'COARTICULATION' IN AN ARTICULATORY SYNTHESIS MODEL OF GERMAN

GEORG HEIKE

Institut für Phonetik
Universität zu Köln
Greinstraße 2, D-5000 Köln 41

ABSTRACT

'Coarticulation' is the main problem in speech synthesis. In the case of German we show that the control of articulatory parameters is dynamic in nature, i. e. depends on effort and time of articulatory gestures.

The purpose of this contribution is the development of a framework in which 'coarticulation' rules for the articulatory synthesis of German can be established. Starting point is the conclusion that the traditional concept of 'coarticulation' must be rejected as inadequate because it presumes discrete phonetic segments as input units into a coarticulatory module. Therefore a target oriented model of articulatory control is proposed. Input units to the control module are labelled by phonetic symbols. They are defined by at least one target value of one parameter (e. g. for bilabials) or more parameters (in the case of most other sounds).

With the exception of the bilabial (rounded) [ʃ] German consonants are defined by one target value only, namely the constrictional position of the lips, the glottis, the anterior part of the tongue or the dorsum. The remaining articulatory configurations, for example the shape of the lips and of the dorsum in the case of an apico-dental consonant, have to be specified according to the syllabic context. This specification, usually termed as 'coarticulation', is language dependent and has to be formalized in an articulatory synthesis model. Since there is a lack of sufficient experimental investigations (especially x-ray studies) in the case of German, our method is restricted to the articulatory interpretation of sonagrams, self-observation, and auditory control of synthesis output.

Preliminary results suggest the hypothesis that the complete articulatory specification of German consonants depends on the factors: vowel context, type of consonant, position within the syllable, speed of articulation. These dependencies will be exemplified in the case of the apico-dental [l] and the dorso-velar (or palatal) consonant [k]. In an initial prevocalic position (e. g. [li:] as in 'Liebe') the most economic (and hence 'coarticulated') position of the tongue would be the same as for [i:] except for the elevation of the tongue tip. This would result in a 'l' with palatalized dorsum. Although German is not said to be characterized by such consonants, the above-mentioned case can be observed in fluent speech, especially in intervocalic position, as e. g. in 'die Liebe'. There is, however, a remarkable difference in the 'coarticulative' effect of the vowel context between slow and fast articulations. In the case of relatively slow articulation - which means slower movement of the tongue tip and greater duration of dental contact - there is enough time for the tongue body to move towards the neutral position. The same principle holds for stop consonants, but since closure and release gestures are clearly separated, relatively fixed in time, and hence independent of speech tempo, a very distinct difference between syllable initial and final position can be observed. The articulatory position of release, e. g. in 'lieg' [li:kʰ], may result from a backwards movement of the dorsum during the closure time interval, whereas in 'Kiel' [kʰi:l] the dorsum must in any case take the appropriate palatal position for the following [i].

Fig. 1a shows the sonagram of the VCV-portion of a relatively careful pronunciation of 'die Liebe'. An appropriate articulatory synthesis of the VCV-gesture can be achieved with a tongue profile of l with central position of the dorsum, as midsagittal tracings (fig. 1c) of i and l and the sonagram of the synthesis output (fig. 1b) show.

Fig. 2 presents, in a case of more rapid pronunciation, both reduction of the unstressed i in 'die' and the palatal position of the tongue dorsum for l, largely identical with that of i. Thus

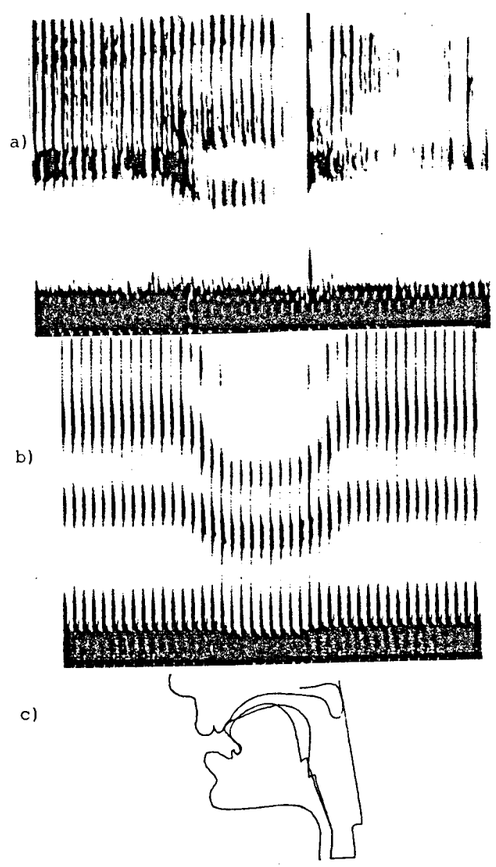

Fig. 1: Spoken (a) and synthesized (b) VCV-portion [ili] of German 'die Liebe', relatively slow speech. Midsagittal computer tracings (c) show corresponding target positions

