PERCEPTION OF COARTICULATED LIP ROUNDING

by

SHARON ADELMAN

B.Sc., McGill University, 1972

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in the Department

of

Paediatrics

Division of Audiology and Speech Sciences

We accept this thesis as conforming to the

required standard

THE UNIVERSITY OF BRITISH COLUMBIA

July, 1974

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Paediatrics

The University of British Columbia
Vancouver 8, Canada

Date August 27, 1974

ABSTRACT

The present study investigates the perceivability of coarticulated lip rounding in French. Nine utterances containing the clusters /kstr/,/rstr/, and /rskr/ followed by one of the vowels /i/, /y/, or /u/ in all possible combinations, were truncated at 4 different points before the vowel. Test items in each of the 4 groups therefore contained different amounts of information regarding the nature of the following vowel, due to coarticulatory influences of the vowel on the preceding consonants. Subjects were asked to predict the identity of the missing vowel on hearing the truncated utterances. Subjects were native speakers of either French or English; some of them had a knowledge of phonetics.

Results show that when segments up to and including at least half of the final consonant of the cluster are present, subjects correctly identify the missing vowel well above chance levels. Several individuals were able to identify the vowel even when presented with shorter versions of the utterances. No significant difference in performance was found between French and English subjects, nor between subjects with and without phonetic training. Perceivability of individual features of the missing vowel is discussed.

It is concluded that coarticulatory effects due to lip rounding (as well as to horizontal tongue position) provide perceivable information at a level significantly above chance, and that this information may be used by the perceptual mechanism as an aid in speech sound identification.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENT

I would like to thank all those who have had a part in this thesis:

- Dr. André-Pierre Benguerel for his guidance during the research and writing of the thesis.
- Dr. Joyce D. Edwards for serving on my committee.
- My subjects for their kind cooperation.
- My parents for their encouragement over the past two years.
- Ingrid, Betty, Lynne, Pat, and Meralin, for much friendship.

CHAPTER 1

INTRODUCTION

The production of speech is a complex process, and its complexities necessitate a unique and equally complex perception process. It would be interesting to know whether the subtleties and variations in the production process are noted in, perhaps even necessary to, the speech perception process.

Speech is not merely a sequence of independent sounds produced by independent gestures. As the motor gestures producing speech overlap in time and change with context, so do the acoustic cues in the speech signal. It is on this ever-changing signal that speech perception is based. The listener must abstract the appropriate cues from the mass of acoustic information to correctly identify the signal, to understand spoken language. How he recognizes the appropriate cues, indeed even what these cues may be, is far from completely understood.

In examining speech production, one sees that if an articulator, such as the tongue tip or the velum, is free to move during production of a particular sound, it may initiate movement towards its target position for the subsequent phone, or for a phone several segments ahead. Also, an articulator may still be moving from its position for the preceding phone while a current phone is already being produced. This overlapping of speech gestures in time is referred to as coarticulation. This characteristic of speech production results in one phoneme being acoustically different virtually each time it is

produced. It also means that the units of production overlap to an extent whereby cues for a phoneme may be found several phones preceding and several phones following the one in question. Does a listener make use of these characteristics of speech production in identifying the speech signal? At least, can he make use of them if required to, for example, when other cues are masked or missing? Or are these cues irrelevant to the perception process, merely a by-product of the complex workings of the articulators, without perceptual correlates?

The present study looks at utterances in which cues for a certain vowel are known to exist several phones preceding the vowel. Subjects are asked to predict the identity of the upcoming vowel after hearing only part of the utterance. This study therefore gives an indication as to how much use a listener makes, or at least can make, of coarticulated information in the speech signal.

CHAPTER 2

REVIEW OF THE LITERATURE

2.1 Introduction

Studies of coarticulation have been carried out on several levels. Section 2.2 discusses coarticulation at the acoustic level. Section 2.3 discusses coarticulation at the articulatory level and outlines several theories that have been proposed to explain the phenomenon. Section 2.4 reviews studies on the perceptual correlates of coarticulation. A discussion of the possible units of speech perception is included in this section.

2.2 Coarticulation: The Acoustic Level

Earliest indications of the phenomenon of coarticulation came from acoustic studies. It has long been known that the acoustic value of a vowel is influenced by the vowel's phonetic context. For example, vowel duration, intensity, and fundamental frequency are known to vary with changes in consonantal environment [House and Fairbanks, 1953].

