Effects of Anticipatory Coarticulation on Vowel Detection in Meaningful Words

V. J. van Heuven and M. Ch. Dupuis
Leyden, the Netherlands

1. Introduction

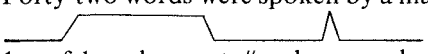
Acoustic and perceptual consequences of coarticulation have been extensively investigated for CV and VC structures in meaningless syllables, using either human or synthetic (pattern playback) speech. Generally, in VC syllables the place of articulation of C is clearly perceptible in the preceding V, and such characteristics as lip rounding, vowel height and backness are audible in the initial C of a CV syllable. Moreover, perceptual effects may not be confined to neighbouring sounds: the formant frequencies and transitions of V_1 in a V_1CV_2 sequence are systematically affected not only by C but also by V_2 (Öhman, 1966; Lehiste and Shockey, 1972). In spite of several attempts, however, it has proven very difficult to ascertain whether such anticipatory coarticulation between such non-adjacent sounds enables the listener to recover the identity of V_2 (or some feature of it) from an earlier portion of the utterance if both V_2 *and* the preceding C are (electronically) removed from the stimulus (cf. Lehiste and Shockey, 1972; Benguerel aud Adelman, 1975). This would seem to suggest that useful coarticulation information does not extend beyond adjacent sounds.

Recently, Martin and Bunnell (1981) were able to show that vowel detection latencies for V_2 were slightly (9 ms) faster with a properly coarticulated V_1 than when V_1 contained acoustically conflicting information after cross-splicing. However their stimuli were meaningless, and phonologically illegal structures of the type /kae'zi:, kae'za:, ku:'zi:, ku:'za:/, with a non-reduced first vowel, and stress on the second. It would be of interest to see if more convincing effects can be found across word boundaries rather than word internally using meaningful and phonologically legal words.

The present experiment assumes that the final portion of a word contains information that enables the listener to predict some or all properties of the beginning of the next word, i.e. the word initial consonant(s) and possibly even the following vowel. We wished to test the hypothesis that such feed-forward information facilitates the identification of the following vowel (V_2) in continuous utterances. Our approach was to compare two types of stimuli: one with the crucial word in its original coarticulated environment, and another in which the vowel immediately preceding the crucial word (V_1) had been replaced by a noise burst.

2. Method

Forty-two words were spoken by a male speaker of Dutch in the fixed carrier



1. ət fəlγəndə wor:t #ɪz də γəwərdə

(Het volgende woord is de ... geworden, 'The next word has become the ...'), i.e. preceded and followed by unstressed syllables containing schwa. The crucial words were disyllabic and with initial stress, which was consistently realised with the same accent-lending pitch movement (see stylized pitch trace in (1)). Six stressed vowels (i:, y:, u:, e:, o: a:) were combined with 6 initial consonants p, t, k, b, d, or no C). Finally, 6 three-syllable words beginning with V were added to the stimulus set.

The 42 stimulus sentences were recorded per target vowel in blocks of 7, interspersed in a quasi-random fashion with an equal number of foils (drawn from a pool of 42 two or three-syllable words containing any of the Dutch non-target full vowels or diphthongs). The schwas immediately preceding the crucial words were excised from the utterances and replaced by pink noise with an amplitude equal to that of the removed vowel (10 ms rise/decay time). In the control condition the schwas were left intact, but now the vowel /I/ in *is* was replaced by noise. Stimuli were sampled from both conditions (coarticulation removed *vs.* coarticulation intact) in equal numbers. Two test tapes were prepared such that when coarticulation intact) in equal numbers. Two test tapes were prepared such that when coarticulation was removed on tape I, it was left intact on tape II, and *vice versa.* As a result each tape contained each word only once.

Forty-one Dutch listeners were instructed to press a button as soon as they heard a particular target vowel. After each block of 14 stimuli they were informed of the change in target.

3. Results

Subjects scoring more than 4 errors (misses or false alarms) were eliminated from further analysis, leaving 2 balanced groups of 14. The results showed that one group (tape I) had significantly faster detection times than the other (tape II) (310 *vs.* 493 ms). To redress subject variability, the data were Z-normalised for individual means and standard deviations. Figure 1 plots normalised vowel detection latency as a function of the initial C (panel A), and as a function of the stressed vowel (panel B). A second abscissa provides a rough translation of Z-scores to milliseconds, on the basis of a grand mean of 401 ms and an average standard deviation (= Z) of 171 ms per subject.

Detection latencies do not differ for any of the 6 target vowels when coarticulation is preserved, $F(5,571)=2.1$ (ins.). Removal of coarticulation has the overall effect of slowing detection down by .16 Z (or 27 ms),

