THE EFFECT OF COARTICULATION ON THE ROLE OF TRANSITIONS IN VOWEL PERCEPTION

By

TERESA A. UKRAINETZ

B.A., University of Calgary, 1984

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

School of Audiology and Speech Sciences

We accept this thesis as conforming

to the required standard

UNIVERSITY OF BRITISH COLUMBIA

July, 1987

46

Department of School of Audiology and Speech Sciences

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date July 29/87

ABSTRACT

The present study examines the effect of context on the use of transitions as cues to vowel perception. Thirty $V_1CV_2CV_1$ utterances were recorded, with V_1 being one of the three vowels /a, i, u/, and V_2 one of ten English vowels (/i, I, eI, E, ae, a, ^, oU, U, u/). After removal of the outer vowels (V_1), three sets of stimuli were created from the CV_2C parts: (1) unmodified controls (CO); (2) V_2 steady-state only (SS); and (3) transitions only (TR). Twenty subjects were asked to identify V_2. Subjects and speaker were matched for dialect and all subjects had some phonetics training.

Results showed significant differences across conditions and contexts. Scores for SS stimuli, for all contexts, were as high as for CO stimuli. Performance on the TR stimuli was as good as on the other two conditions for two of the contexts. However, for the TR condition-/a/ context, performance was considerably worse than for any other combination of conditions and contexts. Possible reasons for this are discussed, and the need for testing of other vowel contexts is emphasized.

It is concluded that, in some $V_1CV_2CV_1$ contexts, transitions can provide information about vowel identity on a level equal to steady-state alone, or to the combined information provided by both transitions and steady-states. This effect, however, is not uniform across contexts. For at

least one context, transitions alone are not sufficient to cue vowel identity at a level comparable to steady-state or combined information. This lack of uniformity suggests that the role of transitions varies with the type of vowel context present, and conclusions about general usefulness await systematic testing of a number of vowel contexts.

TABLE OF CONTENTS

LIST OF TABLES

ACKNOWLEDGEMENT

I would like to express my appreciation to all those people who provided me help and encouragement throughout the writing of this thesis:

- my supervisor, Dr. Andre-Pierre Benguerel;
- my classmates and subjects (who were kindly one in the same);
- Carol and Lisa, without whom, this thesis would have been a much lonelier endeavor;
- and David, whose love and support, as well as unflagging interest in my work, helped keep me going these two years.

Chapter 1

INTRODUCTION

Speech perception has often been examined by use of short, isolated phoneme or syllable-length segments, naturally or synthetically produced. Generalizations were then made to our comprehension of speech sounds in connected discourse.

This approach has tended to assume that speech perception is a sequential process, where the linking of discrete phonemes or syllables in words and sentences has no effect on the nature of those basic speech units, and that the perception of each unit is essentially unaffected by the preceding or following segments.

Descriptions of perception and phoneme identification based solely on isolated productions may be misleading since evidence from studies in speech production suggest that the basic speech units are produced in a temporally overlapping manner. This overlap, or coarticulation, is due to planning and inertial factors, resulting in individual phonemes whose phonetic realizations are quite different in spectral and temporal character from when they occur in isolation.

The coarticulation taking place in the production process is reflected in the resulting acoustic pattern. In some cases, the effect of coarticulation can be seen more than a syllable away on sound spectrograms. If the changes brought about by coarticulation are visible and measurable, it is plausible that they may be used by the perceptual system and they could be reasonably expected to affect the nature of perception.

Evidence suggests that, in perceiving speech, we do seem to make use of many of the cues available, even those relatively distantly located. Identification of speech segments involves a dynamic incorporation of information over a period of time to arrive at a unified percept of a /b/ or an /a/. The "distortion" of the phoneme caused by this temporal and spectral spread of cues would be expected to affect which cues are used and their relative importance for phoneme identification as compared to the ordering of cues established for phone or syllable-length segments. A question arises as to how the changes brought about by coarticulation affect the cues normally available when the phone or syllable is presented in isolation.

The present study looks at the cues used in the perception of vowels, and how the presence of context, through the resultant coarticulation, affects the relative utility of transitions and steady-state information as cues to vowel identity.

Chapter 2

REVIEW OF THE LITERATURE

2.1 Introduction

Modern vowel perception studies date from the invention of the spectrograph nearly forty years ago. Although the number of studies done since then is very large, our understanding of vowel perception is still limited. Section 2.2 summarizes the state of knowledge of vowel perception and the effects of context on perception. Section 2.3 further examines the influence of context and its resultant coarticulation, at the articulatory and acoustic levels. Section 2.4 deals with the perception of coarticulation, and suggests how the perceptual effect of coarticulation on speech segments may affect our current conception of the relative utility of known vowel cues.

2.2 Cues to Vowel Perception

Vowels have been traditionally characterized by the frequencies of their first few formants, especially the first two in the case of English (referred to as F_1 and F_2). When vowels are produced in isolation, the formant frequency values of these vowels, for a speaker, can be plotted as points in a two-dimensional space (F_1 vs. F_2). Formant plotting tends to cluster phonemically equivalent tokens, and can be related to articulatory production patterns. This characterization, however, is insufficient to distinguish vowels in all but the most ideal condition, that of a single speaker producing isolated vowels.

Peterson and Barney (1952) plotted the vowels of 10 /hVd/ words spoken by 76 speakers, in an F_1-F_2 graph. They found that the distributions of formant values did not correspond closely to the distribution of phonemic values, resulting in overlapping vowel loops.

Stevens and House (1963) measured the first two formant frequencies and bandwidths of eight vowels, for three male American English speakers, in #V#, /hVd/, and 14 different /həCVC/ contexts by three male speakers. They found significant differences between their results and the vowel loops plotted by Peterson and Barney (1952), and by Peterson (1961) for vowels in /hVd/ context and in isolation respectively. Some of the difference was systematic and found to be caused by differences in production related to vocal tract length, especially for the vowels in isolation and in /hVd/ context. Consonants were found to exert a large effect on vowel formant values. The consonant effects were systematic and regular, depending on place of articulation, manner of articulation and presence of voicing as well as on the particular vowel involved. The acoustic effects tended to be of movement towards the locus of the consonant or movement towards neutral schwa values. Movement towards the center of the vowel triangle, or centralization, has also been found to occur with increased speech rates and in vowels of weakly stressed syllables, both in British and American English (Lindblom, 1963; Gay, 1978). Individual differences in degree of F_2 undershoot not accounted for by syllabic rate or vocal tract differences have also been documented (Stevens, House & Paul, 1966), but, in general, the differences in formant frequency occurring for the same vowel generated by different speakers can be ascribed to differences in vocal tract size, speaking rate and degree of stress. Stevens and House (1963) concluded that the changes in measured formant frequency values for a given vowel are

> ...in large measure explicable in terms of the dynamic properties of the articulatory mechanism. The inertia and delay characteristics of the articulatory structures and musculature result in undershoot in the motion of the structures from one target configuration to the next, giving rise to corresponding changes in the vocal tract resonances. (p. 126)

Although changes in vowel formant frequency largely follow lawful and predictable patterns based on articulatory dynamics, the problem of how they are decoded by the listener remains. Characterizations based on single formant frequency measures (e.g. F_2 values) result in too great an overlap in categories to adequately account for vowel identification. As a result of this ambiguity, attention was devoted to further specifying the static model of vowel space as well as to investigating other possible cues within the vowel signal.

An early attempt to refine the vowel space model was presented by Peterson (1961). Peterson recorded sustained vowels imitated by men, women and children. The speakers listened to recorded reference vowels and attempted to produce a phonetic match. The vowels spoken included non-English vowels, which Peterson reported the subjects had no difficulty matching. A single frequency measure was taken for each of the first two formants. Various transformations using linear, logarithmic, and mel scales were attempted, so as to produce groupings of phonetically equivalent tokens and differentiation of non-equivalent items. Peterson (1961) concluded that, for any of these transformations, "neither a fixed formant frequency nor a formant frequency ratio hypothesis is adequate to explain vowel perception fully " (ibid, p. 24-25). He also concluded that other aspects of the vowel influence perception, such as the fundamental frequency, amplitude and phonetic environment. Stevens and House (1963) suggested that some of the difficulty specifying a workable model arises from the lack of agreed upon method for sampling vowels and the insufficiency of using a single frequency measure to characterize formants which show continuous movement throughout the syllable nucleus. Lindblom (1963) advocated defining the target not as the measured formant frequency value, but as a potentially unrealized asymptotic value. This value corresponds to the average isolated vowel production value for that speaker. Stevens and House (1963), as mentioned earlier in this section, confirmed the systematic and predictable nature of the changes in vowel formant frequencies. Transformations of the target formant frequencies and the relative relationships between them were developed to account for differences among speakers (e.g. Neary, 1977; Skinner, 1977).

Approaches involving calibrations of a particular speaker's vowel space either from frequencies of the formants of other vowels occurring in the same auditory context, or from high formants (F_4 or F_5) have been postulated, e.g. by Joos (1948). Joos, as reported in Ladefoged and Broadbent (1957), suggested that vowel identification was based on knowledge of that speaker's vowel space. He suggested further that a vowel was identified, not on the basis of its absolute values, but on the relations between those vowel formant frequencies and the general ranges of frequencies which seemed typical for that speaker. Ladefoged and Broadbent (1957) supported this claim by showing, with synthetic speech, that identification of a test word is greatly influenced by the range of formant frequencies in the preceding carrier phrase. Gerstman (1968) was able to successfully develp a computer algorithm using the extreme values of a speaker's vowel space (usually the vowels /a/,/i/,/u/) to scale all the talker's vowels. Whether or not this type of scaling is an actual perceptual strategy is as yet unknown, but other studies, such as Verbrugge, Strange, Shankweiler and Edman (1976), Macchi (1980), and Assman, Neary, and Hogan (1982) have shown that, although speaker familiarity does improve identification, it is not necessary. In these studies, for #V# and /hVd/ productions, with multiple mixed speakers, the vowels were identified well despite the lack of cues for normalization.

It would seem that calibration of the vowel space may be accomplished without the scaling information available in multiple vowel productions by the same speaker. Scaling may be accomplished with information from a single vowel by employment of the values of the formants or their relations between, as cited earlier (i.e. Neary, 1977; Skinner, 1977).

Vowel formant normalization strategies have had some success in accounting for vowel identification over speakers with different size vocal tracts, but they still do not adequately account for listener success with vowels produced in CVC context, where formants are shifted from the position occurring when they are produced in isolation (canonical position). A possibility is that the vowel or syllable may be sufficient in itself to provide vowel normalization information. This information may be quite different from the formant frequency values usually considered. Possible cues include duration differences, intensity differences, degree of diphthongization, and transition rate and direction.

Vowel duration has been shown to separate vowels into two groups: tense and lax, with tense vowels (/i/,/ae/,/a/,/ɔ/,/u/) having longer steady-state segments (Lehiste & Peterson, 1961) and lax vowels (/I/,/E/,/ə/,/U/) having longer offglides. Lack of durational cues can have significantly detrimental effects on vowel identification as Fairbanks and Grubb (1961) found for sustained isolated vowels.

Intensity also varies systematically among vowels. Lehiste and Peterson (1959) found a 5 dB range of intensity for vowels, with /a/ the most intense and /i/ the least.

Vowels also vary in their degree of diphthongal movement, which can be used in identification (Assman et. al, 1982). The diphthongal offglides for the tense vowels tend toward the extremes of the vowel space (/i/ for front, /u/ for back). Conversely, for certain North American dialects, lax vowels may show offglides in the direction of schwa, according to Joos (1948).

Although vowels produced in isolation are considered the ideal or canonical forms, several studies have found high error rates for naturally spoken, isolated vowels (Lehiste & Meltzer, 1973; Strange, Verbrugge, Shankweiler & Edman, 1976) and for vowels excerpted from context (Bond, 1975; Fujimura & Ochiai, 1963), suggesting that more information than that contained in the vowel steady-state may be useful in vowel perception. In connected discourse, vowels occur joined to consonants. The transition intervals occurring between the previous consonant termination and the vowel steady-state, or between the vowel steady-state and the following consonant initiation were initially ascribed the role of characterizing the consonants and were not considered useful for vowel identification:

> It is assumed that such changes are cues for the perception of the consonants rather than linguistically significant components of the vocalic nucleus (in monophthongs). (Lehiste & Peterson, 1961, p. 268)

The contribution of transitions to vowel identification was first indicated by Lindblom and Studdert-Kennedy (1967). They synthesized a vowel steady-state continuum varying from [I] to

[U]. These vowels were both preceded and followed by one of two consonantal frames, [w-w] or [j-j], and presented in CVC form. For the same vowel midpoint frequency, presence of different consonant transitions shifted the category boundary between the vowels. This indicated that the identity of a vowel is determined not only by the formant-frequency pattern of the steady-state, but also by the direction and rate of adjacent "consonant" formant transitions.

