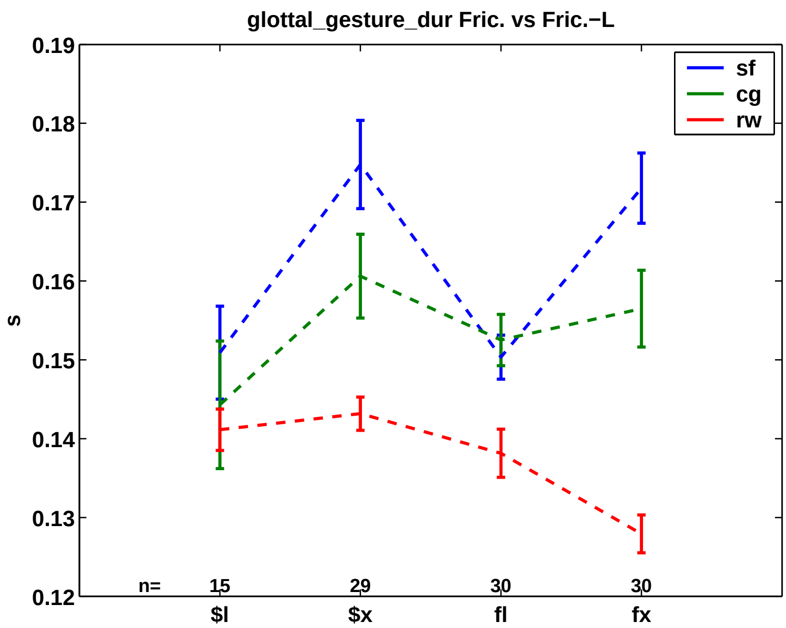


Fig. 6.42: Duration of glottal gesture for single fricative vs. fricative-/**l**/ clusters. Other details as in Fig. 6.41

By comparing the just-discussed figure for fricative occlusion duration (Fig. 6.41) with the one immediately above we see that the results for glottal gesture duration for SF indeed closely parallel her occlusion duration results, with fricative-lateral shorter than single fricative at the same level of significance in the ANOVA (the pairwise comparisons both just miss the criterion level). For CG and RW, however, we find neither a clear pattern nor significant results, so gesture duration does not follow occlusion duration. But it is clear for all three subjects that gesture duration does not increase for the clusters[51].

We will dispense with showing the corresponding gesture amplitude values as they basically show the same picture as gestural duration and add nothing new to the discussion: higher amplitudes for singleton than cluster for SF, but no effects for CG and RW (leaving aside a place of articulation effect for RW). Details are given in the tables (p. 69).

Before turning to detailed examination of the coordination patterns we still need to look at the total duration of voicelessness, as this was the key acoustic parameter that provided much of the impetus for this investigation. This is shown in the next figure.

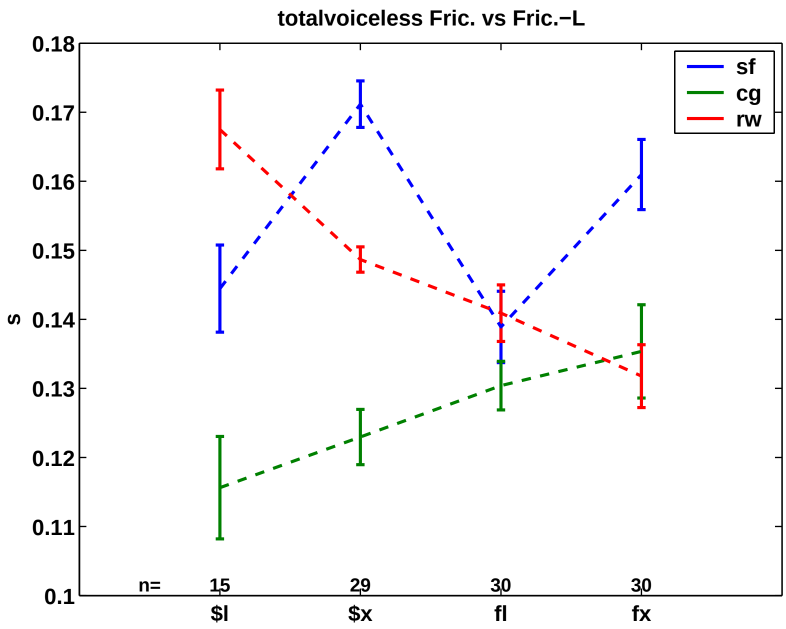


Fig. 6.43: Total duration of voicelessness for single fricative vs. fricative-/**l**/ clusters. Other details as in Fig. 6.41

Results for SF are no surprise since they simply parallel the results for other parameters above, with shorter duration of voicelessness in the case of the clusters ($p<0.001$). For CG there was no significant effect, also as expected from the results already seen. The more surprising case was RW, who actually showed significantly longer voicelessness for the clusters ($p<0.01$), even though gesture duration showed no difference. This leads to the question of whether aerodynamic effects could be playing a role. This is an important aspect for putting purely acoustic investigations into proper perspective and which we can capture to a certain extent here by comparing the length of the interval from end of the glottal gesture to onset of voicing. This is shown in the following figure, and shows a single strongly elevated value for /**ʃl**/ of speaker RW.

Thus in this one specific case one could suspect that in the cluster aerodynamic conditions are less conducive to voicing, and delay its onset relative to the completion of glottal adduction. Why this should happen just in this one specific case is unclear. Apart from this case, one would be inclined to interpret these results - just as for /**p**/ vs. /**pl**/ - as showing that aerodynamic effects

[51]For RW there is, however, a place of articulation effect in both gesture and occlusion duration: labial-dental shorter than postalveolar.

actually have only a weak role to play, at least for laterals relative to high vowels, since no pattern emerges at all.

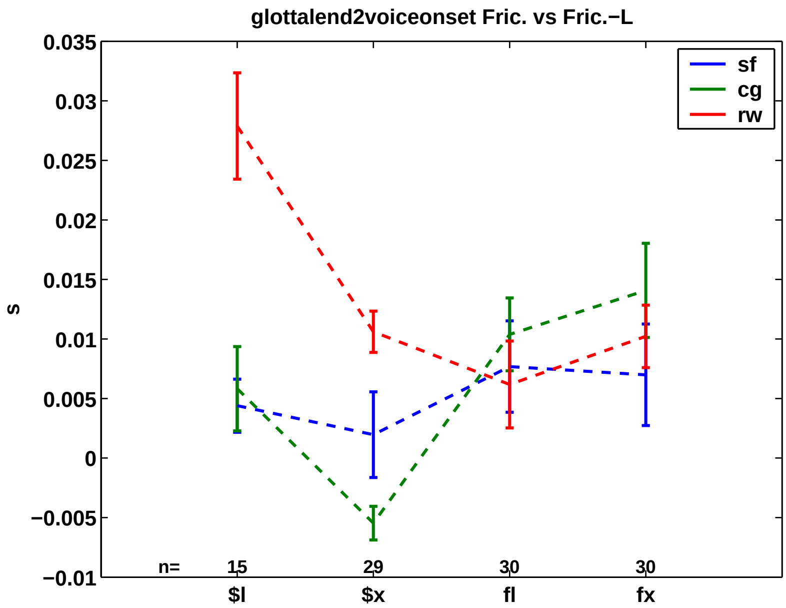


Fig. 6.44: Duration of interval from end of glottal gesture to voice onset for single fricative vs. fricative-/**l**/ clusters (positive values indicate voice onset after end of glottal gesture). Other details as in Fig. 6.41

Before turning to analysis of coordination patterns it should be mentioned in passing that in order to make sure that effects in unexpected locations were not being overlooked, we also checked the interval from onset of fricative occlusion to onset of glottal abduction. No cluster vs. singleton effects were found for any speaker (cf. tables on p. 69).

We will look at the coordination of oral and glottal activity from two complementary perspectives. The first one is the relative location of peak glottal opening within the occlusion interval of the the fricative (see next figure).

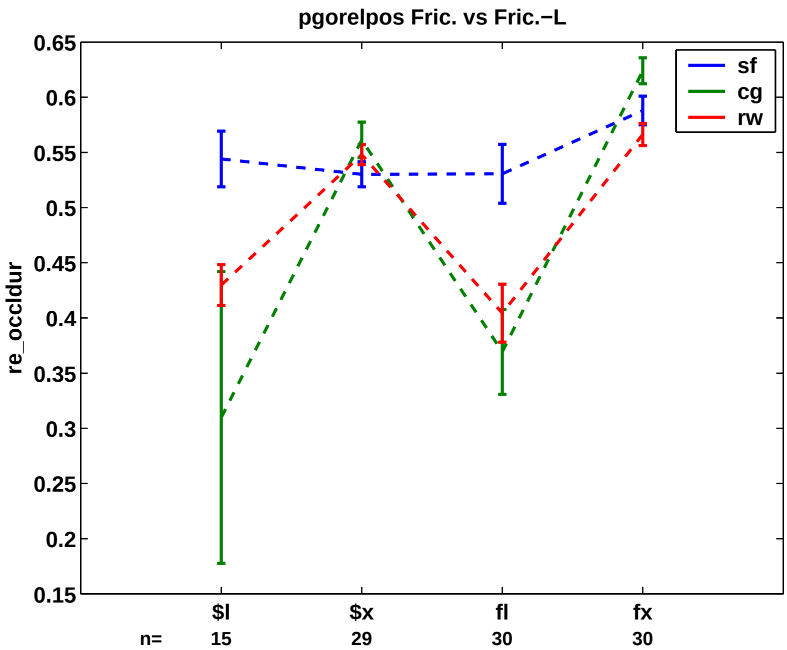


Fig. 6.45: Relative position of peak glottal opening during oral occlusion for the fricative, comparing single fricative vs. fricative-/**l**/ clusters. A value of 0.5 indicates coordination with fricative midpoint. Lower values indicate locations closer to the end of the fricative. Other details as in Fig. 6.41

The results for the relative position of peak glottal opening should come as no surprise given the discussion above. For SF, no differences between clusters and singletons are to be observed, which was to be expected as glottal gesture duration and occlusion duration had reduced more or less in parallel from singleton to cluster. CG and RW showed the contrasting pattern of a fairly constant glottal gesture duration, but shortened occlusion. Accordingly, there are robust differences in position of PGO ($p<0.001$; the /**fl**/ vs /**f**/ comparison reached the nonparametric criterion of $p=0.01$ for both speakers), with PGO occurring relatively later in the occlusion for the clusters.

The second coordination parameter attempts to quantify the effect apparent in the overview figure at the beginning of this section, namely that the glottal gesture terminates further from offset of the fricative occlusion in the cluster. This was done now by expressing the time-point of an oral event relative to a glottal interval (the previous parameter in effect does the reverse : glottal event relative to oral interval), specifically the relative location of offset of the oral occlusion for the fricative within the glottal cycle. This is shown in the next figure.

