Proceedings of The Institute of Acoustics

DIMINUTION OF HIGH FREQUENCY ENERGY AS A CUE TO THE VOICELESSNESS OF FOLLOWING CONSONANTS

Geoffrey Lindsey, Michael Johnson and Adrian Fourcin

SPAR Group, University College London, U.K.

INTRODUCTION

Spectrographic analysis of speech reveals fluctuations in high frequency energy (HFE) -- often quite massive -- which are not generally considered to be invariant properties of the segments with which they are associated. Analysis using an electrolaryngograph signal in addition to standard spectrographic and simple waveform techniques indicates, however, that some of these fluctuations may be associated with rule-governed variations in larynx activity. For example, research using a program which operates on the laryngograph signal to calculate durations of glottal open and closed phases [1] suggests that an increase in the relative open phase duration of the glottal cycle before final and intervocalic voiceless consonants correlates with loss of HFE in this environment.

Figures 1 and 2 show the utterances [ɑ:sɑ:] and [ɑ:zɑ:] respectively, produced by a male speaker with a high fall on the second syllable in each case. Below each spectrogram is the corresponding speech pressure waveform (Sp), the laryngograph waveform (Lx), a row of vertical ticks marking the periods (Tx) and a fundamental frequency contour (Fx) calculated from the inverse of Tx. Below these are plots derived from the Lx waveform showing the cycle by cycle durations of the glottal open phase (OP -- dotted line) and closed phase (CP -- solid line). Of particular interest for present purposes is the marked increase in OP immediately prior to the consonant in [ɑ:sɑ:], which coincides with the HFE diminution visible on the spectrogram above (this diminution reaches as far down as F1 just before the fricative). OP increase and HFE loss are markedly absent in the vowel immediately prior to the voiced fricative in Figure 2.

The present work describes an initial perceptual experiment designed to test the role of HFE loss as a cue to the voicelessness of a following consonant. It was decided to base the experiment on the voicing contrast between the two fricatives [s] and [z], as illustrated by Figures 1 and 2, since plosives and affricates are often glottalized in English, in which case it is closed rather than open phase which increases before the consonant.

METHOD

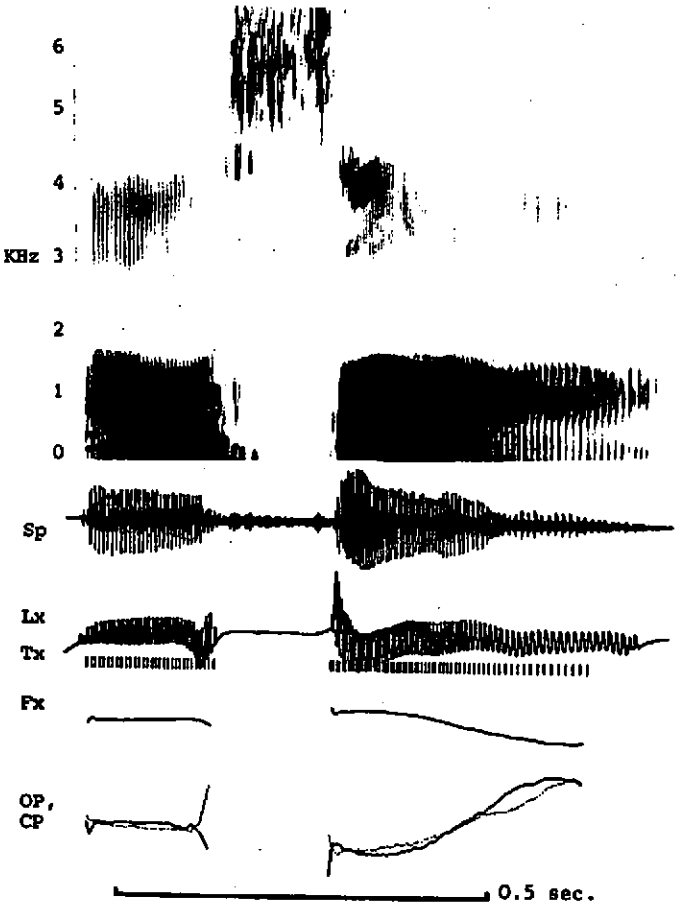
The experiment was conducted in two stages.