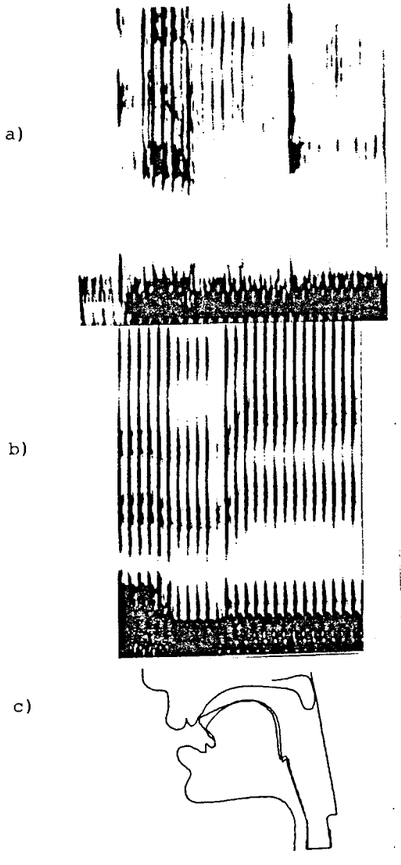

Fig. 2: The same VCV-transitions as in fig. 1, spoken with relatively rapid speech tempo

we may hypothesize that the control of the tongue dorsum as a function of speech tempo works differently for vowels and consonants. We observe reduction (towards neutral) of vowels with high tempo, whereas with consonants a similar effect appears with slow tempo.

But both phenomena can be explained by one principle: the economy of effort as a function of time. Effort may be defined in two respects: as the effort of reaching a target, and as the effort of maintaining a position different from neutral. With high speech tempo reduced effort results in an 'undershot' of the movement towards a target (vowels), with slow tempo the effort of maintaining an extreme position of the dorsum (unnecessary for consonants) is reduced and compensated by a movement to a central position. The same principle holds for the parameter of lip rounding. In the sequence [uli], for instance, we notice (with normal-to-slow speech tempo) a lip spreading gesture from u to l and vice versa. In synthesizing ['uli] ('Uli') we would therefore expect the spreadirg gesture of the l to be continued in the transitional movement towards i (see fig. 3). Comparison with the rather rapidly spoken utterance shows not only a remarkably reduced u (more centralized and less rounded), but also an anticipation of the spreading gesture within the duration of the vowel. This demonstrates

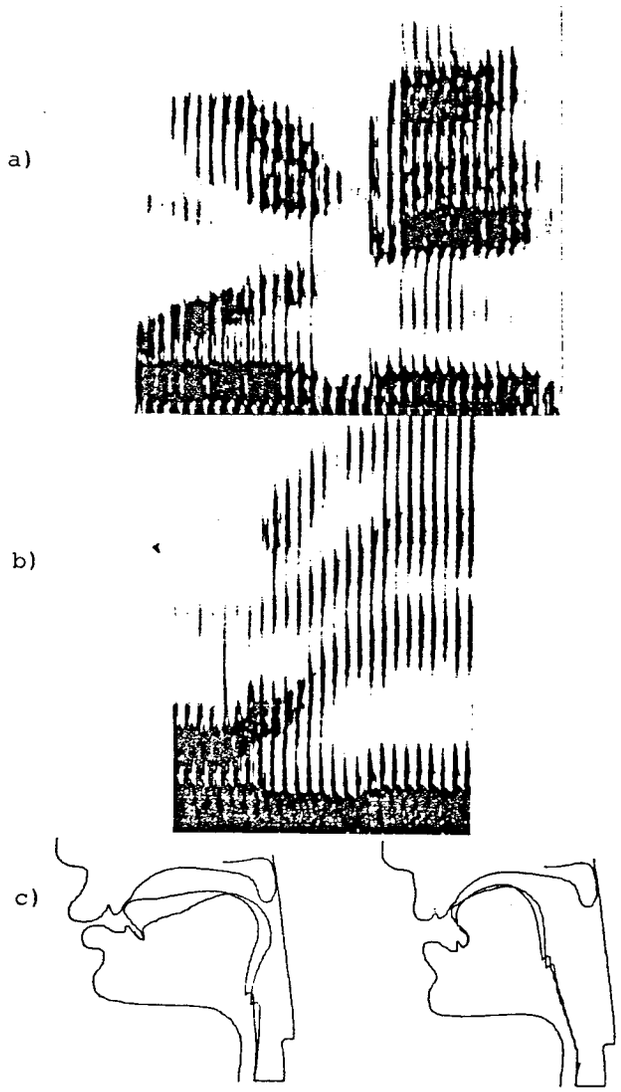

Fig. 3: VCV-transitions of 'Uli';
a) spoken,
b) synthesized without anticipation of lip spreading.

the usefulness of articulatory synthesis for the study of 'coarticulation' phenomena (i. e. articulatory control as a function of time).