Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Perceptual evaluation of coarticulation effects

Fant, G. and Liljencrants, J. and Malác, V. and Borovicková, B.

KTH Computer Science and Communication

journal: STL-QPSR
volume: 11
number: 1
year: 1970
pages: 010-013

http://www.speech.kth.se/qpsr

II. SPEECH PERCEPTION

A. PERCEPTUAL EVALUATION OF COARTICULATION EFFECTS

G. Fant, J. Liljencrants, V. Maláč*, and B. Borovičková*

Abstract

From spectrographic studies of VCV utterances, e.g. those of Öhman (1966), it is apparent that formant transitions and especially the course of F_2 within the final vowel displays considerable variations depending on which initial vowel it is paired with. Symmetric VCV structures accordingly differ systematically from unsymmetrical VCV structures. Thus, the transition after a labial stop [p] or [b] which usually is rising can in some contexts be found to be falling. As a consequence two VCV words with the same vowel frame and differing in terms of the consonant can show the same transitional pattern in the CV part of the word. According to Öhman (1966) the VC part would then provide the discriminatory cues. The purpose of our study is to test this assumption as well as the general perceptual importance of the VCV coarticulation for the identification of the stop sound.

The general explanation of the VCV coarticulation phenomena can be derived from the two-channel model of speech production control postulated by Borovičková and Maláč (1967) and by Öhman (1967) which implies that any formant transition can be regarded as the sum of two superimposed curves, one for the consonant control channel and one for the vowel control. The falling F_2-transition after the release of b in [ibu] is accordingly ascribed to the i-u component being larger and opposite in sign to the b-u component. Fant (1969) has recently pointed out that the time constant of the consonantal component, e.g. that of labial stops, can be much faster than that of the vowel component in the interval of 10-30 milliseconds immediately after the instance of release and that the transition generally observed and traced from a spectrogram is dominated by the course of the vocalic transition well after the critical initial part of the transition. Thus, perceptually important cues may be lost in the overall transition trace. Other cues specific of the first 10-30 msec after the release of a velar stop is the presence of a strong burst in the F_2-domain and a relative lower F1 intensity of the g-burst compared with the corresponding part of the b-release.

---

* At the Dept. of Speech Communication May 16-June 10, 1969.

As a result of the present study it was found that the consonant identification remained 100 % after removal of the initial vowel of real speech VCV test words. In synthetic speech samples the average identification scores were lower and the effects of removing the initial vowel could not be given a simple interpretation since the intact VCV only in some combinations, e.g. voiced labials without burst, gave a higher identification than the CV part alone. It was also found that CV stimuli gated from unsymmetrical VCV words were identified better than CV stimuli from symmetrical VCV words only in case of voiced velars without burst. Our main conclusion is that the release phase of the stop carries the main cues of consonant place of articulation but that these cues were not sufficiently well preserved in the synthetic stimuli. The overall vocalic coarticulation is of secondary importance and contributes to place identification incidentally only.

The experiment

In order to ensure maximally sensitive conditions for consonant confusions and thus of the possibility of the initial vowel acting as a discriminatory cue we selected the test words [ibu] and [igu] which, according to Öhman (1966), have approximately the same CV parts. Since the listening test was to be carried out with Czech subjects we modeled the synthetic spectra after Czech data, Borovičková and Maláč (1967). In all 32 synthetic stimuli words were generated. These included the five binary selections of labial/velar, voiced/unvoiced, burst/no burst, with/without initial vowel, and finally a time course of F_2 typical of that of symmetrical versus unsymmetrical syllables. The latter categorization is illustrated in Figs. II-A-1 and II-A-2. Here "symmetrical" implies that the V_1CV_2 is constructed with formant transitions of V_1C typical of V_1CV_1 syllables whilst the CV_2 part is made to conform with the latter half of V_2CV_2 words.

The "unsymmetrical" or correctly coarticulated V_1CV_2 words were constructed as modifications of the "symmetrical" patterns. A phonetic interpretation of an "uncorrected" or symmetrical [igu] syllable is accordingly a palatal occlusion followed by velar release of the consonant.