Stevens and House [1963] examined changes in vowel formant frequency and formant bandwidth with context. Three speakers produced various /h$\text{ə}C_1VC_2$/ utterances, in which C is a consonant and V is a vowel. In these utterances $C_1 = C_2$. When the first formant frequency (F_1) was plotted against the second formant frequency (F_2) for each

vowel, it was seen that quite appreciable differences occurred with changes in consonantal context. In addition, several of the uttered vowels did not fall within the F_1 vs F_2 contours established by Peterson and Barney [1952]. These contours had been determined using several productions of the utterance /hVd/. Stevens and House showed that the vowel in such an environment is not unlike the vowel produced in a null environment (/#V#/). The major discrepancy between the F_1 - F_2 values noted by Stevens and House, and those found by Peterson and Barney, then, was due to the influence of the consonantal environment imposed by the /hƟCVC/ production. Looking at differences within their own data, Stevens and House found further evidence for the relationship between phonetic context and a vowel's acoustic value. Consonantal context was seen to cause systematic shifts in the vowel's formant frequencies, particularly F_2, depending on the place of articulation, manner of articulation, and the voicing characteristic of the consonant involved. For example, in an environment of labial or post-dental consonants, front vowels showed more of a downward shift of F_2 than they did in a back environment. Fricatives produced greater shifts in interconsonantal vowel formants than did stops. Voiced consonants produced a lowering effect on F_1 of the vowel while F_2 was not as appreciably affected.

These changes in the acoustic value of a vowel are explained by Stevens and House in articulatory terms. In the production of a C_1VC_2 syllable, the structures of the vocal tract assume position for C_1, then maneuver towards position for V. During this movement,

instructions for C_2 are initiated. Vowel modifications are therefore due to overlapping of timing of neural instructions, which may result in anticipation of the upcoming phoneme, and the sluggishness or dynamic constraints (i.e. mass and inertia) of the system. Fricatives, for example, requiring carefully controlled positioning and target approach, would tend to infringe on the neighbouring vowel's articulation more than would a quickly executed stop.

Öhman [1966] looked at the influence of both preceding and following phones on a phoneme. Whereas Stevens and House used symmetrical CVC utterances and were unable to separate the influence C_1 had on the vowel from the influence of C_2, Öhman used C_1VC_2 disyllables. Utterances were spoken by speakers of three different languages and employed vowels particular to each language. Spectrographic analysis yielded measurements of formant frequencies at two points along the VC and CV transitions. Öhman found that, not only did V_1 affect the following CV_2 transition (as might be expected due to mechanoinertial factors), but as well, V_2 influenced the preceding V_1C transition. As noted by previous investigators, it was F_2 that showed the most variation.

Öhman's work yielded results different from that of previous workers, whose studies of CV transitions had led to the formation of the "locus theory" [Delattre, Liberman, and Cooper, 1955]. This theory states that for each consonant there exists a characteristic frequency position (or positions), or locus, from which formant transitions begin or to which they point. Delattre et al. had found fixed loci for the second formant of /b/ and /d/, and two loci for

/g/ (depending on context). Öhman found that the transition loci for /b/ and /d/ are not fixed, but are dependent on context. For example, in a /V_1bV_2/ utterance, the bV_2 transition originates at 500 Hz if V_1 is /u/, but at 1300 Hz if V_2 is /y/. Delattre et al. had postulated a fixed locus for bV transitions at 720 Hz.

The articulatory basis behind the locus theory is that formant transitions are reflections of the change in size and shape of the vocal tract as it moves from one target position to another. Delattre et al. state:

> Since the articulatory place of production of each consonant is, for the most part, fixed, we might expect to find that there is correspondingly a fixed frequency position -- or "locus" -- for its second formant; we could . . . describe the various second-formant transitions as movements from this acoustic locus to the steady state level of the vowel. . . .
> [Delattre et al., 1955, p. 769]

What the theory does not take into account is that if previous and/or succeeding articulations have an appreciable effect on the vocal tract configuration for any given consonant, the locus of a consonant produced in one environment will not be identical to that of a consonant produced in another environment.

Öhman, like Stevens and House, attributes the effect of preceding context on an upcoming phoneme (or in this case, its effect on the upcoming CV transition) to mechanoinertial factors. This type of coarticulation has since been referred to as carryover coarticulation. To explain the influence of succeeding context on preceding events, or anticipatory coarticulation, Öhman points out that speech

gestures are not independent and linearly sequenced. Often the vocal tract can vary a great deal without introducing a phonemic change in the sound produced. For example, the tongue is free to move during the production of a bilabial stop; the lips are free to move during the production of a velar stop or liquid. In general, if an articulator is free to move during production of one phoneme, it will initiate movement toward its target position for the next upcoming phoneme.

> Since traces of the final vowel are observable already in the transition from the initial vowel to the consonant, it must be concluded that a motion toward the final vowel starts not much later than, or perhaps even simultaneously with, the onset of the stop-consonant gesture. A VCV utterance of the kind studied here can, accordingly, *not* be regarded as a linear sequence of three successive gestures. [Öhman, 1966, p. 165]

Öhman also indicates the possible language-dependent nature of coarticulation. Russian stops must be coarticulated with one of only two vowels, whereas American English and Swedish stops enjoy more freedom of coarticulation.

2.3 Coarticulation: The Articulatory Level

With investigations at the acoustic level, the complexity of the coarticulation process began to come to light. Two major approaches to the study of articulatory behaviour, electromyography and cineradiography, began to yield evidence of coarticulation for various articulators, and several models have been advanced to account for the phenomenon. Such models are necessarily related to basic questions of speech organization.

