Journal of Phonetics (1996) **24,** 113–137

Modelling [s] to [ʃ] accommodation in English

Francis Nolan, Tara Holst and Barbara Kühnert

Department of Linguistics, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DA, UK

Received 3rd January 1995, and in revised form 10th August 1995

When adjacent segments compete for the same articulator a window is provided on how conflicting phonological specifications are reconciled in articulation. One possibility is that the phonological specifications of the two segments simply overlap in time, yielding as output a blend of the two segments. This, broadly, is the prediction of Articulatory Phonology. Another possibility is that one phonological specification is changed in a pre-articulatory cognitive representation so as no longer to conflict. This is assimilation in the sense in which it has traditionally been used in phonology. A previous acoustic study of /s#ʃ/ sequences by the first two authors suggested that while some outputs are compatible with gestural blending, there are others where the target for the first segment has been changed from [s] to [ʃ]. Doubts have been raised, however, concerning the validity of inferring articulation from acoustic data, given in particular the possibility of 'quantal' effects in coronal fricatives. The present paper uses electropalatography (EPG) (a) to check the relation between articulation and acoustics in the [s–ʃ] region, and (b) to explore more directly the production of /s#ʃ/ sequences. The findings offer support, but not unequivocally, for the claim that a cognitive phonological process of assimilation may sometimes apply.

1. Background

Just when the field of speech production modelling was getting bogged down in proliferating data on coarticulation, and failing to find general principles governing the translation of discrete phonological units into the articulatory continuum, a radically new perspective on the problem emerged. This perspective said, in effect, that if only phonological units were represented in an appropriate way, the whole problem of "translation" between phonology and articulation might disappear (see e.g., Fowler, Rubin, Remez, & Turvey, 1980; Fowler, 1980). This new perspective has been most comprehensively explored in the development of Articulatory Phonology, associated in particular with the work of Browman and Goldstein (1989; 1990; 1992). Articulatory Phonology uses "gestures" as its primitives. A gesture is "an abstract characterisation of coordinated task-directed movements of articulators.... [and is]...precisely defined in terms of the parameters of a set of equations for a "task-dynamic" model..." (Browman & Goldstein 1989:206).

0095-4470/96/010113 + 25 $12.00/0

Task-dynamics is a general model of skilled movement control based on dynamic equations such as might be used to describe the action of a mass attached to a spring; for an introduction, see Hawkins (1992), and for a more technical account, Kelso, Saltzman, & Tuller (1986).

By using gestures as the primitives of phonological representations, Articulatory Phonology has been able to explain many segmental phonetic phenomena such as (apparent) deletion, coarticulation, assimilation, and weakening, as being simply the natural dynamic response of the articulators to successive phonological specifications. Indeed, Articulatory Phonology has made the strong claim that all "casual speech processes" (as phenomena such as assimilation and deletion are now often known) can be modelled in terms of the overlapping of gestures and changes in their magnitude (Browman & Goldstein 1989: 220): "All changes are hypothesized to result from two simple mechanisms, which are intrinsically related to the talker's goals of speed and fluency—reduce the size of individual gestures and increase their overlap." Phonological rules have no part to play.

An obvious test-bed for such a claim is provided by the variation which occurs in the realisation of one phonological segment dependent on the nature of an adjacent segment. In the present paper, the term "segmental accommodation" will be used. This is intended to be as neutral a term as possible to refer to phenomena which may turn out to be the result of purely articulatory effects—gestural overlapping or blending in the terminology of Articulatory Phonology—or to be the result of a phonological change of the type normally referred to as "assimilation", which presumably involves a cognitive restructuring of the speaker's phonetic plan. It is the question of whether cognitive phonological processes are required to model connected speech processes which is dealt with in this paper. Put crudely, is segmental accommodation all in the mouth, or is some of it in the mind?

1.1. *A previous study of accommodation*

Holst and Nolan (1995) deal with [s] to [ʃ] accommodation at word boundaries, in sequences such as "restocks shelves". They recorded twelve speakers of Southern British English reading seventeen sentence pairs containing a potential [s] to [ʃ] accommodation site. In one member of the pair a clause boundary intervened between the fricatives (+CB), and in the other the two words were part of the same syntactic unit (−CB).

The recordings were analysed by visual inspection of spectrograms, using four categories A, B, C, and D. Schematised examples of these categories can be seen at the top of Fig. 1. The first, A, required there to be two discretely different portions of fricative energy, the first with a higher and the second with a lower low-frequency cutoff to the energy, that is, an [s]-like and an [ʃ]-like spectral pattern abutting. B and C categorised spectra changing from [s]-like to [ʃ]-like spectra, B having an apparent [s]-like steady state. D required there to be a fricative portion in which no spectral change took place (the difference between the shaded and unshaded boxes for D in Fig. 1 will be explained shortly). Intuitively the categories A to D represent increasing degrees of accommodation between the two segments. Briefly, the whole range of degrees of accommodation was found, with an overwhelming tendency for the clause boundary to inhibit Type D. Where no clause boundary intervened at the accommodation site, Type D occurred frequently.

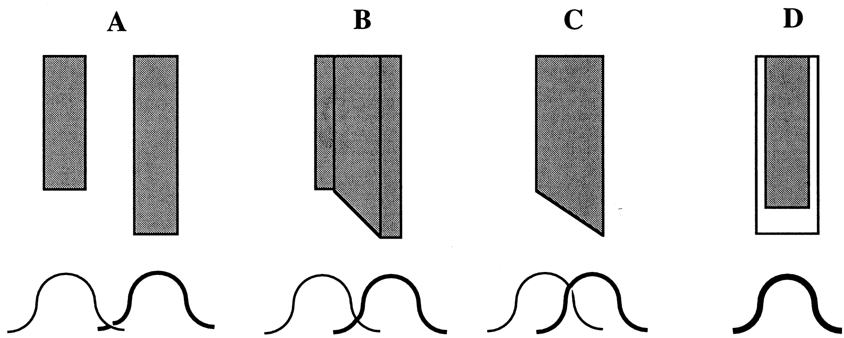

Figure 1. Schematic representation of increasing accommodation between [s] and [ʃ] as spectrographic patterns (above) and gestural overlap (below).

In discussing the findings, Holst and Nolan (1995) consider the probable Articulatory Phonology account. This would be one in which all degrees of accommodation would be purely the result of gestures overlapping. The bottom part of Fig. 1 shows, schematically, the abstract gestures underlying an [s] constriction (thin line) and an [ʃ] constriction (thick line). As the two gestures overlap more and more in time, the resultant articulation results less and less in two clearly distinct fricative steady states, and the fricative event becomes shorter in duration. In the most extreme case, that of complete overlap of the gestures for the two fricatives, the prediction of Articulatory Phonology has apparently been that a single intermediate articulation should result. For instance, in relation to an alveolar preceding a dental, Browman and Goldstein (1989: 220) say that "the location of the constriction should not be identical to that of either an alveolar or a dental, but rather should fall somewhere in between". In Fig. 1 the predicted form for Type D is represented by the shaded spectrographic schema, with short duration and spectral cutoff intermediate between those for [s] and [ʃ].

Holst and Nolan (1995) claim that this prediction of Articulatory Phonology is falsified by their data. The spectrographic pattern actually found for Type D, and represented by the unshaded rectangle in Fig. 1, differs in two crucial ways from the prediction (shaded rectangle). It is spectrally identical with a "canonical" singleton initial [ʃ], and it is longer in duration. The reasoning in more detail is as follows. Speaker-by-speaker comparison showed that the average spectrum of Type D forms for a speaker was identical to that of the speaker's singleton initial [ʃ] (i.e., in a non-accommodatory context). This suggests that in Type D there is no (spectrally detectable) contribution from a simultaneous underlying [s] gesture. The fricative in these cases is not "somewhere in between" an [s] and an [ʃ], but is unequivocally [ʃ]-like. Secondly, a speaker's mean Type D duration is slightly (16%), but significantly, greater than his or her "canonical" initial [ʃ]. This rules out the possibility that the absence of [s] influence in the spectrum is due to the deletion of the [s] gesture because then there would be no straightforward way of accounting for the "extra" duration of the [ʃ]; if the [s] gesture is not there, the duration of the [ʃ] friction should be that of a singleton initial [ʃ]. An account in terms of reduction in magnitude of the [s] gesture is also hard to sustain, because it requires that the gesture can manifest itself durationally but not spectrally (see Section 3 below).

Instead of an explanation in terms of overlapping, or reduction in magnitude, of gestures, Holst and Nolan argued that the first ([s]) gesture (to use the phonological vocabulary of Articulatory Phonology) had been replaced by an [ʃ] gesture before the phonological representation was given its articulatory interpretation—involving, of course, considerable overlapping of the (now) two [ʃ] gestures bringing the duration of the fricative event down to near, but still greater than, that of the singleton [ʃ].[1] Type D then, it was concluded, cannot be accounted for by the process of gestural blending which will account for Types B and C, but argues in favour of the co-existence of an optional cognitive phonological process of assimilation. This might be seen as advance cognitive remedial action to smooth the articulators' task. The predictions of a cognitive approach, and a purely articulatory approach, to connected speech processes will be discussed in more detail in Section 3.