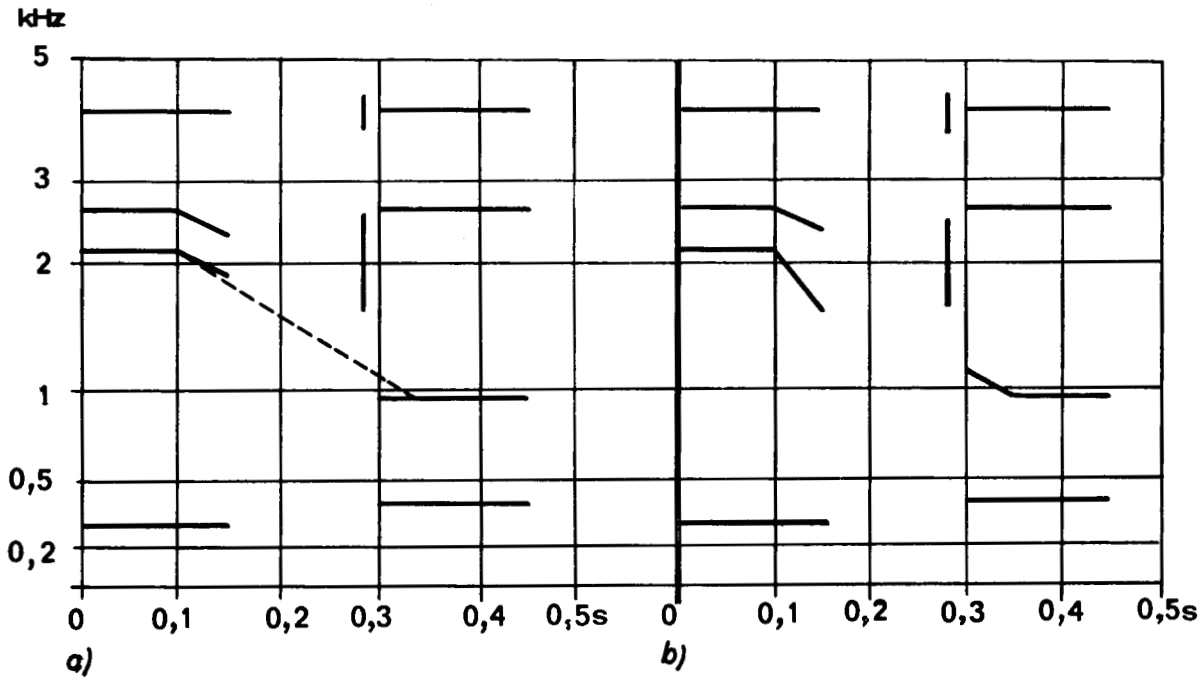

Fig. II-A-1. Simplified spectrum of [ipu]:
a) with separate VC and CV transition,
b) with the correct "asymmetrical" F2-transition.

