INTERARTICULATOR TIMING IN INITIAL CONSONANT CLUSTERS

Phil Hoole [1], Susanne Fuchs [2], & Klaus Dahlmeier [2]

[1] Phonetics Institute, Munich University
[2] Zentrum für Allgemeine Sprachwissenschaft, Berlin

ABSTRACT: The coordination of laryngeal and oral articulations for initial clusters in German was examined by means of transillumination and electropalatography. Both purely voiceless clusters as well as combinations with a sonorant were analyzed. The latter are of interest because it has been reported that introduction of a sonorant can actually increase the total duration of voicelessness. In fact, no consistent evidence for a lengthening of the devoicing gesture itself was found. However, the sonorant did induce a rearrangement of oral articulations relative to the glottal gesture. Over the complete corpus it appeared that the onset of glottal abduction is much more tightly linked to oral articulations than is peak glottal opening (despite the popularity of the latter for formulating principles of coordination). Results also indicated that it is unlikely that clusters like /st/ consist underlyingly of a blending of two gestures. It is argued that speakers learn to plan the glottal gesture to fulfil the aerodynamic requirements of the cluster as a whole.

INTRODUCTION

There is still much remaining to be learnt about the internal articulatory organization of consonant clusters. Even when restricted to a single syllable position, languages such as English and (to an even greater extent) German exhibit considerable variety and complexity. In addition to coordinated activity of multiple supraglottal systems, many clusters also require coordination with the laryngeal devoicing gesture. Work by Löfqvist and colleagues (see e.g Löfqvist, 1990, for overview; Hoole, 1999a, for further discussion) has given substantial information about the basic time-course of glottal movement but little work has directly related laryngeal kinematics to measurements of supraglottal movements. Here, we restrict our attention to word-initial clusters. If we consider first clusters consisting purely of voiceless sounds, it is generally agreed that only single-peaked glottal devoicing gestures can be observed. Browman and Goldstein (1986) believe that it can be taken as a phonological regularity of syllable-initial position in English that only a single devoicing gesture can occur (this explaining the absence of aspiration in clusters like /**st**/). However, Munhall and Löfqvist (1992), on the basis of an experimental exploration of gestural aggregation, believe that there remains merit in what they term the linguisitcally conservative view, namely that such clusters consist underlyingly of two gestures. Due to the dominance of the gesture for the fricative over that for the plosive only a single-peaked movement is actually observable (for further discussion of the concept of dominance in this context see Hoole, 1999a). There is thus an interesting divergence of views about the gestural substrate despite broad agreement as to the observable behaviour. Within Browman and Goldstein's (1986) model the timing of glottal behaviour is formulated in two rules, which we aimed to test over a wide range of consonant sequences:

a) if fricative gesture is present, coordinate the peak glottal opening with the mid-point of the fricative, b) otherwise, coordinate the peak glottal opening with the release of the stop gesture (p. 228)

In order to throw further light on these issues we believed that it could be very revealing to also examine clusters ending in a sound that might seem irrelevant in laryngeal terms (see Hoole, 1999a, for more background). For the present work we have used the sonorant /**l**/. Acoustic measurements (e.g Docherty, 1992) indicate that total duration of voicelessness may even increase after addition of a sonorant. At the same time, it can be expected that the oral occlusion duration of the initial voiceless plosive or fricative elements will shorten. This leads to three possible patterns of coordination in clusters with sonorant, with radically different implications for organization of clusters in general:

(1) The devoicing gesture itself increases in duration when a sonorant is added. This would be sensational, as probably no speech production model would predict this.

(2) The devoicing gesture keeps the same duration but the timing of the oral articulators with respect to it shifts, suggesting that the gesture is in effect a property of the complete cluster

(3) Oral and laryngeal components of the voiceless segments reduce in parallel keeping the same relative timing – in other words the sonorant really is irrelevant to laryngeal-oral coordination.

For the less radical possibilities (2) and (3), increased duration of voicelessness after addition of a sonorant might still be explainable as a passive effect of aerodynamic conditions in the vocal tract.

PROCEDURE, SPEECH MATERIAL, SUBJECTS, PREPROCESSING

Laryngeal behaviour was captured by means of transillumination (photoelectroglottography) combined with synchronized fiberoptic filming (detailed discussion of the procedures followed here, as well as of factors affecting the reliability of transillumination in general is given in Hoole, 1999b).