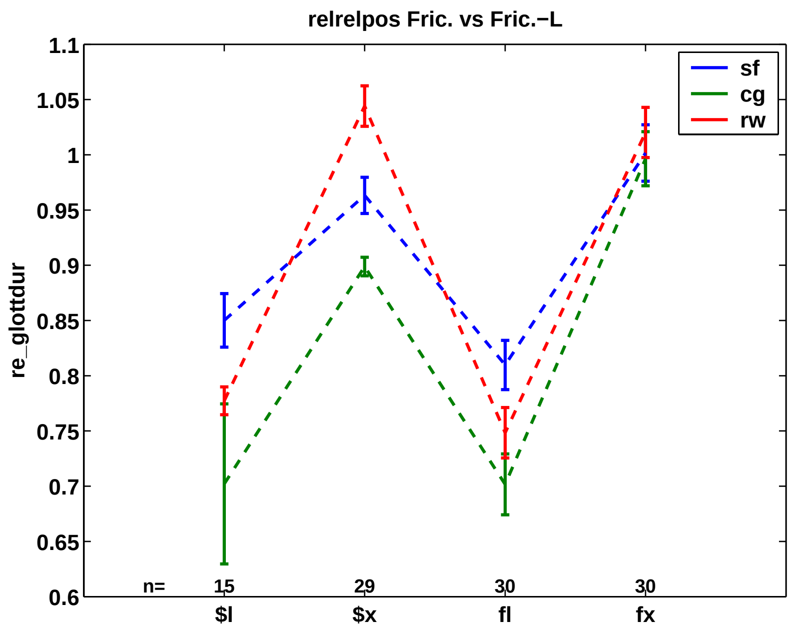


Fig. 6.46: Position of end of fricative occlusion relative to glottal gesture cycle, comparing single fricative vs. fricative-/**l**/ clusters. A value of 1 indicates end of fricative and end of glottal gesture coincide; values below 1 indicate earlier position of fricative offset in the glottal cycle. Other details as in Fig. 6.41

The pattern is very similar for all three speakers. For CG and RW the information is somewhat redundant since it was more or less inevitable that release of the fricative occlusion would occur much earlier relative to the glottal cycle for clusters compared to singletons. The more interesting point is that SF shows basically the same pattern, even if slightly less strikingly than the other speakers. In fact for all speakers the difference was significant at $p<0.001$ in the ANOVA and the /**f**/ vs. /**fl**/ comparison always reached the nonparametric criterion (it might be worth mentioning at this juncture that here as in many other cases in this section the /**ʃ**/ vs. /**ʃl**/ comparison just missed the nonparametric criterion whereas /**f**/ vs. /**fl**/ achieved it; it is to be suspected that this simply reflects the smaller number of tokens for /**ʃl**/ rather than any important difference between the places of articulation). The fact that SF here patterns with the other speakers indicates that although oral and glottal durations for her certainly changed more in parallel than for the other speakers this nonetheless does not mean that phasing of glottal activity relative to the fricative stays absolutely identical over clusters and singletons.

The basically parallel changes for SF do however make clear that this strategy is potentially available to subjects. Thus this increases the significance of the finding for /**p**/ vs /**pl**/ in the previous section that precisely this strategy did *not* occur

One reason why laryngeal-oral phasing differences were in evidence for SF in the vicinity of glottal gesture offset rather than peak glottal opening is probably that the shape of the gesture - i.e the relative duration of abduction and adduction - changed somewhat. This is the last detailed result to be looked at in this section and is shown in the next figure.

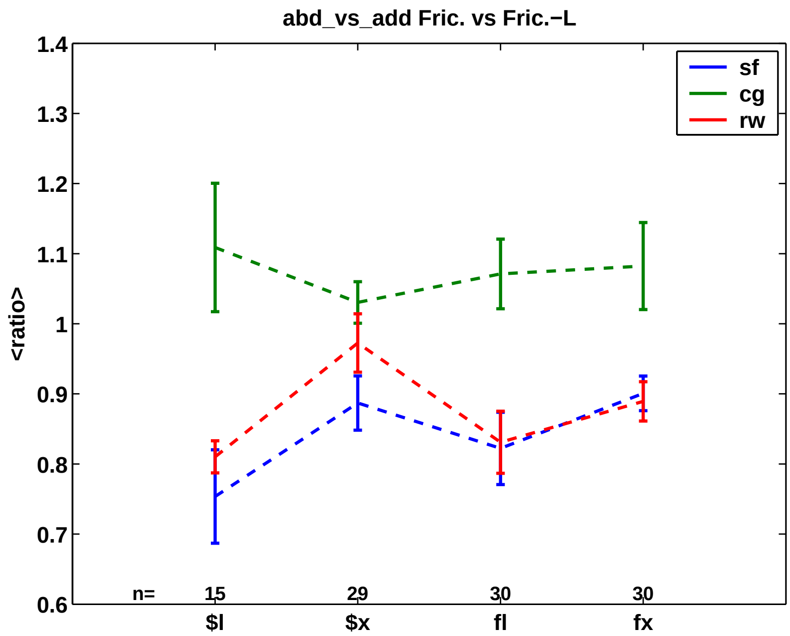


Fig. 6.47: Relative duration of glottal abduction and adduction for single fricative vs. fricative-/**l**/ clusters. Other details as in Fig. 6.41

Both SF and RW show a weakly significant effect for the adduction phase to be relatively long in the cluster (as in previous sections, CG behaves somewhat differently with respect to this parameter than the other two speakers, which we attributed to his particularly early abduction onset). Although the effects here are fairly weak they are nonetheless potentially important for interpretation of laryngeal activity in clusters in general. The direction of the effect is the same as for fricative-plosives versus single fricatives, though of smaller magnitude (refer back to Fig. 6.20 on p. 94). But this means that caution may be necessary in interpreting the effect in fricative-plosives as overlap of two underlying gestures, or as competition between two oral gestures for dominance of a single laryngeal gesture.

Preliminary summary

The baseline finding for fricative-/**l**/, in contrast to /**pl**/, was that the duration of C1 decreased considerably (for both fricative places of articulation) in the cluster. The other common finding was that for all speakers there were changes in the pattern of coordination of the fricative with the glottal gesture; however, the nature of the changes differed over speakers. For CG and RW the relative time-point of peak glottal opening shifted because they kept a similar glottal gesture duration despite shortening of the fricative; for SF, glottal gesture duration reduced more in parallel with the occlusion duration. For all speakers the relative time-point of the end of the fricative occlusion within the glottal cycle was earlier for the clusters. Thus, although the radical scenario of an increase in duration of the glottal gesture in the cluster certainly did not occur, it also does not appear that speakers are particularly concerned to minimize the amount of voicelessness on the sonorant.

This is noteworthy because - unlike the initial stop clusters - there is no constraint imposed by clusters contrasting in voicing (i.e no counterpart to /**pl**/ vs. /**bl**/). Thus, if clarity of C2 were important then one might expect C2 to be realized as fully voiced. Even SF, who heads in this direction, does not go the whole hog. So it actually appears to be communicatively quite acceptable to have rather a loose coordination of the end of the glottal gesture with the end of the fricative. This is probably then a rather different case to that of /**pl**/ discussed above. There we argued for active enhancement of VOT in the cluster; here it is less a case of active enhancement, but rather an absence of active minimization of voicelessness. Why should speakers adopt this pattern (and not scenario 1)? Speculatively, it could be a further case of the principle of "go with the flow" (cf. Part I, General Discussion): Although we have played down the salience of aerodynamic effects when comparing /**l**/ and high vowels following the occlusive, it may well be that speakers are able to learn that re-starting voicing in a sonorant is comparatively difficult (e.g compared to vowels in general). As long as the functional load depending on precise voicing properties of the sonorant is not high, i.e a possibly fluctuating amount of sonorant devoicing in the syllable onset is communicatively acceptable, then it may simply represent an economy of effort to not complete glottal adduction until some time after the start of the sonorant, by which time conditions for voicing will probably have become more favourable - and thus perhaps also allow a sharper transition form voicelessness to modal voicing (cf. discussion of regulation of vocal fold tension in the transition from voicelessness to voiced in the general discussion of Part I)

If this kind of scenario is plausible, then it is tantamount to saying that speakers plan their coordination patterns to take into account the functional demands of the syllable onset as a whole.

6.3.3 /pf/ vs. /pfl/

Introduction

This is a comparison for which no expectations can be derived directly from the literature, as these particular sound sequences are very characteristic of German and appear not to have been previously investigated (except for the small amount of information on plosive-fricative sequences mentioned above; see e.g p. 107). However, the relevant questions are very similar to those in the previous section, and it will be immediately become apparent that there are some strong similarities in the observations that can be made, because here, too, the sound immediately preceding the lateral is a fricative.

Before proceeding to an overview of the temporal structure the following notes on the material should be made:

There are two word-forms each for the /**pf**/ and /**pfl**/ sequences.
All words occur in first position in the carrier phrase.
/**pf**/ words are from Group 1 of the corpus, /**pfl**/ words from Group 2.
One of the /**pfl**/ words was a single-syllable word ("Pflicht")

Overview of temporal structure

The following figure gives an overview of the temporal structure in the usual form:

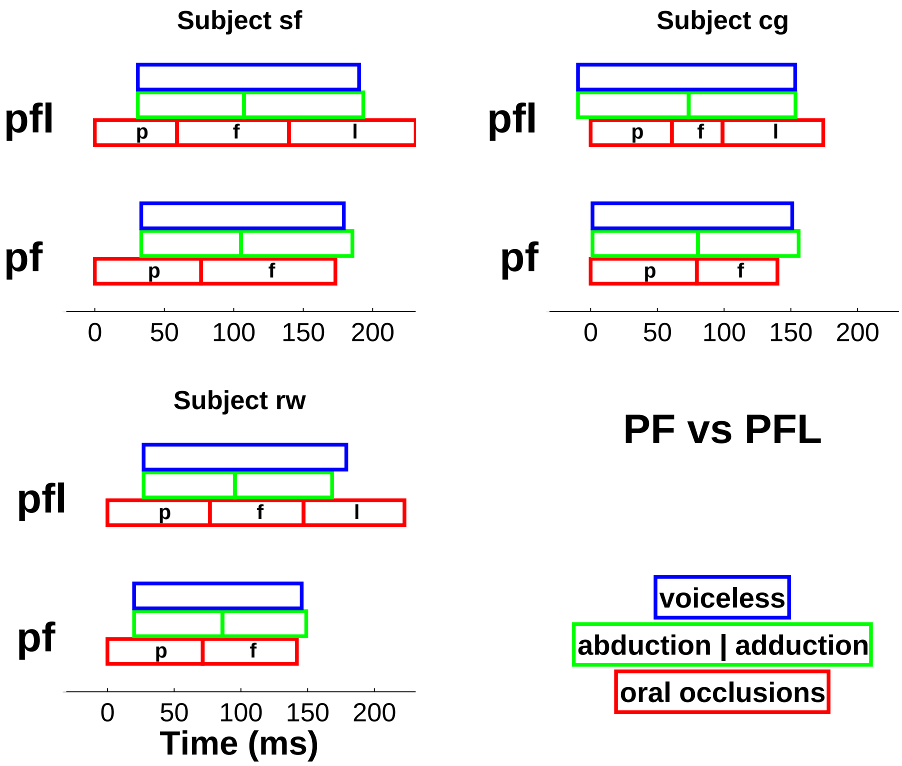


Fig. 6.48: Temporal structure of /**pf**/ and /**pfl**/ sequences, lined up at onset of occlusion for /**p**/

Salient points that can be noted from the above figure include:

- Offset of the glottal gesture appears to occur at a different time relative to the end of the fricative for /**pfl**/ vs. /**pf**/ (refer back to the corresponding figure in the previous section for fricative vs. fricative-lateral).
- Relative location of peak glottal opening within the fricative segment may differ over the two consonant sequences for SF and CG.
- Also for these two subjects the duration of /**pf**/ appears shorter in combination with the lateral than when occurring alone (details of the individual segments will be given immediately below).

Detailed results of individual parameters

Consideration of the detailed results will essentially follow the same path as in the previous section, except that, first of all, possible changes in two rather than one initial segments (here /**p**/ and /**f**/) must be examined. Also, place of articulation does not constitute a factor in the analysis so statistics are based on a straight comparison of /**pf**/ and /**pfl**/ (see p. 70 for the usual table).

The parameters that will be examined in this section are, then:

- Oral occlusion duration of plosive segment /**p**/ (no figure)
- Oral occlusion duration of fricative segment /**f**/
- Glottal gesture duration
- Amplitude of peak glottal opening (no figure)
- Total duration of voicelessness
- Duration of interval from end of glottal adduction to voice onset
- Relative position of peak glottal opening in the fricative segment
- Relative position of end of fricative segment within the glottal gesture

Turning first, then, to the question of the extent to which segment durations decrease in the more complex cluster, it was found that /**p**/ was shorter at $p<0.01$ for SF and CG, but the result for RW was non-significant. The result for /**f**/ was the same, except that the significance level for SF was only $p<0.05$. This is thus a slightly different result from that of the fricative in the previous section where the shortening in the cluster was significant at $p<0.001$ for all speakers. In particular, no obvious reason presented itself for the non-significant differences for RW. We considered, for example, the possibility that the cluster in the single-syllable /**pfl**/-word “Pflicht” might be unusually long, but in fact the tendency was in the opposite direction (i.e it was somewhat shorter than the two-syllable /**pfl**/-word “Pflüge”). Overall, his /**pfl**/-sequences thus pattern more like plosive-/**l**/ than fricative-/**l**/.