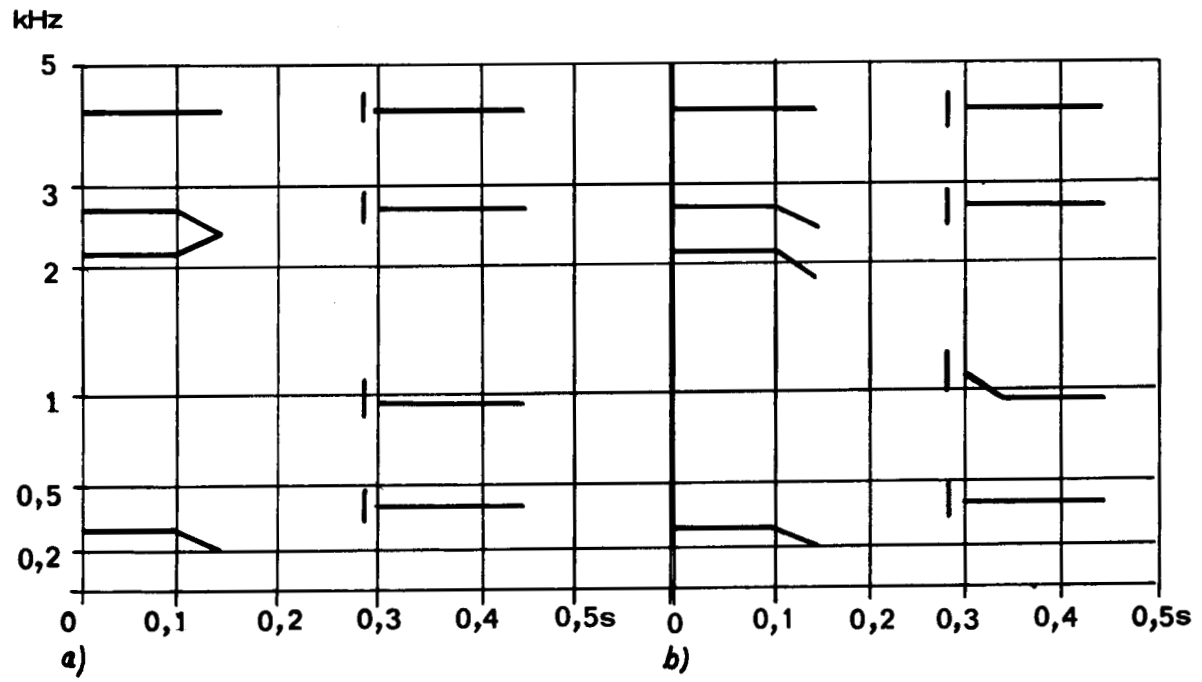

Fig. II-A-2. Simplified spectrum of [iku]:
a) with separate VC and CV transition,
b) with the corrected F2-transition.