In addition, lingual articulation was monitored by means of electropalatography (Reading EPG3). A PC generated synchronization signals that were used to align the acoustic, transillumination and palatographic data.

The speech material consisted of (mostly) real German words with the following word-initial consonants:

Single consonants:	**p, t, f, ʃ**	Fricative plus plosive:	**ʃt, ʃp**
Plosive plus fricative:	**pf, ps, ts, tʃ**	Clusters with / **l** /:	**pl, fl, ʃl, ʃpl, pfl**

The words were mostly two syllables with schwa in the second syllable (and stress on the first syllable) and were embedded in the carrier phrase: “Lese <Word1> wie <Word2> bitte” (‘read <word1> like <word 2> please’). Thus every sentence included two target words. 5 repetitions were recorded of every target word. Most of the consonants and consonant clusters listed above occurred in more than one target word but for the initial analysis reported here one word form for each consonant was selected.

Three subjects with extensive prior experience with electropalatography were recorded, two from eastern Germany, one from southern Germany.

The main pre-processing steps involved determing the onset and offset of the glottal devoicing gesture using a velocity threshold criterion, together with the time-point of peak glottal opening.

EPG has to date mainly been used to supplement the acoustic signal in determining onsets and offsets of oral occlusions.

RESULTS

We will present the results available to date in terms of a series of schematic overviews of the timing of glottal abduction and adduction relative to the timing of the various oral occlusions. No statistics will be reported since, as indicated above, a considerable amount of additional material remains to be analyzed.

1. Clusters of voiceless sounds

1.1 Timing of onset of glottal abduction relative to the onset of the first oral occlusion.

The timing patterns are shown in Fig. 1 for fricative-plosive and plosive-fricative combinations. One important basic point emerges with complete consistency over subjects and clusters, namely that glottal abduction starts earlier relative to the formation of the first oral occlusion for clusters with fricatives in first position compared to clusters with plosives in first position. This characteristic difference between fricatives and plosives is very well-known from earlier investigations (see Hoole, 1999a, for references and discussion; see also fig. 2 below for corresponding data with single fricatives and plosives). The reason for presenting it here at the outset is that it provides a convenient backdrop against which to assess less clear-cut interarticulatory timing patterns and thus to home in on the control strategies being used by the speakers.

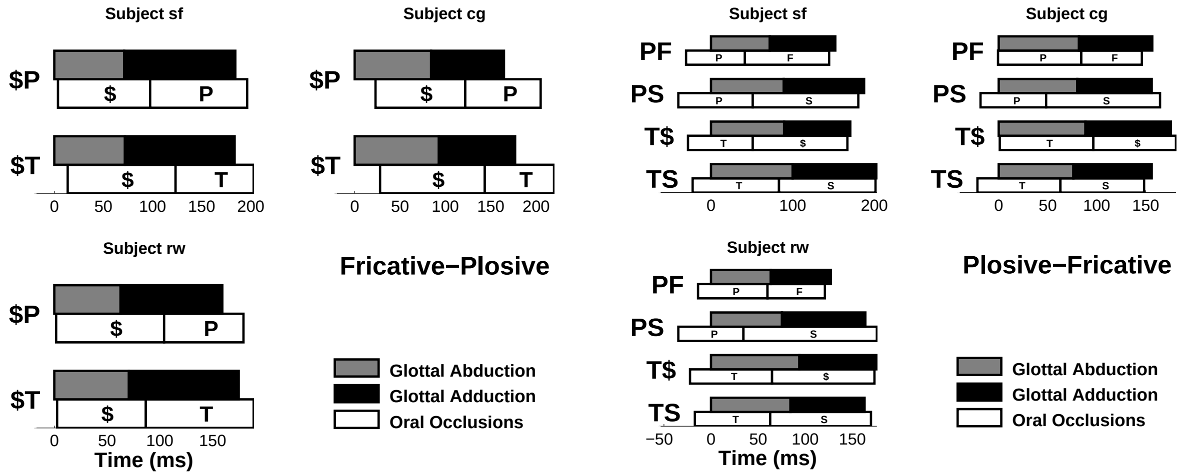

Figure 1. Durations of glottal abduction, glottal adduction and of the oral occlusions for fricative-plosive sequences (left) and plosive-fricative sequences (right). In each pair of horizontal bars the upper bar shows glottal durations, the lower bar the oral occlusion durations. For all utterances the start of glottal abduction has been aligned at 0 on the time axis. ($ stands for the postalveolar fricative /ʃ/