Transitions increased in significance vis-a-vis vowel identification when better vowel identification rates were found to occur in a consonantal context rather than in isolation. Millar and Ainsworth (1972) reported a more reliable and uniform identification of vowels synthesized in a /h-d/ context than when for the same vowels synthesized as isolated steady states. This finding suggested, that, although /h-d/ is sometimes considered a null context, because of its minimal effect on formant steady-state values, the transitions are somehow contributing to identification.

More convincing evidence comes from comparisons of natural vowels spoken in CVC contexts and in isolation where both productions have speaker-controlled onsets and offsets. Strange et al. (1976) compared perception of isolated vowels to vowels in /pVp/ syllables spoken by men, women and children. Speakers were given one practice trial for each stimulus type. The recorded items were presented to listeners in single and mixed speaker conditions. Subjects heard the stimuli over loudspeakers and responded on scoresheets coded with the orthographic equivalent of each vowel in a "pVp" form. Errors for the /pVp/ conditions were 17.0% and 9.5% for mixed and blocked speakers respectively. For the isolated vowels, errors were 42.6% and 31.2%. Similar results were found by Gottfried and Strange (1980) for /pVp/,/bVb/, and /kVk/, but not for /gVg/. These results show a very poor identification performance for isolated vowel productions and a strong advantage for vowels produced in CVC contexts as compared to vowels produced in isolation.

These studies, showing better identification performance for vowels produced in CVC sequences as compared to that of vowels produced in isolation, and high error rates for isolated vowel productions, were criticized on procedural grounds. Criticisms involved the dubious listening quality obtained in free field testing, the lack of checks on vowel quality, and lack of dialect matching. A major source of criticism was the use of CVC orthographically spelled response choices, for both isolated vowels and vowels in CVC frames. It was felt this introduced a response incompatibility for the isolated vowels, thus depressing performance in that condition.

Assman et al. (1982) performed two experiments to evaluate the effect of response types on performance. They attempted to provide high quality listening conditions by use of headphones, monitoring of quality, and dialect matching (Alberta, Canada). Their first experiment examined whether a significant proportion of errors were due to mislabelling rather than perceptual confusions. Subjects were presented /pVp/ stimuli. Responses were /hVd/ keywords and repetition of the sound heard. Errors in the written condition were 17% for #V# and 15% for /pVp/, thus showing no contextual advantage. Errors in the written and spoken condition were 5% for both #V# and /pVp/. The marked decrease in errors for the combined condition as compared to the written condition showed that a large proportion of errors in key word tasks can be the result of mislabelling rather than perceptual confusions. In the second experiment, #V# and /pVp/ stimuli were again presented and keyword (/hVd/) and spelled responses (/pVp/) were compared. The hypothesis was that spelled responses are easier for listeners than keyword responses. Errors were significantly different only for the /pVp/ stimuli, with the keyword task producing more errors. Both these experiments showed response types can inflate error rates and can do so selectively.

Additional studies also showed that when attention is paid to obtaining quality productions and listening conditions, matching regional dialects, and providing compatible response alternatives for both conditions, uniformly low error rates and little contextual advantage is found (Kahn, 1978; Macchi, 1980; Diehl, Buchwald, McCusker, & Chapman, 1981).

Vowels produced in CVC's are not consistently easier to identify than vowels produced in isolation. However, performance on vowels produced in CVC's is no worse than performance on vowels produced in isolation either, which is what might be expected if the presence of transitions served only to "distort" vowels from their canonical steady-state frequencies. Compelling evidence for the involvement of transitions in the vowel identification process comes from two recent studies, by Jenkins, Strange and Edman (1983) and by Strange, Jenkins and Johnson (1983). These studies investigated how strong a cue transitions are when no steady-state information is provided. Results were similar for the two studies. The latter study will be described since it contains a more systematic treatment of the variables in question. Natural /bVb/ utterances were modified by substituting silent intervals for the transitions or the steady-state portions of the vowel. Normal duration differences were maintained to varying degrees: for the steady-state only stimuli (SS), the remaining center was either kept at its natural duration or was cut to a standard length; for the transitions only stimuli (TR), the silent center was either its natural duration or was the standard length, while the transitions maintained their original lengths. The results showed, for stimuli with durational information (both the transition and steady-state durations preserved), error rates for both SS and TR stimuli were very low; 8% and 6% respectively. Additionally, performance was not significantly different between the CO and the TR stimuli. Reduced duration information (standard length center, normal duration transition) did not significantly affect the TR stimuli. Absence of durational information (standard length center, no transitions) did damage the SS sounds.

There is not only an advantage in the identification of vowels produced in a CVC sequence, but transitions without the presence of steady-state or full duration information also provide adequate information for vowel identification. Based on these findings, an approach incorporating dynamic processing into vowel perception was put forward (Strange et al., 1976; Jenkins et al., 1983; Strange et al., 1983) as an alternative to static vowel space characterizations. The dynamic process described identification of a vowel as involving both temporal and spectral information over the course of a complete CVC syllable. Identification could be accomplished through the partial information presented in isolated vowel productions, such as the steady-state formant frequency values, but normal identification involved information integrated over an entire CVC time period.

This dynamic process approach is reminiscent of an acoustic model developed by Stevens, House and Paul (1966) and based on a dynamic articulatory description. These authors described articulation of a CVC syllable as consisting of a superposition of articulatory events sequentially ordered but overlapping. As a result, "the articulatory activity - and the resulting acoustic output - that characterizes a syllabic nucleus depends on the initial and final consonants as well as on the vowel itself" (ibid, p. 123). Their dynamic model calculated vocalic contours representing the first and second formants as a function of time. The parameters used included initial and final frequencies, midpoint frequencies, durations, and measures of transition curvature.

Strange and her colleagues (Strange et al., 1976; Strange et al., 1983; Jenkins et al., 1983) did not specify the details of their model in perceptual terms, but a fairly successful attempt to involve dynamic variables was made by Assman et al. (1982), who developed this dynamic process approach further by combining measures of "steady-state information" (formant frequencies and fundamental frequency), "dynamic information" (formant transition slopes and duration) and "speaker information" (VT length) to achieve a characterization that resulted in relatively little overlap between vowel categories.

2.3 Coarticulation

Evidence suggests that a static conceptualization of vowel identification, involving derivation of formant frequency values, is insufficient to adequately describe vowel perception. A dynamic approach, involving time-varying as well as stationary cues, seems to account for the process of identification better. If a dynamic model is to be postulated, then the presence and utility of phonetic overlap extending further than that occurring for vowels produced in CVC sequences should be specified. How distant can the cues to phoneme identity be? Is the syllable an adequate unit of analysis? Both segmentation studies and coarticulation studies suggest otherwise.

Normal discourse involves the blending of vowels and consonants to such a degree that it is difficult if not impossible to segment the acoustic signal into discrete segments, as Liberman, Cooper, Shankweiler and Studdert-Kennedy (1967) point out in their review of work in speech perception. The sentential context of connected discourse provides cues not only at higher levels of analysis such as syntactic and semantic, but at the relatively primitive phonetic level as well. Ladefoged and Broadbent (1957), as described earlier, showed that identification of a word is greatly influenced by the range of variation of each formant frequency in the preceding carrier phrase. Verbrugge, Strange, Shankweiler, & Edman (1976) compared the identification of three types of stimuli: (1) syllables in their carrier sentences; (2) syllables excised from the same utterance; (3) excised syllables preceded by precursor strings with conflicting tempo information. Errors increased from (1) to (3) and it was concluded that sentence context allows adjustment to speaker tempo, thus aiding in vowel identification.

Connected discourse not only provides additional information, but, by the very nature of its overlapping connections and parallel transmission, distributes phone identification information over a time larger than the duration of the heard phone, thus affecting the realization of the underlying phonemes and changing them from what they are or would appear as in isolation. The process involving this spread of information is called coarticulation and occurs because of the programming and execution properties of the articulatory mechanism.

Coarticulatory effects are of two types, referred to as anticipatory and carryover. Left-to-right or post-articulatory effects, termed carryover coarticulation, effects occur when the articulation of a phone is influenced by the articulation of a previous phone. Carryover coarticulation is the most predictable of the two types, being primarily caused at the execution level by mechanoinertial factors specific to the articulators. Anticipatory effects are right-to-left or pre-articulatory; the articulation of a phone is influenced by that of a subsequent phone. Anticipatory coarticulation originates at the planning level, because the speech gestures are generally not independent, nor strictly linearly sequenced. Anticipatory effects are of greater interest than carryover effects in models of speech production, although both are of concern in perception.

Kozhevnikov and Chistovich (1965), in their model of speech production, postulated an "articulatory syllable" in which commands for the entire articulatory syllable are initiated simultaneously. The commands are also executed simultaneously if the requirements are non-competing, but sequentially if they are competing. For Russian, coarticulation is considered to be maximum over the strongly cohesive CV (or CCV,...) syllable unit and minimum across syllable boundaries. The model was based on their (1965) study of electromyographic activity in the orbicularis oris muscle, for liprounding, during production of various vowel and consonant combinations.

Fromkin (1966) also did an EMG study of the muscle used in liprounding during rounded and unrounded vowel and /b,p/ productions in CVC sequences. She found that different motor commands produce different muscular gestures for initial /b/ or /p/ as compared to final, but vowel context did not significantly affect EMG activity, nor did consonant context seem to affect EMG activity in vowels. Fromkin suggested that some aspects of context somehow restrict or reorganize the neuromuscular commands and gestures for some phonemes, but this effect is not common to all characteristics of context. One explanation given was the possible existence of a feedback system concerning the state of muscle activity, wherein the basic unit of speech production is still a phoneme-size unit. Another explanation was that the minimal linguistic unit at the motor command level may be larger than the phoneme, possibly on the order of a syllable.

Henke's (1966) model gave support for a phoneme-sized execution unit, but added a "look ahead" system. In his model, production is organized phoneme by phoneme, but includes scanning for upcoming features. If a phoneme currently being realized is unmarked for a particular feature, the feature will be set to the value of the next occurrence where it is marked. This value will be determined by means of the look-ahead mechanism, which scans upcoming phonemes to examine the value of each feature.

Öhman (1966) also espoused a scanning approach. He conceived of phoneme production in connected discourse as involving invariant intentions, with variable motoric realizations. These result in variable VT shapes, which are the result of late termination of the previous phoneme (due to inertial factors) and early initiation of the upcoming phoneme, both occurring during the execution of the current phoneme. Planning over at least the interval from one vowel to the next over intervening consonants is thought to occur. Öhman describes the articulation of each phoneme of a VCV sequence as involving separate but overlapping sets of muscles, each set having a separate neural representation in the control networks of the brain. The vocal tract shape at any point in time is a function of messages from multiple channels. Although the phoneme currently in production is the main information contributor to these channels,

> ... a VCV utterance of the kind studied here can, accordingly, not be regarded as a linear sequence of three successive gestures. We have clear evidence that the stop-consonant gestures are actually superimposed on a context-dependent vowel substrate that is present during all of the consonantal gesture. (Ohman, 1966, p. 165)

Öhman (1967) matched lateral x-ray data of VCV productions to his numerical model of coarticulation. He found that actual vocal tract shapes for Swedish VCV utterances compared well with shapes generated by his coarticulation formula. His model was based on spectrographic data concerning transconsonantal coarticulation reported in Ohman (1966). This data were obtained from American, Swedish and Russian speakers producing V_1CV_2 utterances. Transition onset frequency and steady-state frequency were measured for each vowel. A measurable carryover coarticulatory effect from V_1 on the transition portion of the CV_2 and a similar anticipatory coarticulatory effect from V_2 on the transition portion of V_1C were found. The effect on the vowel steady-state portion was reported to be small.