Electromyography (EMG) has been employed to great advantage in coarticulation studies. Electrodes are introduced into the articulator in question, and muscle action potentials are recorded during utterance production. In this way muscle activity during production of any phone can be measured. A major problem in the interpretation of EMG studies is that the activity of one muscle is often closely related to that of others. A given amount of contraction in one muscle may therefore produce different amounts of movement of an articulator, depending on the position and activity of other muscles [MacNeilage and DeClerk, 1969]. Therefore, investigations into the EMG activity of only one muscle do not necessarily reflect all that is happening to the articulator in question. However, EMG studies allow individual muscles to be studied and correlations between neuromuscular activity and linguistic units to be made.

Cineradiography has been used to a great extent as well. Movements of lips, tongue, jaw, velum, and pharynx can be made visible by various methods of cineradiography, and correlated with acoustic output. However the resulting picture is a two dimensional display of the vocal tract and so has limitations. It also can only yield information at the motor level, whereas EMG studies give insight into neuromuscular commands. Perkell states:

> Although a cineradiograph contains a large amount of one type of information, it is obvious that many other types of parameters should be examined and correlated with the cineradiographic data before a comprehensive description of vocal-tract function can be obtained.
> [Perkell, 1969, p. 2]

Kozhevnikov and Chistovich [1965] examined coarticulation of lip movements in Russian, measuring electrical activity of the orbicu-

laris oris muscle and correlating it with utterance production. One speaker produced CV and CCV syllables in which V was a rounded vowel. Results show lip protrusion to begin almost simultaneously with the beginning if the first consonant, even if a word or syllable boundary falls within the CC sequence. Thus lip rounding was found to coarticulate over an entire CCV unit. The authors postulate an "articulatory syllable" model of speech production in which commands for the entire syllable are initiated simultaneously and executed simultaneously as long as they are noncompeting. Competing commands, such as lip retraction vs lip rounding, are executed in sequence. Therefore commands for an /i/ in one environment, would be different from commands for an /i/ in another environment. Coarticulation would be maximum within the articulatory syllable, and minimum across such syllable boundaries. Such a syllable is described by Kozhevnikov and Chistovich as the CC...V unit, which has been found by themselves and others to be a strongly cohesive unit and to exhibit strong coarticulation effects within itself.

Fromkin [1966] used electromyography to study action of the orbicularis oris muscle for production of /b/, /p/, and various rounded and unrounded vowels in English. Her results, obtained from three speakers, show that no simple correspondence exists between phoneme and motor command; different muscle action potentials are responsible for producing an initial /b/ or /p/ and a final /b/ or /p/. However, further contextual aspects have no effect on the muscle gesture for these phonemes, at least as far as this muscle is concerned. Muscle

action potentials are relatively invariant for production of the /b/ in a /bVC/ syllable, regardless of the values of the following phones. Similarly, action potentials for final /b/ are unaffected by preceding phones in a /CVb/ syllable. The same results apply to initial and final /p/.

Looking at EMG activity of the same muscle during vowel production, Fromkin did note influence of adjacent phonemes. The rounded vowels /u/ and /o/ show appreciably lower peak amplitude of EMG activity when following initial /b/, which itself involves contraction of the orbicularis oris muscle, than when following initial /d/. Muscle activity for a rounded vowel is uninfluenced in amplitude or duration by the following consonant of a CVC syllable, be it /b/ or /d/. Thus it seems that some aspects of context somehow restrict or reorganize the neuromuscular commands and gestures for some phonemes, while other aspects do not. Just what the nature of the reorganization is, is not known, Fromkin states. Her findings lead her to put forth two suggestions. Perhaps the minimal linguistic unit at the motor command level is larger than the phoneme, possibly, in her words, of the order of a syllable. This theory agrees with the Kozhevnikov-Chistovich model of speech organization. However, Fromkin does not give any indication of the size or nature of the syllable proposed. The second possibility is that motor commands are altered with context by a feedback system concerning the existing state of muscle position and activity, or by information held in short-term memory. This theory is consistent with the idea that the phoneme is a basic unit of speech production at the neuromuscular level. Both theories

proposed by Fromkin are able to account for the coarticulation effects she observed.

Öhman [1966] describes the coarticulated VCV utterance as follows:

> 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. [Öhman, 1966, p. 165]

Production of the consonant in such a syllable involves three separate, but probably overlapping, sets of muscles in the tongue, each of which has separate neural representation in the motor control networks of the brain. The response of the tongue to articulatory commands coming independently over three different channels is a summation of the components of the instructions. As the tongue is executing commands for one phone, certain subsets of muscles are left free to anticipate the following phone, instructions for which are also coming down independently. Therefore, consonant production is accomplished by articulatory adjustments that partially anticipate the configuration of the succeeding vowel, though certain components of V_2 are inhibited during C production.

Henke [1966] proposes a system whereby production is programmed phoneme by phoneme, but there is a scanning of upcoming feature specifications. If a phoneme has no specification for a particular feature, such as lip rounding, the system looks ahead to the next phoneme for which that feature is specified, and the articulators initiate movement toward that goal.