1.2. *Doubts about the case for a phonological process: quantal effects in acoustic data*

Browman (1995), in her commentary on Holst and Nolan (1995), argues against the case for a phonological process of assimilation. She bases part of her argument on the difficulty of inferring articulation accurately from acoustic analyses. In particular, "Quantal Theory" (e.g., Stevens, 1972; 1989) predicts a non-monotonic relationship between articulatory change and acoustic change. That is, a continuous and smooth articulator movement can result in an acoustic output which changes slowly at some points in the movement and more abruptly at others. In the case of [s] and [ʃ] it might be the case that a small deviation from a canonical [s] constriction in the direction of that of [ʃ] produces a sudden and catastrophic shift to a spectrum appropriate for [ʃ]. Potentially, then, although the Type D found is spectrally [ʃ]-like, the spectrum may be masking an articulation which is in reality intermediate between the two fricatives, and hence in keeping with the prediction of Articulatory Phonology.

Browman supports this doubt about the reliability of inferring articulatory movement from fricative spectra by presenting examples of abrupt spectral change in contiguous [s]–[ʃ] sequences from Perkell, Boyce, and Stevens (1979). Their Fig. 3 is reproduced as Fig. 23.2 in Browman (1995). These examples indeed show a dramatic and abrupt spectral change as the speaker moves from [s] to [ʃ]. Further evidence for a quantal effect in the production of these fricatives can be found in Nguyen and Hoole (1993). They found a step-like change in the main spectral peak in [sʃ] and [ʃs] glides when this was plotted against the gradually changing frontness of the tongue-blade EMA coil.

Since Quantal Theory is an important and influential view in speech analysis, it is of general interest to explore this issue, as well as of specific interest in relation to the reliability of the data in Holst and Nolan (1995). But it is worth first teasing out

[1] The "assimilated" segment should not be expected to be twice the length of the singleton initial [ʃ], since the [s] will be shortened by virtue of being word-final and after a consonant. On the basis of measurements of consonant sequences Umeda (1977: 854) predicts, in [...Cs#s...], contributions of 41 ms and 98 ms from the final and initial fricatives, giving a total of 139 ms. It can be inferred from her Table V that the equivalent values for [...Cs#ʃ...] should be 41 ms and 153 ms, giving a total of 194 ms, 27% longer than the initial on its own. These calculations are not precisely relevant to the present case, but they do illustrate the point that the durations of adjacent segments are not simply additive.

some distinct strands in Quantal Theory. One strand which is not relevant to the present discussion is the existence of non-monotonic relationships between an acoustic parameter and the auditory or perceptual response to it (Stevens, 1989: 29–39). The concern here is with the relationship between articulation and acoustics only. In connection with this latter relationship, there seem to be two observationally distinct cases, the second of which has alternative causal mechanisms.

The first is where gradual articulator movement produces a continuous change in vocal tract configuration and the acoustic change is also continuous but non-linear. Changing the distance along the vocal tract of a vowel-like constriction would seem to be of this kind, as long as the change is not large enough to cause changes in cavity-to-formant allegiances: there are regions where a given formant frequency will change rapidly, and regions of comparative stability. This might be termed a "weak" quantal effect.

The second case is where gradual articulator movement produces an abrupt or catastrophic acoustic change. This might be termed a 'strong' quantal effect. For Perkell *et al.* (1979) the transition from [s] and [ʃ] apparently falls into this category, since their Fig. 3 show a very abrupt change of spectrum in the transition between the two fricatives. There seem to be alternative explanations for this sudden change. Stevens (1989: 25–6) notes that as a constriction moves back in the vocal tract, the formant affiliation of the front-cavity resonance most excited by the fricative energy changes. For instance, for a typical [s], this resonance might correspond to F_4 of an adjacent vowel, but as the constriction moves back, the front-cavity resonance lowers in frequency and crosses a back-cavity resonance, itself rising because of the shortening back cavity. The front-cavity resonance is now in effect F_3, the back cavity resonance having taken on the role of F_4. In the vicinity of such crossovers "there is a rather abrupt jump in the frequency that receives greatest excitation by the source" (1989:25).[2] An alternative mechanism is suggested by Perkell *et al.* (1979): the acoustic discontinuity is associated with the tongue tip breaking away from the lower teeth, on its way from lamino-alveolar [s] to apico-(post)alveolar [ʃ], and abruptly allowing a space below the tongue blade to enlarge the cavity in front of the fricative stricture. This abrupt change in vocal tract configuration resulting from a gradual articulation may well, of course, bring about the kind of formant-affiliation change just described, at the same time augmenting the magnitude of its effect, and Stevens (1989:26) regards it as merely a "special case" of change in cavity-to-formant affiliation.[3]

A partial answer to the "quantal" doubt about the data in Holst and Nolan (1995) lies in the widespread occurrence of tokens classified as B and C in that paper. If the acoustic nature of [s] and [ʃ] is strongly quantal, Types B and C should be non-existent, since a gradual transition from one articulation to the other should result in an abrupt spectral change (hence Type A), and not a glide. A strong version of the quantal theory, then, which predicts two discretely different acoustic outputs, does not seem entirely plausible from the data.

Nevertheless Browman's specific criticism, based on Quantal Theory, of the method used in Holst and Nolan (1995), and the more general concern that acoustic

[2] It may be inferred, however, from Stevens' (1989) Fig. 22 that such jumps are relatively small, apparently in the order of a couple of hundred Hertz.

[3] Though in principle, presumably, an abrupt change in vocal tract configuration could bring about an abrupt change in formant frequencies without affecting cavity-to-formant affiliation.

evidence is yet one step further from underlying gestures and the organisation of speech production than observable articulation, encouraged the authors to carry out two further experiments on [s] and [ʃ]. These both used the technique of electropalatography (EPG), and are reported in Sections 2 and 3 respectively.

1.3. *Doubts about the case for a phonological process*: "*intermediate*" *articulations*

A second issue raised by Browman (1995) is that of what the articulation should be when two gestures (the abstract characterisations which underlie movements) overlap totally, or nearly totally. As noted above, the general view appears to have been that an articulation would result which would be intermediate to those that would result from either gesture in isolation: "the location of the constriction should not be identical to that of either [...], but rather should fall somewhere in between" (Browman & Goldstein, 1989: 220). Indeed Browman (1995: Fig. 23.3) shows how, when gestures for [s] and [ʃ] are overlapped to a high degree, the computational implementation of the gestural model yields a single constriction location intermediate between the distinct constriction locations associated with either fricative when not under the influence of the other, "as must happen given a constraint that the gestures are equally weighted".

But Browman then continues, "Note that this constraint need not be imposed, and indeed may have to be relaxed if articulatory data show unequal influence of [ʃ] and [s]." Such a relaxation might indeed enable the [ʃ]-like fricative constriction to be modelled. The incorporation of context-specific weightings has major implications for the nature of Articulatory Phonology, however, as will be discussed in Section 3.

2. Experiment 1: testing the quantal possibility

In order to test the possibility that either a "strong" or "weak" quantal relationship holds between articulation in the [s] and [ʃ] regions and the acoustic signal, acoustic and EPG recordings were made of the first author (FN) producing a series of glides from [s] to [ʃ] and from [ʃ] to [s]. The Reading "EPG 3" system was used. The speaker's task was similar to one of those performed by the speakers in Perkell *et al.* (1979). Each utterance consisted of a sustained version of the first fricative, a slow and as far as possible smooth transition to the second fricative, and a sustained version of the second fricative. The sequence was started and ended with a homorganic stop (i.e., [tsssʃʃʃt̠] and [t̠ʃʃʃssst]) in order to provide as clear as possible a point of alignment between the EPG and acoustic recordings. The sequences were produced four times each. Lip position was held constant throughout the glides, which is unlikely to have resulted in any unnaturalness since FN is a speaker who does not use appreciable lip-rounding on [ʃ], unlike some speakers of English.

The recordings were analysed in the following way. First, spectrograms were made of each utterance using Xwaves+ speech analysis software from Entropics running on a Silicon Graphics Indigo. Second, the lowest main spectral peak of the fricative energy was measured to quantify the spectral change. This was done in Xwaves+ using a combination of Fourier and LPC analysis. DFT spectra were computed over 20 ms Hanning-windowed frames every 10 ms, and Burg LPC spectra computed for the same frame with the model order set to 20. The frequency of the relevant LPC pole estimate was taken unless visual comparison revealed a poor match of the LPC spectrum to the DFT spectrum, in which case the peak was estimated from the DFT spectrum. Third, for the EPG data, an index was derived which optimally discriminated canonical [s] and [ʃ] palate patterns. The method arrived at was based on the observation that the speaker rarely if ever had contact for [ʃ] in rows 1, 2, and 3 of the palate (at the front), whereas for [s] there was contact in these rows, as seen in the typical example of [s] and of [ʃ] in Fig. 2. This difference results not only from the expectedly more anterior articulation of [s], but also probably from the use by speaker FN of a lamino-alveolar [s], and an apico-postalveolar [ʃ]. Pandeli (1993: Chapter 3) shows that apical articulations tend to have a relatively flat front edge to their EPG contact pattern, while in laminal articulations the contact pattern at the sides tends to extend forward of the narrowest constriction. The use of the tip and blade for these fricatives is probably very variable across English speakers; Bladon & Nolan (1977) found, for eight speakers of British English, that one had an apico-alveolar articulation of [s] and the rest lamino-alveolar, but it is likely from informal observation that the distribution of apical and laminal [ʃ] is more even.

Given the nature of the contact patterns it was possible to calculate an "alveolarity" index as follows:

$$\text{index} = (\text{total of contacts in rows } 1, 2, 3) - (\text{total of contacts in rows } 4, 5, 6)$$

In the case of the tokens in Fig. 2, the index would be $(13 - 12) = 1$ for [s], and $(0 - 12) = -12$ for [ʃ]. On this index [s] normally has a value at or above zero (having contact in front of and behind the division between rows 3 and 4), and [ʃ] has a markedly negative value (having contact only behind the division). It must be stressed that this index is not a characterisation of the two fricatives (for instance as opposed to other sounds), but merely a measure optimised to discriminate between a canonical [s] and a canonical [ʃ].

