On the Perception of Coarticulation Effects in English VCV Syllables

Ilse Lehiste and Linda Shockey

Abstract

Öhman's (1966) investigation of the acoustic correlates of coarticulation in VCV sequences indicates that terminal formant frequency transition values are strongly influenced by the nature of the transconsonantal vowel. This experiment was designed to explore the perceptual correlates of Öhman's spectrographic findings. It was discovered that when a VCV sequence (where C is a voiceless plosive) is cut in two during the period of consonantal closure, there are not enough remaining cues in either the resulting VC or CV sequences to allow for identification of the deleted segment or of its articulatory features. However, it appears that coarticulation effects may hinder recognition of non-final allophones placed artificially in final position: consonants in VC sequences spliced from original VCV utterances are more difficult to identify than unreleased final consonants of the same quality.

Introduction

This investigation was prompted by the observation that while a good deal is known about the acoustic manifestation of coarticulation (with respect to point of articulation), there seemed to exist no published data regarding the perceptibility of the effects of this kind of coarticulation. In 1966, Öhman published the results of an extensive study dealing with coarticulation in English VCV sequences. He found that formant transitions from the first vowel to the intervocalic consonant are strongly influenced by the phonetic quality of the vowel following the consonant. We decided to investigate whether the changes in formant transitions due to the anticipation of the following vowel are perceptually significant.

Method

A set of VCV utterances was constructed, in which the vowels were /i æ a u/ and the consonants /p t k/. The $4 \times 3 \times 4 = 48$ utterances were recorded by one informant (a low-pitched female native speaker of English). In addition, 12 VC and CV syllables were recorded, in which the four vowels were followed and preceded by the three consonants each. The recordings were made in an anechoic chamber, using high-quality equipment. The VCV syllables were cut in two parts, placing the cut in the voiceless plosive gap. Using splicing techniques, four randomized lists were constructed. The first consisted of syllables from which the consonant release and the second vowel were removed. Each stimulus appeared twice on the listening test, for a total of 96 items. The task of the listeners was to identify the missing final vowel. The purpose of the test was to determine whether the transitions from the initial vowel to the consonant carried enough information to make this possible.

The second listening test consisted of syllables from which the first vowel had been removed. There were 96 test items. The task of the listeners was to identify the missing initial vowel.

List three contained 12 syllables produced by removing the initial vowel and consonant transition from symmetrical VCV utterances. Each stimulus appeared twice, randomly mixed with 2 x 12 syllables consisting of the same consonants and vowels, produced as CV sequences. The task of the listeners was to identify the 48 initial consonants. List four was similar, except that the stimuli consisted of VC sequences and the listeners had to identify 48 final consonants.

The listening tests were administered to untrained listeners, who were mostly sophomore-level students at The Ohio State University. 23 listeners took the first test, 36 the second; test 3 was taken by 41 listeners, and test 4 by 50. The data thus consist of 2,208 responses to Test 1, 3,456 responses for Test 2, 1,968 responses for Test 3, and 2,400 responses for Test 4.

Results

The results of the first two listening tests were largely negative, even though the same kind and degree of coarticulation effects reported by Ohman were measured on spectrograms made from our test tape. The listeners were evidently not able to identify the missing vowel. They were told that it was one of the four vowels /i æ a u/, and the results show that they were assigning these four vowels in an essentially random manner.

Percentages of correct responses for Test 1 (identify missing final vowel) ranged from 16 to 30% over all possible VC-combinations. The average of correct responses was 24%. In Test 1, the vowel actually produced on the tape was chosen for an answer in 24.5% of the total responses.

In Test 2 (identify missing initial vowel), percentages correct ranged from 19 to 30% over all possible -CV combinations. The average correct was 24.4%. However, in Test 2 there was a strong tendency among subjects to indicate the missing vowel as being identical with the one following the consonant, i.e. the one plainly audible from the recording. Of the total responses, 45.9% were instances of choosing the vowel heard. Out of the total correct scores, nearly half (42.73%) were due to "correct" identification of formerly symmetrical utterances, e.g. [apa], [iti]. Obviously then, the bias toward selecting the vowel heard is obscuring the number of correct responses. Whether this result is attributable to any sort of coarticulation phenomenon is unverifiable; it may simply be that the subjects were accustomed to identifying the vowel following the consonant after taking Test 1, but since 1) the subjects were given repeated instructions before taking Test 2, and 2) considerably more people took Test 2 than Test 1, this explanation seems unlikely. A satisfactory explanation does not appear to be possible at this time.

Except for the above, there seemed to be no significant trends in the incorrect responses for either test. Incorrect responses did not tend to fall into classes sharing some feature with the correct response, such as high/low or front/back.

The responses to Tests 3 and 4 show a clearer pattern, and will be discussed with reference to Tables 1 - 5.

Table 1 presents summary data about the identification of initial and final consonants.