MacNeilage and DeClark [1969] questionned whether changes in motor gesture with context are due to changes in underlying neurological control or to mechanical constraints and modifications on an invariant phoneme command. Examination of cinefluorograms of the vocal tract and EMG tracings from nine articulatory locations showed that both left-to-right (carryover) effects and right-to-left (anticipatory)effects of adjacent phonemes on each other are present in CVC syllables. They state:

> It is quite clear from these results that the command system responsible for CVC syllables does not consist of a series of context-independent phoneme commands that retain their independence all the way down to the level of muscle contraction.
> [MacNeilage and DeClerk, 1969, p. 1228]

They hypothesize three mechanisms at work to account for these effects. First is an anticipatory mechanism, in which the greater the amount of muscle contraction required for a certain phoneme, the greater the amount of anticipatory contraction of that muscle in the preceding phoneme. An inhibitory component against muscle contraction antagonistic to the muscular movement required for the upcoming phoneme might also be involved in the anticipatory mechanism. Such a system can explain right-to-left coarticulatory effects. The second mechanism at work is a compatibility mechanism. Since more or less contraction is necessary to assume a particular articulatory position, depending on the previous position of the articulator, upcoming commands for contraction might be made compatible with the existing state of muscle contraction. This would be accomplished via a feedback system involving the cerebellum. Such a system is able to account for the strong left-to-right influence imposed by context.

This mechanism is somewhat similar to one proposed by Fromkin [1966]. The third suggested mechanism at work is a gamma-loop mechanism. In this case commands are sent down for a muscle to assume a particular length, regardless of its existing length, by the gamma system of motoneurons which innervate stretch-receptive spindles within the muscles. Thus commands would be invariant, but EMG activity necessary to achieve the specified length would show the context-dependent variety seen in several studies. This model seems appropriate for speech production which involves approximation of target positions regardless of context.

MacNeilage and DeClerk point out that joint action of the three mechanisms outlined above on invariant phoneme commands cannot account for all the coarticulation effects seen. The authors cite two further mechanisms that do not necessitate ruling out invariant phoneme commands as the basis of production. At least they may be present at certain levels of the speech production system. The first possibility is that other modification mechanisms, such as the use of somesthetic information, are at work. The second possibility is that to a certain, maybe considerable, extent, motor commands are organized in units larger than the phoneme; perhaps as suggested by others, commands are issued for a syllable at a time. However, since they were unable to observe effects of initial and final consonants on each other, MacNeilage and DeClerk suggest that the CVC unit does not qualify as the unit of command organization. They feel that the CV segment, which shows more right-to-left coarticulation effects than the VC segment, is a more cohesive unit.

Daniloff and Moll [1968] extended Kozhevnikov and Chistovich's 1965 work on lip protrusion, to the production of strings of one to four consonants followed by the rounded vowel /u/. The sequences were embedded in meaningful English sentences and spoken by three subjects. Though the utterances contained the phonemes /r/ and /l/, which themselves may involve lip protrusion, the authors noted that such an amount of protrusion was small. Cineradiography was used to evaluate articulatory behavior. Findings show that lip protrusion extends over as many as four consecutive consonants before a rounded vowel, and that the extent of coarticulation is not affected by word or syllable boundaries within the consonant string. Results are in general agreement with those of Kozhevnikov and Chistovich. However, Daniloff and Moll observed onset of protrusion before contact for the first consonant was achieved, whereas Kozhevnikov and Chistovich noted protrusion onset at the time of contact for the first consonant. In a number of cases noted by Daniloff and Moll, protrusion began even before movement toward the first consonant was initiated, that is, outside the boundary of the CC...V unit. Cowan [1973] found similar coarticulation effects for lip protrusion in French utterances. Six native French speakers produced utterances containing strings of four and six consonants before a rounded vowel. She found that in almost all cases, protrusion for the upcoming vowel began with production of the first consonant of the cluster, and in approximately half the cases, protrusion began during the production of the vowel preceding the consonant cluster.

Coarticulation effects have been observed in the motion of the lateral pharyngeal wall [Kelsey et al., 1969]. An ultrasonic method of data collection was used, in which a pulsed ultrasonic signal was beamed toward the pharyngeal wall and the time of echo return provided a measure of displacement of the articulator. Three speakers uttered VCV utterances. Data show that displacement during production of /a/ varies as a function of phonetic context.

