European Conference on
Speech Technology
Edinburgh, Scotland, UK
September 1987

10.21437/ECST.1987-107

COARTICULATION AND THE SYNTHESIS OF BRITISH ENGLISH

J B Pickering*

INTRODUCTION

It has been known for some time that the acoustic parameters of vowels and consonants change as a result of the immediate segmental environment (ref 1 & 2); and indeed, it has been claimed that an interconsonantal vowel is imposed on and only slightly modifies a continuous movement from the initial to the final consonants (ref 3). Whatever the extent of such coarticulatory effects, it is reasonable to expect that if we can model them this will contribute to the naturalness of synthesis-by-rule. This paper outlines current work to establish a set of coarticulation rules which are being implemented in the cosegmentation phase of the IBM experimental text-to-speech system (ref 4). We shall concentrate specifically on the analysis of syllable-final stops, which is an extension to the analysis of syllable-initial reported previously (ref 5).

DATA COLLECTION AND ANALYSIS

A single male RP speaker was recorded producing all phonotactically possible /CVC/ combinations. These were digitized and analysed using the speech processing facilities at the IBM UK Scientific Centre (ref 6 & 7). Formant values were tracked across the whole /CVC/ syllable and used to drive a PC/AT based formant synthesizer (ref 8) to check their validity. In an earlier study (ref 4), the initial /CV/s were analysed to establish (i) the duration of the transition from consonant to vowel, (ii) the temporal extent and (iii) the nature of any coarticulation, by picking four frames representing the consonant onset, the consonant offset, the vowel onset and the vowel steady-state. A similar method was used here to analyse the syllable final VOWEL+STOP combinations, again with four frames this time representing <1> the vowel steady-state, <2> the vowel offset, <3> the burst release of the consonant, and <4> the consonant offset. These four frames were located by visual inspection of both formant centre frequencies and formant amplitudes. The temporal separation of frames <2> and <3> was divided into the duration of the transition from the vowel into the stop closure and the duration of the closure itself.

We might not expect there to be any differences between syllable-initial and syllable-final consonants in view of Öhman's study (ref 3), if indeed the interconsonantal vowel is simply superimposed on an essentially diphthongal movement from the first to the second consonant. Further, although the perceptual importance of differences in the duration of the consonant closure is not entirely clear (ref 9 & 10), we would do well to consider this factor in generating synthesis parameters. Finally, we need to be able to model any coarticulatory effects that occur in the syllable-final /-VC/s. We shall, therefore, consider in turn (i) whether the formant values at the burst for syllable-final stops differ significantly from the burst for syllable-initial stops; (ii) whether the duration of the transition into the closure and of the closure itself are affected by the nature of the following consonant; and (iii) how coarticulation manifests itself across the syllable-final /-VC/.

SYLLABLE-FINAL vs SYLLABLE-INITIAL STOPS

Formant values were estimated at the burst release for syllable-final stops across all vowel contexts and compared with the corresponding values for syllable-initial stops using a

*Speech Research, IBM UK Scientific Centre, Athelstan House, St Clement Street, Winchester SO23 9DR.

standard *t-test* for correlated means. For the bilabials, the first formant of the [b] was significantly lower in syllable-final positions ($t[10]=41.645$, $p<0.001$), but no other formant differed significantly between syllable onsets and offsets. By contrast, both the first and third formant values were significantly lower in syllable-final positions for the alveolar stops ($t[10]=6.072$ and 4.836 for F1 of [t,d] respectively, $p<0.001$ in both cases; and 2.529 and 3.940 for F3, $p<0.025$ for [t] and $p<0.005$ for [d]). In the case of the velar stops, F1 is significantly lower for both [k,g] ($t[10]=4.241$ $p<0.001$ for [k] and 3.149 $p<0.01$ for [g]) and additionally the second formant is slightly lower for the voiceless stop only ($t[10]=2.446$ $p<0.025$). In general, therefore, significant differences between formant values for syllable-final stops do occur, usually for F1, but also for F3 in alveolar stops and possibly F2 in velars.

DURATION

Transition durations into the closure for vowels in all consonantal environments were measured. A three-way fixed effects ANOVA (vowel×voicing characteristic×place of articulation) revealed significant main effects for voicing ($F[1,20]=15.381$, $p<0.001$) and for place of articulation ($F[2,20]=4.703$, $p<0.025$), but no significant interactions. The transition into the closure is slightly longer when the vowel is followed by a voiceless stop (51.8 msec) than by a voiced stop (41.8 msec); further, bilabials are associated with the longest transitions (51.8 msec), while alveolars go together with the shortest (42.3 msec) for this speaker.