TABLE 1

Identification of Initial Consonants (all vowels combined)	p	t	k	
#C	89.02%	90.54%	88.10%	89.22%
-C	88.71%	92.68%	86.28%	89.22%
Identification of Final Consonants (all vowels combined)				
Correct scores	p	t	k	Overall correct
C# (Released)	59.25%	90.25%	92.50%	80.67%
C- (Truncated)	54.75%	32.75%	29.50%	39.00%

A first observation is that in initial position, there is no difference between the correct identification scores of initial consonants produced as CV sequences and derived by tape-cutting from VCV sequences. The overall scores are identical and fairly high, 89.22% in both cases. The identification of final consonants is much less reliable. At 80.67%, the overall correct score for released final plosives approaches that of initial consonants; however, the three consonants differ in their relative identifiability, since /p/ has a significantly lower score than /t/ or

/k/. There was no such difference among the initial consonants. The identification of truncated final plosives has an overall score of 39.00%, with /p/ ranking higher than /t/ and /k/.

Table 2 presents a confusion matrix for released and truncated final plosives.

TABLE 2
IDENTIFICATION OF FINAL CONSONANTS (ALL VOWELS COMBINED)

Released	p	t	k
p	59.25	11.75	29.00
t	3.75	90.25	6.00
k	3.00	4.50	92.50
Truncated			
p	54.75	20.50	24.75
t	42.75	32.75	24.50
k	45.00	25.50	29.50

All vowels are combined in these results. It becomes obvious from this table that for released final plosives, the primary confusion was between final /p/ and /k/. As regards the truncated final plosives, the relatively high score of /p/ becomes less striking in view of the fact that /t/ and /k/ were both identified as /p/ far more frequently than they were correctly identified.

Table 3 gives an overview of the effect of different vowels on the identification of final consonants.

TABLE 3
CORRECT IDENTIFICATION OF FINAL CONSONANTS AFTER VARIOUS VOWELS
(ALL CONSONANTS COMBINED)

Preceding Vowel	Released	Truncated
i	87.67	43.0
æ	75.33	42.33
a	73.00	30.00
u	86.67	40.67
Overall correct	80.67	39.00

The highest scores were obtained for /i/ and /u/ for released consonants. In the truncated set, /a/ is associated with a significantly low score, while the other three vowels seem to have had no particular effect on the identifiability of the consonants.

Table 4 presents the data for final consonants arranged in the form of a complete confusion matrix.

TABLE 4
IDENTIFICATION OF FINAL CONSONANTS

Original stimulus (final vowel removed)	Perceived as			Original stimulus	Perceived as		
	p	t	k		p	t	k
ipi	59	23	18	ip	79	5	16
iti	31	42	27	it	2	92	6
iki	36	36	28	ik	6	2	92
æpæ	47	20	33	æp	41	15	44
ætæ	31	34	35	æt	3	93	4
ækæ	14	40	46	æk	4	4	92
apa	48	16	36	ap	45	24	31
ata	60	17	23	at	8	81	11
aka	66	9	25	ak	2	5	93
upu	65	23	12	up	72	3	25
utu	49	38	13	ut	2	95	3
uku	64	17	19	uk	0	7	93

Table 5 gives some results of a spectrographic analysis to which the 24 items of the final consonant test were submitted.

TABLE 5
F_2 - TRANSITIONS AND RELEASES OF FINAL PLOSIVES
FREQUENCIES IN HZ

Preceding Vowel	Final consonant								
	p- F_2	p^h F_2	Release	t- F_2	t^h F_2	Release	k- F_2	k^h F_2	Release
/i/	2800	2900	1600	3000	3000	2500 4500 5250	3000	3000	2500
/æ/	1900	1700	1500	2000	2000	2500 3500 4450	2250	2300	2300
/a/	1450	1500	1400	1500	1750	2500 3500 4250	1400	1450	1650
/u/	1000	950	1450	1250	1250	2500 4000	1050	1050	1500

The table contains terminal values of F_2 transitions toward the final consonant, and center frequencies of energy concentrations observed after the release of final consonants occurring in VC syllables produced as such.

Interpretation of the Data

Let us consider first the differences between the scores for released and truncated plosives. In the case of released /t/ and /k/, the scores are uniformly high. With /t/, the releases always had concentrations of energy at more than one frequency, which distinguishes /t/ releases from other releases following otherwise similar transitions. Compare, for example, the sequences /it/ and /ik/ (Table 5), where both the F_2 terminal frequency and the first energy concentration (and the only one for /k/ visible on the spectrogram) were at the same frequencies. On the other hand, the difference in the releases of /p/ and /k/ after /æ/ evidently was not strong enough to remove the confusion between released /æp/ and /æk/. The confusions between /p/ and /k/ after /a/ and /u/ seem obvious, when the terminal F_2 frequencies and energy concentrations in the release are compared.