The following figure illustrates the results only for the fricative segment, but statistical results for both segments are given in the table on p. 70 above. Note that in this figure, unlike the following ones the material has also been subdivided by the vowel following the consonant cluster (/**i**/ vs /**y**/).

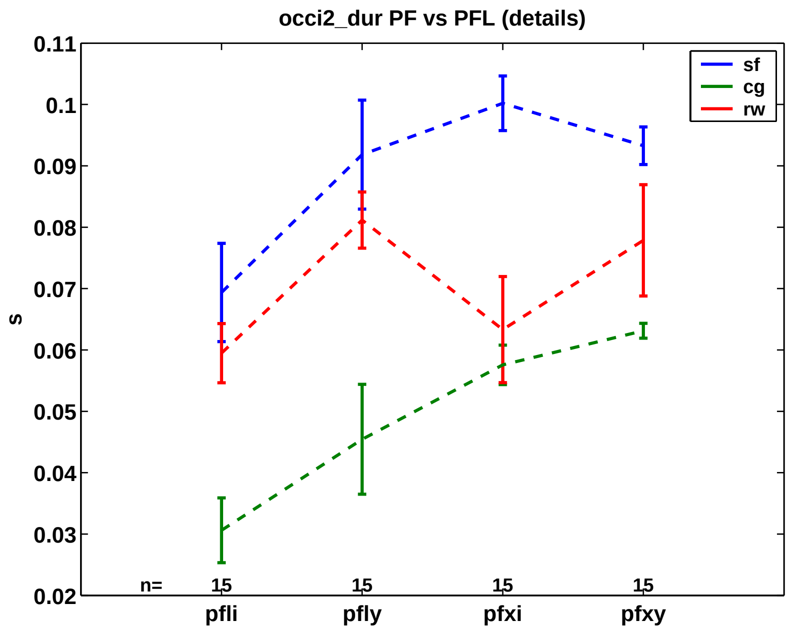


Fig. 6.49: Duration of fricative occlusion for /**pfl**/ vs. /**pf**/ sequences. The abscissa labels 'pfli', 'pfly', 'pfxi', 'pfxy' correspond to the word-forms 'Pflicht', 'Pflüge', 'Pfister', 'Pfütze', respectively (as in previous figures in this section, 'x' stands for the empty position in the consonant sequence).

The result for overall glottal gesture duration (illustrated in the next figure) is also somewhat different from that in the previous section, where duration decreased in parallel with oral occlusion duration for SF, but no clear pattern was found for the other two speakers. In the present case, the pattern is similar for all three speakers, all showing a very weak tendency to *longer* gesture duration in clusters with /**l**/, but none of the results reached statistical significance (except very marginally for RW in the non-parametric test).

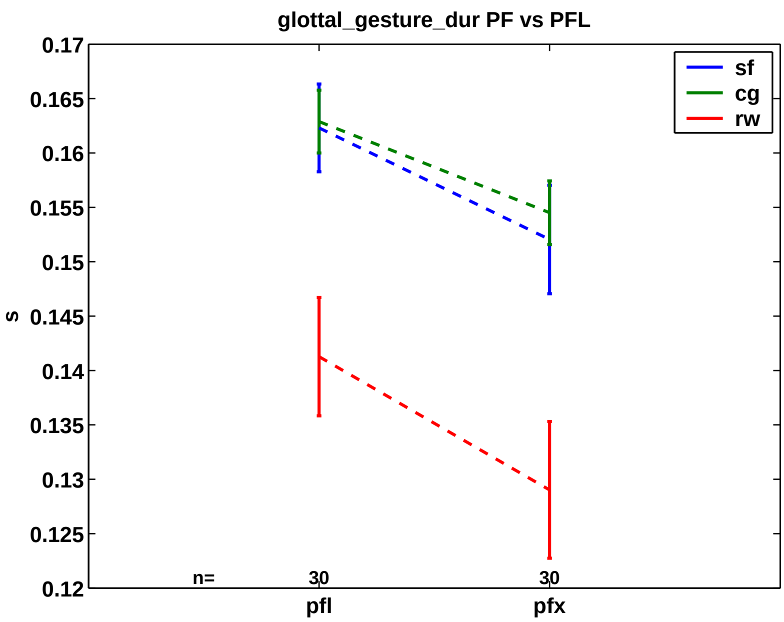


Fig. 6.50: Duration of glottal gesture for /**pfl**/ vs. /**pf**/ sequences (in this and the following figures the abscissa label 'pfx' stands for /**pf**/).

The results for glottal opening amplitude were also completely non-significant, with not even the indication of a tendency in one direction or the other, so once again no further details will be shown.

Regarding the total duration of voicelessness the interesting issue, as with previous combinations with /**l**/, is whether the duration may actually increase. The results just shown for gesture duration would lead one to expect only a weak effect. For CG there was indeed no significant difference but for SF and RW duration of the voicelessness in /pfl/ was actually significantly longer at $p<0.05$. In the figure the magnitude of the effect appears larger for RW (though note the rather high standard errors) and in fact almost reached $p<0.01$.

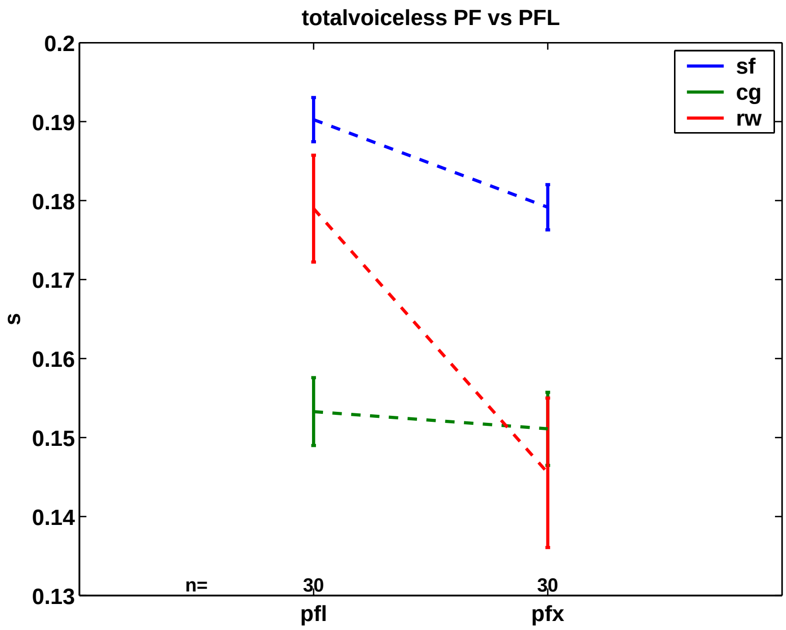


Fig. 6.51: Total duration of voicelessness for /**pfl**/ vs. /**pf**/ sequences. Other details as in Fig. 6.50.

This raises, as before, the question as to whether aerodynamic conditions could be responsible for the increase in voicelessness. Information relevant to this is given by the following figure.

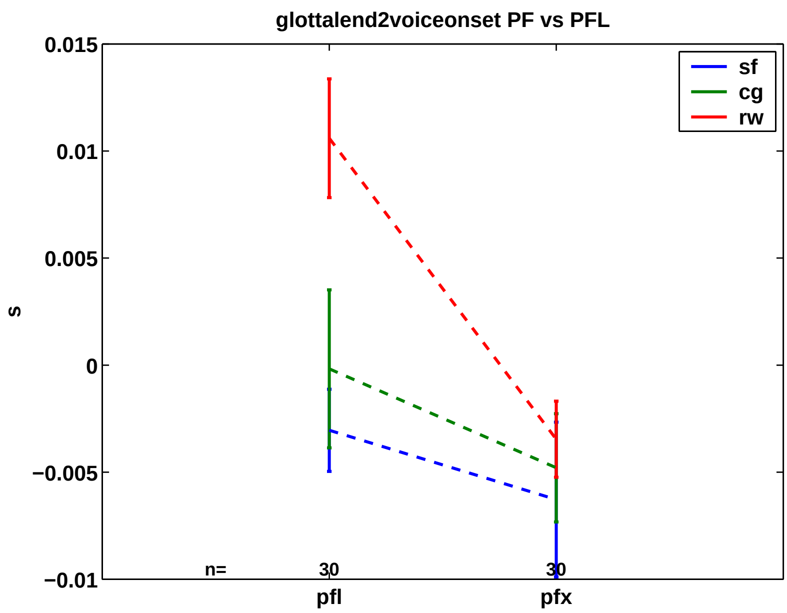


Fig. 6.52: Duration of interval from end of glottal gesture to voice onset for /**pfl**/ vs. /**pf**/ sequences. Positive values indicate voice onset after end of glottal gesture. Other details as in Fig. 6.50.

This shows the interval from completion of glottal adduction to onset of voicing. Unfortunately for a neat interpretation, we encountered a discrepancy in the results between RW and the other two subjects, just as occurred in the previous section (refer back to Fig. 6.44 on p. 124 above): While both CG and SF showed clearly nonsigificant effects (thus speaking against the relevance of aerodynamic effects), RW once again showed clearly delayed onset of phonation relative to the end of the glottal gesture for the cluster with the lateral. This remains a puzzling result because closer inspection showed the effect to be mainly due to the word "Pflüge", and much less clear in “Pflicht”. The fact that in the previous section the effect was only present for /**ʃl**/ (“Schlitze”) but not for /**fl**/ (“Flüsse”, “Flitze”) removes any possibility of an explanation related to some artefact in determining glottal gesture offset related to place of articulation of the fricative, or the nature of the vowel or the medial consonants.

The next two figures provide the details relating to laryngeal-oral coordination, and are set up to directly parallel the results in the previous section on fricative vs. fricative-/**l**/, although the fricative in terms of which the coordination is expressed is now of course the second segment in the cluster rather than the first.

Regarding the relative position of peak glottal opening within the fricative occlusion phase (shown in the figure immediately below), a tendency is observable in all speakers for it to be located later in the /**pfl**/ clusters, but this result is only clearly significant for SF. The fact that the result is non-significant for RW is not surprising given the absence of clear changes in occlusion or glottal gesture duration. For CG the relative position appeared quite distinct in the overview of the temporal structure at the beginning of this section (i.e near /**f**/ onset in /**pf**/, but near the midpoint of /**f**/ in /**pfl**/) but the result only reached significance in the non-parametric test. Probably the very short duration of /**f**/ in /**pfl**/ induced a great deal of variability in this relative measure. It is also worth pointing out here the huge range in values of the relative position of peak glottal opening over the two clusters and over the three speakers - the extremes are of the order of 0.35 and 1. Thus it does not appear that we are hear confronted simply with subtle deviations from an “ideal” value of 0.5 (i.e fricative midpoint).