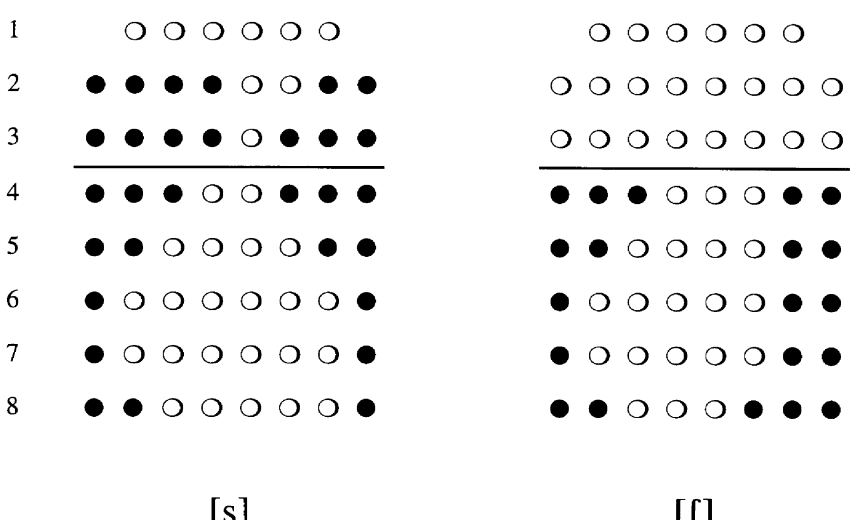

Figure 2. Typical EPG contact patterns for speaker FN.

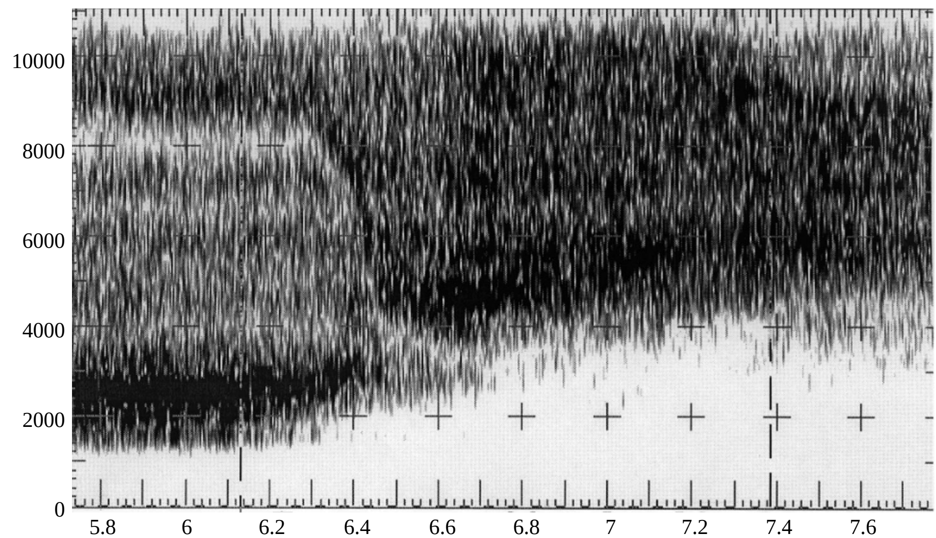

Figure 3. Wideband spectrogram of [ʃ] to [s] transition: token SH-S(4).

At first sight, the spectrograms appear to offer support for a "strong" quantal account. The spectograms in Figs. 3 and 4 show a marked discontinuity in the lowest part of the fricative spectrum during the transition from the [ʃ] target to the [s] target, as if a sudden switch in aerodynamic-acoustic conditions had occurred.

On the other hand, the majority of the spectrograms, as exemplified in Fig. 5 and Fig. 6, show comparatively smooth transitions.

Using the "lowest peak" measurement and the EPG index it is possible to compare the acoustic change in the fricative transitions with the articulatory change

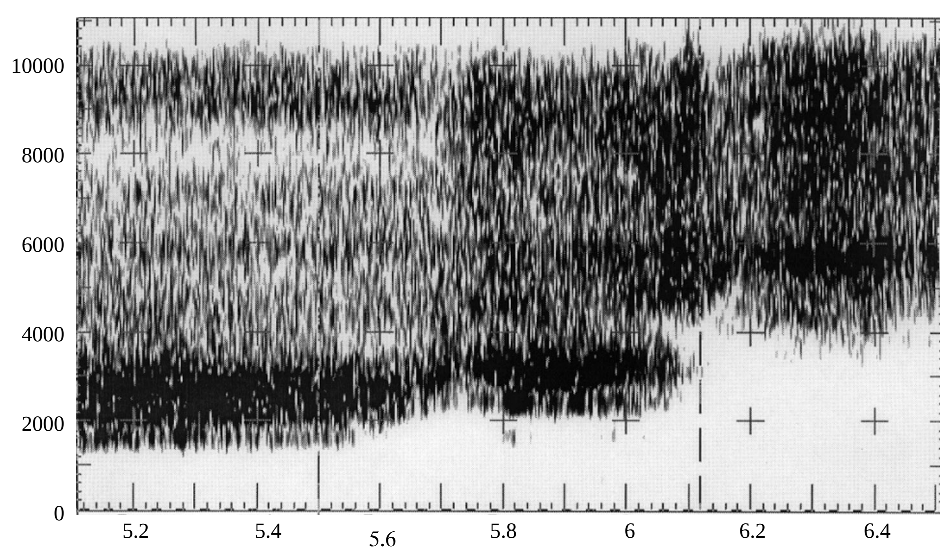

Figure 4. Wideband spectrogram of [ʃ] to [s] transition: token SH-S(3).

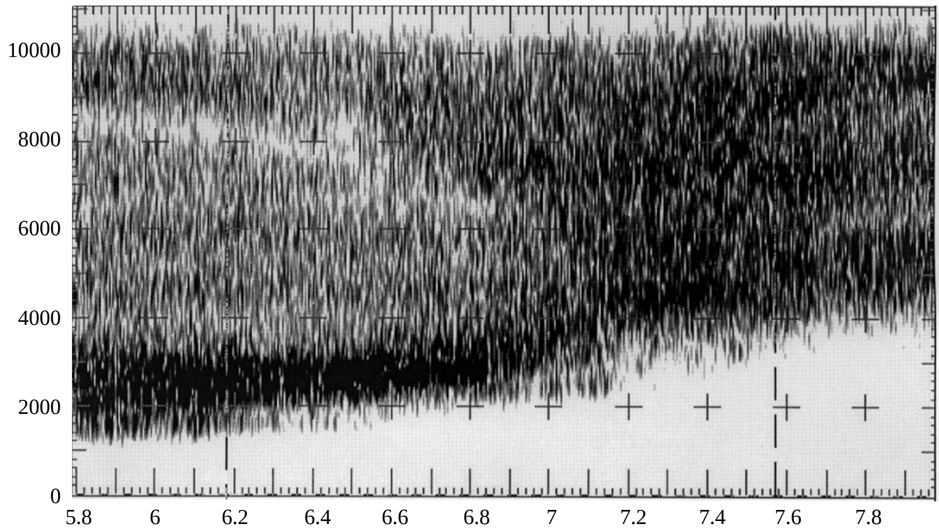

Figure 5. Wideband spectrogram of [ʃ] to [s] transition: token SH-S(1).

in terms of contact pattern. Fig. 7(a–h) presents for each of the eight tokens the time aligned peak frequency (solid line) and the EPG index (dotted line) for the transition between the fricatives. The peak trace is, expectedly, somewhat irregular, since there is moment-to-moment instability in fricative spectra, but overall it reflects the spectral transition. The EPG index is, of course, rather coarsely quantised, since there is a limited number of electrodes involved in its calculation.

A striking finding emerges. In tokens SH-S(4) [Fig. 7(h)] and SH-S(3) [Fig. 7(f)], the spectrograms of which in Figs. 3 and 4 showed apparently quantal effects, the

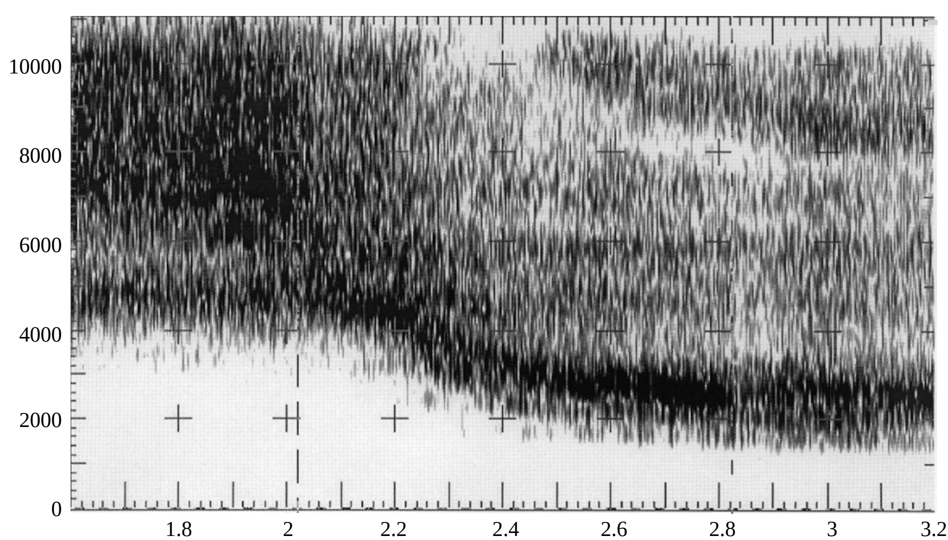

Figure 6. Wideband spectrogram of [s] to [ʃ] transition: token S-SH(1).