Amerman et al. [1970] investigated coarticulation effects jaw and lip movements by cineradiography. Four speakers produced meaningful utterances which included segments of one to four consonants preceding the vowel /æ/. Jaw lowering and lip retraction are two gestures involved in the production of this vowel. Jaw lowering was found to coarticulate over two and sometimes three phones before /æ/, and could presumably extend over all four consecutive consonants, had not one of the consonants consistently been /s/. Amerman et al. found /s/ production antagonistic to jaw lowering; this gesture was never initiated during /s/ production, but began immediately after it. Similarly, lip retraction seemed to be inhibited by /s/ production and was never initiated during it. However, a good /s/ can be produced with retracted lips and the authors suggest that perhaps inhibition of one gesture for /æ/ production facilitates inhibition of another gesture related to /æ/ production. In general, lip retraction was not as extensively coarticulated as jaw lowering. Though it sometimes extended two and three consonants before the vowel, several of the cases showed retraction beginning with the start of the vowel and not

before. However the lip retraction measure was not considered by the authors as reliable a measure as jaw lowering, due for instance to some lip protrusion during /r/ production. The authors feel that inconsistencies in the synchrony and starting points of the two gestures are not predicted by the Kozhevnikov-Chistovich model, which states that commands for the syllable are specified simultaneously and synchronously. The nature of the coarticulatory unit found in this study is in agreement with that model's articulatory syllable, i.e. a CC..V unit. The data fit Henke's model of production equally well.

Carney and Moll [1971] extended Öhman's 1966 study of coarticulation in VCV utterances. Whereas Öhman had examined coarticulation of vowels and stop consonants, Carney and Moll looked at fricative-vowel interactions. MacNeilage [1963] had previously shown acoustic properties of the fricative /f/ to be context dependent; specifically, duration of /f/ in final position was twice as great as for /f/ embedded in a consonant cluster. However electromyograms taken at the lips during /f/ production did not show pattern changes with context, except to some extent for onset of activity. Carney and Moll placed fricatives in a vowel rather than a consonant environment, and looked at effects on the tongue as well as the lips. They analyzed cineradiographs of two speakers producing /hVCV/ utterances, in which C was the fricative /f/, /v/, /s/, or /z/. Unlike MacNeilage, they found muscle gestures for production of fricatives to be influenced by context. Their results agree with Öhman's [1966] description of a consonantal gesture superimposed on a basic vowel-to-vowel diphthongal gesture. The findings show that if an

articulator is free, as the tongue body and tip are during /f/ or /v/ production, then coarticulation is seen in the tongue and in the lips throughout the vowel-to-vowel movement.

Coarticulation effects have been observed in velar movements by Moll and Daniloff [1971]. Four subjects produced English sentences containing various combinations of nasal consonants, non-nasal consonants, and vowels. Examination of cineradiograms showed that movement towards velar opening in a CVN or CVVN (where N = nasal) sequence begins after contact for the initial consonant. Thus nasality is coarticulated over the VN or VVN unit. Similarly, for NVC sequences, movement towards velar closure begins during the approach to the vowel, and sometimes even during the nasal itself. The unit over which coarticulation extends in this case is the VC unit. These results directly contradict Kozhevnikov and Chistovich's hypothesis that CV is the basic unit of production within which coarticulation is strongest. Moll and Daniloff tend to support a model such as Henke's where commands are specified phoneme by phoneme.

Thus three major systems have been put forth to account for coarticulatory behaviour. One is the Kozhevnikov-Chistovich "articulatory syllable" model, in which neural commands are organized in syllable-like units. Though this model accounts for much of the observed data, the articulatory syllable is described as a CC..V group, whereas studies indicate that coarticulation may extend back to encompass a VCC..V group [Daniloff and Moll, 1968] or a VC or CVC group [Moll and Daniloff, 1971]. However, MacNeilage [1972] cites

evidence that, in a CVC syllable, there is weaker coarticulation within the VC segment than within the CV segment, indicating that CV is a strongly cohesive unit. The second major model is that of Henke, whereby a forward scanning system allows a free articulator to begin movement towards position for an upcoming phoneme. Such a system would be operative during anticipatory coarticulation. MacNeilage & DeClerk [1969] point out that such an anticipatory mechanism may be one of several at work during speech production. Öhman [1966, 1967] describes a third model of coarticulation, in which a consonantal gesture is superimposed on a diphthongal vowel-to-vowel movement. The phoneme command for consonant production is invariant, but the vocal tract shape during its production is a result of an overlap of vocal tract shape assumed for the consonant and the varying shape due to vowel environment. Thus contextual modifications take place at the motor level. Carry-over coarticulation is accounted for in most models by mechano-inertial factors, or by the compatibility mechanism [MacNeilage and DeClerk, 1969] described earlier.

2.4 Coarticulation: The Perceptual Level

Recent studies have examined the perceptual correlates of coarticulation. The question asked is, whether the acoustic and articulatory modifications due to coarticulation in an utterance provide information utilizable by the listener. Ali et al. state:

> It is uncertain in most specific cases if coarticulation on the articulatory level results in perceptible differences on the perceptual level. . . . If the answer is affirmative, then it can be said that speech perception 'follows' speech production and makes use of its idiosyncracies. [Ali et al., 1971, p. 538]