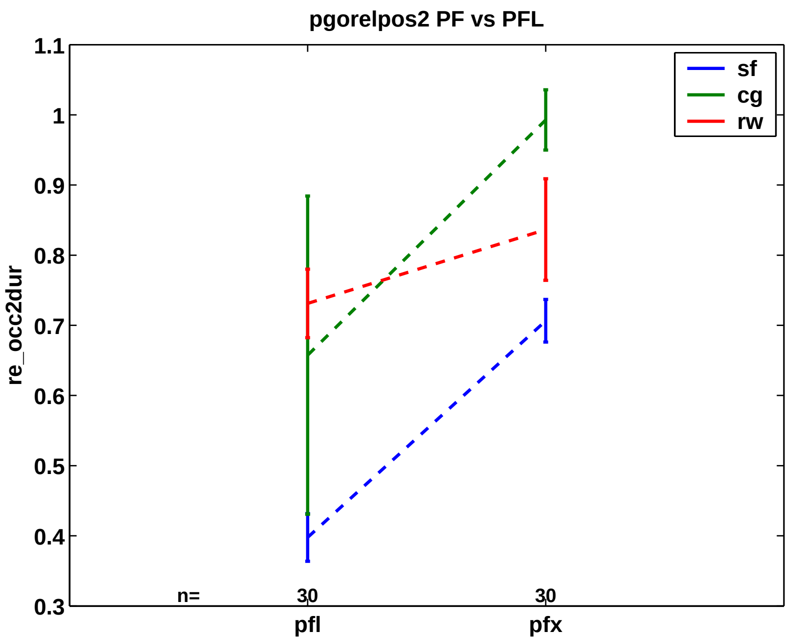


Fig. 6.53: Relative position of peak glottal opening during oral occlusion for the fricative, comparing /**pfl**/ vs. /**pf**/ sequences. 0.5 indicates coordination with fricative midpoint. Lower values indicate locations closer to the end of the fricative. Other details as in Fig. 6.50.

Even though the speaker-specific aspects of the coordination patterns differ between the present and the previous section, one common point is that a more consistent pattern emerges in the second coordination parameter - the relative position of the fricative release - than in the relative position of peak glottal opening. Just as was the case in the previous section, we now find a very similar pattern over all speakers, with offset of the fricative segment coming relatively earlier in the glottal cycle when followed by the lateral ($p<0.001$ for SF and CG, and $p<0.01$ for RW).

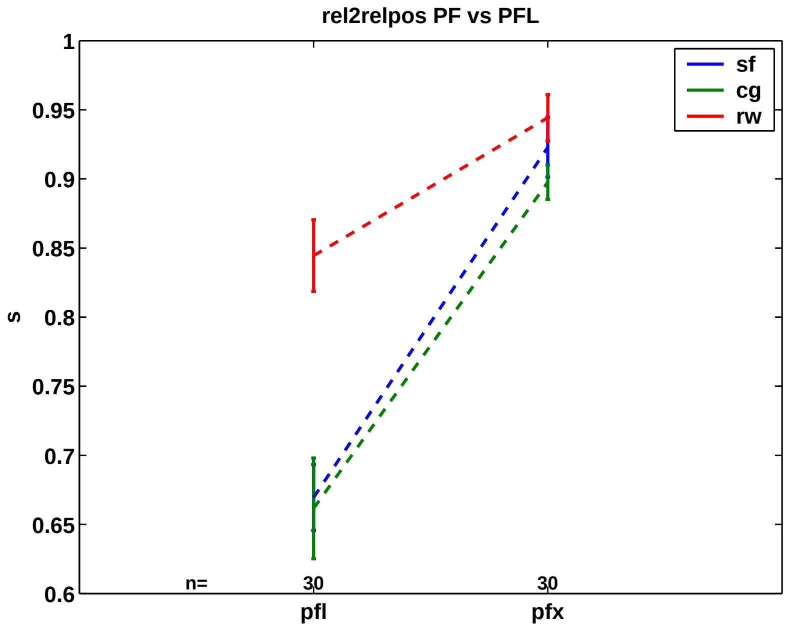


Fig. 6.54: Position of end of fricative occlusion relative to glottal gesture cycle, comparing /**pfl**/ vs. /**pf**/ sequences. Value of 1indicates end of fricative and end of glottal gesture coincide; values below 1 indicate earlier position of fricative offset in the glottal cycle. Other details as in Fig. 6.50.

Preliminary summary

We have just seen that a very clear parallel with the preceding section is the fact that the end of the glottal gesture is not kept closely synchronized with the end of the last underlyingly voiceless segment. Thus a similar question arises as to whether speakers actively enhance voicelessness on the lateral, or whether they simply do not actively try to minimize the voicelessness.

In the present section there was more of a tendency towards longer total voicelessness in the clusters with /l/ than in the previous section (but with only equivocal indications that this could be due to aerodynamic conditions). Also, while in the preceding section one subject showed shorter gesture duration (and shorter voicelessness) for the combination with /**l**/ there was a weak but consistent tendency for longer gestural durations in the present section. Thus the balance is here perhaps more towards active planning of a substantial amount of sonorant devoicing; final interpretation should wait, however, until we can weigh up the complete material involving /**l**/.

There is a final point of similarity between the present and previous section, which might be called the variability of variability: Once again the particular combination of oral and laryngeal adjustments leading in the /l/-clusters to the common outcome of earlier position of the end of the fricative occlusion within the glottal cycle was quite variable over subjects. Moreover, the pattern of variability varied between the two sections. In the present section SF and CG were the two subjects for whom shortening of the voiceless occlusions relative to a fairly constant glottal gesture duration was a relevant effect, whereas in the previous section the two subjects showing this adjustment were CG and RW. This has been characteristic of this whole investigation: The fluctuating patterns can be awkward to analyze, but in the long run they help to bring the small islands of constancy out in sharper relief.

6.3.4 /ʃp/ vs. /ʃpl/

Introduction

The comparison in the present section differs from those in the two previous sections since the segment before the lateral is now a plosive and not a fricative. However, in fact the previous two sections provide a better reference point than the initial section on /**p**/ vs. /**pl**/. This is, firstly, because the cluster nonetheless contains a fricative and - because of the importance that has been assigned to fricatives when present for the devoicing gesture - it is convenient to examine the timing of peak glottal opening relative to the fricative segment. Secondly, the present material is more similar to the previous two sections rather than the first section in the sense that glottal adduction is approaching completion by the end of the last consonant preceding the lateral, i.e quite unlike the large opening at the release of /**p**/ in single initial /**p**/ or /**pl**/. Accordingly, it will again be interesting to examine the relative timing of the release of the second consonant, i.e /**p**/, within the glottal cycle.

Notes on the material:

> There is one word-form for the /**ʃpl**/ cluster ("Splitter").
> There are a total of four word-forms for /**ʃp**/.
> Of these, one ("Spitze") is both from the same part of the corpus (Group 2) and in the same position (second) in the carrier phrase as "Splitter".
> This will be referred to as the "best pair".
> Of the other /**ʃp**/ words two are from Group 2, first position; one is from Group 1, second position.

Overview of temporal structure

Looking now at an overview of the temporal structure (next figure) for a first impression of the actual findings, the following observations can be made:

- Total duration of voicelessness appears similar for /**ʃp**/ and /**ʃpl**/ in all speakers, but there is some indication of a longer glottal gesture for /**ʃpl**/ than /**ʃp**/ in SF and CG.
- Shortening of occlusion durations for /**ʃp**/ in the /**ʃpl**/ cluster are not very salient
- Not easy to determine from this representation the salience of any shifts in the timing of peak glottal opening relative to the fricative.
- But as in the previous two sections the timing of the end of the glottal gesture relative to the end of the second consonant shifts noticeably for two of the three subjects (SF and CG)
- The acoustically defined duration of the /**l**/ segment appears shorter than in previous sections (rightmost segment of red bars). This is probably because of greater overlap with the preceding /**p**/ than, for example, in the /**pl**/ clusters (articulatory onset of /**l**/ determined from EPG marked by left edge of cyan-coloured (light blue) bar; cf. Fig. 6.31 on p. 112 for /**pl**/)

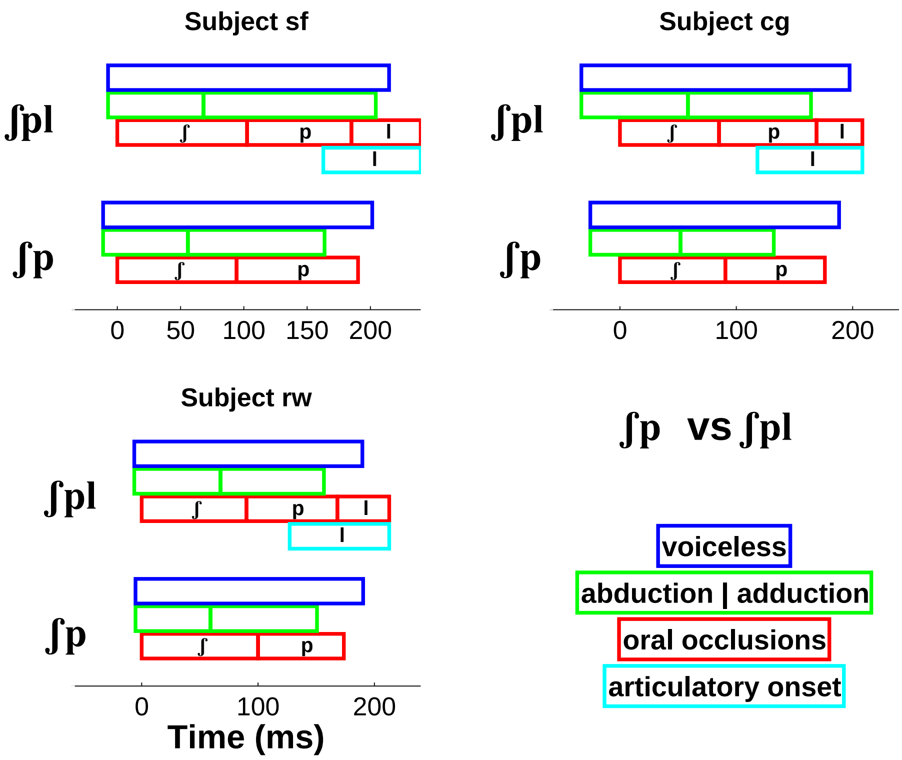


Fig. 6.55: Temporal structure of /**ʃp**/ and /**ʃpl**/ sequences, lined up at onset of occlusion for /**ʃ**/. Left edge of cyan-coloured bar indicates articulatory onset of /**l**/ occlusion determined from EPG (see Fig. 6.31 on p. 112 for further details).

Detailed results of individual parameters

Consideration of the parameters in details will follow basically the same progression as in the previous two sections (the statistical details are to be found in the table on p. 71:

- Oral occlusion duration of fricative segment /**ʃ**/ (no figure)
- Oral occlusion duration of plosive segment /**p**/
- Glottal gesture duration
- Amplitude of peak glottal opening (no figure)
- Total duration of voicelessness
- Relative position of peak glottal opening in the fricative segment
- Relative position of release of plosive occlusion (i.e the second occlusion) within the glottal gesture

Considering first the durations of the /**ʃ**/ and /**p**/ segments it turned out that these changed extremely marginally. Results for both segments are given in in the table on p. 71 but only the result for the /**p**/ segment is shown in the figure below. Note that the table gives two variants of the statistical analysis, one with all available material for /**ʃp**/ word-forms, one with only the word forming the best pair for comparison with “Splitter” (figures will be based on the complete material unless indicated otherwise). The /**p**/ segment was chosen for illustration in the figure because speaker SF gave the only case where the parametric procedure reached significance for either of the two segments. Note that even here the longer duration of /**p**/ in /**ʃp**/ than /**ʃpl**/ is only of the order of 10ms and was no longer significant for the "best pair" material. For the initial /**ʃ**/ segment the only significant result was for speaker SF with the non-parametric test, in the direction of longer duration for /**ʃpl**/ than /**ʃp**/; again this was only for the full material. In short, as a background for consideration of the laryngeal activity, any changes in oral occlusion durations can be considered absolutely marginal. Thus, segmental shortening in the more complex cluster is less apparent than in /**pf**/ vs. /**pfl**/, where in turn it was less apparent than in fricative vs. fricative-/**l**/ clusters.