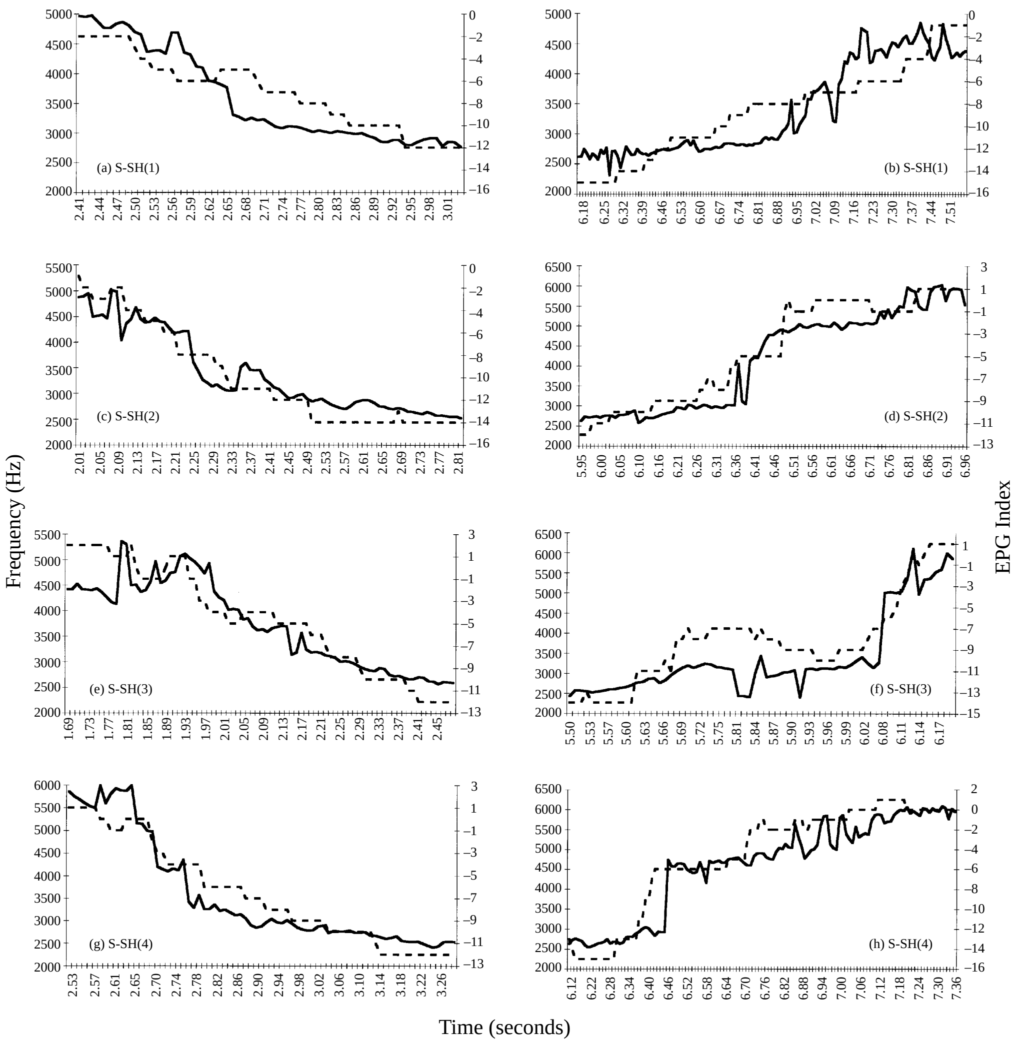

Figure 7. Time-aligned plots of lowest main spectral peak and EPG "alveolarity" index for all eight fricative transitions. —, Peak; - - - -, index.

dramatic shift in spectral peak coincides more or less with an appreciable articulatory discontinuity. This can be found at around 6.40 s in SH-S(4), and around 6.05 s in SH-S(3).[4] This is also true to a smaller extent for token SH-S(2) [Fig. 7(d)] starting around 6.36 s. It seems from these tokens that the speaker did not always succeed, specifically in the [ʃ] to [s] glides, in achieving an articulatorily gradual

[4] It should be borne in mind that the time alignment of the EPG data and the high quality acoustic signal had to be done manually, based on alignment of the frame at which the plosive at the start and end of each utterance was released or made with the acoustic onset or offset of the fricative, and so the alignment may not be perfect.

Table I. Values for Pearson product-moment coefficient of correlation between peak and index measures for the transitional portion of each token (all values significant at $p < 0.0001$)

	S-SH				SH-S				
Token	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	Mean
r	0.877	0.956	0.904	0.941	0.850	0.934	0.873	0.938	0.909

transition. This shows that it is risky to infer a quantal effect from acoustic evidence alone, since to be "quantal" an acoustic irregularity must correlate with a gradual articulation. The remaining five tokens, on the other hand, show a relatively smooth articulatory and acoustic transition, with a good overall correspondence between the EPG index and the peak measurement. Whether the smoother or more abrupt tokens are considered, the match between articulation and acoustics is better than has been supposed in previous work on fricatives in the framework of Quantal Theory.

The discrepancy in findings between the present study and those of Perkell *et al.* (1979) and Nguyen and Hoole (1993), which claimed a quantal effect for these fricatives, is not surprising. Although Perkell *et al.* (1979) did try to monitor articulation by means of two electrodes at the bottom of the lower front teeth, whose purpose was to monitor when tip contact there was broken and the sublaminal cavity opened up, they were not able to monitor for overall smoothness of articulatory transition. As seen from Fig. 7, if the articulatory change between the fricatives is not smooth, the acoustic change will not be smooth. It is therefore quite possible, though not inevitable, that the acoustically most "quantal" tokens in previous studies involved "quantal" articulatory change. In the case of Nguyen and Hoole (1993) the effect is only clearly evident for one of three vowel environments. Even there, it is possible that a single EMA coil might not always reflect change in the constriction.[5] The conclusion of the present study is that there is no support for a "strong" quantal effect, whether caused by resonance crossover, or sublingual cavity coupling (or both).

Assessing whether a "weak" quantal effect is operative or not is more difficult, because it is not clear for a particular phonetic transition what the appropriate function relating an articulatory measure and an acoustic measure should be, except that they should not be related in a highly linear fashion. To test the data with respect to linear correlation, the Pearson product-moment coefficient of correlation (r) between the lowest acoustic peak and the EPG index was calculated for the transitional part of each token. The values are given in Table I. The mean of the correlations is $r = 0.909$, which is rather higher than the value of $r = 0.694$ found by Nguyen and Hoole (1993) between their lowest main spectral peak and position of the tongue-tip EMA coil.

[5] For instance, the change between [s] and [ʃ] might to some extent involve a wave-like rolling of the upper surface of the tongue without much change in the position of the tongue, in which case the coil would (in the manner of a cork on the surface of water) move largely orthogonally to the plane in which the constriction moves.

The rather high value for linear correlation makes it tempting to dismiss the "weak" quantal account, but it might be premature to do so. What is needed is a clear prediction of the non-linear function expected, so that it can be tested whether the data matches that function better. There is clearly more work to be done, both in terms of clarifying what counts as evidence for and against Quantal Theory, and in increasing the quantity of data. The indication of the present small experiment, however, is that in transitions between [s] and [ʃ] there are unlikely to be articulatory movements which are not reflected in at least some acoustic change. Inferences from acoustic patterns about articulatory change, position, and stability in [s] and [ʃ] are not invalidated by the quantal effect, and the conclusions of Holst and Nolan (1995) need not be unsound. Nonetheless it is clearly desirable to compare the findings from that paper based on acoustic inference with findings obtained by more direct monitoring of articulatory behaviour. The second experiment reported here addresses that need.

3. Experiment 2: articulatory data on [s] to [] accommodation

The purpose of Experiment 2 is to assess whether the articulatory activity associated with Type D accommodations is compatible with the mechanisms of Articulatory Phonology (AP), or whether the observable articulations require the postulation of cognitive phonological processes. Experiment 2 is essentially a replication of the experiment in Holst and Nolan (1995), summarised in Section 1, but with the major addition of simultaneous EPG recordings. It also includes, through hindsight, better-matched control utterances designed to provide durations for singleton initial [ʃ]. Before describing the experiment, it will be helpful to review and elaborate on the rationale behind it.

The view that phonological processes sometimes play a part in segmental accommodation will be referred to as "Cognitive Phonology" (CP). This name is not meant to point to a specific version of phonological theory, nor to make any strong claim about the representation of linguistic phonological knowledge in the brain. Its sole purpose is to act as a cover term for a view—common to much of phonology prior to AP—that allocates some of the phenomena of connected speech to a level which is not that of articulation, a level in the "mind" rather than the "mouth". There can be little doubt that some such operations are needed at the more grammatical end of phonology. For example, if the relationship between *electric* and *electricity* is indeed real for speakers, there is no plausible purely articulatory account for the relationship between [k] and [s] at the end of the root morpheme. But the point at issue is whether the changes which happen to pronounceable words when they occur in different rates, styles, and contexts, involve cognitive changes to the speaker's articulatory intention, or are merely effects in the implementation of that intention.