A point to keep in mind when studying the perceptual correlates of coarticulation is that the subject is being asked to make subphonemic discriminations, subtle distinctions that do not affect the value he assigns to a phone. To what extent can we realistically expect him to do so? It is known that subphonemic detail (one form of which is allophonic variation) can be distinguished within a single phoneme category, even though speech perception is itself to some extent a categorical process. For example, Liberman et al. [1957] showed that listeners can make subphonemic distinctions when they are presented with synthetic speech sounds varying along an acoustic continuum. Stimuli were produced by a pattern playback, consisted of first and second formant patterns, and varied in direction and extent of the second-formant transition. This variable is a cue which has been found to be instrumental in making /b,d,g/ distinctions. Fourteen different stimuli were produced, and presented to subjects in an ABX arrangement. In a separate test, subjects were asked to make phonemic judgments of the same stimuli, that is, to state whether each was /b/, /d/, or /g/. Comparing the results of both studies, the authors determined that (1) phonemic distinctions along the continuum are categorical, the point at which a response changes from one phoneme to another being abrupt and consistent, (2) subphonemic discriminations across phoneme boundaries are able to be made to some extent, and (3) discriminations across phoneme boundaries are better and more consistently made than discriminations within a phoneme category.

Fry points out that

> . . . a pair of utterances may appear indistinguishably the same to a listener of one nationality and indisputably different to a listener of another nationality. . . . [Fry, 1964, p. 60]

This is another point to consider in evaluating perceptual studies of coarticulation. Fry cites work by Lotz et al. [1960] on phonemic labelling of the same set of stimuli by different language groups. Fortis aspirated, fortis unaspirated, and lenis unaspirated stops were presented to speakers of various languages. The stimuli were placed into phonemic categories as follows: by English speakers, into /p,t,k/, /b,d,g/, and /b,d,g/ groups respectively; by Hungarian and Spanish speakers, into /p,t,k/, /p,t,k/ and sometimes /b,d,g/, and /b,d,g/ groups; by Thai speakers (in whose language aspiration is phonemic), into /p,t,k/, /p^h,t^h,k^h/, and /b,d,k/ groups. For the velar case, Thai speakers assigned the lenis unaspirated stop to the /k/ category, there being no /g/ in Thai, though the possibility of the /g/ label was available to them.

Thus it seems that perceptions are influenced by language learning. In considering this point in relation to coarticulation studies, one might ask whether French listeners, for example, make finer judgments regarding lip rounding than do English ones. It has already been seen that coarticulation on the articulatory level may be language dependent [Öhman, 1966].

Findings on phonemic labelling opposite to those described above emerged in a study of cross-language vowel perception carried

out by Stevens et al. [1969]. Thirteen unrounded and thirteen rounded vowels were synthesized on the OVE II speech synthesizer, with the first three formants varying along an acoustic continuum. Two ABX discrimination tests were administered, one for the unrounded and one for the rounded vowels, to a group of Swedish and a group of American English speakers. Two phonemic identification tests were administered for the same stimuli to the same subject groups. The rounding feature is phonemic in Swedish, but not in English. Results show that for vowels presented in isolation, the listener's linguistic experience has essentially no effect on his ability to make subphonemic discriminations, nor does it appreciably affect his identification of phonemic categories. Little difference was seen in the phoneme boundaries determined by the Swedes and those determined by the Americans. The boundaries assigned by these groups differed by no more than one step along the acoustic continuum for the unrounded vowel series, and one to two steps for the rounded vowels.

These findings in a sense do not contradict the language-dependence found by Lotz and his colleagues [1960]. Subjects were presented with different tasks in these two studies. There is no reason to assume that, given the same series of fortis aspirated, unaspirated, and lenis unaspirated stimuli, and asked to place each into one of *three* categories (a situation similar to the identification task presented by Stevens et al.) speakers of all languages investigated by Lotz et al. would not be able to assign each phone to its appropriate category. For some of these speakers, some of the category assignments would be based on a phonemic distinction,

and some would be based on a subphonemic distinction. English speakers involved in the experiment by Stevens et al. placed rounded vowels into phoneme categories not appreciably different from (although somewhat less consistent than) those chosen by the Swedes, though for the English speakers the placements were based on subphonemic discriminations. What Lotz's experiment does show, is that depending on his linguistic experience, a listener may chose to ignore some of the distinctions he is capable of making.

In addition to their study of vowel discriminations described above, Stevens et al. [1969] replicated the experiment on consonant discrimination done by Liberman et al. [1957, also described above]. Synthetic stop consonants, for which the first three formants varied along an acoustic continuum, were presented in an ABX situation. Stevens and his coworkers found that subphonemic discriminations along a physical scale were better made for vowels than for stop consonants. For example, correct discrimination could be made within a vowel phoneme category 80-90% of the time (depending on how far along the acoustic continuum they differed), but within a consonant phoneme category only 60-65% of the time. The authors cite the suggestion that different mechanisms may be involved in vowel and consonant perception. In addition, investigators have found that vowels are not perceived as categorically as consonants [Kozhevnikov and Chistovich, 1965; Liberman et al., 1967], also suggesting that separate perceptual processes may be at work for these two classes of phones. However, Liberman et al. point out that vowels studied in isolation, or the "unencoded" state, as in the above studies, may not trigger perception

in the speech mode, and that evidence exists that vowels embedded in phonetic context are more nearly categorically perceived than are unencoded vowels.