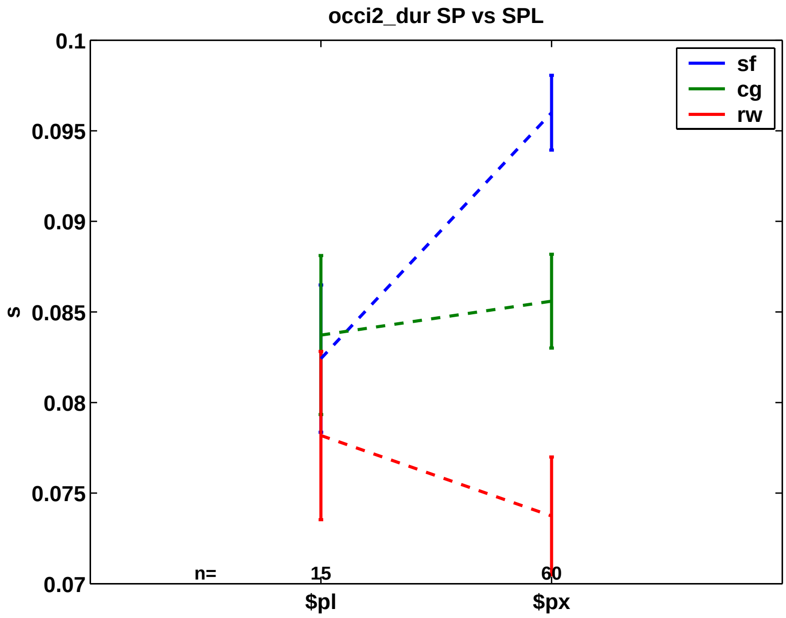


Fig. 6.56: Duration of /**p**/ occlusion in /**ʃp**/ and /**ʃpl**/ clusters. Complete /**ʃp**/ material. The abscissa labels '$pl' and '$px' stand for /**ʃpl**/ and /**ʃp**/, respectively (as in previous sections the 'x' in the label stands for the unoccupied consonant position in the syllable onset)

Turning to the overall duration of the glottal gesture, it will be recalled that in none of the previous sections involving combinations with /**l**/ did we find robust evidence for the radical hypothesis that the glottal gesture is longer in clusters with /**l**/ than without. The present section provides the clearest indication so far that this possibility cannot be completely discounted (see figure below).

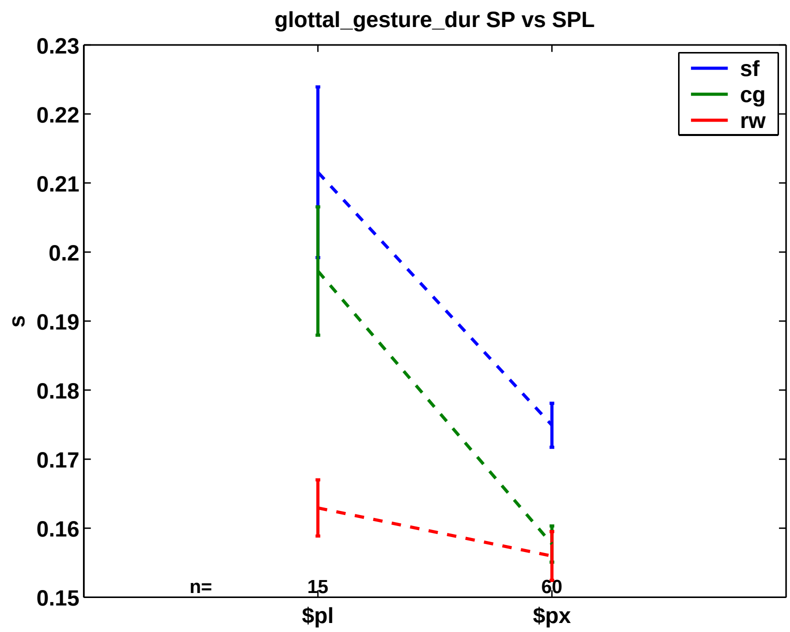


Fig. 6.57: Duration of glottal gesture in /**ʃp**/ and /**ʃpl**/ clusters. Other details as in Fig. 6.56

As seen in the figure, for both SF and CG there was a difference of 30ms or more, and the difference appeared statistically quite robust (see table on p. 71 for details). For RW, on the other hand, there was clearly no significant effect. (By way of comparison with /**pf**/ vs. /**pfl**/ in the previous section, we there found a consistent trend in all three subjects for longer gestural duration in the /**l**/-cluster, but only on the order of 10ms difference, and not significant.) We will return at the end of this section to the question of whether there can be any functional significance to a lengthening of the glottal gesture, and, if so, why the strategy is not followed by all subjects.

As in most previous sections there was little evidence of any robust tendencies regarding glottal opening amplitude, so no detailed results will be shown. For CG and RW the trend was in the direction of larger opening for /**ʃpl**/, but the only case to reach even $p<0.05$ in the nonparametric test was the best-pair material for speaker RW.

Regarding the total duration of voicelessness the results can also be quickly stated: The tendency was in the direction of slightly longer durations for /**ʃpl**/ for SF and CG (of the order of 10ms; see next figure), but no results at all reached significance. Given this, it is not necessary in the present section to consider what contribution aerodynamic conditions would make to the duration of voicelessness[52].

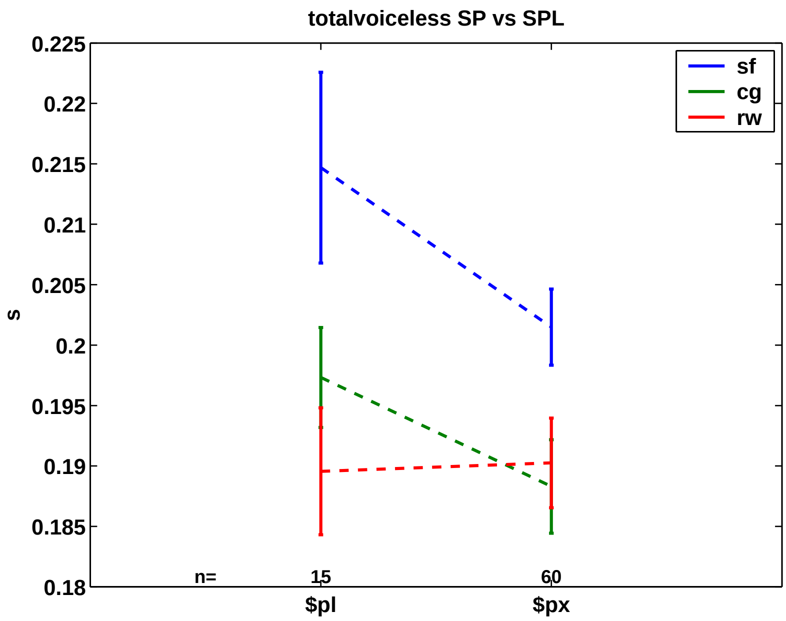


Fig. 6.58: Total duration of voicelessness in /**ʃp**/ and /**ʃpl**/ clusters. Other details as in Fig. 6.56

As we turn to the specific measures of laryngeal-oral coordination it should be noticed that there is no contradiction in the finding of no significant differences in voicelessness duration on the

[52]Estimating any contribution of aerodynamic conditions is also more problematic in this section because it is not clear that glottal conditions are comparable at the release of /**p**/ for /**ʃp**/ and /**ʃpl**/, at least for speakers SF and CG. For speaker RW, however, glottal adduction is essentially completed before release of /**p**/ in both cases, so for him it might be possible to assume that the delay from /p/ release to voice onset does reflect supraglottal influences on decay of intraoral airpressure. And indeed, little difference between /**ʃpl**/ and /**ʃp**/ is apparent, fitting in with the general impression that for the material used in this study aerodynamic conditions have a relatively minor role to play.

one hand, and significant differences in glottal gesture duration (for two speakers) on the other hand: This is because of the rather early completion of glottal adduction during the /**p**/ of /**ʃp**/ already discussed above in the section on fricative-plosive combinations (p. 95f and Fig. 6.21), leading to a delayed onset of voicing relative to completion of glottal adduction simply because the oral occlusion is still in place.

The relative position of peak glottal opening within the initial fricative segment shows a common pattern over all three subjects for lower values in the cluster with /**l**/, i.e a relatively later position. The statistical significance of the result varies somewhat over the subjects in terms of whether the complete or best-pair material shows the higher level, but the basic pattern is the same regardless of the material. Taking all the comparisons involving /**l**/ together (/**p**/ vs. /**pl**/, fricative vs. fricative-/**l**/, /**pf**/ vs. /**pfl**/ and /**ʃp**/ vs. /**ʃpl**/) the shift of peak glottal opening to a relatively later position can thus be considered a common pattern, even if a statistically significant shift is not found for every speaker and sound sequence (see Figs. 6.35, 6.45 and 6.53 on p. 115, 124 and 132 respectively for the first three comparisons).

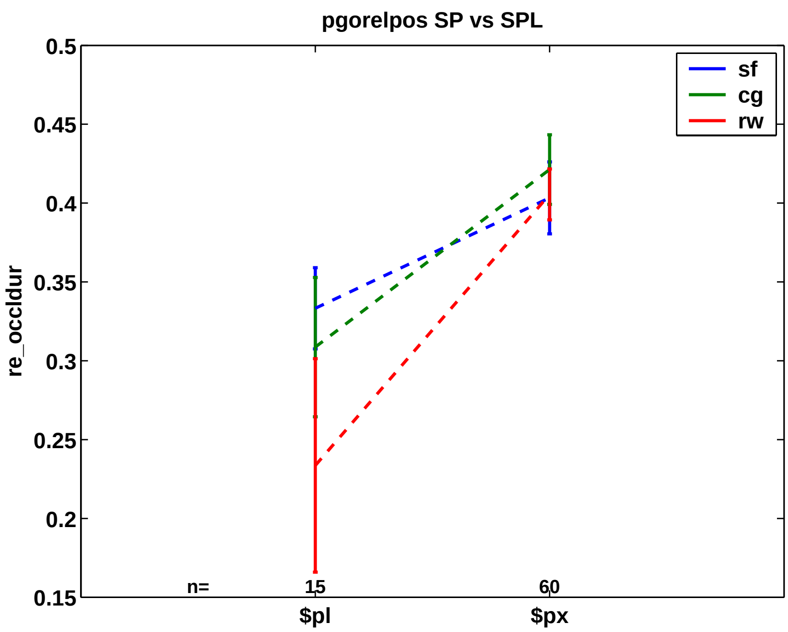


Fig. 6.59: Relative position of peak glottal opening in the fricative segment of /**ʃp**/ and /**ʃpl**/. Other details as in Fig. 6.56

The result for the second main coordination parameter, the relative position of the release of the second oral occlusion (i.e of the /**p**/) within the glottal cycle, also continues a pattern found with the /**pfl**/ and fricative-/**l**/ sequences, regarding the last non-lateral consonant in the examined sequences. For two speakers, SF and CG, there is a very pronounced effect in the direction of an earlier end of the /**p**/ when the lateral follows (the values often well above 1 for the non-lateral context /**ʃp**/ indicate that /**p**/ release is there often located after the end of the glottal gesture). For RW there is a very weak tendency in the same direction, but it is nowhere near reaching statistical significance. Looking back over the results for fricative vs. fricative-/**l**/ and /**pf**/ vs. /**pfl**/ (Figs. 6.46 and 6.54 on p. 125 and 133) it would appear that in fact the relative position within the glottal cycle of the final release before the lateral actually shows stronger evidence of shifts in timing between lateral and non-lateral contexts than does the timing of peak glottal opening since the result here for RW is the one major exception to an otherwise consistent pattern.