A curious finding is the fact that releases did not improve the scores of /æp/ and /ap/ at all (Table 4). In fact, the release in /æp/ seems to have increased the tendency of listeners to identify this stimulus as /æk/. This is strange, since the release of /æk/ has a high frequency concentration as compared to the release of /æp/. Evidently in this case, the contribution of the release toward differential identification was negligible.

Discussion

The purpose of this investigation was to study the effects of coarticulation on perception. The results turned out to be essentially negative. Whatever the effects of coarticulation in terms of their influence on formant transitions, these effects are not sufficient to have an influence on perception. Thus the anticipation of a following vowel may result in a modification of the transition from a preceding vowel to the intervocalic consonant; but this modification is apparently not sufficient to enable the listeners to identify the following vowel from stimuli from which the following vowel itself was deleted. Likewise, whatever the lingering effects of a preceding vowel on the intervocalic consonant, a deleted initial vowel cannot be identified by listeners on the basis of effects that may have been physically present in the transition from the intervocalic consonant to the second vowel.

There was also no difference between the identification scores of initial consonants produced as such and consonants that became initial after the first vowel was deleted from a VCV sequence.

Only final consonants produced some differences between allophones produced as final and allophones produced originally as medial. Here the allophones preceding final silence are clearly much more easily identified than allophones placed into final position by tape-cutting. It is not immediately obvious how much of that difference is due to the effects of coarticulation. Wang (1959), in an experiment which was in part similar to ours, studied the relative contributions of releases and formant transitions to the correct identification of final plosive consonants. He found little difference between identification scores of released final /p t k/ and final /p t k/ whose releases had been eliminated by tape-cutting. The former ranged between 90-98%, the latter between 73-85%. On the basis of these data, it would seem that the contribution of releases was approximately 15% and the contribution of transitions was approximately 85%. Our listeners achieved overall scores for released final /p t k/ of 80.67% and 39.00% for unreleased final plosives. If the contribution of the releases was approximately 15%, there is still a difference of 25% to be accounted for. We conjectured that the anticipation of another vowel may affect the characteristic transitions to pre-silence final consonants to such an extent that listeners make additional errors in identification.

In order to investigate this hypothesis, we conducted an auxiliary experiment. Using the same speaker, equipment and splicing techniques, we prepared a randomized listening test composed of 12 VCV sequences from which the consonant release and second vowel were removed and 12 VC syllables in which the final consonant was unreleased, i.e. was produced by the speaker as an unreleased plosive. Twenty listeners were asked to identify the final consonants. These subjects were also given the original listening test for the identification of final consonants, as described above. The results are presented in Table 6.

TABLE 6
PER CENT CORRECT IDENTIFICATION OF FINAL CONSONANTS

Stimulus type	Per cent correct
Truncated VC- (re-test)	34.0
Released VC^h (re-test)	91.3
Truncated VC- (auxiliary test)	55.4
Unreleased $VC^{\urcorner}$ (auxiliary test)	73.7
Released VC^h (Wang, 1959)	95.3
Truncated VC- (Wang, 1959)	77.6

For these 20 subjects, identification of truncated plosives on the original test was about 5% lower than for the 50 sophomores, but identification of final released plosives was over 10% higher. The latter result may be attributed to the facts that, first, the re-tests were given to subjects singly or in groups of two, whereas the larger group was tested in a single session; thus the conditions for the re-test were more conducive to producing higher scores. Second, the subjects for the re-test were both more mature and more highly motivated than the 50 students. This, of course, does not explain the lower score in the identification of truncated plosives, but this difference is hardly significant.

On the auxiliary test, there was a higher correct score for the consonants whose releases were eliminated by tape cutting (55.4% as compared to 39.0%). A possible reason is the slightly slower rate of speech which the speaker chose for this recording session. Most significantly, the naturally unreleased consonants show a higher identification score (by nearly 20%) than the consonants placed in final position by tape-cutting. The 73.7% correct score falls within the lower range of Wang's results for final unreleased consonants.

It would thus appear that our hypothesis that coarticulation effects reduce intelligibility in the event that they are found in an environment where they do not occur naturally is supported by these additional data.

It is difficult to say what the actual physical cues or miscues were that caused the lowering of identification scores for plosives which had been placed in final position by the elimination of the second vowel from a VCV sequence. The only obvious case would be the sequence /aka/, in which the second formant transition to medial /k/ has a low terminal frequency, while the corresponding transition from the medial /k/ to the final /a/ has a high initial frequency (cf. also Green (1959), esp. pp. 50-52). It might be expected that the anticipation of the following high frequency would result in a raising of the terminal frequency of the transition from initial /a/ toward medial /k/. As Table 4 shows, no such raising occurred in the utterance produced by our informant.

References

Green, Peter S. (1959) Consonant-Vowel Transitions. Travaux de l'Institut de phonétique de Lund. Lund.

Öhman, S. (1966) "Coarticulation in VCV utterances." JASA 39.1.151-168.

Wang, William S-Y. (1959) "Transition and release as perceptual cues for final plosives." Journal of Speech and Hearing Research 2.66-73.