Major evidence for anticipatory coarticulation in the speech production process comes from articulation studies, e.g. Őhman (1966; 1967). Carney and Moll (1971) extended these findings with a cinefluorographic investigation of fricative consonants in VCV context. They found coarticulation effects similar to those found by Ohman (1967).

Coarticulatory effects have also been investigated for velar movements by Moll and Daniloff (1971). Four subjects produced English sentences containing various combinations of nasal consonants (N), consonants (C), and vowels (V). These were examined cinefluorographically. The results indicated extensive anticipatory coarticulation of velar movement in CVN and CVVN sequences, beginning during the approach to the initial vowel.

Coarticulation effects across longer segments have also been shown to occur (Benguerel and Cowan, 1974). Benguerel and Cowan examined the timing of upper lip protrusion in French, particularily the onset of protrusion in a consonant cluster followed by a rounded vowel (e.g. /kstry/ in "la dextre universelle"). Results showed that rounding movement could start as early as four to six consonants before the vowel.

The studies cited show that coarticulation can be observed at both the articulatory and the acoustic level to varying degrees depending on the feature involved. The models of speech production described differ in the degree to which they admit coarticulation, but all those cited offer plausible explanations for coarticulation in production of speech.

2.4 Perceptual Effects of Coarticulation

Coarticulation is acoustically present, but whether it is perceivable, and if and how it is involved in the normal process of phoneme identification is still to be determined.

Perception of coarticulation could occur in a number of ways. A strong coarticulatory presence would allow prediction of the identity of an upcoming phone. A weaker manifestation would be sufficient to allow prediction of features (e.g. liprounding). Another possibility is that coarticulation may affect the nature of the phoneme it overlays, resulting in a change in identity or at least a change in quality.

Ali, Gallagher, Goldstein and Daniloff (1971) tested the perceivability of the nasal coarticulation described by Moll and Daniloff (1971). They spliced out the final consonant (including its VC transition) from English CVC and CVVC utterances, where the final C was sometimes a nasal consonant, and sometimes a non-nasal consonant. 22 subjects were asked to identify whether the missing consonant was nasal or non-nasal. Nasal consonants were predicted correctly at better than chance levels ($p<.001$). Stops were perceived as nasals more frequently than fricatives. Consonants following low vowels were perceived as nasals more often than those following high vowels. Ali et al. reported all of these findings to be consistent with previous studies on nasal identification such as Lintz & Sherman (1961).

Benguerel and Adelman (1977) examined the perceptual significance of lip rounding coarticulation in French vowels described by Benguerel and Cowan (1974). They truncated -CCCCV utterances (or parts thereof) at four different points before the vowel and had French and English subjects predict the identity of the missing vowel. French and English subjects were used to examine the possibility of differential sensitivity related to linguistic experience since in French, liprounding on vowels is contrastive, while in English it is not. The results showed that when segments up to and including at least half the final consonant of the cluster are present (i.e. truncation occurs before the CV transition), subjects correctly identified the missing vowel well above chance level. There were no significant differences between the French and English subjects, suggesting that this coarticulatory feature can be perceived subphonemically.

Another study looking at the perceptual effects of liprounding and nasality was that of Sharf and Ostreicher (1973). They constructed utterances of the form təNC_{0-2}V where N is a nasal consonant (/n,m,ŋ/), C_{0-2} is a consonant cluster consisting of zero, one or two non-nasal consonants and V is an unrounded or a rounded vowel (/i,u/)(e.g. /təmi/, /təmti/, /təmsti/). The postnasal segment was then spliced out of the utterance. 37 subjects were asked to identify the nasal consonant, in noise and in silence, and the final (deleted) vowel in silence. Identification of the nasal consonant was better when all the postnasal sounds were retained than when they were deleted, suggesting that the carryover coarticulation normally present in the second syllable aids identification. When no consonant intervened between the nasal and the vowel, identification of the missing vowel was significantly better than chance, demonstrating the presence and utility of anticipatory coarticulation.

The consequences, at the perceptual level, of Öhman's (1966) findings on the acoustic presence of transconsonantal coarticulation were examined by Lehiste and Shockey (1972). VCV utterances, composed of the consonants /p/,/t/,/k/ and the vowels /i/,/ae/,/a/, and /u/, were cut in two parts, at the voiceless plosive gap, leaving either a VC- or a -CV syllable. Subjects were asked to identify, from a choice of four vowels (/i/,/ae/,/a/,/u/), the missing vowel. The results showed that the remaining information was insufficient to cue the identity of the missing vowel and responses did not fall into classes sharing some feature with the correct response, such as high/low or front/back.

Clark and Sharf (1973) examined the coarticulation effects on short-term memory recall. V_1CV_2, V_1C- (V_2 deleted), and V_1C utterances were composed, where V_1 was one of six lax vowels, C was a /t/, and V_2 was an /a/ or an /I/. These utterances were presented to subjects who were asked to recall the first vowel. The V_1C- stimuli resulted in better recall scores than the other two conditions only when the missing vowel (V_2) was the same as the vowel present (V_1). The authors concluded that the presence of coarticulation influenced recall under memory-taxing conditions: coarticulation of the deleted vowel appeared to facilitate recall of V_1 if V_2 was the same vowel, and to hinder recall if it were different. The lack of facilitation in the V_1CV_2 stimuli was attributed to perceptual overload from hearing two vowels.

Coarticulation effects are not sufficient to predict phoneme identity. They are sufficient to predict certain phonetic features, both contrastive and non-contrastive, such as rounding and nasality, but not others, such as vowel height. The presence of coarticulation can aid identification and recall, although the presence of the underlying elements can be distracting.

An aspect of coarticulation that has not been examined is its effect on the underlying phoneme. Does the presence of coarticulation change the way in which an affected phoneme is perceived? A possible candidate for this is the identification of vowels in the presence of the transconsonantal coarticulation described by Öhman (1966).

Öhman reported that, in VCVs, transitions were affected to a much greater degree than steady-states by coarticulation from the transconsonantal vowel. The recent work by Jenkins et al. (1983) and Strange et al. (1983) suggest that, in a CVC utterance, transitions carry sufficient information to allow vowel identification scores comparable to those for the entire

CVC. Neither Strange et al. (1983), nor Jenkins et al. (1983) excised the experimental stimuli out of a larger utterance, consequently, the effect of non-adjacent elements could not be examined. It is a well supported fact that items removed from context are less intelligible than isolated productions, due to lost information about context and rate, but the effect of the coarticulation resulting from producing them in context is not usually considered. The relative importance of the various cues used in vowel identification, as determined when produced in isolation, may change when the same vowels are taken out of a larger context, where the forces of coarticulation selectively affect aspects of these cues.

Chapter 3

AIMS OF THE EXPERIMENT

Cues to vowel perception include formant frequencies, duration, diphthongal movement and consonant-vowel transitions. Recent work (Strange et al., 1983; Jenkins et al., 1983) has suggested that consonant-vowel transitions alone contain sufficient information to permit vowel identification. The speech segments used by these authors have been produced as CVC syllables and have not been spliced out from longer utterances. Studies have shown that the surrounding phonetic context influences the nature of the consonant-vowel unit. This influence is strong enough to permit identification of certain features of the source of influence. Accordingly, its effect on the intelligibility of the underlying segment could be expected to be reasonably strong, at least in some situations. Since this effect has been shown to differentially alter parts of the vowel, the question asked will be whether it alters the relative utility of the various cues to vowel identification.

This study examines the effect of context on the cues to vowel perception. The contextual effect to be examined is transconsonantal coarticulation, which is known to affect the transition part of the vowel more than the steady-state part (Ohman, 1966). It is expected that the changes to the transitions caused by transconsonantal coarticulation will reduce the intelligibility of utterances which have transitions as the primary cues (i.e. the steady-state portion of the vowel removed). The degree of intelligibility should be less than that of vowels containing steady-state and transition information, or steady-state information alone. The degree of intelligibility will be measured in a forced choice identification task. It is expected that correct identification rates will be lower for the transition only stimuli spliced out of longer utterances than for unmodified stimuli, or steady-state only stimuli, both spliced out of longer utterances.

Starting from recorded utterances of the form $V_1CV_2CV_1$, the outer vowels are removed and the sounds altered following the Strange et al. (1983) procedure for creating steady-state only and transition-only stimuli. The experiment to be described examines

(1) whether consonant-vowel transitions are sufficient to cue vowel identification in the presence of coarticulatory effects.

(2) whether different vowel contexts differentially affect vowel cues (steady-state and transitions).

Chapter 4

METHODOLOGY

4.1 Preparation of Test Tapes

Speech Materials

The speech materials from which the stimuli were extracted were natural utterances of the form $V_1CV_2CV_1$. The consonant (C) was always the voiced plosive /b/. V_1 was one of the three vowels, /a/, /i/, /u/, all found by Ohman (1966) to have a transconsonantal coarticulation effect. V_2 was one of ten vocalic nuclei; /i/, /I/, /eI/, /E/, /ae/, /a/, /^/, /oU/, /U/, /u/.

A male adult speaker from the same region as the subjects (Western Canada), with some phonetics experience, produced the utterances. He read them from a list written in phonetic symbols, with careful monitoring to ensure correct pronunciation. Instructions to the speaker were that he say each utterance with even stress on each syllable or, failing that, with slightly increased center syllable stress, at a comfortable rate. He could practice beforehand, repeat the utterance, and pause as needed. Each token was recorded a number of times. The tokens selected for further evaluation were chosen for the accuracy of production, relative evenness of pitch and stress, and lack of extraneous noise such as paper rattling. The tokens were then transcribed by three listeners trained in transcription but uninformed as to the nature of the experiment. They were told that the tokens were of the form $V_1CV_2CV_1$ (see above). They were encouraged to transcribe using whatever labels they felt necessary. They were also asked to make qualitative comments on pitch, stress and noise, whenever possible.

Editing

Editing was done on a PDP-12 computer with WAVES, a set of programs written by Lloyd Rice at UCLA.

The audiotaped speech was played back at half-speed, low pass filtered at 2.5 kHz and transferred to the PDP-12 computer and digitized with a 10-bit analog-to-digital converter at 6 kHz sampling frequency.

The $V_1CV_2CV_1$s were reduced to $-CV_2C-$'s and three sets of stimuli were generated from these with WAVES. The first set of stimuli consisted of the unaltered $-CV_2C-$'s (control or CO condition), the second set had the transitions replaced with silence (Steady-State Only or SS condition) and third set had the steady-state portion replaced with silence (Transitions Only or TR condition).

The cutpoints for elimination of the outer vowels were determined on spectrograms and corroborated with voicing information obtained from the speech-signal display used for editing. The boundary of the segment retained ($-CV_2C-$) was considered to be from the initial consonant release to the final consonant closure.

The criteria used in selecting cutpoints were based on the procedure used by Strange et al. (1983). They reported three sets of cutpoints, arrived at by examination of Lehiste and Peterson's (1961) data on vowel transition and center durations.

In Lehiste and Peterson (1961), there were two groups of simple syllabic nuclei: short and long. The two types were differentiable, not by absolute duration, but by the relative duration of their centers and offglides. Long vowels had longer centers and shorter offglides, and short vowels had shorter centers and longer offglides. A third group of single vocalic nuclei, composed of /ei/ and /ou/, were described as long, complex nuclei, with an extremely long onglide /ei/ or offglide /ou/.

Strange et al. (1983) used a number of different cutpoints for the onglide and offglides. They set the initial portion, or onglide, at a constant 15% of the total duration for all vowel types. For the offglide portion, they used three cutpoints. Long vowels were cut at the last 20% and short vowels at the last 35%. The third cutpoint, for a group termed the intermediate vowels, /i/ and /u/, was at 25% of the total duration. There is no discernable reason why this group was created: /i/ and /u/ are not intermediate in offglide length relative to total duration (Lehiste & Peterson, 1961), nor in absolute duration (Peterson & Lehiste, 1960). It is possible that the goal of the manipulation was to prevent steady-state information from occurring in the transitions, due to the individual variation occurring for the particular vowels used by Strange et al. (1983), but this is not actually specified in the paper.