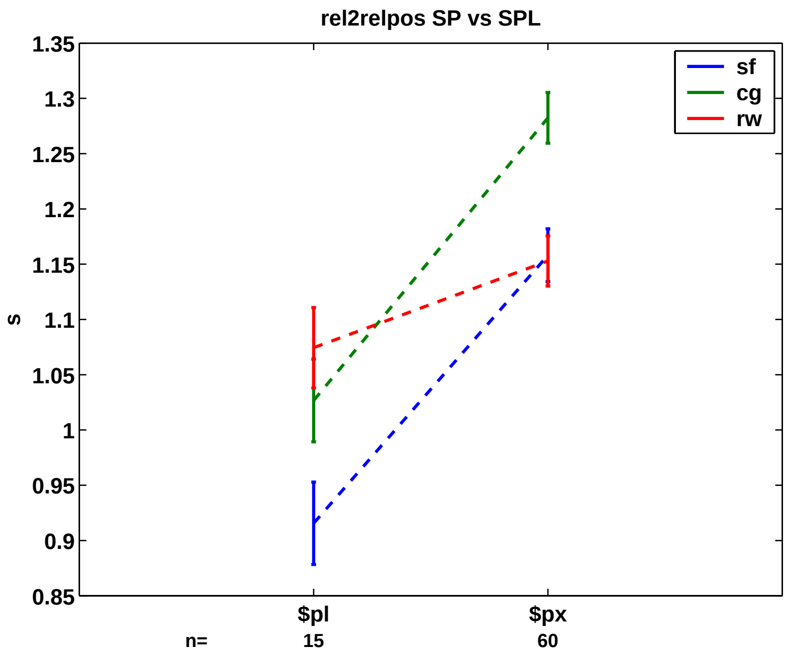


Fig. 6.60: Relative position of release of /**p**/ occlusion within the glottal cycle for /**ʃpl**/ and /**ʃp**/. Other details as in Fig. 6.56

Preliminary summary

We have now presented the bare bones of the results for the last in the series of comparisons of syllable onsets with and without /**l**/. Throughout this series we have noted that shifts in coordination almost invariably occur and are most consistently observable in terms of the point of the glottal cycle coinciding with the end of the last non-sonorant occlusion. Before going on to the general discussion of the whole investigation there remains one point specific to the /**ʃpl**/ vs. /**ʃp**/ comparison to deal with. This is the intriguing question as to why the glottal gesture should lengthen noticeably for two subjects in the lateral context, and lead to equally noticeable shifts in relative timing. In the section on /**ʃ**/ vs. /**ʃp**/ we argued that the early completion of the glottal gesture in the /**ʃp**/ sequences argued against the presence of two underlying laryngeal gestures (or competition between two voiceless segments for dominance of a single laryngeal gesture). The present results strengthen this conclusion, in the sense that even if, for example, a lengthening of the glottal gesture occurred for /**ʃp**/ vs. /**ʃ**/ then this could not be unambiguously interpreted as overlap or competition of two voiceless segments, since lengthening of the glottal gesture can apparently occur under the influence of a segment that would be assumed to be underlyingly voiced. However this may be, it still does not answer the question as to why lengthening occurs: one would assume that the pattern of coordination found in /**ʃp**/ with early completion of the glottal gesture would be perfectly acceptable also for /**ʃpl**/. Unlike /**pl**/ clusters there is no obvious functional reason to delay voicing onset in the /**l**/ since there are no contrasts for voicing in the segments preceding /**l**/ that could be enhanced. Perhaps the crucial constraint is that the sound following simple /**ʃp**/ is the vowel, i.e part of the syllable nucleus, whereas in /**ʃpl**/ the sound following /**ʃp**/ is still part of the syllabic onset. Thus, spreading of voicelessness beyond the /**ʃp**/ segments may be strongly dispreferred in the first case, but not the second. In fact, as discussed in the section on fricative-/**l**/ clusters it may simply be physiologically economical to tolerate devoicing of the sonorant. For /**pfl**/ vs. /**pf**/ in the preceding section we were hesitant as to whether the results were better viewed as active enhancement of voicelessness for /**pfl**/ or just absence of active minimization of voicelessness. In the present case of /**ʃpl**/ vs.

/**ʃp**/ there is probably even slightly more justification for active support of voicelessness in the sonorant case, firstly because of a clearer tendency towards lengthening of the glottal gesture, and secondly because ‘passive’ emergence of sonorant devoicing from shortening of oral occlusion durations of the pre-sonorant segments was extremely restricted compared to the situation in earlier sections. Once again the most parsimonious overall interpretation appears to be that speakers plan the laryngeal-oral coordination in terms of the constraints obtaining for the syllable onset as a whole, rather than assembling the movement pattern segment by segment on the basis of pre-defined coordination relations.

7 General Discussion

The first main focus of this discussion will be on consonant sequences with /**l**/, and will essentially consist in weighing up again the evidence in favour of the three potential scenarios outlined at the end of Chapter 2 (p. 23f). These sequences will be looked at before the purely voiceless sequences (the second main focus) because, as also pointed out in Chapter 2 and elsewhere, the results for the mixed voice sequences can have implications for the interpretation placed on the results for the purely voiceless clusters.

Before turning to these central topics of the investigation let us recap briefly the main findings for single consonants, since clearly they provide a convenient framework for viewing the more complex sequences.

7.1 Single consonants

Regarding the comparison of plosives and fricatives, the clearest findings were that fricatives showed a somewhat longer glottal gesture and, in particular, an earlier onset of glottal abduction relative to the formation of the oral occlusion. Neither of these results were unexpected, but the present investigation made very clear that the latter point constitutes probably the most stable of all the laryngeal-oral coordination relations that were looked at: The distinction between fricative-onset and plosive-onset was completely clear-cut over all the syllable onsets looked at in the present material. Slightly unexpected was the fact that no consistent evidence was found for a greater glottal opening in fricatives, nor for higher abduction velocity. This may just reflect the uncertainties in estimating the magnitude of glottal opening using transillumination, so more carefully controlled material or simply a larger number of repetitions might give a different result (throughout this investigation hardly any robust differences in glottal opening amplitude were found - leaving aside the gross effect of loud vs. normal speech). However, at least for languages with clearly aspirated plosives, and at least when the latter occur in prosodically strong positions, the differences between fricatives and plosives may not be as clear-cut as the literature has hitherto suggested.

The second issue relating to single consonants was that of place of articulation in plosives. The point of departure was the common finding that VOT is longer in /**t**/ than /**p**/, which raises the question as to what pattern of laryngeal-oral coordination is responsible for this. One scenario that appeared plausible was that the glottal gesture is essentially constant in duration, and the difference in VOT simply emerges from the shorter oral occlusion duration for /**t**/ than /**p**/. This is of interest in the context of our overall concern with consonant sequences because a similar mechanism could explain, for example, a longer VOT in /**pl**/ than /**p**/. Thus, as a preliminary step it seemed worth investigating whether a preferred scenario would emerge for the single-consonant question. In fact, the results were rather messy: Only two of the three speakers showed longer aspiration for /**t**/ in the first place. Of these two, one roughly fitted the above scenario, whereas for the other a longer gestural glottal gesture duration for /**t**/ appeared more important. Unexpectedly, for both speakers, the time instant of the start of glottal abduction relative to oral occlusion onset was shifted towards later values for /**t**/ and thus could also have contributed to later completion of the gesture and longer VOT. The latter finding regarding the time-point at which glottal abduction is initiated is probably not of great functional significance in itself, but it prefigured a finding in the first cluster with /**l**/ that was looked at, namely /**pl**/, and thus reinforces the impression that speakers may exercise quite fine-grained control over the initiation of the glottal gesture above and beyond the basic difference between plosives and fricatives.

7.2 Clusters with /l/

As just mentioned, three basic scenarios were envisaged for the laryngeal-oral coordination relations (enumerated on p. 23f).

Under the first scenario the sonorant is basically irrelevant: In cases where the voiceless segments shorten in the cluster with /**l**/ the glottal gesture would shorten in parallel, maintaining the same coordination relations with the voiceless segment(s). The second scenario has just been recapped with respect to place of articulation effects in single plosives, i.e oral articulations shift with respect to a constant glottal gesture. The third scenario, nicknamed the 'radical' one, was that the glottal gesture might actually lengthen as the syllable onset becomes longer with the addition of the sonorant.

The task now is to weigh up the extent to which each scenario is compatible with the empirical observations and consider the implications.

For scenario 1 virtually no evidence was found. This lack can be seen from two points of view: In cases where voiceless segment durations decreased in the the sonorant clusters only one case was found where the glottal gesture duration decreased in parallel (fricative-/**l**/ for speaker SF), and even then the coordination relations did not remain completely fixed (specifically, the end of the fricative occlusion was located earlier in the glottal cycle for fricative-/**l**/). The other perspective involves those cases where voiceless occlusion duration actually did not change much (e.g /**pl**/, /**ʃpl**/); here, too, the nature of the coordination relations generally changed. Scenario 1 was in fact the one for which least prior evidence was available in the literature. Nevertheless, we think it still worth emphasizing the implication, namely that coordination may not be planned purely at the level of the individual voiceless segments, but at a higher level of the syllabic structure. Accordingly, this scenario has a much more serious status than simply that of a cheap target that is brought into contention purely so that it can be shot down with a flourish.

Having used scenario 1 to make the point that coordination relations do shift when an /**l**/ is added to the syllable onset one then has to ask whether this can be directly attributed to one of the remaining two scenarios. The answer, as the reader will hopefully have been able to pick up in the course of negotiating the winding path through the results section, is a resounding 'no'.

Let us briefly list the cases that fit either of scenarios 2 or 3 in a straightforward way.

Scenario 2: As just seen this was effectively short-circuited by the cases where occlusion duration showed little change. Of the remaining possible cases, the pattern of shorter occlusion with constant glottal gesture was observable for /**pl**/ for one subject (CG), fricative-/**l**/ for two subjects (CG and RW), and /**pfl**/ for two subjects (CG and SF).

Scenario 3: Clear glottal gesture lengthening was only found for /**ʃpl**/ (speakers SF and CG). A weak effect was found for /**pl**/ (speaker RW), while for /**pfl**/ the tendency was consistent for all three subjects but did not reach significance.

We thus find ourselves in the quandary that neither coordination strategy is preferred, but, equally, neither is so far outside the ball-park that we can be sure that subjects never use it.

The way out of this situation - as already indicated in the course of presenting the results - is to consider where a variety of different movement patterns leads to a consistent outcome, i.e to consider where there may be a similar effect not only of shortening of occlusion duration and lengthening of glottal duration but also other features we no longer need to consider in detail, such as changes in the time at which glottal abduction is initiated, and in the ratio of abduction to adduction duration.

In fact, a fair amount of consistency was found. To review this, it is convenient to consider /**pl**/ sequences separately from the other three categories (all containing a fricative), even though the underlying phenomenon is probably the same.

For /**pl**/ a variety of movement patterns were unified by the fact that they led to longer VOT in /**pl**/ compared to /**p**/. This is where variability is actually useful, since it makes it more plausible that the increased VOT corresponds to a relevant component in terms of which the utterances have been planned. If increased VOT had simply resulted in a mechanical way from reduced oral occlusion durations then it might represent no more than a side-effect of preferred oral timing patterns in clusters[53].

Regarding the other sequences (fricative-/**l**/, /**pfl**/, /**ʃpl**/) the most obvious common feature was that the end of the last non-sonorant occlusion (i.e the fricative in the first two of these cases, the plosive in the last one) was located earlier within the glottal cycle compared to the corresponding sequence without /**l**/ [54]. This can be seen as a parallel to increased VOT in the /**pl**/ sequences in the sense that speakers do not appear to be trying to minimize the amount of voiclessness on the /**l**/.

Why should speakers adopt this pattern of coordination? For /**pl**/ sequences it seemed possible to a certain extent to appeal to the necessity to maintain distinctiveness between clusters with voiced and voiceless initial plosives on the one hand, and clusters ending in different sonorants on the other hand. For the clusters including fricatives this functional explanation is less clear. This is where the question of aerodynamic effects as a possible influence on the duration of VOT or total voicelessness comes in. It may have been noticed that there is a touch of ambivalence in the way we have approached this topic up to now.

---

[53]Here I am reminded of a dictum attributed to Shinji Maeda: “Variability: Awful or Lawful?”