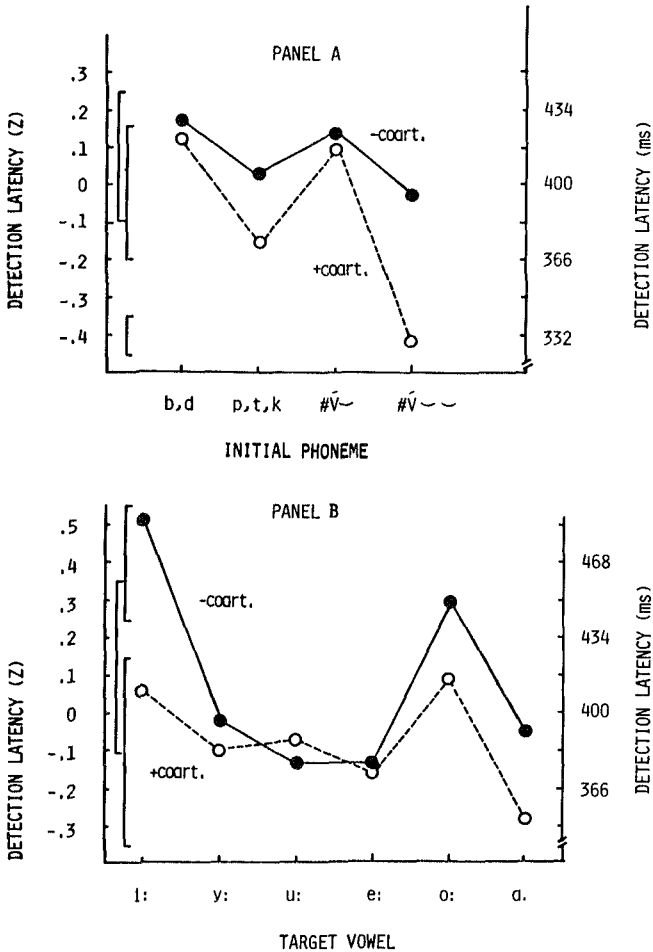


Figure 1 Z-transformed detection latency (normalised per subject for individual mean and standard deviation) as a function of the target vowel (panel B) and type of initial phoneme (panel A), separated out for removed vs. intact coarticulation with the preceding word. Z-scores have been reconverted to milliseconds along the right-hand axis. The ordinate does not represent continuous variables Braces enclose means that do not differ significantly from each other by a Newman-Keuls test with $p < 05$.

$F(1,1155)=7.9$ ($p=.005$), but not all the vowels are affected to the same extent. Specifically, /i:/ and /o:/ are now significantly slower than the other vowels in the same condition, $F(5,574)=6.8$ ($p<.001$), and in fact slower than any of the other vowels in either coarticulation condition, $F(11,1145)=4.8$ ($p<.001$). For /a:/, /y:/, and /e:/ the effect of coarticulation is in the predicted direction, but it is too small to reach significance. Finally, for /u:/ the effect is in the wrong direction, but again insignificantly so.

It is apparent from panel A that vowels are detected sooner after voiceless plosives or initially in three-syllable words than after voiced plosives or initially in two-syllable words, $F(3,1153)=6.5$ ($p<.001$).

4. Conclusions and Discussion

First and foremost, our data support the hypothesis that target vowels are detected earlier when anticipatory coarticulation is provided in the preceding syllable (word), even across an intervening consonant.

One might object, of course, that the longer latencies in the coarticulation removed condition are simply due to the disturbing influence of the noise burst immediately preceding the target word. Had this been the case, however, the delay should have been the same for each of the 6 target vowels, which clearly it was not. Similarly, its effect should have been more disruptive for targets in initial position (i.e. abutting the noise) than for targets separated from the noise by a consonant. Again this effect does not obtain (cf. panel B).

As a final point we want to speculate on the origin of the difference in detection latency for initial vowels in two as opposed to three-syllable words. Why would targets in the longer words be detected sooner than in the shorter words? Explanations on the basis of word-frequency differences can be ruled out, as both types were selected from the low frequency brackets in the Dutch lexicon. Neither do the two types differ with respect to the point at which the word can be uniquely distinguished from all other words in the lexicon: both two- and three-syllable words have their theoretical recognition points (cf. Marslen-Wilson and Welsh, 1978) after 4½ phonemes, on average. Therefore, we reason that the effect must have a phonetic origin. We know that in Dutch the duration of a long vowel in an initial stressed syllable is 20 to 30 ms shorter in a three-syllable word than in a two syllable word (Nooteboom, 1972). It might thus be reasonable to assume that our subjects delay their decision until they have heard the end of the target vowel, which comes earlier in the three-syllable words (all else being equal). We have re-analysed our data, using target offset rather than onset as the reference point for the latency measurements, and again normalising the new results for individual means and standard deviations. Generally, having eliminated an uncontrolled source of error, residual variance in the data should now decrease, the statistical significance of all effects should go up, except for those that are contingent on vowel duration. Error variance does indeed go down, viz. from 1079 to 1035, and the differences between the target word onset conditions are substantially reduced, but fall short of statistical significance, $F(3,1153)=2.4$ (ins.). Counter to the prediction, however, the differences between the 6 target vowels increase, especially between the phonetically long vowels /e:/, /o:/, /a:/ on the one hand, and the phonetically shorter vowels /i:/, /y:/ and /u:/. Possibly, therefore, identification of the shorter vowels is delayed until the end of the vowel, but may take place at some earlier point for the longer vowels.

Acknowledgement

Research supported by a grant from the Foundation of Linguistic Research, funded by the Netherlands Organisation for the Advancement of Pure Research (ZWO) under project nr. 17-21-22. The experiment was carried out by D. Lemoine and P. van den Mortel.

References

Benguerel, A P., Adelman S. (1975) Coarticulation of lip rounding and its perception In. A Cohen, S Nooteboom (eds.), *Structure and process in Speech perception,* Berlin. Springer, 283-293.

Lehiste, I., Shockey, L (1972) On the perception of coarticulation effects in English VCV syllables *Journal of Speech and Hearing Research,* **15**, 500-506

Marlsen-Wilson, W.D., Welsh, A. (1978). Processing interactions and lexical access during words recognition in continuous speech. *Cognitive Psychology,* **10**, 29-63

Nooteboom, S G. (1972) Production and perception of vowel duration. a study of durational properties of vowels in Dutch. Doct. Diss, University of Utrecht

Martin, J.G , Bunnell, H T. (1981). Perception of anticipatory coarticulation effects *Journal of the Acoustical Society of America,* **69**, 559-567.

Öhman, S E.G. (1966) Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America,* **39**, 151-168