The *duration of the stop closure* was also measured and subjected to statistical analysis. A similar three-way fixed effects ANOVA yielded a significant main effect for voicing only ($F[1,20]=113.0$, $p<0.001$), and a significant interaction between voicing and place of articulation. Not surprisingly (cf. ref 9; see also ref 10), the voiced stops occurred together with a shorter closure duration (57.6 msec compared with 91.8 msec for the voiceless stops). But somewhat strangely, among the voiceless stops, the duration of the closure increased from bilabial via alveolar to velar consonants (78.2 msec, 95.5 msec and 101.8 msec respectively), whereas it decreased for the voiced consonants (70.91 msec for the bilabial, 52.7 msec for the alveolar, and 49.1 msec for the velar). It is not yet clear whether this second finding is a feature of the particular speaker used here. Further analyses will help clarify the matter.

COARTICULATION

Using a similar analysis to that in the earlier study on syllable onsets (ref 5), the following general tendecies emerged. At vowel offset, all three formants are the same as the steady-state vowel formants in velar environments. For the bilabials, the influence of the following consonant can be seen in the first formant and to some extent in the third, but the second formant is the same as the vowel steady-state. However, there are significant differences in all three formants at vowel offset in alveolar contexts: alveolar consonants influence the formant structure of preceding vowels. At the burst release, there are poor correlations between formant values for the bilabials and the vowel steady-state suggesting that the vowel may still influence the formant structure of the consonant. For the velars, F1 and F3 show no influence from the perceding vowel. However, F2 shows a strong dependence on the vowel. This is not surprising given that the place of articulation of these consonants is said to be particularly sensitive to the preceding vowel. In the case of the alveolar consonants, there is no apparent influence from the preceding vowel. Finally, after the release of the consonant, all three consonant types show some influence from the preceding vowels, particularly in the second formant. Inasmuch as these were fully released, it may be that the release burst is superimposed in Öhman's terms (ref 3) onto a diphthongal movement from the syllable nucleus to the aspiration after the burst release.

CONCLUSION

In this paper, we have considered some of the coarticulatory effects in syllable final VOWEL+STOP combinations in British English. Assymmetries were found in the formant parameters for syllable-final and syllable-initial stop consonants at the burst release, which usually resulted in the lowering of one or more formant values at the end of the syllable. But a similar pattern of coarticulatory effect was found for syllable-final stops as occurred for syllable-initial stops: the vowel offset was rather unsensitive to influences from bilabial or velar stops, but in turn the burst release for these consonants showed the effects of the preceding vowel. By contrast, the alveolar stops had some influenced on vowel parameters at vowel offset, and were less sensitive to the influence of the preceding vowel at the burst release. This was also the case for syllable-initial alveolar stops. Finally, the duration of the transition from the vowel into the stop closure and of the stop closure itself were highly sensitive to the voicing characteristics and place of articulation of the following consonant, with voiced consonants being associated with shorter transition durations and shorter closures. These f&i.ndings are now being implemented as part of the IBM experimental text-to-speech system (ref 4) with encouraging results.

REFERENCES

[1] STEVENS, K.N. & HOUSE, A.S. (1963) 'Perturbation of Vowel Articulation by Consonantal Context: An Acoustical Study', *J. Speech Hearing Res.* (6) 111-128

[2] LINDBLOM, B.E.F. (1963) 'Spectrographic study of vowel reduction', *J. Acoust. Soc. Am.* (35) 1773-1781

[3] ÖHMAN, S.E.G. (1965) 'Coarticulation in VCV Utterances: Spectrographic Measurements', *J. Acoust. Soc. Am.* (38) 151-168

[4] SHARMAN, R.A. (1986) 'Designing an Experimental Text-to-Speech Synthesiser', *Proc. Int. Acoustics* Vol. 8 Part 7 355-362

[5] PICKERING, J.B. (1986) 'Cosegmentation in the IBM Text-to-Speech System', *Proc. Int. Acoustics* Vol. 8 Part 7 387-392

[6] ALDERSON, P.R., KAYE, G., LAWRENCE, S.G.C. & SINCLAIR, D.A. (1985) 'SAY - Speech Analyser', *UKSC Report 135*

[7] 'Speech Processing at the UKSC', *Internal IBM UKSC Report*

[8] SINCLAIR, D.A. (1987) 'A Versatile Speech Synthesiser based on an IBM Personal Computer', *Inst. Acoustics Speech Group Meeting* Applications of Speech Technology, May 1987.

[9] RAPHAEL, L.J. (1981) 'Durations and Contexts as Cues to Word-Final Cognate Opposition in English', *Phonetica* (38) 126-147

[10] LUCE, P.A. & CHARLES-LUCE, J. (1985) 'Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production', *J. Acoust. Soc. Am.* (78) 1949-1957

[11] FISCHER-JORGENSEN, E. (1969) 'Untersuchungen zum sogenannten festen und losen Anschluß', *Kopenhagener germanistische Studien I (Festskrift Peter Jorgensen)*, 138-164