[54]Considering all the material with /**l**/, a consistent trend was also found for the other main timing parameter, namely the position of peak glottal opening. This shifted to a relatively later position in the clusters with /**l**/ (expressed relative to the /**p**/-occlusion in /**pl**/ and relative to the fricative segment in the other cases). However, the statistical significance of the effect fluctuated somewhat over cluster type and speakers.

On the one hand, it was an important feature of this investigation that the direct registration of glottal activity made it possible to give some estimate of these effects. Specifically, it was assumed that if the duration of the interval from completion of glottal adduction to onset of voicing was longer in the /l/-clusters, then aerodynamic conditions were less conducive to voicing in this case. In fact, however, very little evidence for this was found in our material (although there were a couple of largish exceptions for speaker RW). This fact, together with the directly observable changes in the coordination patterns themselves, means that we are in a position (which would not be the case in a purely acoustic investigation) to assume that the increased voicelessness in sonorant contexts represents planned behaviour on the part of the speaker. Nevertheless - this is the other hand - in the search for an explanation as to why speakers should favour this pattern of organization, particularly in clusters with fricatives where there is no obvious functional advantage to it, we suggested that aerodynamic conditions may still represent a physical bias to the system, pushing it in the direction of greater voicelessness in the sonorant contexts, and that speakers where possible exploit these physical forces in planning their behaviour. We encounter here a style of explanation that also appeared attractive in Part I of this monograph. Do these two hands represent a contradiction? Not necessarily. There were hints of a tendency in the direction of delayed voicing onset in sonorant contexts in our data; presumably, the difference versus high vowels is simply too slight to give a statistically significant result, but this need not be an insuperable barrier to speakers learning to associate sonorants as a whole and vowels as a whole with different aerodynamic behaviour. Moreover, there is probably also a functional boundary condition that needs to be taken into account: We assume that speakers are only able to exploit a pattern of organization that follows the physical bias in the system (thus representing in some sense a physiological economy of effort) because the propagation of voicelessness in the syllable onset to segments that one might still like to regard as underlyingly voiced does not compromise the basic pattern of sonority modulation on which syllable structure is built up.

This account can thus perhaps be seen as a kind of extrapolation from Browman & Goldstein's original contention that it is regularity of English that words can begin with at most one glottal gesture. It is also interesting to refer in this connection to the recent work of Kehrein & Golston (2004). Coming from a completely different perspective, namely an examination of the laryngeal contrasts attested in a very wide range of languages, they conclude that "*laryngeal features are properties not of segments, but of the onsets, nuclei and codas that dominate them*" (p. 325).

7.3 Voiceless clusters

Regarding purely voiceless clusters, we were interested in basically two questions: Firstly, to assess whether the glottal movement pattern in sequences of two voiceless segments is more plausibly modelled as consisting underlyingly of one gesture (Browman & Goldstein) or as the overlap of two gestures (Saltzman & Munhall, Munhall & Löfqvist); secondly, to assess whether Browman & Goldstein's two rules give an adequate account of the coordination relations.

Regarding the first question, and restricting consideration to fricative-plosive vs. single fricative, we think there are three reasons why the evidence comes down in favour of a single underlying gesture:

1. No clear evidence for increased gestural magnitude or duration in the longer sequence (in this study, glottal gesture duration really only separated single plosives from the rest)

2. Early completion of glottal adduction in the fricative-plosive case is difficult to reconcile with an underlyingly active glottal gesture for the plosive (even if it is assumed to be weak in non-initial position), but is perfectly natural in terms of the aerodynamic requirements of the cluster.

3, Shifts in the timing of peak glottal opening are potentially quite similar for /**ʃt**/ vs. /**ʃ**/ as for /**ʃl**/ vs. /**ʃ**/ (compare Figs. 6.18 and 6.45 on p. 93 and 124), so at least we must conclude that overlap of two underlying gestures cannot be the only possible explanation for such shifts.

Regarding the second question: This revolves to a large extent around the status that is to be accorded to the timing of peak glottal opening in formulating the coordination relations. Clearly it is a physiologically attractive time-point since it marks the boundary between abductory and adductory activity. However, considering the wide range of sequences investigated in this study indicates that its role may have been overstated. Looking at the sequences involving fricatives there is no evidence that its location is linked particularly closely to mid-frication. A particularly drastic case is plosive-fricative, where peak glottal opening is generally located very early in the fricative (and occasionally even in the plosive). Browman & Goldstein do not consider this sequence specifically, so it is perhaps not fair to use it as evidence against their rules. Perhaps they could be expanded along the lines of the formulation of Saltzman & Munhall to account for plosive-fricative as a sound sequence that favours a 'compromise' location of peak glottal opening particularly strongly (see p. 109)[55]. No doubt it would be possible to find some specification of the dominance relations (for fricatives vs. plosives and for initial vs. non-initial position) that causes peak glottal opening to emerge in the right location. But it is not clear how far one can push such a weighting scheme without incurring a degree of abitrariness. As an alternative approach we feel it could be more parsimonious (involve fewer ad hoc stipulations) if the coordination relations are captured in terms of the fulfilment of a set of constraints given by the aerodynamic and functional demands of each specific syllable onset, rather than in terms of a pre-defined set of dominance relations[56].

This can be observed quite neatly for the purely voiceless sequences (but we have tried to argue for its basic plausibility in the sonorant combinations as well): comparing fricative-plosives and plosive-fricatives there is a clear constraint in the former case that glottal abduction should start early (as repeatedly emphasized), while the time-point at which glottal adduction is completed is relatively unconstrained. For plosive-fricatives there is in effect the opposite constraint: completion of glottal adduction has to be fairly closely coordinated with end of fricative occlusion, while start of glottal abduction can use the standard plosive timing pattern. The great advantage of a formulation in terms of constraints is that it provides a framework within which

---

[55]The approaches of Browman & Goldstein and Saltzman & Munhall are here fairly similar in terms of the idea of dominance: For the former it is a question of which segment 'wins' when two compete for a single glottal gesture; for the latter it is question of how two gestures are weighted when they blend together.

[56]It is nevertheless pertinent to note that Browman & Goldstein's rules would probably actually not result in aberrant output if used as rules for articulatory synthesis. The point we are trying to make is that they may, however, not correspond to how human speakers represent the coordination relations.

interspeaker variability can readily occur, rather than just being the bane of analysis. For example, in fricative-plosive sequences speakers can presumably make different choices regarding, for example, the precise length of the glottal gesture or the ratio of abduction to adduction duration. For /**pl**/ we suggested that the basic constraint is to achieve a VOT that is longer than that of single /**p**/. Several different coordination patterns can achieve this goal.

7.4 Outlook

As pointed out in the introduction, the basic kinematic properties of the glottal abduction-adduction movement are now quite well-understood. And at least since the seminal work of Lisker & Abramson on VOT it has been clear that understanding coordination relations is crucial to the understanding of linguistic distinctions. Nevertheless as Gafos (2002, p. 10) remarked, "*the study of temporal relations among gestures is in its infancy*". The purpose of this work has been to get us closer to at least being on speaking terms for one specific circumscribed area. The point of departure was that it would be premature to assume that we really understand the coordination principles even for such simple and completely unexotic sound sequences like /**pl**/, /**ʃl**/, /**ʃp**/.

In closing, it now remains to consider how the additional knowledge that we have gained in the course of this investigation would benefit from further refinement.

Two areas suggest themselves.

The first one involves plosive-fricative sequences. The sounds we included in this category (e.g /**pf**/, /**ps**/ etc.) are certainly not linguistically homogeneous, and there are almost certainly differences even in the purely supraglottal organization of these sounds. Whether there are accompanying differences in the laryngeal-oral coordination did not seem feasible to determine on the basis of the present material because of the rather uneven nature of the target words. So further analysis with a different corpus could be worthwhile (perhaps extended to include sequences like /**ks**/).

The second area has potentially wider ramifications, namely closer consideration of further clusters with sonorants. Now that we have seen that clusters with /**l**/ can potentially lead to interesting shifts in the coordination patterns it would be interesting to expand the material in several directions. The simplest extension would be to include /**k**/ as initial consonant, since we saw right at the beginning in the literature review that occlusion duration of /**k**/ and /**p**/ may not behave in quite the same way in clusters. A more illuminating direction could be to now take the risk of including those sonorants that are potentially somewhat problematic with the fiberoptic technique: In particular, clusters with nasals would be interesting because of the probably differing aerodynamic conditions between laterals and nasals; furthermore, clusters with /**r**/ would be interesting in the light of Jessen's striking results that they can attract very noticeable devoicing activity. This is also a reason why extending the investigation to additional languages (or dialects of German) where /**r**/ cannot be realized as a uvular fricative could be worthwhile. In any case, reproducing this finding with additional speakers seems important because of the intriguing possibility that even closely related sounds such as /**l**/ and /**r**/ may require different laryngeal specifications (apparently with a stronger tendency to devoicing on the sound that is traditionally assumed to have higher sonority), and that some German speakers (in sequences like /**ʃr**/) may actively aim to produce a highly marked sequence of two fricatives. A further argument for a cross-language perspective is that even in closely related languages like German and

English the functional load of these clusters certainly need not be identical; for example, the status of /**pl**/ undoubtedly differs between these two languages because of the presence of /**pfl**/ in German. Finally, if we are correct in assuming that voicelessness of the sonorants at least to a certain extent represents planned behaviour on the part of the speaker, then the question arises as to whether this then represents a feature that can be further enhanced, for example when the target words occupy prosodically prominent locations.

8 References

Baer, T., Löfqvist, A., & McGarr, N.S. (1983). "Laryngeal vibrations: A comparison between high-speed filming and glottographic techniques". JASA, 73, 1304-1308.

Benoit, C. (1986). "Note on the use of correlations in speech timing". J. Acoustical Soc. America, 80, 1846-1849.

Borden, G. & Gay, T. (1979). "Temporal aspects of articulatory movements for /s/-stop clusters". Phonetica, 36, 21-31.

Browman, C.P. & Goldstein, L. (1986). "Towards an articulatory phonology". Phonology Yearbook, 3, 219-252.

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

Butcher, A. (1977). "Coarticulation in intervocalic plosives and fricatives in connected speech". Arbeitsberichte Kiel, 8, 154-213.

Chasaide, A N. & Gobl, C. (1993). "Contextual variation of the vowel voice source as a function of adjacent consonants". Language and Speech, 36 (2-3), 303-330.

Cho, T. & Ladefoged, P. (1999). "Variation and universals in VOT: evidence from 18 languagess". J. Phonetics, 27(2), 207-229.

Cooper, A.M. (1991). "Laryngeal and oral gestures in English /p, t, k/". Proc. XIIth Cong. Phonetic Sciences, Aix, 2, 50-53.

Dent, H. (1984). "Coarticulated devoicing in English laterals". Reading WP 4, 111-134.

Docherty, G.J. (1992). "The timing of voicing in British English obstruents". Berlin, Foris.

Döllinger M, Braunschweig T, Lohscheller J, Eysholdt U, Hoppe U (2003). "Normal Voice Production: Computation of Driving Parameters from Endoscopic Digital High Speed Images Methods". Inf Med, 42, 271-276.

Fowler, C. & Smith, M. (1986). "Speech perception as 'vector analysis': An approach to the problems of invariance and segmentation". Perkell, J. & Klatt, D. (eds.), Invariance and variability in speech processes, pp. 123-136.

Frøkjær-Jensen, B., Ludvigsen, C. & Rischel, J. (1971). "A glottographic study of some Danish consonants". In Hammerich, L., Jakobson, R. & Zwirner, E (Eds.), Form and Substance. Copenhagen: Akademisk Forlag.

Frøkjær-Jensen, B. (1967). "A photo-electric glottograph". ARIPUC, 2, 5-19.

Fuchs, S. (2005). "Articulatory correlates of the voicing contrast in alveolar obstruent production in German". ZASPiL, 41, 1-238.

Fujimura, O., Baer, T. & Niimi, A. (1979). "A stereoscopic fiberscope with magnetic interlens bridge for laryngeal observation". JASA, 65, 478-480.

