Forschungsberichte des
Instituts für Phonetik und
Sprachliche Kommunikation
der Universität München
(FIPKM) 35 (1997) 89-99

LARYNGEAL COARTICULATION
Section 1: Coarticulatory investigations of the devoicing gesture[1]

Philip Hoole

Institut für Phonetik und Sprachliche Kommunikation
Ludwig-Maximilians-Universität München
Schellingstr. 3
D-80799 Munich, Germany
hoole@phonetik.uni-muenchen.de

ABSTRACT

This paper reviews the current state of our knowledge about the organization of the devoicing gesture in speech. First of all, the influence of place and manner of articulation on the devoicing gesture in single voiceless consonants is discussed. This provides the background for consideration of coarticulatory effects in two main classes of consonantal sequences. The first class involves sequences of voiceless stop (or fricative) plus sonorant (e.g /pl/). It is well-known that the sonorant can undergo devoicing, induced by the coarticulatory effect of the adjacent voiceless consonant. It is much less clear if and how laryngeal-oral interarticulatory coordination in modified with respect to the pattern found for single voiceless consonants. This could be of great theoretical interest since there are reports that *total* duration of voicelessness in e.g /pl/ is longer than in /p/. It would be intriguing if the devoicing gesture, accordingly, were longer in the former case, as it is not clear what current theory of coarticulation could handle this. The second class involves sequences of purely voiceless sounds. Here the view is of coarticulation as coproduction, i.e what sounds in consonant sequences are associated with a separate laryngeal gesture, and how multiple gestures blend. While a considerable amount is known about the laryngeal movements *per se*, it is argued (as for the first class of sequences) that the laryngeal findings need to be linked more closely to improved knowledge of the organization of the relevant oral gestures.

INTRODUCTION

At one level, voicing coarticulation has frequently been studied as a temporal phenomenon, i.e. as the extension of ±periodicity from one segment to another adjacent one (differently specified phonologically). To take a simple example, consider the devoicing of /l/ in English words such as "plead". There is universal agreement that this can be explained by the influence of the adjacent voiceless /p/. In other words, minimal pairs such as "plead" and bleed" have not been used to suggest that English has a phonemic contrast between voiced and voiceless laterals. Such combinations of voiceless obstruent + sonorant sequences are probably the consonant clusters that have been most extensively investigated (see below and especially Docherty, 1992, for further references), at least in terms of acoustic measurements of the timing of voicing - and they represent one of the two main topics covered in the section on coarticulatory investigations of the devoicing gesture. Of course, there are many other potential sequences of consonants in which the individual members of the sequence differ with respect to voicing specification, and can accordingly influence one another (Westbury, 1979; Docherty, 1992).

Yet, consideration of even apparently simple coarticulatory

[1]This work is to appear as a contribution to "*Coarticulation: Theoretical and Empirical Perspectives*", W.H Hardcastle & N. Hewlett (eds.), CUP. We here only reproduce Section 1, "*Coarticulatory investigations of the devoicing gesture*". Section 2, entitled "*Voice source variation in the vowel as a function of consonantal context*", was contributed by Ailbhe Nì Chasaide and Christer Gobl, Dublin. However, the Introduction pertains to both sections of the chapter. See also the companion chapter on "*Techniques for Investigating Laryngeal Articulation and the Voice-source*" also reproduced in this volume of FIPKM.

phenomena such as the /pl/ case leads inevitably to a widening of the perspective to include questions of interarticulator coordination, in other words the formation and the release of oral constriction for /p/ and /l/, and their temporal relationship to the devoicing gesture. In this respect, coarticulation with respect to the laryngeal system requires an approach somewhat different from that traditionally followed, for example, for labial and velar coarticulation. In the latter cases it is legitimate to examine issues such as relative extent of carryover and anticipatory effects, explanatory adequacy of feature-spreading vs. time-locking models etc. (for example for liprounding), without explicit consideration of details of *inter*articulator coordination (implicitly, all the above concepts of course require a further articulatory system as reference for the analysis of the coarticulatory behaviour of the targeted articulatory sub-system). The situation can be summarised by observing that while simple alternating CV sequences have provided a point of departure for coarticulatory studies in almost every other articulatory system, for an understanding of oral/laryngeal coordination one would need to look at more complex environments.

The phonatory system differs from other subsystems used in speech production (such as the lips, velum and tongue) in that the acoustic output is less directly/transparently determined by the articulatory (laryngeal) gestures. Simply knowing the precise configuration of the glottis at one instant in time does not in itself permit us to specify whether the acoustic signal at that point is voiced or voiceless. The initiation, maintenance or cessation of phonation depend on the interplay of a number of factors: muscularly controlled adjustments of the vocal folds which determine whether and to what degree they are abducted or adducted as well as the precise configuration of the glottis when adducted; aerodynamic conditions at the glottis, and in particular, the transglottal pressure, which is sensitive not only to respiratory factors, but also to the degree and duration of any supraglottal occlusion; the intrinsic elasticity of the vocal folds and the muscularly controlled tension in the folds. During phonation, variation in either of these factors will also affect their mode of vibration, and hence the auditory quality of the voice produced.