The cutpoints for the stimuli of the present experiment were also based on Lehiste and Peterson (1961). The steady-state portion was defined as that time period during which the formants (F1,F2,F3) are steady (i.e. parallel to the time axis on a spectrogram). Transitions were measured from the onset of voicing to the beginning of steady-state and from the end of steady-state to the conclusion of voicing at the point of consonant closure. When the cutpoints described above were applied to the vowels for the present experiment, they did not match the actual spectral boundary between the transition and steady-state. The greatest concern was for those vowels where the relative length of transitions was shorter than the length specified by Strange et al. (1983). In those vowels, steady-state information is present in the transition sections, thus providing information known to be useful in identification (i.e. target information). The vowels causing the major concern were /u/, /U/, /oU/, and /^/. These vowels had near-horizontal transitions for either or both the onglide and the offglide. As well, the vowels /eI/ and /oU/ presented the added problem of being placed in the simple nuclei, long vowel category, although their gliding nature did not conform to the cutpoints for this category. Finally, variation in total vowel, steady-state and transition durations was much greater in this experiment than in Strange et al. (1983), as seen in Table I.

Table I

<u>Range of Vowel and Vowel Component Durations (in msec.)</u>

<u>for the Present Data and</u>

<u>for the Strange et al. (1983) data</u>

	Present Data	Strange et al. (1983)
total vowel	142 - 248	114 - 202
onglide	15 - 82	22 - 30
target	45 - 142	57 - 127
offglide	22 - 98	33 - 42

Some of this additional variation was probably due to the multiple contexts and/or to transconsonantal coarticulation. All of these factors contributed to making the application of the cutpoints laid out by Strange et al. (1983) problematic.

In determining how to resolve these difficulties, increased attention was paid to the method of determining boundaries between transitions and steady-state. Most studies do not explain how exactly the boundary is determined, so two methods were compared to see whether the choice of method would significantly affect duration values. Each vowel was divided into transition and steady-state portions in two ways, based on spectrographic measurements: (1) the division point was chosen where a line drawn in the center of the formant band departed by more than 1 mm from its average position, and; (2) the division point was chosen at the intersection of a horizontal line with a line extending the transition slope. Table II shows the results obtained with these two methods. These two methods were not significantly different ($p<0.05$) when measured with a t-test for dependent means. The difficulty of determining the division between transition and steady-state increased as the slope of the transition decreased.

Table II

Relative Duration (in % of total vowel duration)
of Initial and Final Transitions
obtained with Two Methods

	INITIAL						FINAL					
	abVba		ubVbu		ibVbi		abVba		ubVbu		ibVbi	
V	#1	#2	#1	#2	#1	#2	#1	#2	#1	#2	#1	#2
a	19	14	18	20	31	17	19	14	12	12	19	14
i	31	21	39	46	20	16	23	28	27	32	18	18
ae	20	20	16	22	20	11	20	20	32	32	17	11
E	29	20	23	27	27	31	40	40	27	31	42	46
I	16	11	26	21	15	15	47	48	36	36	36	40
u	13	13	20	12	8	8	38	38	37	29	53	36
U	16	11	24	10	9	11	18	16	16	14	9	9
^	26	19	24	21	12	8	14	12	16	19	18	10
eI	37	29	52	62	69	59	17	20	14	7	14	21
oU	8	8	14	10	11	6	62	50	21	18	72	57

Methods #1 and #2 are described in the text. The items underlined are those for which both methods result in relative durations at least 2% less than the relative durations specified in Table III.

An approach considered was to tailor-make the cutpoints to fit each vowel. This would insure that no steady-state information was present in the transition segments. This approach was not used for several reasons. A major concern was the reliability of the measures of the near-horizontal transitions. In several case, F2 was horizontal, and movement measures were based on F1 or F3. This would mean that the major cue, F2, was already at its target. As well, the brevity of the durations of these transitions was such that less than 20% of the total stimulus would be presented in some cases, thus possibly making the task too difficult. Finally, it was felt that since the effects of the subgroup of vowels (/U,u,ou,^/) could be separated and examined afterwards, using different cutpoints for each vowel was unnecessary. Two changes were made though. Firstly, since the rationale for the intermediate grouping (/i,u/) used in Strange et al. (1983) was not apparent, its members were maintained as an intermediate group for the sake of comparison but also included in their correct grouping (based on Lehiste & Peterson, 1961), with the other long simple nuclei vowels, resulting in a double grouping for them. Secondly, the long complex nuclei were treated as a separate group. The divisions for /eI/ had a 30% onglide portion and a 10% offglide portion, while for /oU/, it was the reverse: 10% for the onglide and 30% for the offglide. These percentages were chosen with reference to the values obtained from Lehiste and Peterson (1961) as well as to those obtained from the present speaker's vowels across contexts.

In summary then, the cutpoints that were chosen were based on Lehiste and Peterson (1961), Strange et. al (1983), and the present speech materials. The vowels were separated into four groups; short nuclei (/U,^,E,I/), intermediate nuclei (/i,u/), long simple nuclei (/a,ae,i,u/), and long complex nuclei

(/eI,oU/). The percentages used are detailed in Table III. These percentages are not perfect matches for the actual relative durations of each vowel stimulus. The mismatches that result in some steady-state information being included in the transition portions are underlined in Table II.

Table III

Relative Durations (in % of total vowel duration) of Initial and Final Transitions Chosen for Each Vowel

VOWEL GROUP	MEMBERS	INITIAL	FINAL
short	/U/, /^/, /E/, /I/	15	35
"intermediate"	/i/, /u/	15	25
long simple	/a/, /ae/, /i/, /u/	15	20
long complex	/eI/	30	10
long complex	/oU/	10	30

The groups and durations shown here, other than the intermediate group, are based on descriptions by Lehiste and Peterson (1961). The intermediate group is based on Strange et. al (1983).

The $V_1CV_2CV_1$'s were reduced to $-CV_2C-$'s and each vowel (V_2) was then cut, according to its category. The stimuli for the Control condition (CO) remained unaltered after being reduced to /bVb/. The stimuli for the Steady-State Only condition (SS) was composed of only the center portion of the vowel. The stimuli for the Transitions Only condition (TR) were made up of the initial and final transitions, separated by a silent interval, whose duration was equal to the duration of the absent center portion. The composition of the stimuli were such that any timing cues available from the relative transition and target durations were still present all the experimental conditions.

The relative intensity of the vowels, as measured on the control $-CV_2C-$'s, was then compared to the values found in Lehiste and Peterson (1959). Several recorded vowels showed significantly atypical amplitudes, possibly as a result of the speaker's efforts to maintain distinctions in the face of the articulatory constraints that had to be met. Those vowels not falling into a high, medium, low intensity classification according to Lehiste and Peterson (1959) were adjusted by replacing the original amplitude with the group mean, thus achieving a better agreement with Lehiste and Peterson's data, as seen in Table IV.

Table IV

Vowel Intensities (in dB) after adjustment according to Lehiste and Peterson (1959)

INTENSITY	V	CONTEXT		
		aCVCa	uCVCu	iCVCi
High	a	77.1	77.1	79.1
	oU	78.2	77.6 (75.8)	79.6
	^	77.6 (80.0)*	77.1	77.6
	ae	78.7	77.6	77.6
	E	76.5	74.1	77.1 (80.0)
Med.	eI	75.8	74.6	74.1
	U	74.6	76.5	77.6
	u	74.6	71.6	72.6
	I	73.6	74.6 (78.7)	74.1 (80.0)
Low	i	71.6	71.6 (73.6)	72.6

The values in parentheses are the original intensities of the vowels not fitting into the high, medium, low distribution of Peterson and Lehiste (1959). Their adjusted values are listed in the regular columns.

The edited stimuli were lowpass filtered at 2.5 kHz and recorded onto audiotapes on a reel-to-reel Revox A77 tape recorder for presentation to the subjects.

4.2 Experimental Conditions

There were three experimental conditions, corresponding to the three sets of stimuli described above: 1) Control (CO); 2) Steady-State Only (SS); 3) Transition Only (TR). Within each condition, each of the ten vowels occurred in three contexts, resulting in a total of 30 stimulus types. In addition, for the second and third conditions, two of the vowels (/i,u/) were modified in two ways (see Table III), resulting in twelve vowel types and three contexts, for a total of 36 stimulus types.

Each stimulus type was repeated five times for each condition. Each condition was preceded by five buffer items which were not scored. This resulted in a total of 155 items for the Control Condition (CO) and 185 items for the Steady-State Only (SS) and the Transitions Only (TR) conditions.

For each condition, presentation order of the stimuli was randomized with the constraint that each vowel had to occur once in each block of 30 (or 36 items, depending on condition), and no vowel (V_8) could occur twice consecutively.

4.3 Subjects

Listeners were 20 graduate students, 19 women and one man, from the speech science department. They were not paid for their time. Prior to the experiment, each was questioned on his or her language background and hearing ability.

All participants (speaker, transcribers, and listeners) had some phonetics training and were native speakers of (standard) Canadian English, most being from regions of Western Canada.

4.4 Procedure

All subjects were administered all the conditions. Tapes were presented to the subjects individually over Beyer DT48 headphones in an IAC soundproof booth, at 50-60 dB SPL. The subjects were told that the stimuli were electronically modified vowels. They were asked to identify the vowels heard by circling the appropriate keyword. The keywords were: beeb, bib, babe, beb, bab, bob, bub, bobe, buub, boob. A list of corresponding phonetic symbols occurred after every ten items on the answer sheet to aid the subjects in interpreting the orthographic representations. The ten choices were reviewed once before beginning testing. A copy of the instructions to the subject can be found in the appendix.

The testing period, consisting of a practice period and three test periods, with short breaks between, was approximately one and a half hours long. The practice items were always presented first. Presentation order for the test conditions was balanced across subjects.

The ten item practice period was composed of a subset of the control stimuli. Each vowel type occurred once, but the choice of context was random. Feedback was not provided.

Items were presented at an ISI of 4 seconds, with an 8 second gap occurring after every block of ten. A 1000-Hz beep was inserted in the center of the gap to help subjects keep their place on the answer sheets, since the stimuli were not preceded by any spoken numbers.

Chapter 5

RESULTS

5.1 Organization of the Data

A de-randomization program was written to sort the subjects' responses. After the data was entered and de-randomized, it was collapsed into a number of confusion matrices to allow examination of the pattern of responses, using subject, context and condition as the independent variables. The matrices were initially organized as 10x10 grids (ten stimuli by ten responses (10x10) for the CO condition, and 12x10 for the SS and TR conditions. The 12x10 matrices were actually the result of collapsing two 10x10 matrices, each identical except for the presence of a different version (intermediate or long grouping) of /i/ and /u/. Because the two versions of /i/ and /u/ were found to yield identical results, the intermediate grouping of Strange et al. (1983) was eliminated, resulting in 10x10 matrices for the SS and TR conditions as well as for the CO condition. The 10x10 matrices will be referred to as the unmodified matrices in descriptions to follow. In addition, for reasons to be stated in following sections, /a/ and /^/ results were collapsed, and data regarding /U/ were eliminated. The resulting matrices are referred to as the modified matrices, and are considered to present a clearer summary of the subjects' perceptions than the 10x10 matrices. ANOVAs and Newman-Keuls were then performed on the total correct responses for both the unmodified and the modified matrices to determine any significant effect(s) due to condition, context, interaction or subject. Post-hoc transcriptions (described in the discussion) by four of the subjects were performed later in an attempt to explain some of the error patterns encountered.

The following results and discussion sections involve error rates rather than correct performance rates. Error rates were used because the studies on which aspects of this experiment are based (Strange et al., 1983; Jenkins et al. 1983) reported their findings in terms of error rates. In the confusion matrices of Table V, correct responses are located on the diagonal, and errors are responses occurring anywhere outside of the diagonal.

5.2 Performance over Conditions and Contexts

Table V shows the nine unmodified confusion matrices for condition and context, each summated over all the subjects. The total correct responses, calculated by summing the responses from the diagonals of the confusion matrices in Table V were subtracted from the total possible. The values remaining are the error rates and are shown for each condition and context in Table VI.