A further quite consistent finding probably related to the first one is that the ratio of abduction duration to adduction duration is smaller for fricative-initial clusters. Both these findings are usually interpreted with respect to the aerodynamic demands of fricatives, i.e early, vigorous abduction ensures cessation of voicing and adequate pressure build-up to drive the frication source.

1.2 Timing of peak glottal opening

The main aim here is to determine how well Browman and Goldstein’s rule (a) above holds over quite a wide range of voiceless clusters containing fricatives. In the figures, peak glottal opening (PGO) corresponds to the point where the bars for glottal abduction and adduction meet. Refering again to Fig. 1, it appears that the location of PGO is quite variable over subjects and certainly does not appear to be tightly linked to the midpoint of the fricative. About the only generalization that can be made is that PGO is more centrally located within the fricative for fricative-plosive rather than plosive-fricative sequences. Tsuchida et al. (2000) have also pointed out (based on one AmEng speaker) that PGO is not necessarily closely aligned with midpoint of the fricative. Thus Browman and Goldstein’s original rule does not appear to be completely accurate. In fact, in more recent remarks Goldstein (1990, pp.446-447) clearly entertains the possibility that the rule is not to be viewed in a hard and fast sense, and that for example PGO may be later in the fricative in fricative-plosive combinations than in single fricatives. The rule is probably a useful generalization that would work well in an articulatory synthesis framework (i.e it would probably never result in aberrant output being generated) but it may not reflect the organisational principle on which speakers base their motor behaviour.

2. Combinations with /**l**/

2.1 Plosive plus /**l**/

Timing patterns for single /**p**/ vs. /**pl**/-clusters are compared in the left part of Fig. 2. Somewhat unexpectedly there was no tendency for the p-occlusion to be shorter when occurring in a cluster than when occurring on its own. There is also no change in the timing of PGO relative to the release of the p-occlusion for /**pl**/ vs. /**p**/. Regarding the overall duration of the glottal gesture it certainly does not become shorter when the plosive combines with a sonorant. For subject RW it appears to become longer (but the p-occlusion also lengthens).

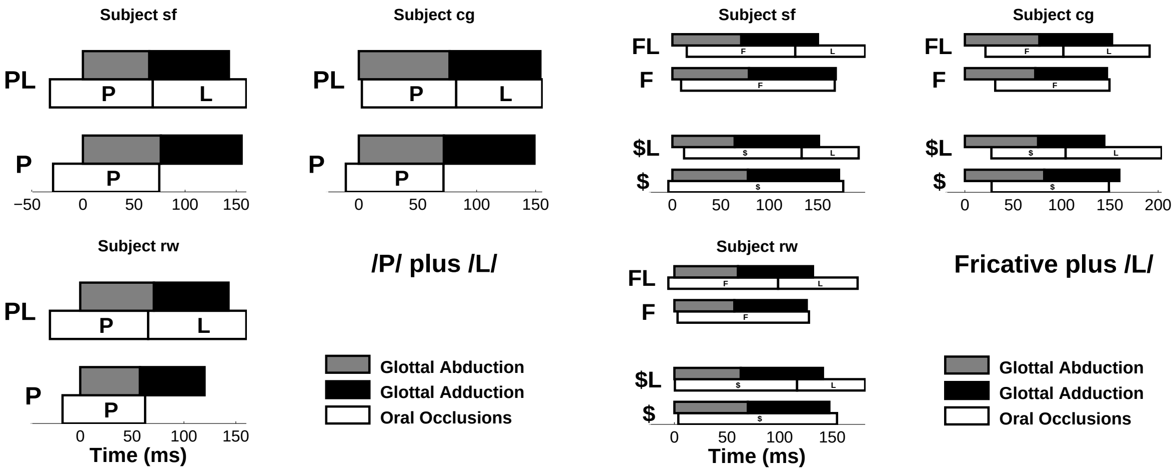

Figure 2. Laryngeal-oral timing patterns for combinations of plosive (left) and fricatives (right) with sonorant /**l**/. Other details as for Fig. 1