(1) Firstly, it was necessary to synthesise a token which would be perceived as intermediate between [s] and [z], as a basis for testing the status of HFE diminution as a cue. The utterances [ɑ:sɑ:] and [ɑ:zɑ:] illustrated in Figures 1 and 2 underwent larynx-synchronous analysis with ILS software (extracting four peaks) and a formant-tracking program due to Mark Huckvale. The parameter values estimated in this way were input to a parallel formant synthesiser developed at University College, which has six formant filters (one "nasal" formant, which we did not use, and five others with higher frequency ranges) and variable bandwidth control.

We then created six syntheses which were intermediate between the synthesised [ɑ:sɑ:] and [ɑ:zɑ:]. All six intermediate tokens had identical (neutralised)

DIMINUTION OF HIGH FREQUENCY ENERGY AS A CUE TO THE VOICELESSNESS OF FOLLOWING CONSONANTS

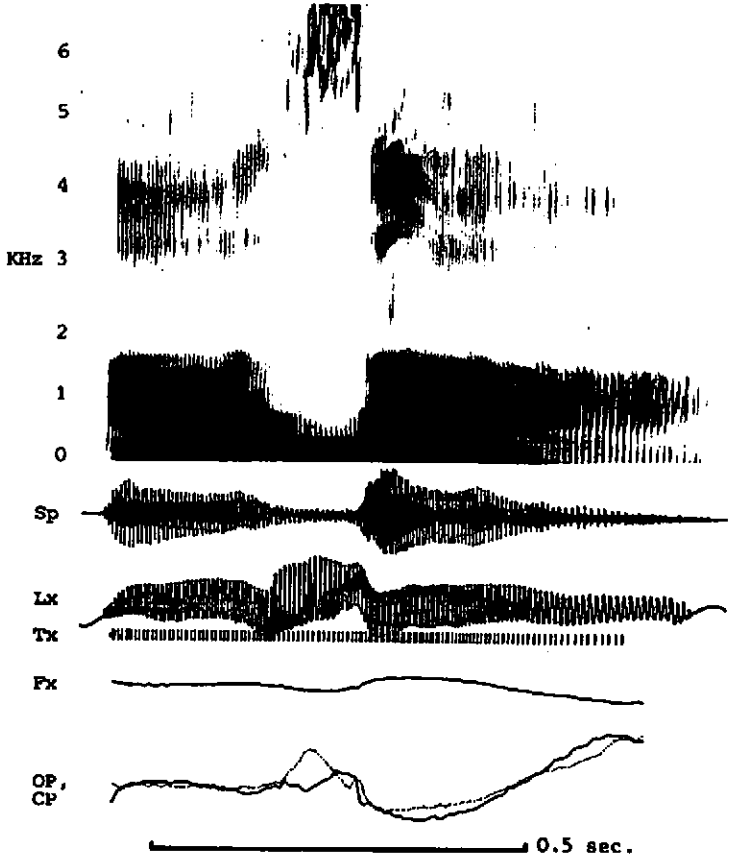
Figure 1. [ɑ:sɑ:] produced by a male speaker.



durations for the first two segments, Fx contours and formant transitions. Constraints on our time, however, meant that we were not able to vary other parameters independently, so various combinations were tried: two tokens had the [s] from [ɑ:sɑ:], one of these with added voicing; four had the [z] from [ɑ:zɑ:] with decreased voicing; five had the final vowel of [ɑ:sɑ:], one of these with greater amplitude on the first syllable than the second; one token had the final vowel of [ɑ:zɑ:]. The microintonational Fx contour at the onset of the second vowel was varied between the rather flat shape associated with a preceding voiceless consonant and the slight rise associated with a preceding voiced consonant. All six intermediate tokens, however, were synthesised without appreciable HFE diminution prior to the fricative.