Liberman et al. [1967] discuss subphonemic perception as being essential to speech perception:

> That subphonemic features are present both in production and perception has by now been quite clearly established . . . we must deal with the phonemes in terms of their constituent features because the existence of such features is essential to the speech code and to the efficient production and perception of language. . . . high rates of speech would overtax the temporal resolving power of the ear if the acoustic signal were merely a cipher on the phonemic structure of the language.
> [Liberman et al., 1967, p. 446]

It should be noted that the "features" discussed above are not the distinctive features discussed by Jakobson and his colleagues, but are constituent motor gestures and neural commands of phonemes. These researchers support the motor theory of speech perception, which states that speech is perceived in reference to the motor gestures that can produce it. They showed that acoustic signals may vary greatly and still produce the same perceptual effect. For example, the frequency of the starting point of the second formant transition from /d/ to a following vowel can vary by as much as 1000 Hz, depending on the vowel, yet a /d/ is perceived in all cases. Since a phoneme's acoustic signal varies not only with context but also from speaker to speaker, it is necessary to explain how the listener identifies the phoneme each time. Liberman et al. propose that the listener traces the variable acoustic signal back to the less

variable articulatory gestures with which he himself would produce the signal. He then identifies the signal in reference to these motor gestures. Since the motor gesture for a particular phone can be broken down into several elements (e.g. raising the velum, raising or lowering the tongue, initiating vibration of the vocal cords), then perception of the phone's constituent features can in some manner occur.

To what extent the listener may perceive subphonemic, or allophonic, variations, has been examined by Wickelgren [1969]. He cites the context-sensitive allophone as a unit of perception. Such a unit is one which specifies its right and left-hand neighbours. Thus the word "tap" would be coded as $/_{\#}t^{æ}/$, $/_{t}æ^{p}/$, $/_{æ}p^{\#}/$. The input to the perceptive mechanism could thus be an unordered set of symbols, the coding system allowing correct order to be recovered from such a set. The context to which such allophones are sensitive is limited to one preceding and one following phone, in Wickelgren's model. As we have seen, such is not the case in production, where a phoneme such as a rounded vowel may exhibit an effect on another phoneme as many as six sounds removed from itself [Cowan, 1973]. Perhaps though, an allophone is sensitive to an extent which is perceivable only to adjacent phonemes. A major problem with Wickelgren's hypothesis is the extremely large number of neural units that must be available and through which all acoustic input must be channeled, for it is assumed that each context-sensitive allophone has its own neural representation.

It may be appropriate here to point out the arguments that

exist for various other perceptual units. Speech perception may take place on several levels. Though subphonemic distinctions can be made, the fact that consonants show strong and definite categorical perception [Liberman et al., 1957], and that the same is true of vowels to a lesser extent [Stevens et al., 1969], provides evidence for the phoneme as a basic speech perception unit. Savin and Bever [1970], however, believe that individual phonemes are identified only after perception on yet another level has been carried out. They asked subjects to monitor a speech sample for a particular unit, either a syllable or a phoneme within a sylable. Results showed that response times for syllable identification were faster than for identification of a particular phoneme, suggesting the syllable was first perceived as a unit, before the phoneme itself was identified.

Certain syntactic sequences may be perceived as units. By presenting extraneous sounds (clicks) during sentences, Ladefoged and Broadbent [1960] found that listeners tend to locate the clicks far removed from their actual location. They argue that subjective displacement of clicks is towards boundaries of perceptual units. Several further studies on click displacement, outlined by Lehiste [1972], have been carried out with inconsistent results. Subjective location of extraneous sounds is also related to stress, intonation, and other surface phenomena. However, it is clear that acoustic cues alone do not determine the boundaries of perceptual units, and that higher level sequences are somehow perceived as units.

Lehiste [1972] sums up a discussion on perceptual units by saying that two basic steps exist in speech perception: primary processing, consisting of auditory and phonetic processing, and linguistic processing, consisting in part of phonological and syntactic processing. Though the auditory level must precede other levels of processing, it is possible that phonetic and linguistic processing may proceed concurrently. Units at different levels differ in size.

Thus we see that, though perception is primarily a categorical process on one level, and that higher level sequences may act as units in perception, listeners are indeed capable of making subphonemic distinctions. It is this type of discrimination that subjects are asked to make in the coarticulation studies outlined below. It is possible that many of the large number of distinctions a listener can make when hearing a speech sample are ignored, in favor of grouping several different, but somehow similar, sounds into a single category for quicker processing. Whether subphonemic perception is of primary importance in the speech perception process is not clear, since discrimination is consistently poorer within a phoneme category than across its boundaries. However, in times of unfavorable conditions, for example a noisy environment, or a large amount of information having to be processed quickly, it may be that subphonemic nuances are used by the perceptual mechanism to provide additional cues. Perceptual reality of coarticulatory effects would mean that, on hearing one sound, the listener not only has acoustic information on *its* value, but has information to verify the value he has assigned

to the preceding phone(s), and to tentatively anticipate the value of the upcoming phone(s). Such a process would facilitate correct identification of any one speech sound. Let us now examine the few studies that have been done on the perception of coarticulatory effects.