Table V

Summated Confusion Matrices for All Subjects

(1a) CO - /a/

stim \ resp	i	I	eI	E	aE	a	ʌ	oU	U	u
i	79	13	8	-	-	-	-	-	-	-
I	-	83	-	17	-	-	-	-	-	-
eI	-	-	96	4	-	-	-	-	-	-
E	-	1	-	97	2	-	-	-	-	-
aE	-	-	-	-	74	15	11	-	-	-
a	-	-	-	-	-	3	97	-	-	-
ʌ	-	-	-	-	-	36	63	-	1	-
oU	-	-	-	-	-	-	-	100	-	-
U	-	-	-	-	-	-	-	-	100	-
u	-	-	-	-	-	-	-	-	2	98

(2a) SS - /a/

stim \ resp	i	I	eI	E	aE	a	ʌ	oU	U	u
i_1	81	10	9	-	-	-	-	-	-	-
i_2	79	19	2	-	-	-	-	-	-	-
I	-	91	-	9	-	-	-	-	-	-
eI	1	17	72	10	-	-	-	-	-	-
E	-	-	-	100	-	-	-	-	-	-
aE	-	-	-	-	90	8	2	-	-	-
a	-	-	-	-	1	14	85	-	-	-
ʌ	-	-	-	-	-	42	58	-	-	-
oU	-	-	-	-	-	-	-	100	-	-
U	-	-	-	-	-	-	6	-	94	-
u_1	-	-	-	-	-	-	-	1	16	83
u_2	-	-	-	-	-	-	-	2	8	90

(3a) TR - /a/

stim \ resp	i	I	eI	E	aE	a	ʌ	oU	U	u
i_1	4	9	85	-	-	-	-	-	-	-
i_2	4	13	82	-	-	-	-	-	-	-
I	-	13	-	87	-	-	-	-	-	-
eI	-	2	72	25	-	-	-	-	-	-
E	-	-	-	50	49	1	-	-	-	-
aE	-	-	-	2	56	17	25	-	-	-
a	-	-	-	-	-	6	92	-	-	-
ʌ	-	-	-	-	-	48	51	1	-	-
oU	-	-	-	-	-	-	-	100	-	-
U	-	-	-	-	-	-	2	4	94	-
u_1	-	-	-	-	-	-	1	3	32	64
u_2	-	-	-	-	-	-	-	1	29	70

(1b) CO - /u/

stim \ resp	i	I	eI	E	ae	a	ʌ	oU	U	u
i	98	2	-	-	-	-	-	-	-	-
I	-	99	-	1	-	-	-	-	-	-
eI	-	-	99	-	1	-	-	-	-	-
E	-	-	-	100	-	-	-	-	-	-
aE	-	-	-	-	100	-	-	-	-	-
a	-	-	-	-	-	22	74	-	-	-
ʌ	-	-	-	-	-	39	61	-	4	-
oU	-	-	-	-	-	-	-	100	-	-
U	-	-	-	-	-	5	90	-	5	-
u	-	-	-	-	-	-	-	-	5	95

(2b) SS - /u/

stim \ resp	i	I	eI	E	ae	a	ʌ	oU	U	u
i_1	98	1	1	-	-	-	-	-	-	-
i_2	100	-	-	-	-	-	-	-	-	-
I	-	98	-	2	-	-	-	-	-	-
eI	-	-	100	-	-	-	-	-	-	-
E	-	-	-	100	-	-	-	-	-	-
aE	-	-	1	-	97	2	-	-	-	-
a	-	-	-	-	1	21	77	-	-	-
ʌ	-	-	-	-	2	37	61	-	-	-
oU	-	-	-	-	-	-	-	100	-	-
U	-	-	-	-	-	3	94	-	3	-
u_1	-	-	-	-	-	-	-	1	4	95
u_2	-	-	-	-	-	-	-	1	1	98

(3b) TR - /u/

stim \ resp	i	I	eI	E	ae	a	ʌ	oU	U	u
i_1	66	7	27	-	-	-	-	-	-	-
i_2	60	13	26	1	-	-	-	-	-	-
I	-	70	-	30	-	-	-	-	-	-
eI	-	-	99	1	-	-	-	-	-	-
E	-	-	-	77	23	-	-	-	-	-
aE	-	-	-	5	89	3	3	-	-	-
a	-	-	-	-	2	14	84	-	-	-
ʌ	-	-	-	-	-	38	61	-	1	-
oU	-	-	-	-	-	-	-	91	-	-
U	-	-	-	-	-	20	77	-	3	-
u_1	-	-	-	-	-	-	-	-	-	100
u_2	-	-	-	-	1	-	-	-	-	99

Table V (cont.'d)

(1c) CO - /i/

resp / stim	i	I	eI	E	ae	a	ʌ	oU	U	u
i	100	-	-	-	-	-	-	-	-	-
I	2	97	-	1	-	-	-	-	-	-
eI	1	-	99	-	-	-	-	-	-	-
E	-	-	-	100	-	-	-	-	-	-
aE	-	-	-	-	100	-	-	-	-	-
a	-	-	-	-	-	89	9	2	-	-
ʌ	-	-	-	-	-	9	91	-	-	-
oU	-	-	-	-	-	-	-	100	-	-
U	-	-	-	-	-	-	-	-	100	-
u	-	-	-	-	-	-	-	-	1	99

(2c) SS - /i/

resp / stim	i	I	eI	E	ae	a	ʌ	oU	U	u
i_1	98	2	-	-	-	-	-	-	-	-
i_2	100	-	-	-	-	-	-	-	-	-
I	-	100	-	-	-	-	-	-	-	-
eI	24	-	76	-	-	-	-	-	-	-
E	-	-	-	100	-	-	-	-	-	-
aE	-	-	-	-	99	1	-	-	-	-
a	-	-	-	-	-	98	2	-	-	-
ʌ	-	-	-	-	-	10	90	-	1	-
oU	-	-	-	-	-	-	-	99	-	1
U	-	-	-	1	-	-	9	-	90	-
u_1	-	-	-	-	-	-	1	3	-	96
u_2	-	-	-	-	-	-	-	2	-	98

(3c) TR - /i/

resp / stim	i	I	eI	E	ae	a	ʌ	oU	U	u
i_1	90	1	9	-	-	-	-	-	-	-
i_2	86	-	13	-	-	-	1	-	-	-
I	-	95	-	5	-	-	-	-	-	-
eI	-	-	98	2	-	-	-	-	-	-
E	-	-	-	84	15	1	-	-	-	-
aE	-	-	-	2	97	1	-	-	-	-
a	-	-	-	-	-	66	34	-	-	-
ʌ	-	-	-	-	1	22	77	-	-	-
oU	-	-	-	-	-	-	-	99	-	1
U	-	-	-	1	-	-	6	-	93	-
u_1	-	-	-	-	-	-	-	-	2	98
u_2	-	-	-	-	-	-	-	-	6	94

(1d) CO - 3 contexts

resp / stim	i	I	eI	E	ae	a	ʌ	oU	U	u
i	277	15	8	-	-	-	-	-	-	-
I	2	279	-	19	-	-	-	-	-	-
eI	1	-	294	-	-	-	-	-	-	-
E	-	1	-	297	2	-	-	-	-	-
aE	-	-	-	1	273	15	11	-	-	-
a	-	-	-	-	-	114	180	2	4	-
ʌ	-	-	-	-	-	84	215	-	1	-
oU	-	-	-	-	-	-	-	299	-	-
U	-	-	-	-	-	5	90	-	205	-
u	-	-	-	-	-	-	-	-	8	292

(2d) SS - 3 contexts

resp / stim	i	I	eI	E	ae	a	ʌ	oU	U	u
i_1	277	13	10	-	-	-	-	-	-	-
i_2	279	19	2	-	-	-	-	-	-	-
I	-	289	-	11	-	-	-	-	-	-
eI	25	17	248	10	-	-	-	-	-	-
E	-	-	-	300	-	-	-	-	-	-
aE	-	-	1	-	286	11	2	-	-	-
a	-	-	-	-	2	133	164	-	-	-
ʌ	-	-	-	-	2	89	209	-	-	-
oU	-	-	-	-	-	-	-	299	-	-
U	-	-	-	1	-	3	109	-	187	-
u_1	-	-	-	-	-	-	1	5	20	274
u_2	-	-	-	-	-	-	-	5	9	286

(3d) TR - 3 contexts

resp / stim	i	I	eI	E	ae	a	ʌ	oU	U	u
i_1	160	17	121	2	-	-	-	-	-	-
i_2	150	26	121	2	-	-	1	-	-	-
I	-	178	-	122	-	-	-	-	-	-
eI	-	2	269	27	-	-	-	-	-	-
E	-	-	-	211	87	2	-	-	-	-
aE	-	-	-	9	242	21	28	-	-	-
a	-	-	-	-	2	86	210	-	1	-
ʌ	-	-	-	-	1	108	189	1	1	-
oU	-	-	-	-	-	-	-	290	-	10
U	-	-	-	1	-	20	85	4	190	-
u_1	-	-	-	-	-	-	1	3	34	262
u_2	-	-	-	-	1	-	-	1	35	263

The summated confusion matrices for the twenty subjects are printed above. The matrices are coded by condition and context. The SS and TR conditions are asymmetrical as a result of collapsing the 10x10 confusion matrices occurring for each of the two groupings of /i/ and /u/ (#1 = intermediate and #2 = long) into 10x12 matrices for each condition and context combination.

Table VI

Error Rate (in %) across Conditions and Contexts

	CONTEXT		
CONDITION	/a/	/u/	/i/
CO	20.4	22.0	3.2
SS	21.2	22.0	5.0
TR	48.4	33.4	11.6

The values from Table VI show error rates as high as 20% for the CO condition and over 33% for two of the three contexts in the TR condition. This is actually misleading, because, on examination of the confusion matrices of the control condition, it is apparent that the low scores are primarily caused by two factors: (1) confusion of /a/ and /^/ with each other (264/271 or 97% of the total misclassifications for both, from matrice (1d) of Table V) and; (2) misclassification of /U/ in the /u/ context as /a/ or /^/ (95/100 or 95% of the total presentation times for /U/, from matrice (1b) of Table V).

The confusion of /a/ and /^/ appears to result from a similarity in production, arising from the lack of contrastivity in the dialect of speaker and subjects, since the confusion was consistent over subjects, few other confusions occurred, and spontaneous subjective judgements of difficulty in differentiating /a/ and /^/ were reported. Assman et al.

(1982), who drew subjects from the same dialectal region as those in the present study, used the /^/-/ɑ/ contrast instead.

The extremely regular misclassification of the stimulus /U/ in the /u/ context as an /a/ or an /^/ suggests that this vowel type was poorly produced, although it had been transcribed correctly when presented in its original $V_1CV_2CV_1$ form.

Based on these two rationales, certain adjustments to the matrices were made. The /a/ and /^/ stimuli were collapsed into a single stimulus type. The number of combined responses was divided by two to ensure equivalent weighting compared to the other vowels. All confusions, then, of /a/ and /^/ with each other were counted as correct, and all other errors were maintained (with the absolute scores for both correct and incorrect halved). An example of a result of this manipulation is score of 98 correct and 2 incorrect obtained for /a/+/^/ in the matrice (1b) of Table V.

Next, the vowel type /U/ was eliminated from the analysis, due to the low performance on this vowel type in the /u/ context, even in the control condition. Although performance in the other two contexts was good, /U/ was also eliminated for those contexts. This was done to keep the contexts equivalent in terms of the number of vowel choices thereby reducing the number of statistical asymmetries to be dealt with.

The values remaining yielded the modified totals. The sum of these totals for all subjects for each combination of context and condition was subtracted from the total responses possible. The resulting error rates are shown as percentages in Table VII. They are considered to present a clearer summary of the performance in each of the conditions and contexts than the unmodified matrices and their corresponding scores and error rates.

Table VII

Modified Error Rates (in %) across Conditions and Contexts

	CONTEXT		
CONDITION	/a/	/u/	/i/
Control	9.2	1.5	1.0
Steady-State	10.2	1.9	3.5
Transitions	41.6	14.8	5.2

The modified values show high performance (less than 10% errors) in the CO condition for all contexts: the error rate is 9.2% for the /a/ context, as compared to less than 1.5% for the other two contexts. Performance is similar for the SS condition, with 10.2%, 1.9%, and 3.5% errors in /a/, /u/ and /i/ contexts, respectively. The TR condition shows a marked drop in one of the contexts, with error rates of 41.6% for the /a/ context, as compared to 14.8% for the /u/ context and 5.2%

for the /i/ context.