Articulatory Phonology is more constrained than Cognitive Phonology in what it predicts for connected speech. In the following, an attempt is made to clarify the AP position on connected speech by examining what its predictions on accommodation seem to be. Taking sequences such as "...claps Shaun..." and "...clap Shaun..." it seems to predict the following two (provisionally named) possibilities for accommodated forms of the first sequence:

either

(AP1) "GESTURAL BLENDING"

"...claps Shaun..." will have some trace of [s] articulation as a result of gestural blending (even in tokens acoustically classified as Type D) which is lacking in "...clap Shaun...". Either there will be a changing articulatory stricture, as presumably underlies Types B and C in Fig. 1, or a static stricture at an intermediate location. The restriction "at an intermediate location" will need to be discussed further below (and cf. section 1.3), but for the time being let us accept it as being in accord with the intention of Articulatory Phonology, and at odds with the claimed existence of Type D accommodations in Holst and Nolan (1995), which, remember, are wholly [ʃ]-like.

or

(AP2) "GESTURAL DELETION"

"...claps Shaun..." and "...clap Shaun..." will be identical in articulatory location *and* duration as a result of reduction is magnitude of the [s] gesture to a point where it no longer influences articulation. It is a moot point whether Articulatory Phonology allows a gesture to be deleted (gestures not in the lexical representation certainly can't be added); but reduction in magnitude of a gesture to zero must surely be equivalent to deletion in its effect, even if it is ontologically distinct.

At this point we should examine in more detail why GESTURAL BLENDING does not predict a fricative episode which is both identical in articulation to one of the fricatives involved, and longer than it (as Type D is claimed to be by Holst and Nolan (1995)), since this point is crucial to the testability of AP. Figure 8 addresses this issue. Combinations of completely overlapping gestures are seen in the middle

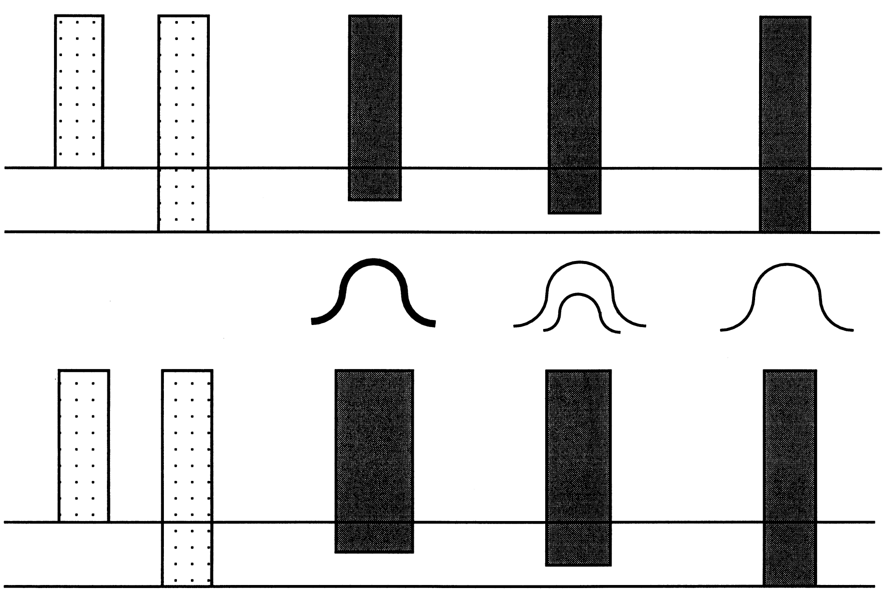

Figure 8. Schematic representation of the effect of reducing an [s] gesture assuming (top) no reinforcement and (bottom) reinforcement of articulatory movement.

row of the figure, with the magnitude of the [s] gesture reducing from, in the leftmost schema, equality with the gesture for [ʃ] (the two overlaid gestures indicated here by the thick line), to, in the third one, zero (equivalent to DELETION). The top and bottom rows show at the left, for reference, schematic spectrograms for [s] and for [ʃ], their cut-off frequencies being marked by lines across the other spectrograms. The rest of the top row shows the three resultant fricative events based on the assumption that, since the two gestures create different articulations, they will not reinforce each other to increase their mutual duration. The bottom row shows the three resultant fricative events based on the alternative assumption that the gestures create articulations similar enough that they will mutually reinforce each other, and create a fricative event more extensive than that of a singleton fricative.

On the first of these assumptions (no reinforcement), which was adopted in Holst and Nolan (1995), completely overlapping gestures never produce a fricative longer than a singleton fricative, as seen in the top schemata. With equal magnitudes the gestures compete on equal terms for shared articulators and (as in Type D in Fig. 1) a fricative intermediate between [s] and [ʃ] results, as shown by the first schematic (shaded) fricative in Fig. 8. When the magnitude of the [s] gesture is reduced, the place of articulation and hence the spectrum should move nearer to that of [ʃ], as in the second one. Zero [s] magnitude, of course, yields a pure [ʃ]. On the second assumption (reinforcement), the simultaneous combination of (non–zero) gestures results in a more extensive articulatory trajectory, and hence a longer fricative, as in the bottom schemata. When, however, the magnitude of the [s] gesture reduces to the point where it has no effect on place of articulation (as in the rightmost schematic fricative), its durational contribution is also lost. Under neither assumption, however, is there any *a priori* justification to assume that a reduced gesture can contribute duration without an effect on the place of articulation, hence the claim that Type D fricatives are not accounted for by AP.

So far we have taken for granted that the relative contributions of gestures can vary as necessary to produce the different forms in Fig. 8. Likewise, as noted in Section 1.3. above, having shown that through its default of weighting gestures equally the computational gestural model produces an intermediate stricture for (almost) totally overlapping [s] and [ʃ] gestures, Browman (1995) points out that the constraint of equal weighting "may have to be relaxed if articulatory data show unequal influence of [ʃ] and [s]". However, this raises interesting difficulties. The weighting would presumably be in favour of [ʃ] to account for data of the "...claps Shaun..." type. But in "fish soup", it seems that [fɪʃʃuːp] is not what people say (cf. Shattuck-Hufnagel, Zue, & Bernstein, 1978; Cruttenden 1994: 259–60), so such a weighting is not a generalisation about [ʃ]. Is it then a generalisation about word or syllable final position? [fɪs suːp] is also not possible, so this hypothesis does not hold. There are clearly complexities here which are unlikely to find their explanation in vocal tract dynamics. The use of weighting to resolve gestural interactions is a powerful tool, but if the claim of AP is now no longer that connected speech processes "...are hypothesized to result from two simple mechanisms... reduce the size of individual gestures and increase their overlap" but instead "...reduce the size of individual gestures, increase their overlap, and, in the case of overlapping gestures, apply weighting functions which are specific to phonological context, the identity of the segments involved, and their order", it is clear that AP has made a major step towards admitting the need for that device which traditionally models arbitrary facts about the sound patterns of a language—by name, the phonological rule. Further, if weighting is freely allowed, the version of AP which results becomes so powerful in terms of the variety of outputs it can generate that it becomes far less testable than a more constrained version.

Let us now consider what the equivalent predictions of Cognitive Phonology might be:

either
(CP1) "ARTICULATORY BLENDING"
"...claps Shaun..." will have some trace of [s] articulation as a result of articulatory blending (i.e., specifically in tokens classified as Type B and C) which is lacking in "...clap Shaun..." This situation is, of course, equivalent to "GESTURAL BLENDING" as in (AP1), and the name differs only because not all Cognitive Phonologists might want to adopt the Task Dynamic, gestural model as a model of phonetic implementation.
or
(CP2) "PHONOLOGICAL DELETION"
"...claps Shaun..." and "...clap Shaun..." will be identical in articulatory location *and* duration as a result of the operation of a phonological rule deleting the [s] completely (i.e., its featural content and its timing slot). This might be termed "PHONOLOGICAL DELETION"
or
(CP3) "PHONOLOGICAL ASSIMILATION"
"...claps Shaun..." and "clap Shaun..." will be identical in articulatory location (i.e., Type D), but "...claps Shaun..." will be longer than "...clap Shaun..." because a phonological rule has applied changing the phonological specification of [s] into that for [ʃ]. The resultant fricative will be longer than a single fricative,[6] reflecting the two timing slots in the underlying specification, but there will be no trace of [s] activity. The obvious name for this prediction, of course, is "PHONOLOGICAL ASSIMILATION".

CP1 is the same as AP1 because Cognitive Phonology does not deny the necessity of some mechanism of articulatory blending, whether it be precisely that of AP or some other. It is clear that many gradient phenomena of accommodation occur which are not appropriately modelled in terms of phonological rules (see e.g., Nolan, 1992). For the present, in the absence of other comparably explicit candidates, we will assume the mechanism of AP. For concision, AP1/CP1 will be referred to as BLENDING.

CP2 and AP2 are not testably distinct. If all traces of an underlying segment disappear, then it is impossible to know whether this is because of a cognitive change in the speaker's phonological intention, or because it has simply been submerged in the articulatory process. Whether or not AP allows, in some discrete sense, the deletion of a gesture, if its magnitude can reduce effectively to zero the case is empirically indistinguishable from deletion, and we can merely recognise a category of articulatory event termed DELETION and which is predicted by either model.

It is ASSIMILATION which is the crucial case for distinguishing AP and CP.

[6] Though not necessarily much longer—see footnote 1.

"Assimilation" is often used to cover a wide range of phenomena, including those referred to above as BLENDING. Here, ASSIMILATION (with PHONOLOGICAL omitted for brevity) will be used exclusively for cases where the identity of a whole segment or a feature of that segment appears to have accommodated completely to an adjacent segment, but evidence remains (i.e., from duration or other features) that the segment was nonetheless present in the speaker's articulatory plan (i.e., had not been subject to DELETION).