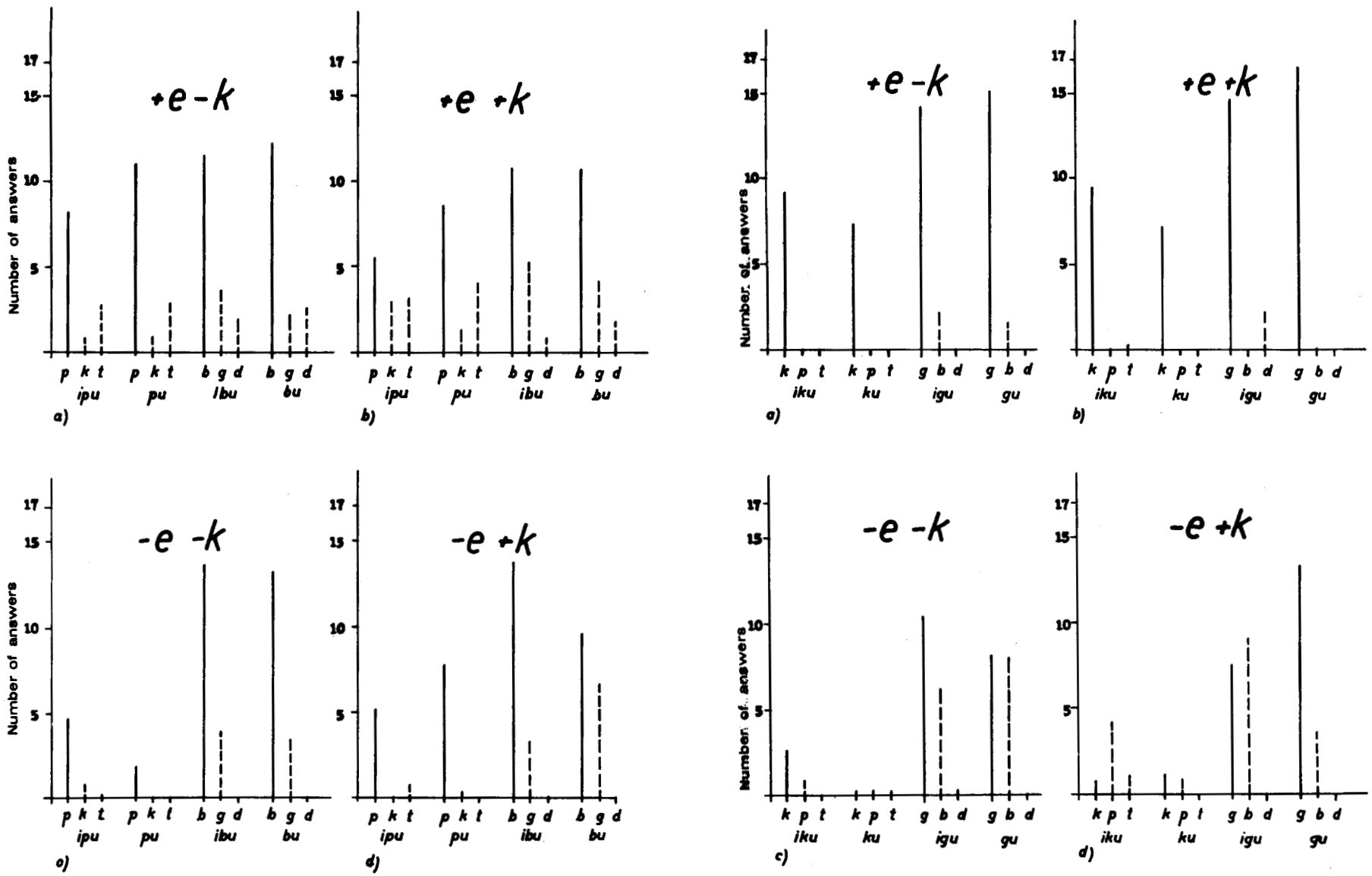

Fig. II-A-3 and II-A-4. Histograms of average response of listeners to the synthetic sound samples (the second row under abscissa). Type of individual stops are marked in the first row under abscissa. Full lines show the correct answers, dashed lines the confusions. The letters above the histograms mark the type of sample variant. The notation +e indicates presence of burst, +k indicates corrected, "unsymmetrical" F_2-pattern.