Some recent investigations have suggested that the mode of phonation of a vowel may indeed be affected in rather subtle ways by adjacent consonants (specifically voiceless consonants). To the extent that these effects appear, the simple measure of the timing of voice onset or offset (±periodicity in the acoustic signal) underestimates the coarticulatory influence of consonants on vowels (and presumably on adjacent voiced segments generally). They further provide insights into the control mechanisms that may be involved in the regulation of voiceless consonants.

The two sections of this chapter deal with very different aspects of laryngeal coarticulation. The first subsection deals with the spatial and temporal organisation of the laryngeal devoicing gesture, dealing in particular with interarticulator coordination in single voiceless consonants as well as in clusters (the relevant instrumental techniques are presented in the section on investigation of the devoicing gesture in the chapter on techniques for investigating laryngeal articulation and the voice-source. The second deals with the variations which may be found in the vowel's mode of phonation in the vicinity of certain consonants. In this subsection, a number of illustrations are presented, based on a rather fine-grained analysis of the voice source (this methodology is described in the section on techniques for analyzing the voice source in the chapter just mentioned). The implications of these data for our understanding of laryngeal control mechanisms are discussed.

SECTION 1

Coarticulatory investigations of the devoicing gesture

1. Introduction

As just discussed in the general introduction to this chapter, investigation of coarticulation with respect to the devoicing gesture almost inevitably requires consideration of the interarticulatory organisation of consonantal sequences, rather than of simple alternating sequences of single consonants and vowels. The first main topic to be discussed under this heading - as already mentioned - will be the organisation of sequences of voiceless obstruent plus sonorant. The second main topic follows on from observations made by Yoshioka, Löfqvist & Hirose (1981), who point out in a paper examining consonant clusters in American English that in terms of coarticulation at the laryngeal level it can be revealing to examine sequences of purely *voiceless* consonants, since this can provide insight into the organisational principles according to which individual gestures blend and mutually influence one another (coarticulation as coproduction). This topic will accordingly provide the second main area to be addressed below. Again, as with the first topic, we will be arguing that it is difficult to divorce the question of laryngeal coarticulation and coproduction from the question of laryngeal-oral interarticulator coordination. Having brought the question of interarticulator coordination into play we must now also make clear what aspects of this all-embracing topic we will *not* be considering here: in fact, we will be ignoring what has been probably the major topic in laryngeal-oral coordination over the last 30 years, namely the role of laryngeal timing in voicing contrasts in stop consonants, developed under the influence of the work of Lisker & Abramson (e.g. 1964) and with its important implications for the status of time and timing in phonological representations (cf. Löfqvist, 1980). There is now a very substantial literature on the physiological and kinematic aspects of the mechanisms of stop consonant voicing control employed in many different languages (e.g for laryngeal-oral timing patterns ranging from preaspiration, via unaspirated and postaspirated to voiced aspirated). Nonetheless, we will preface the discussion of the two main topics in this sub-chapter with consideration of some of the basic kinematic properties of laryngeal articulation in single voiceless plosives and fricatives to form a 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 in section 3 with respect to these longer sequences. A useful group of illustrations showing some of the interarticulatory relationships discussed below for single consonants and clusters in terms of transillumination and EMG signals is to be found in Löfqvist (1990), Figures 5 to 7.

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. (Refer to chapter XXX, Fig.1, for an illustration of the time-course of laryngeal abduction and adduction in an utterance containing several voiceless sounds.)

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 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".

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[2]. 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; 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." 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).

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 not ideally suited material) that the timing of glottal abduction could be identical for plosives and fricatives when viewed from the onset of the previous vowel. 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 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

[2]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.