The statistical analyses were performed on both the modified and unmodified totals for comparison. The ANOVA and the Neuman-Keuls tests had similar results (concerning significance) for condition and interaction effects. This should be expected since the difficulty with /a/ and /^/ occurred across all the contexts and conditions, resulting in generally depressed performance scores. The major difference between the unmodified and the modified data arose in comparisons of performance across contexts; it was caused primarily by the apparent selective difficulty in perceiving the /U/ stimuli. Since the difficulty originated most certainly in the production, not the perception of /U/, the results for the modified totals, where /U/ is not included, will be the ones discussed in detail.

A treatment-by-treatment-by-subject ANOVA was performed on the total correct responses for each condition and context.

The ANOVA showed a highly significant condition effect, a highly significant context effect, and a highly significant interaction effect with F values of 115.42, 107.64 and 49.51 respectively. The results are displayed in Table VIII.

Table VIII

ANOVA Results for both Unmodified Totals

and the Modified Totals

UNMODIFIED TOTALS

SOURCE	SS	df	ms	F	p
Total	9433.20	179	-	-	-
Subjects	376.76	19	-	-	-
Treatment 1 (Conditions)	2408.13	2	1204.07	85.85	p<.01
Treatment 2 (Contexts)	4669.20	2	2334.60	323.88	p<.01
Tr. 1 x Tr. 2	761.47	4	190.37	35.22	p<.01
Error Treatment 1	532.98	38	14.03	-	-
Error Treatment 2	273.91	38	7.21	-	-
Error Tr. 1 x Tr. 2	410.76	76	5.40	-	-

MODIFIED TOTALS

SOURCE	SS	df	ms	F	p
Total	5270.95	179	-	-	-
Subject	238.95	19	-	-	-
Treatment 1 (Conditions)	1723.60	2	861.80	115.42	p<.01
Treatment 2 (Contexts)	1565.63	2	782.82	107.64	p<.01
Tr. 1 x Tr. 2	854.67	4	213.67	49.51	p<.01
Error Treatment 1	283.73	38	7.47	-	-
Error Treatment 2	276.37	38	7.27	-	-
Error Tr. 1 x Tr. 2	328.00	76	4.32	-	-

A Neuman-Keuls analysis was used to examine which pairs were causing the significant differences.

The pair-wise comparison showed that the condition effect, with context collapsed, was caused by significant differences

TR and the SS conditions at $p < .05$ (degrees of freedom (df) beginning at 59 and 3). The CO and the SS conditions were not significantly different up to $p < .05$.

The context effect, with conditions collapsed, was found to be significant for /a/ versus /u/ and /a/ versus /i/, again at $p < .05$, (df beginning at 59 and 3). The contexts /u/ and /i/ were not significantly different.

The interactions were also tested with the Neuman-Keuls (df beginning at 19 and 3). Performance for the TR condition-/a/ context combination was found to be significantly different from every other total at $p < .01$. None of the other comparisons, including those with the /a/ context in the other conditions, were significant at $p < .05$.

From these analyses, it can be seen that TR stimuli are identified significantly more poorly than SS or CO stimuli. The /a/ context was found to be a significant detriment to vowel identification as compared to the /i/ and /u/ contexts. The combination of condition and context resulted in significantly poorer performance in the /a/ context for the TR condition as compared to every other combination of context and condition.

The /a/ context yielded the lowest score in every condition, producing a significant context effect, but not producing significant interaction comparisons except in the TR condition. Based on this, an additional Newman-Keuls was performed to compare conditions, without the influence of the

/a/ context on the scores.

Without the /a/ context, the TR condition was not found to be significantly different from the SS or CO conditions up to $p < .05$ (df beginning at 39 and 3). This indicates that the TR stimuli are perceived as well as the SS or the CO stimuli for the /u/ and /i/ contexts.

The results not only support the sufficiency of transitions as cues to vowel identification, but show, for two out of three contexts, performance level on transitions only stimuli to be at a level comparable to that of control and steady-state only stimuli. A marked context effect, however, qualifies this conclusion by depressing performance in the TR condition for the context /a/.

5.3 Comparison of the two medial /i/ and /u/ groupings

Two different cutpoints had been used for the /i/ and /u/ vowel nuclei: one conforming to the Strange et al. (1983) and the Jenkins et al. (1983) intermediate grouping, and one placing these two vowels in Lehiste and Peterson's (1961) long simple nuclei grouping. Performance was very similar for the two versions of /i/ and /u/ in each condition (see Table IX). Comparison of performance on these items with a t-test for dependent means found no significant differences at $p < .05$ for either the SS or the TR conditions.

Table IX

Comparison of Error Rates (in %) for

Two Groupings of /i/ and /u/

	SS condition		TR condition	
VOWEL	intermed.	long	intermed.	long
/i/	7.7	7.0	46.7	50.0
/u/	8.7	8.0	12.7	12.3

Based on these findings, the stimuli from the intermediate grouping were dropped from further analysis.

5.4 Item Analysis

The error distribution was not uniform over the ten vowel types as can be seen in the confusion matrices of Table V. Errors tended to be one to two vowels away around the vowel "loop", when plotted in an F1/F2 graph, with confusions occurring for spectrally similar members. Errors were not limited to confusions within a single vowel category (e.g. tense vowels). Error rates for each vowel type are displayed in Table X.

Table X

Error Rates (in %) as a Function of Vowel Type for All Conditions and Contexts

CO condition

V	/a/	/u/	/i/
i	21	2	0
I	17	1	3
eI	4	1	1
E	3	0	0
ae	26	1	0
a	97(1)	78(2)	11(1)
^	37	39	9
oU	0	0	1
U	0	95	0
u	2	5	1

TR condition

V	/a/	/u/	/i/
i	96	40	14
I	87	30	5
eI	28	1	2
E	50	23	16
ae	44	11	3
a	94(2)	86(2)	33(1)
^	49	39	23
oU	0	9	1
U	6	97	7
u	30	1	6

SS condition

V	/a/	/u/	/i/
i	21	0	0
I	9	2	0
eI	28	0	24
E	0	0	0
ae	10	3	1
a	86(1)	79(2)	2(0)
^	42	39	10
oU	0	0	1
U	6	97	10
u	10	2	2

The numbers in parentheses the error rates after collapsing /a/ and /^/ into a single vowel type. The original scores are in the regular columns.

The error pattern varied somewhat over conditions and contexts. Consistently good performance was obtained only for the /oU/ and /u/ vowels. For the TR condition-/a/ context combination, where the majority of errors occurred, the highest error rates were for /i/ (96%), /I/ (87%), and /E/ (50%). Of these errors, the majority of confusions were /i/ to /ei/ (82%, with 14% other errors and 4% correct), /I/ to /E/ (87%, with 13% correct), and /E/ to /ae/ (49%, with 1% other errors and 50% correct). These errors, as well as those for the other vowels, are quite consistent, suggesting that the elimination of steady-state information does not result in ambiguous stimuli, but in categorical changes in vowel quality, for example, /i/ in the /a/ context becomes /ei/ upon removal of the steady-state portion of the vowel.

Chapter 6

DISCUSSION

6.1 Vowel Cues Available in TR Stimuli

Performance in two of the three contexts of the TR condition was as good as performance in the CO and SS conditions. If good performance resulted in the TR condition, it was originally hypothesized to be due to the sufficiency of transition information as cues to vowel identity. Examination of other possible reasons should occur before concluding that high performance on stimuli without steady-states is actually due to the presence of the transitions. Sources of information still remaining in the edited stimuli other than transitions could play a significant role in the identification process. This information includes duration, diphthongization (formant movement), and intensity differences.

If duration is an important cue, then its maintained presence should aid differentiation of two groups of vowels: long (tense) and short (lax). Examination of the TR matrix (3d) of Table V reveals that the confusions of the vowel types with the highest error rates, namely /i/ to /ei/ and /I/ to /E/, are both within the short vowel group, or within the long vowel group, but matrix (3d) also shows that many of the other confusions are between these two groups. Duration may be a useful cue, but its presence here does not prevent tense-lax confusions. This finding is further supported by the finding of Strange et al. (1983) that partial neutralization of duration does not significantly affect perception of vowels with the steady-state removed.

Formant movement resulting from diphthongization is difficult to separate from that resulting from consonant-vowel transitions. The obvious diphthongs (/ai,oi,aU/), described originally by Lehiste and Peterson (1961), were not included in the present vowel set. Glided vowels (/ei,oU/), however, were used. For the these vowels, the glide and the consonant-vowel transition were not separated; the cutpoints were adjusted to encompass the entire period of movement. As a result, these stimuli had the additional initial or final gliding cue available. Examination of performance (see matrice (3a), Table V) shows that the vowels /ei/ and /oU/ had among the best identification scores in the critical combination (where overall performance was worst): TR condition-/a/ context. The gliding information present in the transition segments may, therefore, have been useful in identification of the TR stimuli.

Normal intensity cues were maintained and even provided wherever the typical relative intensity did not occur at production of the vowels. This information is present in all the conditions, and could then be used to identify the TR vowels. Whether it is used or not cannot be determined in this design, and no published work on its utility as a cue could be found.

In summary, the TR stimuli have the additional cues of duration, glide, and intensity. Duration information does not appear to have affected the error patterns. Formant movement information contained in the glides present may have aided differentiation of glided and non-glided vowels. Intensity cues are present, but whether or not they are used is not discernable in this experimental design.

Although possibly useful, the presence of these cues (duration, glide and intensity), does not provide an adequate explanation of performance in the TR condition, especially for the contexts showing performance equal to that in the other conditions. Therefore, the remaining cue, transitions, already shown to play a role in vowel identification (Lindblom & Studdert-Kennedy, 1967; Strange et al., 1976; Strange et al., 1983; Jenkins et al., 1983), may reasonably be invoked as an explanation for the good performance in the TR condition.

Transitions, then, could be said to be the major cues available in the TR stimuli. They provide sufficient information to cue vowel identity and are, most of the time, sufficiently robust to do so even in the presence of the transconsonantal coarticulation arising in a $V_1CV_2CV_1$ utterance. The utility of transitions, however, is not uniform over all vowel contexts. In the presence of coarticulation from surrounding /a/'s, the indentification of vowels cued by transitions alone is considerably impaired, whereas for vowels embedded in /i/ or /u/ contexts, the transitions succeed equally well in cueing the identity of the vowel as the information available in the unmodified vowel (V_2) excised from its $V_1CV_2CV_1$ utterance.

6.2 Problem Vowels

Performance on two of the three contexts in the CO condition was high (less than 2% errors). This can be attributed, in part, to experimental design factors such as the monitoring of production quality, dialectal matching, good listening conditions, and the use of subjects with transcription experience, all factors cited as important in previous studies (e.g. Macchi, 1980; Assman et al., 1982). Reasons for the decreased performance on the /a/ context in the control condition are not obvious, but parallel decreases in performance for this context in the other two conditions. Possible explanations for this effect will be dealt with later. Difficulty occurred with several of the V_2 vowels (those to be identified) in the control condition, despite the precautions taken. Possible reasons for this shall be discussed in this section.

The monitoring of production through speaker and experimenter judgements, as well as through unanimous agreement from three transcribers was still not sufficient to insure the intended responses for all vowel types in the CO condition from all the subjects. Consistently differing perceptions from those of the evaluators were obtained for the /a/ and /^/ stimuli in all contexts, and for the /U/ stimulus in the /u/ context. Possible explanations for the fact that these items were not judged as anomalous initially include low reliability of the transcriptions and changes in perception from the case where the vowel (V_2) is in the $V_1CV_2CV_1$ context to the case where it is excised from its context.

The possibility of low reliability for the transcriptions was examined first. Four additional transcriptions of the entire $V_1CV_2CV_1$ utterances were made, this time by four listeners who had participated in the forced choice $-CV_2C-$ experiment. The subjects were chosen on the basis of availability, not on their particular performances in the earlier experiment.

Good agreement was obtained for the $V_1CV_2CV_1$ transcriptions and the $-CV_2C-$ forced choice results in that, high error rates occurred in the transcriptions for those vowels designated problematic in the forced choice procedure, and low error rates occurred for the others. Examination of the individual listeners' responses for the problematic vowels (/a/ and /^/ in all contexts, /U/ in context /u/) showed, however, little consistency in their perception over the two situations (see Table XI).