DIMINUTION OF HIGH FREQUENCY ENERGY AS A CUE TO THE VOICELESSNESS OF FOLLOWING CONSONANTS

Figure 2. [a:za:] produced by a male speaker.



Twelve repetitions of each of the eight synthetic tokens (the two extremes and six intermediates) were played in random order to eight subjects, who had to label each as [a:sa:] or [a:za:]. Most of the intermediates were labelled systematically as [a:sa:] or [a:za:]; we selected for the remainder of the experiment the token which was labelled most randomly. This token had the fricative and final vowel of [a:za:], with full devoicing of the fricative and flattening of the microintonational rise at the start of the second vowel.

(2) In the second stage of the experiment, eight versions of the selected intermediate token were synthesised, with graded degrees of HFE loss in the 20 ms. preceding the fricative (spreading downwards from the top of the frequency range). Since the intermediate token on which these eight were based had no HFE loss, it was felt that the eight new tokens would show an overall [a:sa:] bias; therefore the flattened Fx contour at the start of the second

vowel was slightly re-lowered (in all eight tokens). Figures 3 and 4 are diagrams of the parameter values of the new synthetic tokens with most and least (i.e. no) HFE diminution, respectively. The vertical lines crossing the formant plots represent the relative amplitude of the formant parameters.

Figure 3. Diagrammatic representation of a synthetic token intermediate between [ɑ:sɑ:] and [ɑ:zɑ:], with amplitude diminution before the fricative.

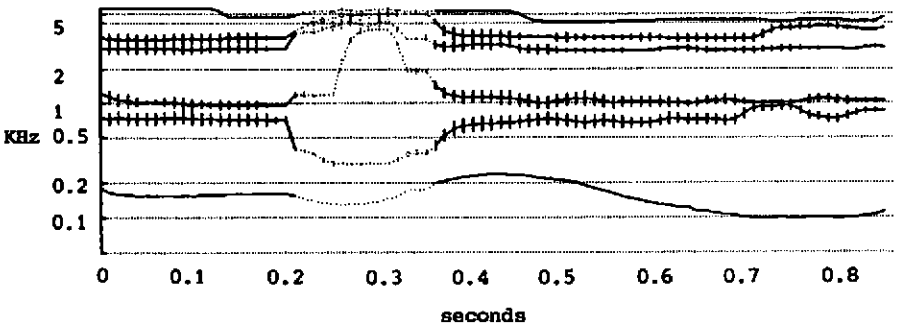
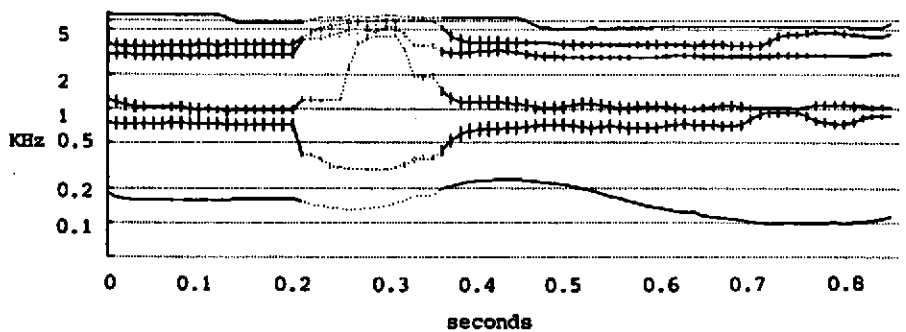


Figure 4. Diagrammatic representation of a synthetic token intermediate between [ɑ:sɑ:] and [ɑ:zɑ:], with no amplitude diminution before the fricative.