One might thus take these results as evidence that the /**l**/ is simply irrelevant to the organization of laryngeal-oral coordination. Other studies have shown different results, however. Tsuchida et al. (2000) found the expected slight shortening of the oral occlusion in stop+/**l**/ sequences and indeed found a lengthening of the total glottal gesture duration, mainly due to an increase in the adduction phase. (Just to complicate matters further, this effect did not generalize in their study to fricative+lateral clusters, which tended to show shorter glottal gesture duration than single fricatives. It is also not known whether place of articulation is a relevant effect in their data, which included /**k**/ as well as /**p**/.) Jessen (1999), in his very extensive material for one German speaker, found no change in occlusion duration for /**p**/ vs. /**pl**/ but did find a significant shortening from /**k**/ to /**kl**/. There were no differences in the glottal gesture duration for /**k**/ and /**kl**/; however /**pl**/ proved non-significantly shorter than /**p**/ in one of two sessions, but significantly longer in the other session.

2.2 Fricatives plus /l/

Unlike for the plosive /**p**/ there is quite a consistent reduction in the occlusion duration for the fricatives (/**f**/ and /**ʃ**/) when followed by the lateral (see right part of Fig. 2). On the other hand, there is no evidence for a consistent change in the duration of the glottal gesture (both slight increases and slight decreases are found). Taken together, this means that the timing of PGO tends to shift to a relatively later point in the fricative when it is followed by a lateral than when it occurs on its own.

2.3 Two occlusives plus /**l**/

Our corpus provided one example of a plosive-fricative sequence to which /**l**/ can be added (/**pf**/ > /**pfl**/ (note that this kind of sequence cannot occur in English), and one example of a fricative-plosive sequence (/**ʃp**/ > /**ʃpl**/). The timing patterns for all these sequences are assembled in Fig. 3 (/**pf**/ and /**ʃp**/ are repeated from Fig. 1). Once again there is no evidence that addition of an /**l**/ leads to a reduction in the duration of the glottal gesture; if anything, there is a slight tendency for longer gestures. Overall occlusion duration for the underlyingly voiceless elements of the clusters (/**pf**/ and /**ʃp**/) is somewhat shorter when combined with /**l**/ (exception: /**pfl**/ for RW). Taken together these two rather weak effects do lead to one quite consistent pattern, which is that glottal adduction is completed later relative to /**pf**/ and /**ʃp**/ when /**l**/ is added to the cluster. As already indicated above, these results are somewhat different from those found by Tsuchida et al. (2000), who found a slight reduction in the duration of the glottal gesture (especially the adduction phase) when /**l**/ is added to fricative or fricative+plosive.

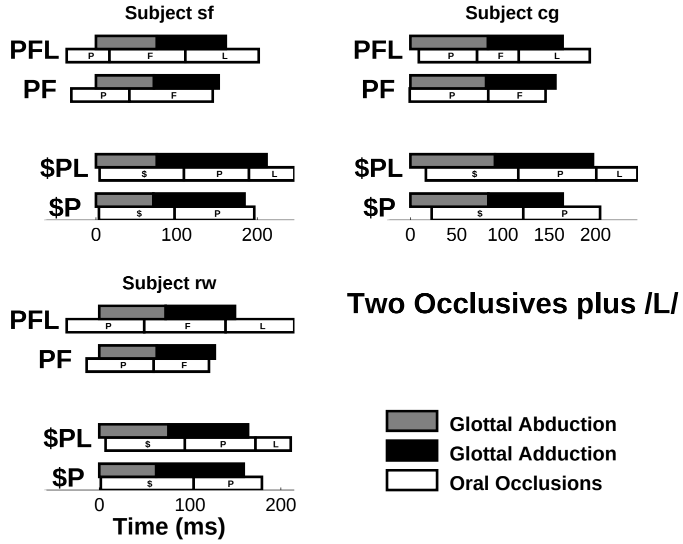

Figure 3. Laryngeal-oral timing patterns for combinations of two occlusives with /**l**/. Other details as for Fig. 1

DISCUSSION

What conclusions can now be drawn about laryngeal-oral coordination in purely voiceless clusters and in clusters also containing a sonorant?

