The Role of Coarticulation in the Identification of Consonants

L.E.O. Klaasen-Don and L.C.W. Pols
Soesterberg and Amsterdam, the Netherlands

The study of the articulatory and acoustic properties of speech has failed to reveal basic segments at the phoneme level. On the contrary, every phonemic realisation reflects the influence of properties of preceding and/or following phonemic environments. Thus on the one hand there is the notion that 'pure', 'basic form' speech segments may exist (e.g. Daniloff and Hammarberg, 1973), whereas on the other hand it is found that information present in the speech sound is coarticulated (e.g. Bell-Berti and Harris, 1982). The concept of articulation assumes that, during (the process of) speech, basic forms, usually phonemes, are transformed in such a way that their articulatory target values become modified due to the interaction with properties of contiguous phonemes. Of course, some overlap in articulatory movements is inevitable, given that the speech organs are not capable of infinite acceleration. However, what makes coarticulatory influences interesting is that their explanations go beyond simple inertial factors, although inertia must play some role. Coarticulatory influences have been found that reflect planning in motoric programming. It seems that the speech apparatus can make preliminary adjustments for different phonemes. Obviously, the system of motor control has information about several phonemes at once. Perhaps the range of phonemes over which simultaneous information is available defines the organisational scheme of speech articulation.

Research concerning coarticulation has mostly been concentrated on the articulatory aspects of the phenomenon. Nowadays, several studies also deal with its perceptual aspects. In analogy with studies on the articulatory aspects the perceptual studies are particularly interested in the range of speech over which simultaneous information about different phonemes is available, related sometimes to perceptual units. Usually, these studies are confined to a few sounds, like plosives, and use synthetic speech or carefully pronounced utterances.

The present study investigates the relative contribution of vowel transitions to the identification of *all Dutch* consonants, both in utterances pronounced in isolation and in excerpts taken from conversational speech. By means of speech editing, parts of these utterances are deleted and listeners have to identify the (absent) consonant on the basis of information in the remaining part of the utterance. Figure 1 presents an example of the relevant CV part of the waveform of a CVt utterance /na:t/ pronounced in isolation,

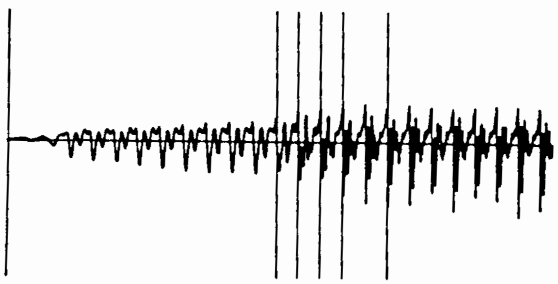

Figure 1. Example of the relevant CV part of the waveform of /na:t/. The vertical lines mark the segmentation points which define the 6 segments.

with the vertical lines marking 6 segmentation points by means of which 6 segments could be defined. Segmentation point 1 left the whole utterance intact and 5 other segments could be made from this CVt utterance by removing more and more of the vocalic transition.

Analogous to spectral analyses, the identification results show that presentation of the stationary vowel part alone does not contain enough information to identify adjacent consonants. However, the initial part of the vowel transition in CV syllables and the final part in VC syllables do contain some information about the adjacent consonant. The amount of information and its extension into the vowel transition differ widely for the various articulatorily defined groups of consonants. In utterances spoken in isolation voiced plosives, liquids and semi-vowels can be identified rather well on the basis of the vowel transition only, for nasals and unvoiced plosives this is more difficult, and for fricatives it is impossible. Thus it seems that Dutch initial voiced plosives on the one hand and fricatives on the other hand reflect extreme groups in this respect. Results of experiments (Don and Pols, 1983) using fast reaction times to detect and classify the consonant and vowel in CV syllables, reveal that this may be related to the fact that voiced fricatives are identified *earlier* than the following vowel, whereas voiced plosives are identified later than the following vowel. Moreover, fricatives are identified *earlier* than plosives, since in CV syllables containing a fricative, identification can start right from the physical beginning of the utterance, whereas in CV syllables containing a plosive identification is hindered by the vowel murmur preceding the burst.

A very important aspect in the study on coarticulatory effects is the direction of the perceptual influence. Is it more advantageous to have additional cues about a consonant before than after it is heard? The literature is not clear on this point. Most studies show that anticipatory influences are either dominant or at least equal to carryover influences. Results of the present study show that Dutch voiceless plosives are slightly better identified from VC than from CV vowel transitions. For voiced plosives and nasals the situation is reversed, with nasals being very much better identified from CV

than from VC vowel transitions. Broadly speaking, we can say that some consonants are better identified on the basis of vowel transitions in CV syllables, whereas others are better identified from vowel transitions in VC syllables. No universal dominance can be found of anticipatory or carryover influences.

Information about the consonant in CV vowel transitions did not differ according to whether the segments were isolated from CVC or VCV utterances. However, some consonants are better identified from vowel transitions in VC segments excerpted from CVC utterances, whereas others are better identified from vowel transitions in VC segments excerpted from VCV utterances.

For excerpts from *running* speech the relative behaviour is similar, but the absolute scores are much lower.

There are various explanations for the observed perceptual interactions between vowels and consonants. One has to do with a more or less fixed time span over which the perceptual mechanism integrates. Another supposes a functional symbiosis between production and perception in such a way that whenever segments are coarticulated they are also perceived in a dependent way. The actual confusions made, which are often related to place of articulation, point in this direction.

For utterances pronounced in isolation, results of the present study show that perceptual information in the vowel transition about the contiguous consonant is in fact redundant since the consonant segment alone already causes almost perfect consonant identification.

In conversational speech consonant parts are often not well articulated or masked. Moreover, the durations of the steady-state vowel segments become shorter and the durations of the vowel transitions become relatively longer than in utterances pronounced in isolation (Schouten and Pols, 1979). For these reasons, we may expect the vowel transition to become relatively more important for consonant identification in utterances excerpted from running speech. The fact, however, that vowel transitions do not seem to contain much information about the adjacent consonants in conversational speech rejects this hypothesis and is an indication of reduced redundancy at the acoustic-phonetic level, which is probably compensated for by additional information at the lexical, syntactic, and semantic levels (Marslen-Wilson and Tyler, 1980).

Future research on 'coperception' should concentrate, more than has often been done in recent studies, on fundamental units of speech perception. If e.g. a specific form of a syllable (VCV,CV or VC) is found to be basic to speech perception, perceptual influences among speech segments can then be explained in terms of the perceptual organisation of the syllable cycle. Besides, studies on units of speech perception also allow one to study the relation between perceptual units and units of speech production.

Acknowledgement

This research was supported by grant 17-21-11 from the Foundation for Linguistic Research funded by the Netherlands Organisation for the Advancement of Pure Research (Z.W.O.).

References

Bell-Berti, F. and Harris, K.S. (1982). Temporal patterns of coarticulation. *J. Acoust. Soc. Amer.*, **71**, 449-454.

Daniloff, R.G. and Hammarberg, R.E. (1973). On defining coarticulation. *J. Phon.*, **1**, 239-248.

Klaassen-Don, L.E.O. and Pols, L.C.W. (1983). Dependencies of processing of consonants and vowels in CV syllables. Paper presented at the 11th International Congress on Acoustics, Toulouse.

Marslen-Wilson, W. and Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, **8**, 1-71.

Schouten, M.E.H. and Pols, L.C.W. (1979). Vowel segments in consonantal contexts: A spectral study of coarticulation - Part I. *J. Phon.*, **2**, 1-23.