Fujino, A., Kaburagi, T., Honda, M., Murano, E. & Niimi, S (2000). "Relative timing between articulatory and glottal motions in voiceless consonant production.". Proc. 5th Speech Production Seminar, 89-92.

Gafos, A. (2002). "A grammar of gestural coordination". Natural language and linguistic theory, 20(2), 269-337.

Goldstein, L. (1990). "On articulatory binding: Comment's on Kingston's paper". In J. Kingston and M.E. Beckman (eds.), Papers in laboratory phonology: 1. Between the grammar and physics of speech, pp. 445-450. Cambridge, CUP.

Haggard, M. (1973). "Abbreviation of consonants in English pre- and post-vocalic clusters". J. Phonetics, 1, 9-24.

Hamlet, S.L. (1981). "Ultrasound assessment of phonatory function". In C.L. Ludlow & M. O'Connell Hart (eds.), Proc. of the conference on the assessment of vocal pathology = ASHA reports 11, pp.128-140).

Hanson, H.M. & Stevens, K.N. (2002). "A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn". J. the Acoustical Society of America, 112(I), 1158-1182.

Hawkins, S. (1979). "Temporal coordination in the speech of children: further data". J. Phonetics, 7, 235-267.

Hirose, H., & Niimi, S. (1987). "The relationship between glottal opening and the transglottal pressure differences during consonant production". In Baer, T., Sasaki, C. & Harris, K.S (Eds.), Laryngeal function in phonation and respiration, pp. 381-390 Boston: College Hill.

Hirose, H. (1971). "Electromyography of the articulatory muscles: Current instrumentation and technique". Haskins SR 25/26, 75-86.

Hirose, H. (1975). "The posterior cricoarytenoid as a speech muscle". ABRILP, 9, 47-66.

Hirose, H. (1986). "Pathophysiology of motor speech disorders (dysarthria)". Folia Phoniatrica, 38, 61-88.

Honda, K., Kiritani, S., Imagawa, H. & Hirose, H. (1987). "High-speed digital recording of vocal fold vibration using a solid-stae image sensor". In Baer, T. et al (Eds.), Laryngeal function in phonation and respiration, pp. 485-491. Boston: College Hill.

Hoole, P., Pompino-Marschall, B. & Dames, M. (1984). "Glottal timing in German voiceless occlusives". In Van den Broecke, M.P.R. & Cohen A (Eds.), Proceedings 10th Int. Cong. Phonetic Sciences, pp. 399-403. Dordrecht: Foris.

Hoole, P., Schröter-Morasch, H. & Ziegler, W. (1997). "Patterns of laryngeal apraxia in two patients with Broca's aphasia". Clinical Linguistics and Phonetics, 11, 429-442.

Hoole, P. (1987). "Velar and glottal activity in a speaker of Icelandic". Proc. 11th Int. Cong. Phonetic Sciences, 3, 31-34. Tallinn: Estonian Academy of Sciences..

Hoole, P. (1996). "Theoretische und methodische Grundlagen der Artikulationsanalyse in der experimentellen Phonetik". FIPKM, 34, 3-173.

Hoole, P. (1999a). "Laryngeal Coarticulation. Section A: Coarticulatory investigations of the devoicing gesture". In: W.H Hardcastle and N. Hewlett (eds.), Coarticulation: Theory, Data and Techniques, pp. 105-121. Cambridge University Press.

Hoole, P. (1999b). "Techniques for investigating laryngeal articulation. Section A: Investigation of the devoicing gesture". In: W.H Hardcastle and N. Hewlett (eds.), Coarticulation: Theory, Data and Techniques, pp. 294-300. Cambridge University Press..

Hutters, B. (1976). "Problems in the use of the photoelectric glottograph". ARIPUC, 10, 274-312.

Hutters, B. (1984). "Vocal fold adjustments in Danish voiceless obstruent production". ARIPUC, 18, 293-385.

Hutters, B. (1985). "Vocal fold adjustments in aspirated and unaspirated stops in Danish". Phonetica, 42, 1-24.

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

Jessen, M. (1995). "Glottal opening in German obstruents". Proc. XIIIth Int. Cong. Phon. Sci., Stockholm, 3, 428-431.

Jessen, M. (1998). "Word boundary marking at the glottal level in the production of German obstruents". ZAS Papers in Linguistics 11, 147-166.

Jessen, M. (1999). "Redundant aspiration in German is primarily controlled by closure duration". Proceedings XIVth ICPhS 2, 993-996.

Kehrein, W. & Golston, C. (2004). "A prosodic theory of laryngeal contrasts". Phonology 21(3), 325-357.

Keller, E., & Ostry, D.J. (1983). "Computerized measurement of tongue-dorsum movement with pulsed-echo ultrasound". JASA, 73, 1309-1315.

Kingston, J. & Cohen, A.H. (1992). "Extending articulatory phonology". Phonetica, 49, 194-204.

Kingston, J. (1990). "Articulatory binding". In J. Kingston and M.E. Beckman (eds.), Papers in laboratory phonology: 1. Between the grammar and physics of speech, pp. 406-434. Cambridge, CUP.

Kiritani, S., Imagawa, H. & Hirose, H. (1992). "Vocal-cord vibration during consonants - High-speed digital imaging using a fiberscope". Proc. 1992 Int. Conf. on Spoken Language Processing, 1, 1661-1664.

Kiritani, S. (1971). "X-ray monitoring of the position of the fiberscope by means of computer-controlled radiography". ABRILP, 5, 35-39.

Klatt, D. (1973). "Duration of pre-stress word-initial consonant clusters in English". MIT Research Lab of Electronics Quarterly Progress Report, 108, 253-260.

Klatt, D. (1975). “Voice onset time, frication, and aspiration in word-initial consonant clusters”. J. Speech and Hearing Research, 18, 686-705.

Larson, J. (2003). "A Role for the Larynx in Contextual VOT Variation in American English?". Proc. XVth Int. Cong. Phonetic Sci., Barcelona, 3217-3220.

Lisker L., Abramson, A.S, Cooper, F.S & Schvey, M. (1969). "Transillumination of the larynx in running speech". JASA, 45, 1544-1546.

Lisker, L. & Abramson, A. (1964). "A cross-language study of voicing in initial stops: acoustic measurements". Word, 20, 384-422.

Löfqvist, A. & McGarr, N. (1987). "Laryngeal dynamics in voiceless consonant production". In Baer, T., Sasaki, C. & Harris, K.S (Eds.), Laryngeal function in phonation and respiration, pp. 391-402. Boston: College Hill.

Löfqvist, A. & Yoshioka, H. (1980a). "Laryngeal activity in Icelandic obstruent production". Haskins SR, 63, 275-292.

Löfqvist, A. & Yoshioka, H. (1980b). "Laryngeal activity in Swedish obstruent clusters". J. Acoustical Soc. America, 63, 792-801.

Löfqvist, A. & Yoshioka, H. (1984). "Intrasegmental timing: Laryngeal-oral coordination in voiceless consonant production". Speech Communication, 3, 279-289.

Löfqvist, A. (1980). "Interarticulator programming in stop production". J. Phonetics, 8, 475-490.

Löfqvist, A. (1990). "Speech as audible gestures". In: W.J. Hardcastle & A. Marchal (eds.), Speech Production and Speech Modelling, pp. 289-322. Dordrecht: Kluwer Academic Publishers.

Malécot, A. and Peebles, K. (1965). "An optical device for recording glottal abduction-adduction during normal speech". Zschr. f. Ph., 18, 545-550.

McGarr, N. & Löfqvist, A. (1988). "Laryngeal kinematics of voiceless obstruents produced by hearing-impaired speakers". J. Speech and Hearing Research, 31, 234-239.

Munhall, K. & Löfqvist, A. (1992). "Gestural aggregation in speech: laryngeal gestures". J. Phonetics, 20, 111-126.

Munhall, K. G., and Ostry, D. J. (1985). "Ultrasonic Measurements of laryngeal kinematics". In Titze. I., and Scherer, R (Eds.), Vocal fold physiology: Biomechanics, acoustics and phonatory control, pp. 145-162. Denver: Denver Center for the performing arts.

Munhall, K.G., Ostry, D.J. and Parush, A. (1985). "Characteristics of velocity profiles of speech movements". J. Exp Psychology, Human Perception and Performance, 11, 457-474.

Murano, E., Imagawa, H., Niimi, S., Kumada, M., Kobayashi, T. & Kiritani, S (2000). "Vocal fold vibration patterns in spasmodic dysphonia: observation from a high-speed digital imaging system". Proc. 5th Seminar on Speech Production, 317-319.

Ní Chasaide, A. & Gobl, C. (1993). "Contextual variation of the vowel voice source as a function of adjacent consonants". Language and Speech, 36 (2-3), 303-330.

Ohala, J. (1966). "A new photoelectric glottograph". UCLA Working Papers in Phonetics, 4, 40-54.

Pétursson, M. (1977). "Timing of glottal events in the production of aspiration after [s]". J. Phonetics, 5, 205-212.

Ridouane, R., Fuchs, S. & Hoole, P. (2006). "Laryngeal adjustments in the production of voiceless obstruent clusters in Berber". in J. Harrington & M. Tabain (eds.), "Speech Production: Models, Phonetic Processes, and Techniques", Psychology Press: New York, pp.275-297.

Saltzman, E. & Munhall, K. (1989). "A dynamic approach to gestural patterning in speech production". Ecological Psychology, 1(4), 333-382.

Sawashima, M. & Hirose, H. (1968). "A new laryngoscopic technique by means of fiberoptics". JASA, 43, 168-169.

Sawashima, M. & Miyazaki, S. (1974). "Stereo-fiberscopic measurement of the larynx: a preliminary experiment by use of ordinary laryngeal fiberscopes". ABRILP, 8, 7-12.

Sawashima, M. and Hirose, H. (1983). "Laryngeal gestures in speech production". In MacNeilage, P.F (Ed.), The Production of Speech, pp. 11-38. N.Y: Springer.

Sawashima, M., Abramson, A.S., Cooper, F.S. & Lisker, L. (1970). "Observing laryngeal adjustments during running speech by use of a fiberoptics system". Phonetica, 22, 193-201.

Sawashima, and Ushijima, T. (1971). "Use of the fiberscope in speech research". ABRILP, 5, 25-34.

Shipp, T., Fishman, B.V., Morrissey, P. & McGlone, R.E. (1970). "Method and control of laryngeal EMG electrode placement in man". J. Acoustical Soc. America, 48, 429-430.

Shipp, T. (1982). "Aspects of voice production and motor control". In S. Grillner, B. Lindblom, J. Lubker & A. Persson (eds.) Speech motor control, pp. 105-112. Pergamon.

Sonesson, B. (1960). "On the anatomy and vibratory pattern of the vocal folds". Acta Otolarynologica, Supplement, 156, 1-80.

Suomi, K. (1980). "Voicing in English and Finnish stops". Publications of the Department of Finnish and General Linguistics of the University of Turku, 10.

Tsuchida, A., Cohn, A., Kumada, M. (2000). "Sonorant devoicing and the phonetic realization of [spread glottis] in English". Working Papers of the Cornell Phonetics Laboratory 13, 167-181.

Weismer, G. (1980). "Control of the voicing distinction for intervocalic stops and fricatives: some data and theoretical considerations". J. Phonetics, 8, 427-438.

Yoshioka, H. Löfqvist, A. & Collier, R. (1982). "Laryngeal adjustments in Dutch voiceless obstruent production". ABRILP 16, 27-35.

Yoshioka, H., Löfqvist, A. & Hirose, H. (1980). "Laryngeal adjustments in Japanese voiceless sound production". Haskins SR, 63, 293-308.

Yoshioka, H., Löfqvist, A. & Hirose, H. (1981). "Laryngeal adjustments in the production of consonant clusters and geminates in American English". J. Acoustical Soc. America, 70, 1615-1623.

Yoshioka, H. (1984). "Glottal area variation and supraglottal pressure change in voicing control". ABRILP, 18, 45-49.