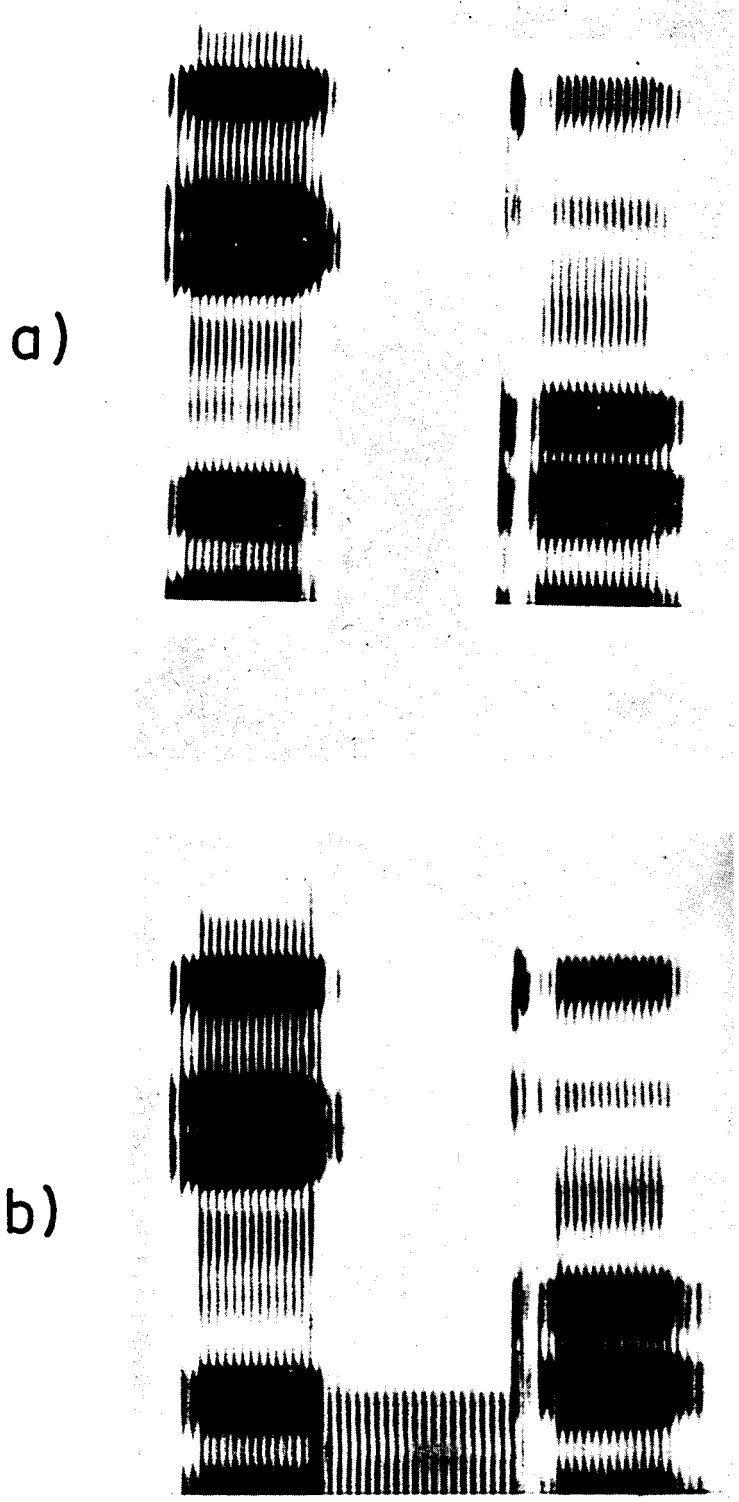

Fig. II-A-5. Sonagrams of synthetic OVE III, a) [iku] b) [igu]. Note that the formants between F_2 and F_3 are artefacts due to some overloading phenomenon in the spectrographic analysis.

The burst was generated with the A_h control of noise into the F-filter of the synthesizer. This is representative of the aspirative rather than of the transient velar burst phase. It would probably have been more representative to produce the velar burst as a single formant A_c-controlled K-filtered noise.

Voicing was introduced by means of a low F1 "voice bar" in the occlusion. The length of the occlusion as well as of the initial and final vowel was 150 msec in all combinations.

The set of stimuli from natural speech comprised the syllables [ipu][ibu][iku][igu] and the [pu][bu][ku] and [gu] parts of these syllables plus [pu][bu][ku][gu] out from [upu][ubu][uku][ugu] syllables, in all 12 test items.

The listening tests were carried out with a crew of 17 Czech subjects seated in a special test room. A high quality Tesla loudspeaker was used for the stimuli presentation. The 44 samples were randomized and mixed with 45 samples of another series and presented in subsets of ten stimuli with four seconds stimuli spacing.

Results and discussion

The general conclusions that can be drawn from the test results, Figs. II-A-3 and II-A-4, are the following:

(1) The complete unsymmetrical coarticulation provides an improved consonant identification compared with the symmetrical items only in the [gu] syllables without burst, whilst the opposite effect is found in [bu] syllables.

(2) The unsymmetrical VCV (+k) words were not better identified than the symmetrical VCV (-k) syllables. With the CV parts of the burst-free samples the [gu] cut from the +k words received a higher score than the [gu] from the -k words, whilst the opposite was found for [bu].

(3) The removal of the initial vowel impaired the identification of the consonant from [b] only whilst the consonant [g] was better identified in the CV than in the VCV context providing the coarticulation was made asymmetrical (+k).

(4) The presence of a burst enhances the [g] identification.

(5) All test stimuli derived from natural speech, including those with initial vowel removed were correctly identified.

The above-mentioned trends can be interpreted as a lower F_2-locus favoring [b] responses and a higher F_2-locus favoring [g] responses, everything else being equal. There is no general effect of overall transconsonantal vocalic coarticulation improving the consonant identification. Some of the trends observed might be explained by incidental differences in the synchrony between voice amplitude control and the time of arrival of the first pitch pulses of the final vowel.

A closer study should be made of burst and transition cues in the release phase of the stop as suggested by Fant (1969) and the synthesis strategy and program should be modified accordingly in further tests.

References:

BOROVIČKOVÁ, B. and MALÁČ, V. (1967): "The spectral analysis of Czech sound combinations", ČSAV, Vol. 77, No. 14.

FANT, G. (1969): "Stops in CV-syllables", STL-QPSR 4/1969, pp. 1-25.

ÖHMAN, S.E.G. (1966): "Coarticulation in VCV utterances: Spectrographic measurements", J. Acoust. Soc. Am. 39, pp. 151-168.

ÖHMAN, S.E.G. (1967): "Numerical model of coarticulation", J. Acoust. Soc. Am. 41, pp. 310-320.