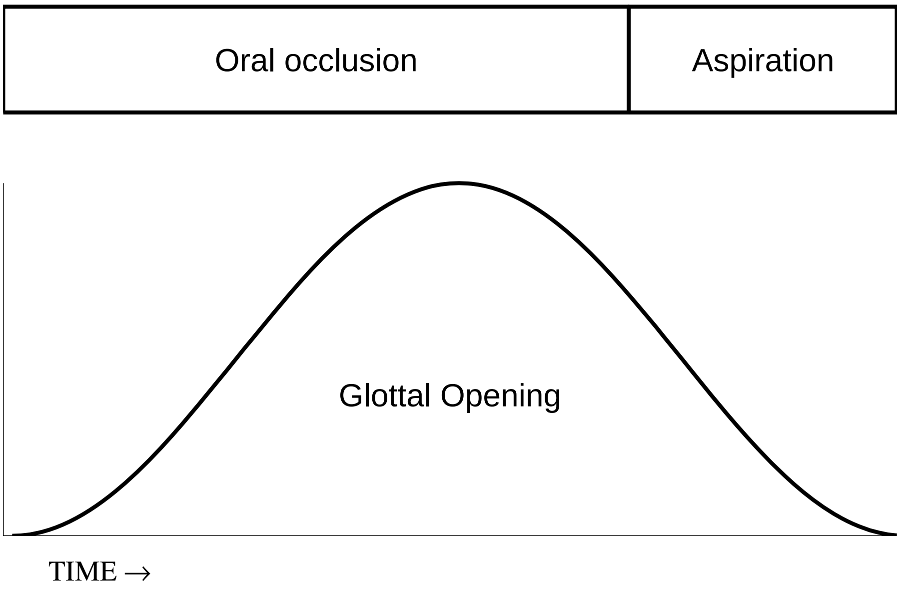

Fig. 1: *Typical pattern of laryngeal-oral coordination in aspirated stops*

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.

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 Fig. 1 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.

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 (see e.g Jessen, 1995). 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[3].

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. 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. The most direct test of this question is to be found in Cooper, where /p/, /t/ and /k/ were compared. 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

[3]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).

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)

3. Devoicing organisation in consonant sequences

3.1 Coarticulatory devoicing in stop-sonorant and fricative-sonorant sequences

As outlined in the introduction we will move here from consideration of the coarticulatory effects themselves to discussion of the implications of the available data for more general issues of interarticulator coordination. The most accessible source 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[4]: 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; Haggard, 1973; Hawkins, 1979). Docherty notes 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)[5] that the above two findings can be simply related in a manner entirely analogous to the attempt (discussed above) to explain place of articulation differences in VOT 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 an essentially voiceless /l/. In terms of the schematic illustration given in Fig. 1 one can think of the devoiced sonorant replacing the phase labelled "aspiration", this phase being proportionally longer and the preceeding 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: 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). Thus the acoustically measured period of voicelessness may not be an accurate reflection of the duration of the laryngeal gesture itself. Further articulatory data may thus still save the invariant laryngeal gesture hypothesis, although Docherty seems to be of the opinion that the magnitude of the effects makes this rather unlikely.

Even if it remains an open issue whether devoicing duration is genuinely longer in stop-sonorant clusters, it does seem to be clear that devoicing duration is not *shorter*. This is

[4]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.

[5]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 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.

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 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[6].

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), and that in /#sl/ the manifestation of /s/ is truncated by the /l/ competing for the tongue-tip articulator. In both cases the underlying lingual input for /s/ may have remained constant, together with the glottal timing with respect to this input.

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[7]. This idea 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. The devoicing gesture may then lengthen as the syllable onset becomes longer. With regard to their two rules, there is, however, the possibility that they might not be strictly correct, because of the change of the temporal relationship between peak glottal opening and oral occlusion in clusters (but note the distinction just made between surface manifestation and underlying input).

An alternative, more output-oriented style of explanation might be that it is perceptually important to have a substantial amount of devoicing on the second element in a cluster (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 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)

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/, but total devoicing duration increased, 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. 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.

In conclusion to this subsection, it can safely be said that some fairly straightforward transillumination/fiberscopic data on clusters with mixed voicing characteristics (in plentiful supply in languages such as English and German) could swiftly resolve some of the speculative discussion above and already prove illuminating for our understanding of laryngeal-oral coordination. The more demanding task will be to link the laryngeal findings to improved understanding of the organization of supraglottal gestures in clusters.

3.2 Devoicing patterns in voiceless clusters

As outlined in the introductory section, clusters of voiceless consonants provide one of the most suitable fields for examining processes of coarticulation or coproduction at

[6] It is also unclear whether current formulations of the dominance and the binding concepts can handle the different timing of peak glottal opening relative to release for different places of articulation (see section 2.2).

[7] The relevant laryngeal-oral coordination patterns are captured in two rules (op. cit. p.228):
a. If a fricative gesture is present, coordinate the peak glottal opening with the midpoint of the fricative.
b. Otherwise, coordinate the peak glottal opening with the release of the stop gesture.

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 information on this topic is a series of articles published some 10 years ago 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 briefly in turn.

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. 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[8]. 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.

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 (though note also that haplology is a productive phonological process). However, here we reach the limits of the interpretability of this group of papers 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 Fig.2.