3.1. *Experimental design*

One female speaker (LA) and one male speaker (JR) were recorded, both of them naive to the purpose of the experiment. As well as an acoustic recording, an EPG recording was made using the Reading system with EPG 3 software. The subjects read four (LA) or six (JR) sets of sentences[7] of the following type, randomly ordered among a larger set of material, the sentences in each set being repeated once during the recording:

Cs#ʃ Although the crowd claps Shaun, he isn't very good
C#ʃ Although the crowds clap Shaun, he isn't very good
Cs#C Although the crowd claps Paul, he isn't very good

Thus for each each utterance potentially containing an accommodation (e.g., "clap[ʃʃ]aun"), there was a control with a singleton (word-initial) [ʃ] and another with a singleton word-final [s]. The choice of verbs ending in voiceless stops, which undergo the addition of [s] in the third person singular, was determined by syntactic considerations: these sentences correspond to the −CB context (no clause boundary at the accommodation site) of Holst and Nolan (1995), and have +CB parallels in sentences such as "Although the crowd claps, Shaun isn't very good". The rest of the material recorded did include parallel +CB sentences, but given the tendency (Holst & Nolan, 1995) for the crucial Type D accommodation to be inhibited by a clause boundary, and the fact that durational data from +CB and −CB conditions would not be directly comparable because of the effect of a prosodic boundary, it was decided to concentrate here on the −CB context. Nonetheless it should be noted here that the present two speakers seemed more ready to produce Type D accommodation across a clause boundary than were the speakers in the earlier experiment. This may be a result of the sentences in the present experiment being shorter, which might favour weaker prosodic boundaries at the clause division.

3.2. *Analysis*

The acoustic and EPG recordings at the accommodation site were analysed as follows. The acoustic signal was analysed spectrographically using Xwaves+. As described in Section 1.1 the fricative portions of the utterances were allocated on the basis of visual inspection of their spectrograms to one of four categories A, B, C, and D. Those with discretely different [s]-like and [ʃ]-like portions of fricative noise were allocated to Type A, those including a continuously changing portion of

[7] An additional two sets of sentences of this type were added before JR's recording session.

friction to Type B or C, and those which manifested no change in the distribution of spectral energy through their duration, other than that attributable to the coarticulatory influence of adjacent vowels, and which were not spectrally distinct from their C#ʃ control, to Type D. The categorisation provides an estimate of the "spatial" accommodation between segments. The duration of each fricative event was measured from the spectrographic display, the beginning and end points being defined as the onset and offset of high frequency aperiodic energy.

In the case of the EPG recording, the same method was used as described in Section 3. That is, for each speaker a dividing line near the front of the palate was chosen which allowed for optimal discrimination of that speaker's canonical [s] and [ʃ] articulations. It should be noted that it was less easy to achieve this discrimination than in the case of FN. This was in part because both fricatives of LA and JR appeared to be articulated further forward than those of FN (although the placement of electrodes in the artificial palate may contribute part of the apparent difference), which in turn meant that some of the contact for [s] may have been in front of the first EPG row. For the female speaker (LA) and for the male speaker (JR) the optimal division turned out to be between rows 2 and 3. Figure 9 shows examples of typical [s] and [ʃ] for LA, and Fig. 10 for JR. The dividing line was therefore placed between rows 2 and 3, and the index for these speakers calculated as follows:

$$\text{Index} = (\text{total of contacts in rows } 1, 2) - (\text{total of contacts in rows } 3, 4)$$

The index was calculated for 100 frames (i.e., one second) from the start of the first word of the pair containing the accommodation site, for instance from the start of the word "claps" in "...claps Shaun...". The span of 100 frames was chosen to be long enough to include both words adjacent to the accommodation site. The trajectory of the index allows a comparison of accommodated fricatives with control environments. It also allows a measure of duration to be made which may have a

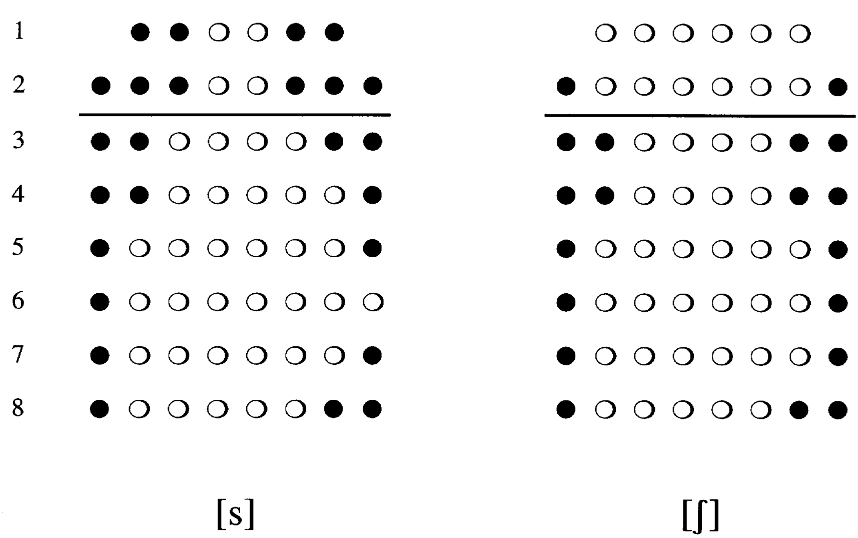

Figure 9. Typical EPG contact patterns for speaker LA.

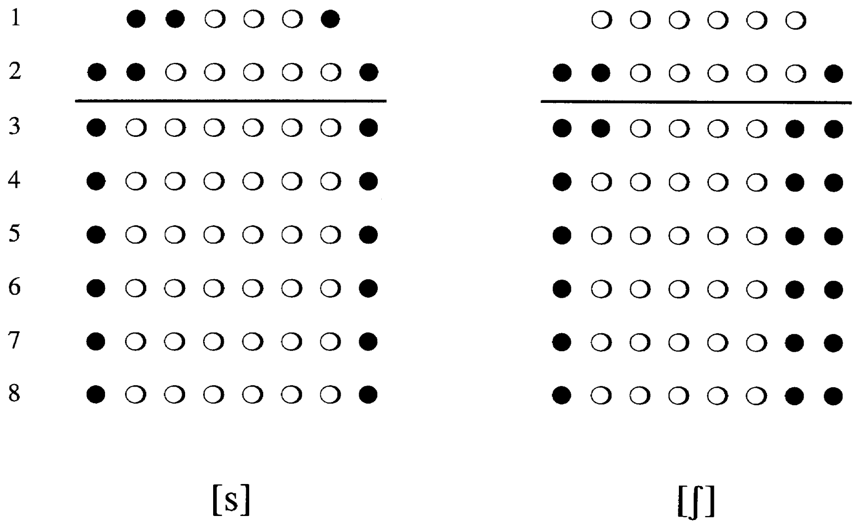

Figure 10. Typical EPG contact patterns for speaker JR.

closer relation to the lingual gesture than the acoustic duration of friction. The nature of the index means that [ʃ]-type fricatives involve an excursion of the index below the zero line, whereas [s]-type fricatives yield index values at or above the zero line.

Figure 11 illustrates the trajectory of the index, over the fricative episode, for the mean of four tokens from JR which were classified spectrographically as either B or C. These tokens are not included in the discussion which follows, since they were not of the crucial Type D, and since they were taken from +CB sentences; but because the tokens begin with a more [s]-like fricative they illustrate well the information available from the index. Figure 11 and the subsequent two figures show

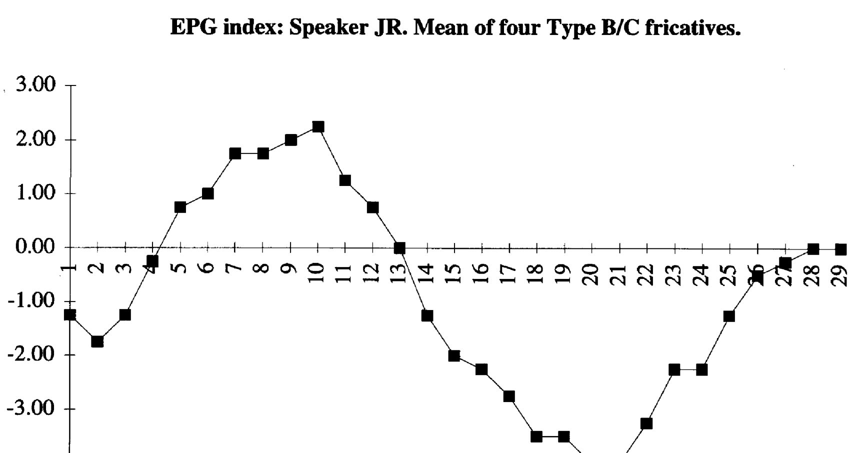

Figure 11. EPG "alveolarity" index for the mean of four tokens classified spectrographically as either B or C (speaker JR).

TABLE II. Mean durations (in ms) of post-consonant word-boundary fricative events (Type D realisations of Cs#ʃ, e.g., "...claps Shaun...", and realizations of C#ʃ, e.g., "...clap Shaun...") measured acoustically and from the EPG index. The significance of the difference between the conditions is assessed by a one-tailed paired T-test. Compatibility (√) or incompatibility (x) of the results with various hypothesis is also shown

Spkr	n	Measure	Type D Cs#ʃ	Control C#ʃ	(Cs#ʃ-C#ʃ)	Significance	DELETION	BLENDING	ASSIMILATION
LA	8	Accoust.	140	131	+9	ns	√	√	×
	8	EPG	205	194	+11	ns	√	√	×
JR	9	Acoust.	135	113	+22	$p < 0.01$	×	√	√
	9	EPG	211	187	+24	$p < 0.05$	×	√	√

29 frames of the EPG index, this being the maximum required to encompass the fricative event in all the data. It will be observed that even for the [s]-like fricative there is a negative excursion of the index at the start; this is because contact is built up from the back of the pattern. It is, therefore, normally possible to define the articulatory starting point of each fricative episode as the first negative value of the index after the zero values of the vowel, and this is the point of alignment of different tokens for averaging. The duration of the articulatory event corresponding to a fricative gesture can then be estimated as the number of EPG frames from the first to the last frame inclusive below the zero line—from frame 1 to frame 27 in Fig. 11 (although since Fig. 11 is the mean of four trajectories, it would not be meaningful to take a duration value from it). Each EPG frame is equivalent to 10 ms.