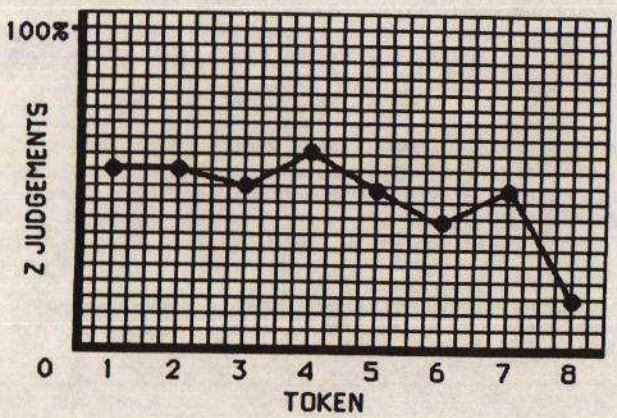
Two labelling tests were now carried out, with nine subjects:
(i) a repeat of the first test, replacing the two tokens which had been most randomly labelled with the two extremes of the eight new syntheses, i.e. the two syntheses represented in Figures 3 and 4;
(ii) a new test presenting twelve repetitions each of the eight new syntheses, in random order.

RESULTS

Those subjects who in test (i) perceived the token represented in Figure 3 mainly as [ɑ:sɑ:] and the token represented in Figure 4 mainly as [ɑ:zɑ:]

also produced appropriately sloping labelling curves on test (ii). A curve showing the average response of these six subjects is shown in Figure 5.

Figure 5. Labelling curve for 8 tokens with graded HFE diminution, averaged from 6 subjects who in a previous test labelled token 1 mainly as [ɑ:zɑ:] and token 8 mainly as [ɑ:sɑ:].



Those subjects who in test (i) heard both the tokens represented in Figures 3 and 4 (i.e. tokens 8 and 1 in Figure 5, respectively) as [ɑ:sɑ:], or as [ɑ:zɑ:], reacted randomly to test (ii).

With respect to the six subjects who did distinguish the two intermediate tokens in test (i), those who showed a bias towards [z] or [s] in test (i) did so in test (ii) as well.

CONCLUSIONS

We may conclude that, for some speakers at least, HFE diminution may be a sufficient cue to the voicelessness of a following fricative. It is important to note that those subjects who responded randomly to test (ii) did not respond randomly to the two extremes in test (i), but rather heard both as [s] or both as [z]. This indicates that the intermediate token on which test (ii) was based simply was not sufficiently intermediate for all subjects.

It can be seen from Figure 5 that the greatest perceptual break was between the token with maximum HFE diminution (token 8) and all the tokens with less diminution (tokens 1-7). Only in token 8 did the diminution stretch down into F1 (see Figure 3).

FUTURE WORK

We feel that the work described here can be extended in several ways:
(1) Stage 1 of the experiment should ideally be carried out more systematically,

in order to arrive at a token which is truly intermediate between voiceless and voiced, as a basis for the second part of the test. Since different subjects often use different cues in perceiving contrasts, more than one intermediate token (and therefore more than one test) will probably be necessary.

(2) Related to (1), the goal is to arrive at an intermediate token with is neither voiced nor voiceless. In the work described here, the intermediate tokens were biassed towards [s] by the HFE diminution which we imposed on them, and were biassed towards [z] by the rising Fx at the start of the following vowel. Consequently, these tokens were felt to contain both [s] and [z], sequentially: [ɑ:szɑ:]. Our results might be improved if the rising-Fx cue were to be replaced with, say, a slight degree of voicing during the fricative, or if the tokens had graded HFE reduction at both onset and offset of the fricative.

(3) The eight tokens in test (ii) were graded linearly with respect to the frequency range over which pre-fricative amplitude was reduced. Since the categorial break was between tokens 1-7 (less diminution) and token 8 (most diminution), the experiment might be improved by biassing the gradation of amplitude-diminution towards the lower end of the frequency range.

(4) Stimuli based on real words may well produce better responses from at least some subjects. The tokens used here were nonsense phonetic sequences, with the result that the subjects were not being asked to assign the stimuli to discrete lexical categories.

REFERENCE

[1] G. Lindsey, P. Davies and A. Fourcin, 'Laryngeal coarticulation in VCV sequences', Proceedings of the IEE Conference on Speech Input/Output, March 1986.