Table XI

Comparison of Identification Responses for

$V_1CV_2CV_1$ Transcriptions and $-CV_2C-$ Forced Choices

for the Problem Vowels

subj.	utt.	/abaaba//ab^ba/		/ubabu//ub^bu//ubUbu/			/ibabi//ib^bi/	
#1	$-CV_2C-$	4/^/	4/^/	4/^/	4/^/	4/^/	3/a/	4/^/
	$V_1CV_2CV_1$	/a/	/a/	/a/	/a/	/U/	/a/	/a/
#2	$-CV_2C-$	5/^/	3/^/	5/^/	5/^/	5/^/	5/a/	5/a/
	$V_1CV_2CV_1$	/a/	/a/	/a/	/a/	/U/	/a/	/a/
#3	$-CV_2C-$	4/^/	3/a/	3/a/	5/^/	4/^/	4/a/	4/^/
	$V_1CV_2CV_1$	/a/	/^/	/a/	/^/	/U/	/a/	/^/
#4	$-CV_2C-$	4/^/	4/a/	5/a/	4/^/	4/^/	4/a/	3/a/
	$V_1CV_2CV_1$	/^/	/ae/	/a/	/^/	/U/	/a/	/U/

The underlined vowel in the first line is the vowel to be identified. The forced choice $-CV_2C-$ situation involved five repetitions of each vowel type in this position. Only the modal response is recorded here. In almost all cases, the remainder was either /a/ if the modal was /^/, or /^/ if it were /a/. The $V_1CV_2CV_1$ stimuli were only presented once each and the transcriptions for each are recorded here.

The subjects involved in this comparison all had training in transcription. Although the /a/ and /^/ vowel pair is not contrastive in the dialect of the region the subjects were drawn from, it is expected that, because of their training, they should have been able to reliably transcribe the phones. The presence of disagreements for the /a/ and /^/ transcriptions across subjects, and the inconsistent nature of the choices for the five repetitions of each of the $-CV_2C-$'s by each subject suggest that these vowel types were not produced contrastively by the speaker, but were produced sufficiently differently to result in uncertainty and the use of more than one symbol by some of the listeners. The tendency to attempt to fit the vowel productions in more than one category was probably exacerbated by the bias existing from the listeners' knowledge that both /a/ and /^/ were available as possible responses. A number of the twenty subjects in the $-CV_2C-$ experiment spontaneously commented that they had difficulty differentiating these two vowel types, and that the /a/ seemed to lack "openness".

For the /U/ in the /u/ context, the subjects had previously responded with an /a/ or an /^/ to presentations of the vowel in the $-CV_2C-$ form. In the additional $V_1CV_2CV_1$ transcription task, this vowel was transcribed correctly by all four listeners. These four transcriptions, plus those of the original evaluators, provide reliable evidence that this vowel actually had the quality /U/ in the full $V_1CV_2CV_1$ form. If so, this makes the suggestion of a change in vowel quality for V_2 upon removal from the surrounding vowel (V_1) context more tenable.

If vowel quality changes when the vowel is removed from its surrounding utterance, that would suggest that necessary vowel identification information is spread over a segment larger than a CVC syllable. This possibility is consistent with findings on the perception of other types of coarticulation (e.g. Ali et al., 1971; Benguerel & Adelman, 1977) and the well established difficulty encountered in identifying vowels excised from connected discourse as compared to identification of those produced in isolation.

For nine of the ten vowels in the present study, identification of the medial vowel was not affected by the removal of coarticulatory information available in the surrounding vowel context of a $V_1CV_2CV_1$ utterance. For the vowel /U/, however, the transconsonantal coarticulation information provided by the frame /uCV_2Cu/ is apparently critical to its perception in the V_2 position.

6.3 The Effect of Cutpoint Placement

The hypothesis to be tested in this study was that the information contained in transitions would not be as effective as that contained in the steady-state in cueing vowel identity. Transition only, steady-state only and unmodified vowel stimuli were used to test this hypothesis. A crucial factor in drawing conclusions from the identification of these stimuli is whether or not steady-state information was present within the "transition only" vowels.

In chapter four, it was stated that, for /U/, /u/, /oU/, and /^/, difficulty was encountered in determining the boundary between steady-state and transition because of the nearly horizontal nature of the transitions. For those four vowels, the cutpoint placement resulted in steady-state information being present in the transition segments of the TR stimuli. For the glided vowel /oU/, the low percentage used (10%) for the near-flat transition reduced this concern, but some steady-state information was still present. Presence of this "target" information in the transition segments could be expected to lower error rates in vowel identification performance in the TR condition.

Examination of the performance of /u/,/U/,/^/, and /oU/ in the TR condition in Table X, reveals that these vowels do, in fact, show lower error rates than do the other vowel stimuli. The vowel /^/ did not show good performance, and had to be collapsed with /a/, but their combined error rate was low (2% errors over all contexts). The vowels /U/ (in contexts /i/ and /u/), /u/, and /oU/ are among the best identified vowels, although the presence of the additional glide cue for /oU/ must be included in the explanation of good performance.

Good performance on the vowels described, however, may be explained by reasons other than the presence of "target" information, such as the glide cue mentioned, the salient position of /u/ in the vowel triangle (as an extreme value), and the lack of close neighbours for /^/ because of the collapsing of /a/ and /^/. If the presence of steady-state information is the critical factor in aiding identification in the TR vowel stimuli, then any vowels displaying steady-state patterns within the designated transition segments should be expected to show correspondingly lower error rates, while those not possessing this "target" information should show higher rates of error.

The vowels showing the highest error rates in the TR condition, and contributing most to the poor performance were /i/,/I/,/E/, and /ae/ with error rates of 19% or above. All showed a clear increase in errors from the CO and SS conditions to the TR condition. If these particular vowels are examined across contexts, it can be seen that in the /i/ context, performance is comparable to the other vowels and to the other conditions. In the /u/ context, performance is poor for /i/,/I/, and /E/ at error rates of at least 23%. Almost all the vowels (/i/,/I/,/E/,/ae/,/ei/,/u/) in the /a/ context are poorly identified (errors at or above 28%).

If one examines the degree of match between the actual relative durations of the transitions in Table II and the relative durations chosen to base the cutpoints used on (Table III), the /u/ context shows the greatest number of mismatches (items underlined in Table II), with five vowels (/a/,/E/,/U/,/^/,/oU/) having some steady-state in their transition segments. These particular vowels are not, however, the vowels with the lowest identification error rates. As well, despite containing a large number of vowels with mismatches between actual and chosen relative durations (where steady-state information is present in the transition segments), /u/ is not the best perceived context. Furthermore, in the /a/ and /i/ contexts, the particular mismatches occurring do not match up with the pattern of performance in the identification task. As well, the mismatches in the /a/ and /i/ contexts are similar in number and pattern, yet performance for each differs greatly.

An additional source of support for transitions as cues to vowel identity was the identification performance for those vowels for which the mismatch went in the opposite direction, resulting in incomplete transitions for either the initial or final segments. A comparison of the actual relative durations of Table II and the relative durations chosen (Table III) show almost all the vowels having one or both of the transition segments cut too short. Performance in the contexts /i/ and /u/, for the TR condition, shows no decrease in identification rates as a result of this early truncation; performance is as good as in the SS and CO conditions.

Another point of comparison is the identification performance for the two versions of /i/ and /u/. Two different sets of cutpoints were used for the vowels /i/ and /u/, based on how the vowels were grouped. The intermediate grouping resulted in a mismatch for the vowel /i/ in the /i/ context, with steady-state information present in the transition segment, but the error rates for the two versions were similar (10% and 14%). As well, differing degrees of transition present in the TR condition for the two versions of /i/ and /u/ in the other condition-context combinations, as a result of the

20% and 25% final relative durations used, do not result in different performance. The presence of completed transitions therefore, does not appear to be crucial for identification.

The lack of relation between the location used for the cutpoints and identification performance suggest that the presence or absence of steady-state information does not greatly affect vowel perception, and that completed transitions are not required for identification.

6.4 Poor Performance in the /a/ Context

Performance in the /a/ context is consistently below performance in the other contexts. The difference is not statistically significant for the CO or the SS conditions, but is still present. For the TR condition, a significant increase in errors from those in the CO and SS conditions occurs. Possible factors to be considered include poor production quality, coarticulatory constraints, cutpoint placement and subject-specific performance.

The production of /a/ in the V_2 position was shown earlier to have been non-contrastive with /^/. It is possible that the /a/ was poorly produced in the outer vowel position too. Only one of the original three evaluators took exception to the choice of /a/, /u/, or /i/ as a transcription of the outer vowel (she transcribed the outer /a/ as /^/), but an /a/ bias was introduced when a set of the intended productions was provided to the three evaluators. If the /a/ vowel was actually produced as an /^/, it still does not explain the high error rate of the "/a/" context since the transconsonantal coarticulation would occur regardless of whether the original context was that of an /a/ or an /^/. Öhman (1966) did not specifically examine /^/ coarticulation, but there is no reason to think that it would not occur. If, for some reason, it should not occur, the effect would be better performance in the TR condition - /a/ context (actually /^/ context) since potentially less "distortion" would be introduced. Quality of production should not, therefore, be a factor in the poor identification performance for vowels produced in the /a/ context.

Place of production affects the spectral pattern of the vowels. There is the possibility that /a/, as the only low vowel tested in the outer position of the $V_1CV_2CV_1$ utterance, has an effect on the transconsonantal vowel (V_2) and its transition that is somehow more damaging than the effect of the high vowels. This assumes, of course, that the /a/ was actually produced as a low back vowel, not as the more central vowel /^/. If the /a/ was actually articulated, greater coarticulatory effects might be expected if the coarticulation was extreme, for example, from a low to a high and back to a low vowel, like /abiba/. However, the error pattern in this context does not reveal difficulty only with high vowels (see Table IX); in the TR condition, the highest error rates occur for the high front vowels /i/ and /I/, but high error rates also occur throughout the vowel set. In the other conditions, the errors are less widespread; /i/ and /I/ are still among the highest, but /ae/ (low front) in the CO and /ei/ (mid front) in the SS conditions also show error rates as high or higher. Articulatory movement from low to high to low does not, therefore, show a consistent correlation with poor identification. Coarticulatory difficulties stemming from the reverse pattern (high-low-high) do not result in higher error rates either, as evidenced by the good identification performance obtained for low vowels produced in /i/ and /u/ contexts. Production constraints do not appear to result in perceptual ambiguity, and do not account for the high error rates for the vowels in the /a/ context.

Articulatory factors do not provide clear evidence to account for the performance obtained in the /a/ context. It is possible that articulatory patterns may result in subtle acoustic changes that result in specific identification difficulties. The acoustic patterns of the medial vowels were, therefore, examined for unusual characteristics. The spectrograms for each vowel type varied slightly across contexts, presumably as a result of the transconsonantal coarticulation from the outer vowels. Measures of transition onset and offset frequencies, and transition slope angles were taken. There were no large or consistent differences for any of the vowel types, taken individually or as a whole, across contexts. Acoustically, then, the context /a/ does not appear to stand out from the other two contexts.

Performance in the /a/ context was most notably impaired in the TR condition. These results might be an artifact of cutpoint placement caused by inflated rates for the /i/ and /u/ contexts due to the presence of steady-state information in the transition segments. This possibility was dealt with earlier (section 6.3) in an attempt to explain the pattern of performance over vowel types. Contexts /a/ and /i/ had very similar patterns of mismatch between actual relative durations and chosen relative durations, yet performance for these two differed greatly. As well, the actual error patterns over the vowel types did not match up with the pattern of mismatches over the vowel types. The presence or absence of steady-state information thus does not explain the performance variation across contexts.

The detrimental effect of the /a/ context on medial vowel identification is strong, but the possibility that the pattern of errors may have been caused by a few subjects and be, in fact, idiosyncratic, was examined. Subjects were rank ordered by the average performance across conditions and separate ANOVAS on the top and bottom ten subjects were performed. The results from this analysis were the same as that for all the subjects together, showing a context effect for /a/. Thus, the lower performance on vowels in the /a/ context is a consistent effect over the twenty subjects tested.

None of the factors considered account for the poor performance of subjects in the /a/ context, or, more specifically, in the TR condition for this context. This effect, although robust, is unexplainable by the information examined so far. Until explained, the performance in the TR condition-/a/ context presents a constraint on concluding that transitions are sufficient to cue vowel identity. This constraint becomes more severe with the possibility that the intended low back /a/ was actually produced as the more central /^/; the likelihood of other contexts showing similar patterns increases as the vowel in question becomes less articulatorily "extreme". However, the fact that performance in the /a/ context was also impaired in the other conditions, albeit to a lesser and non-significant extent, suggests that the identification difficulty may not be due to the insufficiency of transitions as cues but rather to some essential ambiguity about vowels produced in the /a/ context, which is increased when contiguous chunks as large as the steady state (50%-65% of the total duration) are removed from the vowel.