3.3. *Results and interpretation*

Duration will be dealt with first. Table II summarises the durations in ms for Type D realisations of Cs#ʃ (e.g., "...claps Shaun...") and their paired singleton-fricative C#ʃ controls ("...clap Shaun..."). Durations are given as measured acoustically and from the EPG index. Acoustic durations are regularly shorter than the corresponding durations measured from the EPG index, which is entirely to be expected since the approach and release of the articulatory constriction will extend beyond the acoustic segment produced.

The potential total number of utterance sets available from the recordings was 8 from LA, and 12 from JR, who recorded some additional tokens. Three Cs#ʃ utterances (and therefore their controls, too) had to be rejected from JR's data because of difficulty defining the start or end of the fricative event in the EPG index, since occasionally the index did not achieve zero either before or after the fricative event, thus making the definition of the fricative duration impossible; or, in one case, because of difficulty classifying the acoustic pattern. Since the number of tokens is relatively small, the results must be treated as suggestive rather than conclusive.

Let us first consider speaker LA. Her Type D tokens show only a slight and non-significant trend towards greater length than the controls: +9 ms and +11 ms, respectively, on acoustic and articulatory measures. If the slight trend is ignored, as it probably has to be, LA's durational results are compatible with DELETION or

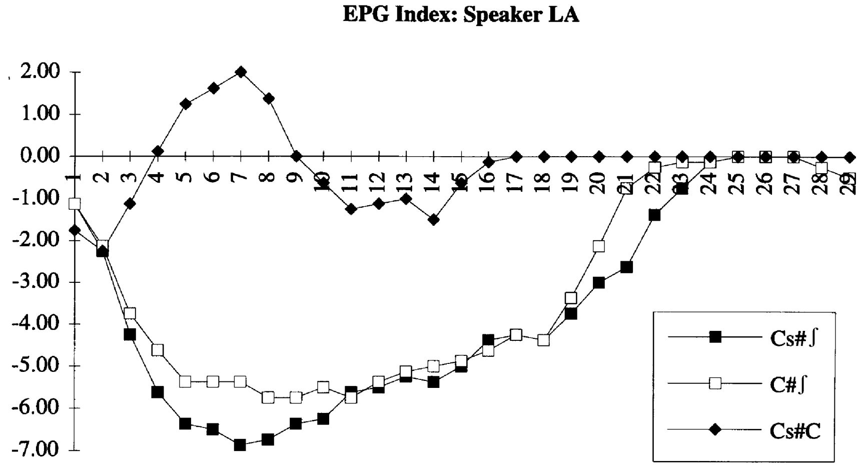

Figure 12. Mean EPG "alveolarity" index for fricative events in Type D Cs#ʃ, and the two matched control contexts C#ʃ and Cs#C (speaker LA).

BLENDING, and the ambiguity will have to be resolved by reference to the articulatory (EPG) trajectory.[8]

JR appears to behave differently. For him, the Type D tokens are significantly longer, as seen both from the acoustic signal (+22 ms, $p < 0.01$) and from the EPG index (+24 ms, $p < 0.05$), than the singleton-fricative controls. It is interesting to note that these figures conform fairly well to the findings of Holst and Nolan (1995: Table I). In that study, the equivalent of the Type D Cs#ʃ items in the present experiment were found to be 16% longer than singleton initial [ʃ].[9] In Table II above, JR's Type Ds are either 20% or 13% longer than the control initial [ʃ], as estimated from the acoustic signal or from the EPG index, respectively.

The greater duration of Cs#ʃmakes it unlikely that a segment (i.e., [s]) has been subject to DELETION. The durational finding is, however, compatible with both BLENDING and ASSIMILATION, at least if (despite the indications to the contrary in Experiment 1) it is still assumed that the "Type D" spectrographic pattern might be concealing articulatory evidence of an [s]-gesture. To resolve this ambiguity it will be necessary to look at the articulatory trajectory in the form of the EPG index to see whether there is evidence or not of an [s]-gesture.

Turning to the articulatory trajectory, Fig. 12 shows the mean EPG index for LA's fricative events in three contexts: Type D Cs#ʃ, and the two control contexts C#ʃ and Cs#C (for instance "...claps Paul..."). The indices are lined up at the

[8] There is a way in which LA's Type D durations could arise from ASSIMILATION, namely if (a) after the process applied the two [ʃ] gestures overlapped completely and (b) if overlapping identical gestures do not "reinforce" each other to yield a more extensive movement and hence greater duration. However, since the interest here is to find phenomena which give unambiguous evidence for one theory or the other, this interpretation will not be pursued.

[9] This value was measured acoustically and calculated over 117 tokens from the productions of twelve speakers. The "control" items had not been designed into the experiment, and hence were not from matching environments. They consisted merely of tokens of singleton word-initial [ʃ] occurring fortuitously in the material. Because different speakers were involved, durations were normalised for each speaker with respect to that speaker's mean duration for the singleton word-initial [ʃ], the latter being taken as 1.0.

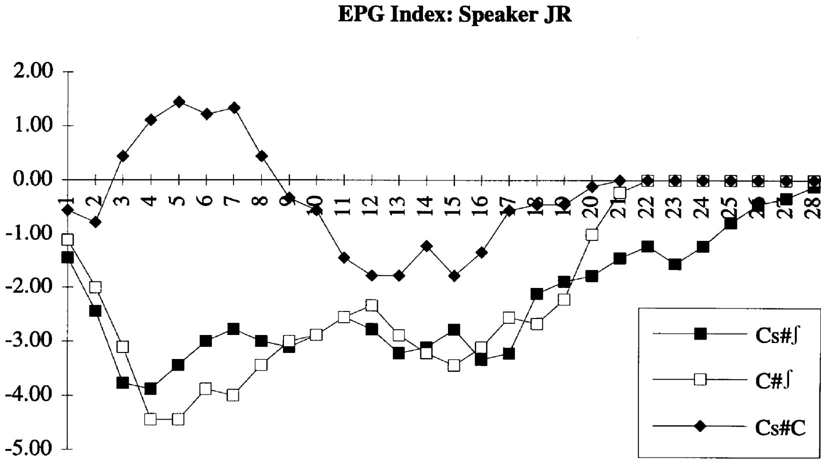

Figure 13. Mean EPG "alveolarity" index for fricative events in Type D Cs#ʃ, and the two matched control contexts C#ʃ and Cs#C (speaker JR).

first frame which deviates below the zero line.[10] This can be done for [s] (filled diamonds) as well as [ʃ] because (as noted above) although the target articulation for [s] yields a clearly positive index, the contact pattern builds up from the back, and decays towards the back. In the first few frames of contact, the contact is, therefore, behind the dividing line for the index, and only when the contact spreads forward does the index achieve its characteristic positive value for [s]. Note that the return of the EPG trace to the zero line in Fig. 12 (and later in Fig. 13) corresponds to the longest token, since averaging over any sub-zero values will bring the mean index below zero. The fricatives, therefore, "look" longer in the figures than their mean durations in Table II, since the values in Table II are the mean durations calculated over the time taken for each individual token to return to the zero line.

Recall that the durational data for LA (Table II) pointed to LA being a speaker for whom there is no evidence of ASSIMILATION, and who, therefore, exhibits either DELETION or BLENDING. Figure 12 resolves clearly which it is. Her mean Type D Cs#ʃ trajectory on the index (filled squares) is completely [ʃ]-like. In fact, the curve for the mean Type D is *lower* than that for her matching canonical C#ʃ control items, although the distinction is small by comparison with the range available between [s] and [ʃ], and unlikely to be meaningful. What is clear is that there is no evidence of an [s] gesture, and therefore her behaviour appears to be DELETION and not BLENDING.

Fig. 13 shows the mean EPG index for JR's fricative events in the same three contexts. It can be seen that the trajectory for Cs#ʃ (filled squares) is broadly very similar to that for C#ʃ, being at all times within approximately one index unit, and shows no excursion to the positive ([s]) side of the zero line. To this extent, Fig. 13 seems to support the view that JR is producing only an [ʃ] gesture, but (see Table II, as well as Fig. 13) one which is marginally longer than his realisation of a single

[10] This alignment is operationally convenient, and probably justifiable in the present context where the early part of the fricative episode is of most interest, but it should be kept in mind that other alignments between the traces are also possible in principle, such as aligning peak displacement, or aligning the traces symmetrically with respect to start and end.

[ʃ] in C#ʃ. These findings speak in support of ASSIMILATION on the part of this speaker.