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

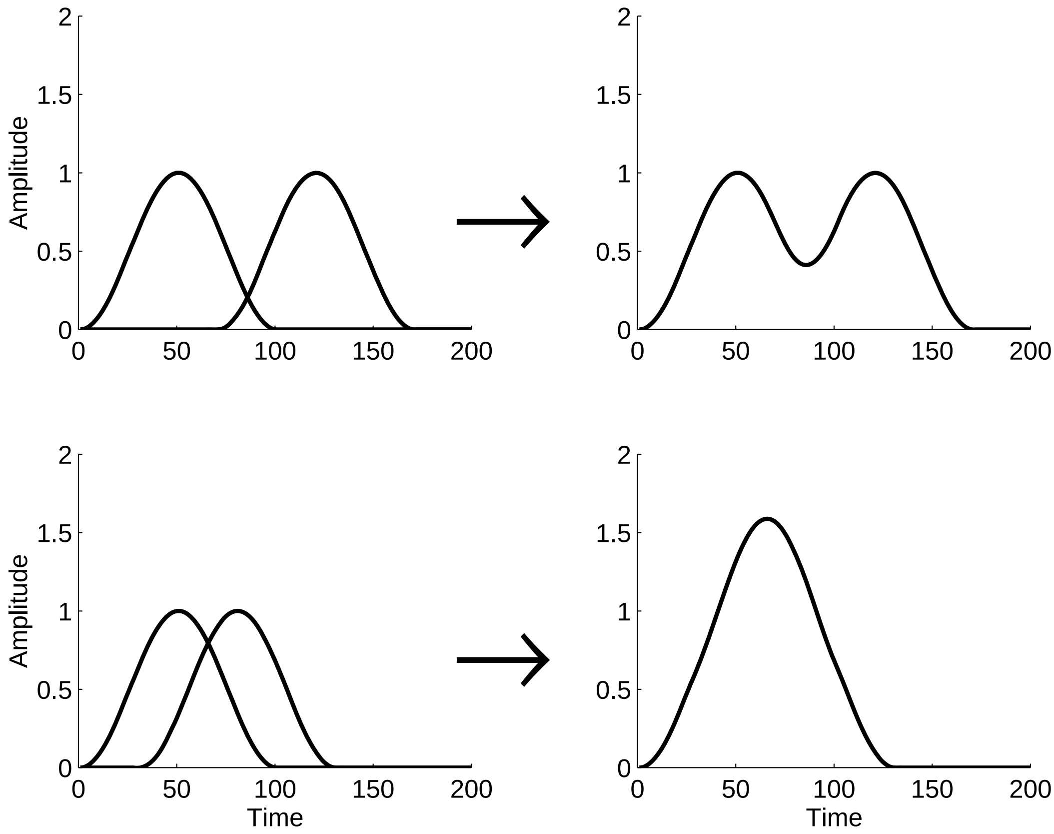

Fig. 2: Simplified illustration of relationship between hypothesized underlying gestural input (left panels) and observable behaviour (right panels) at different degrees of gestural overlap (top vs. bottom). Amplitude of the basic laryngeal gesture has been arbitrarily set to 1, and duration to 100. See Munhall & Löfqvist (1992, Fig. 5) for a wider range of more realistic simulations.

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. 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); 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).

This remains a significant experiment for coarticulatory studies as a whole (one might even say that it was long overdue, following the pioneering studies discussed at the beginning of this section): 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). In particular this determines the order of the two rules given above in Section 3.1.

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).

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[9]. 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. 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 1 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.

Concluding this topic, the aim of this review is not so much to arrive at a conclusion as to what is the more persuasive analysis of fricative-stop clusters, on which there is a substantial further literature of both phonological and phonetic orientation (see references in Browman & Goldstein, 1986; Ewen, 1982; Pétursson, 1977), but rather to highlight the still existing gaps in our knowledge about the relevant articulatory interrelationships.

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 principle since it is quite clear that the devoicing

[9]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 can probably be assumed that two peaks in the glottal abduction will be observed.

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.

At the conclusion 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.

4. Conclusion

This review of laryngeal coarticulation has shown that we have quite a good understanding of the organisation of the laryngeal devoicing gesture both in simple and more complex sequences of sounds. However, some gaps remain. Some should be easy to fill, for example with respect to the amplitude and duration of laryngeal activity in voiceless-voiced consonant sequences. Others, particularly those relating to the details of laryngeal-oral coordination will be more difficult; in fact what is required is not so much better knowledge of laryngeal behaviour per se, but rather improved insight into the organisation of labial and lingual gestures in consonant sequences. In view of the considerably greater complexity of supraglottal articulations compared to laryngeal articulations this will be no mean task.

References

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.

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.

Ewen, C.J. (1982) The internal structure of complex segments. In: H. van der Hulst and N. Smith, *The structure of phonological representations, Part II*, pp.27-67. Dordrecht, 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.

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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.

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

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

Westbury, J. (1979). *Aspects of the temporal control of voicing in consonant clusters in English*. Unpublished dissertation, University of Texas at Austin.

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