Regarding combinations with /**l**/, the most radical hypothesis, i.e lengthening of the glottal gesture, received little support. (Interestingly, data in Jessen, 1998, 1999, indicate that combinations with /**r**/ may be a different story - unfortunately, these sounds are tricky to capture with transillumination.) On the other hand, it was also not the case that the timing of the glottal gesture remained constant relative to the voiceless elements, i.e it did not seem to be case that the presence of the sonorant was simply irrelevant (in our data the case of /**p**/ vs. /**pl**/ seemed to be an exception).

Thus, hypothesis (2) - changing patterns of oral articulation are overlaid on a fairly constant glottal articulation - seems to be the most plausible account. The fact that timing shifts can take place under the influence of an /**l**/ weakens for us the attractiveness of the linguistically conservative hypothesis that clusters of the /**st**/ kind may underlyingly contain two glottal gestures: in these clusters the PGO may be shifted away from the position found in a single fricative, but in the light of the present results it is difficult to attribute this unequivocally to competing, separate gestural input for /**t**/. The most parsimonious account seems to be that initial clusters are organized around a single glottal gesture that cannot be assigned to a specific sound in the conventional sense. The idea that oral articulations shift relative to the underlying glottal articulation fits in quite well with various findings in the literature. For example, it has been observed quite often that the interval from PGO to oral release in aspirated plosives may be related to place of articulation, and, in turn, to differences in oral closure duration (Hoole, 1999a; Jessen, 1999).

Since all clusters in the corpus contained fricatives one specific aim of the present study was to examine Browman and Goldstein's timing rule (a). About the only generalization that could be made is that if a fricative is present PGO almost always occurs within the fricative: clearly a very weak statement. It was not obvious from our material how one might formulate a better prediction of the timing of PGO (and also of the time of completion of glottal adduction) that would hold over all the subjects and sound-sequences examined. This contrasts markedly with the timing of the start of glottal abduction which shows a completely clear-cut distinction depending on whether the first consonant is fricative or plosive, and is stable over subjects and independent of the number and nature of following consonants in the sequence. This suggests that although PGO is a very salient feature of the laryngeal gesture, and thus attractive for descriptive purposes and for the formulation of rules, it may not be the feature that subjects are at most pains to control in terms of timing relative to salient events in the oral articulations (recall also the just-mentioned dependency of PGO to Release on place of articulation)

Even if precision in control centres around the time of glottal abduction, we hasten to concede that this does not mean that, once initiated, the movement is completely ballistic and stereotypic. In common with other investigations it can be observed in our data that the glottal adduction phase is longer for /**ʃt**/ compared to /**ʃ**/ or /**ʃl**/ (compare Figs. 1 and 2). This long adduction phase could be taken as evidence for the presence of a weak glottal gesture for the plosive riding on the skirts of the dominant initial fricative. However, the account that we prefer is that speakers are simply able to learn the pattern of glottal movement that fulfils the aerodynamic requirements of the cluster as a whole. The crucial phase is at the start of the first fricative; after that, the constraints on the movement pattern are very weak. All that is required is that adduction is completed around the time of the plosive release. Adduction could probably be completed more or less any time during the plosive closure with negligible difference in acoustic output. In practice, speakers simply learn to make use of the time available.

For aspirated plosives, and plosives plus fricative, the main constraint is that the adduction phase should not start too soon; accordingly abduction can be relatively late and slow.

Within a couple of constraints of this kind, oral organization can shift around somewhat freely relative to the underlying glottal gesture without compromising the sound output.

In conclusion, this work on laryngeal-oral coordination is for us just one part of arriving at a better general understanding of the articulatory organization of complex clusters. One shortcoming of the present work is that the timing of oral occlusions may not be perfectly reflected in what can be gleaned from the acoustic signal and EPG. Accordingly, EMA experiments with an expanded corpus are planned (quite a few interesting types of cluster are not well-suited for study with transillumination).

REFERENCES

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

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

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.

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.

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

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

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.

Munhall, K. & Löfqvist, A. (1992) “Gestural aggregation in speech: laryngeal gestures” *Journal of Phonetics* **20**, 111-126.

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