Closer examination of the EPG indices in Fig. 13 shows, admittedly, that the Cs#ʃ trajectory does not fall to as low a value as that for C#ʃ. This tantalising discrepancy might be taken as indicating that there is indeed some trace of an [s] gesture in the early part of the fricative. Such a trace would, of course, speak in favour of BLENDING and against ASSIMILATION. However, the difference is in the order of one index unit, and it should be borne in mind that a difference of one index unit corresponds to a change in the contact status of only one electrode. A difference of similar magnitude, for instance, exists between Type D Cs#ʃ and control C#ʃ for speaker LA (Fig. 12) in the early part of her fricatives—but for LA it is the Type D which has the greater negative excursion. There is no question for LA of interpreting the difference as caused by the presence of an [s] gesture perversely in "...clap Shaun..." but not "...claps Shaun", and so there seems no justification to interpret a difference of similar magnitude as the presence of an [s] gesture in the case of JR where the difference is in the expected direction. We may conclude, but only tentatively, that JR is operating with a process of ASSIMILATION.

3.4. *Discussion*

The picture which emerges, then, is not straightforward, and the experiment cannot claim to have resolved the debate between Articulatory Phonology (AP) and Cognitive Phonology (CP). Speaker JR has Type D fricative events which are longer than single fricatives, and which appear [ʃ]-like (but not unequivocally so), and his data, therefore, offer qualified support for the view of Holst and Nolan (1995) that in some cases a speaker applies a cognitive phonological rule of ASSIMILATION to the phonetic plan, and then articulates a cluster of two homorganic fricatives. This is the prediction labelled CP3 above.

The other speaker (LA) appears to reject ASSIMILATION in favour of DELETION. As noted at the start of Section 3, DELETION (AP2/CP2) is evidently compatible with both Articulatory Phonology and Cognitive Phonology. In AP it is the result of reducing the magnitude of a gesture to the point where it has no detectable influence on observable articulation. In CP it is the result of a symbolic phonological process, the removal of a segment. It is unlikely that an empirical method could be found to test which of these is happening in a particular case, or in general. The issue might have to be decided derivatively, on the basis of the broader adequacy or inadequacy of AP. If it were the case that all other connected speech phenomena were modellable by the mechanisms of AP, it would clearly be unparsimonious to set up a mechanism for the symbolic processing of pre-articulatory representations just to handle DELETION. On the other hand if, as is claimed here, that symbolic mechanism is already required for ASSIMILATION, then it cannot be ruled out from having a role for DELETION, too.

It will be apparent that before a definitive answer emerges to the question of whether we need Cognitive Phonology, a number of shortcomings of the present experiment will have to be rectified. It yielded a relatively small amount of data, and further studies will have to be on a larger scale. If such studies use EPG, thought will need to be given to the relation of palate patterns to gestures. The EPG index used here was chosen because it successfully discriminates [s] and [ʃ], but it is not the only possible characterisation of these sounds, and others should be explored, as should the relation of such indexes to other measures (e.g., of tongue displacement) more frequently used as reflexes of articulatory gestures. And an objective metric will need to be devised for deciding when two index (or other) trajectories are the same and when they are different.

4. Conclusion

This paper has continued the exploration of articulatory accommodation as a window on speech production. It has presented two experiments, both of which address cogent criticisms levelled at Holst and Nolan (1995), which investigated [s] to [ʃ] accommodation on the basis of purely acoustic data, and which claimed that the most radical degree of accommodation could not be modelled by Articulatory Phonology. These criticisms are to be found in Browman (1995).

The first experiment has shown, for one speaker at least, that there is a close correlation between spectral change in coronal fricatives and articulatory change as reflected in an "alveolarity" index derived from EPG contact patterns, contrary to the prediction of a strong interpretation of "Quantal Theory". This suggests that the findings of Holst and Nolan's study, which are based on a relatively large amount of data from twelve speakers, need not necessarily be vulnerable to attack simply because they are not based directly on articulatory data. Nonetheless, the point is accepted that acoustic data are yet one further step away from the "gestures" hypothesised by Articulatory Phonology to underlie observable articulatory movements, and so ideally would be supplemented by or replaced by articulatory movement data.

The second experiment provides such supplementary data, albeit on a small scale. The findings are not unambiguous. The two speakers analysed appear to differ in behaviour in tokens showing the greatest degree of accommodation, one speaker apparently applying a phonological assimilation rule in the way suggested by Holst and Nolan (1995), and the other behaving in a way which more nearly suggests deletion of the first segment. In neither case can it be argued convincingly that there is evidence of an [s] gesture, and so both speakers contradict a "strong" version of Articulatory Phonology which maintains that gestures in underlying phonological representations will always be present. The predictions of Articulatory Phonology are still somewhat equivocal, however, on whether gestures can be deleted, whether the magnitude of gestures can be reduced to zero, and whether these two locations refer to distinct phenomena or are merely paraphrases of each other. Further empirical testing of AP is certainly required, but also needed is an exact statement of the predictions of AP with regard to segmental accommodation, so that crucial tests of the predictions can be formulated.

Articulatory Phonology already goes impressively far towards accounting for the phenomena of connected speech by providing an implementable description of the process of articulatory blending. It is fully acknowledged by the authors here and in Holst and Nolan (1995) that very many observed outputs of segmental accommodation are appropriately described in gestural terms, that is, as mechanical consequences of the dynamics of the articulators. It would then be extremely elegant and economical if AP were able to go the whole way, and account for all accommodatory phenomena, making Cognitive Phonology redundant, at least in the arena of connected speech processes. Elegance and economy are not enough, however, and the scope of AP must be tested by seeking phenomena it cannot handle. Such a phenomenon appears to be the case of [s] to [ʃ], if the resultant fricative is homogeneously [ʃ]-like, yet longer than a singleton [ʃ]. But it must be admitted that the demonstration of this phenomenon in the present study hangs by rather fine threads, and it may yet turn out that Cognitive Phonology has no role in the production of connected speech.

The authors are grateful to the two reviewers, Noel Nguyen and D. H. Whalen, and to Daniel Recasens, for their detailed comments and suggestions on an earlier version. Responsibility for remaining infelicities lies, of course, solely with the authors. The experiment reported here formed part of a project "Post-lexical and prosodic phonological processing", held at the University of Cambridge by Paul Warren and Francis Nolan (Linguistics Department) and Ted Briscoe (Computer Labotarory), funded by the Joint Research Councils' HCI initiative under grant SPG 9030657.

References

Bladon, R. A. W. & Nolan, F. (1977) A video-fluorographic investigation of tip and blade alveolars in English, *Journal of Phonetics,* **5**, 185–93.

Browman, C. P. (1995) Assimilation as gestural overlap: comments on Holst and Nolan. In *Phonology and phonetic evidence: papers in laboratory phonology IV* (B. Connell & A. Arvaniti, editors), pp. 334–342. Cambridge: Cambridge University Press.

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

Browman, C. P. & Goldstein, L. (1990) Tiers in articulatory phonology, with some implications for casual speech. In *Papers in laboratory phonology I: between the grammar and physics of speech* (J. Kingston & M. E. Beckman, editors), pp. 341–376. Cambridge: Cambridge University Press.

Browman, C. P. & Goldstein, L. (1992) Articulatory Phonology: an overview, *Phonetica,* **49**, 155–80.

Cruttenden, A. (1994) *Gimson's Pronunciation of English* (5th edition). London: Arnold.

Fowler, C. (1980) Coarticulation and theories of extrinsic timing, *Journal of Phonetics,* **8**, 113–133.

Fowler, C., Rubin, P., Remez, R. E., & Turvey, M. T. (1980) Implications for speech production of a general theory of action. In *Speech production, Vol. 1* (B. Butterworth, editor), pp. 373–420. New York: Academic Press.

Hawkins, S. (1992). An introduction to task dynamics. In *Papers in laboratory phonology II: gesture, segment, prosody* (G. J. Doherty & D. R. Ladd, editors), pp. 9–25. Cambridge: Cambridge University Press.

Holst, T. & Nolan, F. (1995) The influence of syntactic structure on [s] to [ʃ] assimilation. In *Phonology and phonetic evidence: papers in laboratory phonology IV* (B. Connell & A. Arvaniti, editors), pp. 315–333. Cambridge: Cambridge University Press.

Kelso, J. A. S., Saltzman, E. L., & Tuller, B. (1986) The dynamical perspective on speech production: data and theory, *Journal of Phonetics,* **14**, 29–59.

Nolan, F. (1992) The descriptive role of segments: evidence from assimilation. In *Papers in laboratory phonology II*: *gesture, segment, prosody* (G. J. Docherty & D. R. Ladd, editors), pp. 261–280. Cambridge: CUP.

Nguyen, N. & Hoole, P. (1993) Frequency variations in the lowest main spectral peak in sibilant clusters. In *Proceedings of the Eurospeech conference 1993,* pp. 81–84.

Pandeli, H. (1993) The articulation of lingual consonants: an EPG study. Unpublished PhD Dissertation, University of Cambridge.

Perkell, J. S., Boyce, S. E. & Stevens, K. N. (1979) Articulatory and acoustic correlates of the [s − š] distinction. In *Speech communication papers presented at the 97th meeting of the ASA, Cambridge MA 12–16 June, 1979* (J. J. Wolf & D. H. Klatt, editors). New York: Acoustical Society of America.

Shattuck-Hufnagel, S., Zue, W. V., & Bernstein, J. (1978) An acoustic study of palatalisation of fricatives in American English, *Journal of the Acoustical Society of America,* **64**, S92(A).

Stevens, K. N. (1972) The quantal nature of speech: evidence from articulatory-acoustic data. In *Human communication: a unified view* (E. E. David & P. B. Denes, editors), pp. 51–66. New York: McGraw-Hill.

Stevens, K. N. (1989) On the quantal nature of speech, *Journal of Phonetics,* **17**, 3–45.

Umeda, N. (1977) Consonant duration in American English, *Journal of the Acoustical Society of America,* **61**, 846–858.