6.5 Summary and Conclusions

This study has examined the effect of transconsonantal coarticulation on the usefulness of transitions as cues to the perception of the medial vowel (V_2) in $V_1CV_2CV_1$ utterances, where V_1 was one of three vowels (/a/,/u/, or /i/), C was the plosive /b/, and V_2 was one of ten vowels (/i/, /I/, /eI/, /E/, /ae/, /a/, /^/, /oU, /U/, /u/).

For two of the three V_1 contexts, the vowels in $-CV_2C-$ stimuli with the V_2 steady-state removed (TR condition) were identified equally as well as vowels in $-CV_2C-$ stimuli with the transitions removed (SS condition) and those in unmodified-CV_2C- stimuli (CO condition). For these contexts, information from transitions alone is sufficient to cue vowel identity at a level equal to that of the vowel cue formly considered to be the most important, steady-state information.

A major limitation on concluding that transitions cue vowel identity equally well as steady-state or combined information results from the uneven performance occurring across contexts. Vowels in an /a/ context were identified consistently more poorly than those in /u/ or /i/ contexts. This effect was not statistically significant for the CO or SS condition, but was highly significant for the TR condition. Various explanations of the effect, including quality of production, cutpoint placement, and articulatory constraints, were considered. None were found to be adequate. A major difficulty in accounting for this effect was the small number of contexts used. Further understanding of this pattern of performance requires a more systematic, comprehensive examination of the effect of coarticulation by other vowel contexts.

It can be concluded that, in $V_1CV_2CV_1$ utterances, where V_1 is /u/ or /i/, the medial vowel (V_2), with steady-state information removed, can be identified well. In this situation, transitions can, depending on the nature of the transconsonantal coarticulation present, cue identity equally well as steady-state information alone or even information from steady-state and transitions combined. The usefulness of transitions as cues to vowel identity in the presence of transconsonantal coarticulation, however, cannot be generalized to all vowel contexts without further examination of the relative uniqueness of the effect of the /a/ context. The reasons for this effect are yet to be explained and the range of effects of vowel context are yet to be charted, so definitive claims about the sufficiency of transitions alone as cues to vowel identity cannot be made. Whether or not transitions are sufficient in themselves to cue vowel identity, it can still be concluded, with some surity, that the performance on transition only stimuli strongly supports the potential usefulness of information from transitions in the normal identification situation, where steady-state and transition information are both available.

REFERENCES CITED

Ali, L., Gallagher, T., Golstein, J., and Daniloff, R. (1971). Perception of coarticulated nasality. Journal of the Acoustical Society of America, 49, 538-540.

Assman, P.F., Nearey, T.M., and Hogan, J.T. (1982). Vowel identification: orthographic, perceptual and acoustic aspects. Journal of the Acoustical Society of America, 71, 975-989.

Benguerel, A-P., and Cowan, H.A. (1974). Coarticulation of upper lip protrusion in French. Phonetica, 30, 41-55.

Benguerel, A-P., and Adelman, S. (1977). Coarticulation of liprounding and its perception. In Structure and Process in Speech Perception, A. Cohen and S.G. Nooteboom (eds.), NY: Springer Verlag.

Bond, Z.S. (1975). Identification of vowels excerpted from context. Journal of the Acoustical Society of America, 62 (Suppl. No. 1), S101.

Carney, P.J. and Moll, K.L. (1971). A Cinefluorographic investigation of fricative consonant-vowel coarticulation. Phonetica, 23(4), 193-200.

Clark, M. and Sharf, D.J. (1973). Coarticulation effects of post-consonantal vowels on the short-term recall of pre-consonantal vowels. Language and Speech, 16, 67-76.

Diehl, R.L., Buchwald McCusker, S., and Chapman, L.S. (1981). Perceiving vowels in isolation and in consonantal context. Journal of the Acoustical Society of America, 69(1), 239-248.

Fairbanks, G. and Grubb, P. (1961). A psychophysical investigation of vowel formants. Journal of Speech and Hearing Research, 4, 203-219.

Fromkin, V. (1966). Neuro-muscular specifications of linguistic units. Language and Speech, 9, 170-199.

Fujimura, O. and Ochiai, K. (1963). Vowel identification and phonetic contexts. Journal of the Acoustical Society of American, 35, 1889(A).

Gay, T. (1978). Effect of speaking rate on vowel formant movements. Journal of the Acoustical Society of America, 63(1), 223-229.

Gerstman, L.J. (1968). Classification of self-normalized vowels. IEEE Trans. Audio Electroacoustics, AU-16, 78-80.

Gottfried, T.L. and Strange, W. (1980). Identification of coarticulated vowels. Journal of the Acoustical Society of America, 68, 1626-1635.

Henke, W.L. (1966). Dynamic articulatory model of speech production using computer simulation. Doctoral Thesis, M.I.T.

House, A.S. and Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. Journal of the Acoustical Society of America, 25, 105-113.

Jenkins, J.J., Strange, W., and Edman, T.R. (1983). Identification of vowels in vowel-less syllables. Perception & Psychophysics, 34 (5), 441-450.

Joos, M. (1948). Acoustics Phonetics. Language, 24, 1-136.

Kahn, D. (1977). Near-perfect identification of speaker-randomized vowels without consonantal transitions. Journal of the Acoustical Society of America, 62, Suppl. No. 1, S101.

Kozhevnikov, V.A. and Chistovich, L.A. (1965). Speech, Articulation and Perception (translated from Russian), Joint Publication Research Service, U.S. Dept. Commerce No. 30 (Washington).

Ladefoged, P. and Broadbent, D.E. (1957). Information conveyed by vowels. Journal of the Acoustical Society of America, 29(1),k 98-104.

Lehiste, I. and Meltzer, D. (1973). Vowel and speaker identification in natural and synthetic speech. Language and Speech, 16, 356-364.

Lehiste, I., and Peterson, G.E. (1959). Vowel amplitude and phonemic stress in American English. Journal of the Acoustical Society of America, 31(4), 428-435.

Lehiste, I. and Peterson, G.E. (1961). Transitions, glides and diphthongs. Journal of the Acoustical Society of America, 33(3), 268-276.

Lehiste, I., and Shockey, L. (1972). On the perception of coarticulation effects in English VCV syllables. Working Papers in Linguistics No. 12, The Ohio State University, 78-86.

Lindblom, B.E.F. (1963). Spectrographic study of vowel reduction. Journal of the AcousticalSociety of America, 35, 1773-1781.

Lindblom, B.E.F., and Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. Journal of the Acoustical Society of America, 42, 830-843.

Lintz, L.B. and Sherman, D. (1961). Phonetic elements and perception of nasality. Journal of Speech and Hearing Research, 4, 381-396.

Macchi, M.J. (1980). Identification of vowels spoken in isolation versus vowels spoken in consonantal context. Journal of the Acoustical Society of America, 68(6), 1636-1642.

Millar, J.B., and Ainsworth, W.A. (1972). Identification of synthetic isolated vowels and vowels in h-d context. Acoustica, 27, 278-282.

Moll, K.L. and Daniloff, R.G. (1971). Investigation of the timing of velar movements during speech. Journal of the Acoustical Society of America, 50(2), 678-684.

Neary, T.M. (1977). Phonetic feature systems for vowels. PhD thesis, University of Conneticut, 1977 (reproduced by the Indiana University Linguistics Club, 1978).

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

Öhman, S.E.G. (1967). A Numerical model of coarticulation. Journal of the Acoustical Society of America, 41, 310-320.

Peterson, G.E. (1961). Parameters of vowel quality. Journal of the Acoustical Society of America, 4(1), 10-29.

Peterson, G.E. and Barney, H.L. (1952). Control methods used in a study of the vowels. Journal of the Acoustical Society of America, 24(2), 175-184.

Peterson, G.E. and Lehiste, I. (1960). Duration of syllable nuclei in English. Journal of the Acoustical Society of America, 32 (6), 693-703.

Sharf, D.J. and Ostreicher, H. (1973). Effect of forward and backward coarticulation on the identification of speech sounds. Language and Speech, 16, 196-206.

Skinner, T.E. (1977). Speaker invariant characterizations of vowels, liquids, and glides using relative formant frequencies. Journal of the Acoustical Society of America, 62, Suppl. No. 1, S5.

Stevens, K.N. and House, A.S. (1963). Pertubation of vowel articulations by consonantal context: An acoustical study. Journal of Speech and Hearing Research, 6, 111-128.

Stevens, K.N., House, A.S., & Paul, A.P. (1966). Acoustica description of syllabic nuclei: An interpretation in terms of a dynamic model of articulation. Journal of the Acoustical Society of America, 40(1), 123-132.

Strange, W., Jenkins, J.J., and Johnson, T.L. (1983). Dynamic specification of coarticulated vowels. Journal of the Acoustical Society of America, 74 (3), 695-705.

Strange, W., Verbrugge, R.R., Shankweiler, D.P., and Edman, T.R. (1976). Consonant environment specifies vowel identity. Journal of the Acoustical Society ofAmerica, 60, 213-224.

Verbrugge, R.R., Strange, W., Shankweiler, D.P., and Edman, T.R.(1976). What information enables a listener to map a talker's vowel space? Journal of the Acoustical Society of America, 60(1), 198-212.

Appendix

Instructions to Subject

We're looking at how people perceive vowels, and what parts of the sound are used in vowel identification.

You are going to be presented with electronically modified speech sounds. Listen to the sound carefully, and circle the keyword on this answer sheet corresponding to the vowel you think you heard. There are no right or wrong answers in this task, just circle the keyword that you think best approximates the vowel you heard. Guessing is fine, just be sure to choose one answer every single time. Do not take too long to decide because there is only four seconds between sounds. I will go over the coding system with you before we start.

There will be three sessions of approximately fifteen minutes each. One will be 155 items long, and the others will be 185 items long. There is a pause after every tenth stimulus at which time you will hear a beep. The beep should coincide with the gap on the paper. If it doesn't at any time, say "stop" and I will stop the tape and figure out where you have gone wrong.

Before we start, I'd like to ask you some questions and have you read and sign this consent form. The only risk involved in the experiment is possible boredom.

- go to Questionnaire and Consent form (pen) -

I am now going to play ten sounds to familiarize you with the task. I am not scoring these; it is just to get you used to moving across the sheet and locating the answers. Here is the answer sheet. The practice items are on the back page. The response shoices are written orthographically for those who do not know the phonetic alphabet. For those of you wh do,k you will probably find the orthographic representation confusing initially. That is why I have written the corresponding phonetic symbols at intervals on the response sheet. The vowel key is as follows: /bib/, /bIb/, /beib/, /baeb/, /bab/ as in "baaa" that a sheep would say, /b^b/, /boUb/, /bUb/ as in could, /bub/. Say the sounds to yourself, then tell me when you are ready.

- Do practice items -

Any questions? O.K., now we'll start. The sounds you hear may not sound exactly like those you heard on the practice tape.

Appendix (cont.'d)

Questionnaire

Subject Name Date

Assigned Number Order of Conditions

code answer sheets

1. Have you taken any courses in phonetics or have you experience in transcription?

2. Is your native tongue English?

3. What other languages do you speak fairly fluently?

4. What region were you brought up in, or spent a major part of your life?

5. Do you have normal hearing?

6. Consent form

7. Note exceptional occurrences

Appendix (cont.'d)

Response Sheet* Cond. 1 2 3 No. ______

	/i/	/ɪ/	/eɪ/	/ɛ/	/ae/	/a/	/ʌ/	/oʊ/	/ʊ/	/u/
1.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
2.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
3.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
4.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
5.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
6.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
7.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
8.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
9.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
10.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob

	/i/	/ɪ/	/eɪ/	/ɛ/	/ae/	/a/	/ʌ/	/oʊ/	/ʊ/	/u/
11.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
12.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
13.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
14.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
15.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
16.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
17.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
18.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
19.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob
20.	beeb	bib	babe	beb	bab	baab	bub	bobe	buub	boob

*The spacing of the ten choices has been reduced from that presented to the subjects to fit margin demands.