Moll and Daniloff [1971] had shown that velopharyngeal opening in CVN and CVVN sequences (where N is a nasal consonant) almost always begins during the CV transition. To test the perceivability of this coarticulated nasality, Ali et al. [1971] spliced the final consonant and its VC transition from English CVC and CVVC utterances, in which the final consonant was sometimes a nasal and sometimes not. Twenty-two subjects were presented with the spliced utterances and asked to identify the missing consonant as nasal or non-nasal. Results show that nasal stimuli were correctly identified significantly above chance level. There was no significant difference between correct perception of /n/ and /m/. Stop consonants were identified as nasals more frequently than were fricatives. Consonants following the vowel /a/ were perceived as nasal more often than consonants following other vowels. Significant individual subject differences were found. The authors believe that in the case of nasality, the perceptual mechanism does make use of coarticulated information.

Lehiste and Shockey [1972] tested the perceivability of vowels removed from a VCV utterance (where C is a stop consonant). Öhman [1966] had previously shown that the VC and CV transitions in such an utterance are influenced by the transconsonantal vowel. For the perceptual test, VCV utterances were cut in two during the

consonant closure. Over twenty subjects were asked to identify the missing initial or final vowel. Though Lehiste and Shockey noted the same coarticulation effects spectrographically for their utterances as did Öhman, they found that these contextual effects are not sufficient for identification of the deleted segment. Nor was enough information present in the spliced utterances to identify a feature of the deleted phone, such as high/low or front/back; incorrect responses did not tend to share a feature with the correct response. The authors conclude that "whatever the effects of coarticulation in terms of their influence on formant transisitions, these effects are not sufficient to have an influence on perception" [Lehiste and Shockey, 1972, p. 84]. Lehiste [1972] cites these results as evidence against Wickelgren's [1969] model of speech perception, which involves coding of context-sensitive allophones.

> The physical modifications are undoubtedly there, but if the context of a context-sensitive allophone is not perceptible, it seems unjustified to assume that context-sensitive allophones are the basic units of speech perception. [Lehiste, 1972, p. 5]

Lehiste and Shockey's [1972] findings are contrary to those of Kuehn [1970], as cited by Carney and Moll [1971], who found that listeners were able to predict V_2 of a V_1CV_2 utterance above chance level, when they were given the initial segments of the utterance. However, Carney and Moll do not discuss the test situation used by Kuehn, and therefore strict comparisons between the two studies cannot be made.

In comparing the Ali et al., and Lehiste and Shockey studies, we see that context of the CV- and CVV- units was recoverable, but

that context of the VC- or -CV unit was not. It may be noted that in the first case, the subphonemic cues relating to context must be elicited from the preceding vowel, and in the second case, from the preceding or following VC or CV transition. It has already been seen that subphonemic discriminations are more easily made for vowels than for consonants [Stevens et al., 1969], and if we for a moment consider the CV or VC transition as part of the consonant, or at least as behaving as a consonant in this respect, then we may adduce an explanation for the above findings: coarticulation effects on a vowel are more easily perceived than those on a consonant. However, it must be kept in mind that there is indication that vowels in phonetic context are not as differently perceived from consonants as data on isolated vowels suggests [Liberman et al., 1967]. Also, the motor gestures involved in the coarticulatory effects of the two cases described above are different -- the first involves lowering of the velum, the second involves tongue movement. It may be that the effects of these two motor gestures are perceived to different extents. Human listeners may be inherently more aware of slight changes in one type of gesture than in another.

Clark and Sharf [1973] looked at coarticulatory effects of V_2 on short term recall of V_1 in V_1CV_2 utterances. By presenting lists of VC/V (final vowel deleted), VCV (final vowel retained), and VC# (no final vowel produced and thus no coarticulation present) utterances to subjects, they found that the presence of coarticulation influenced the % correct recall of the initial vowel. They determined that the coarticulation effects in question are perceived by the

listener and registered in short term memory. Previous investigators have suggested that the listener remembers for a certain time the spectral characteristics of the phone he hears, and on identifying it as a phoneme, uses the necessary information and discards the rest [Lehiste, 1972]. In other words, he retains subphonemic information in his memory for some unspecified length of time. Whether the process as described by Clark and Sharf is naturally operative in speech perception is not clear, since, though recall for the VC/V condition was facilitated over the VC# condition, the VCV condition did not have the same facilitative effect. The authors attribute this to a possible perceptual overloading, the subject hearing twice as many vowels in the VCV than in the VC/V condition. They suggest that even in the VCV condition, the effects may be registered but ignored.

Sharf and Ostreicher [1973] looked at the effects of coarticulation on identification of nasal consonants in noise. Using utterances of the form C_1VNC_2V, where C_2 consists of 0, 1, or 2 non-nasal consonants, they found that identification of N was significantly better when all the post-nasal sounds were retained than when they were deleted. That is, when the carryover coarticulation effects present in the post-nasal sounds were available, subjects scored better in nasal identification in noise than when these effects were removed. By asking subjects to identify the final vowel from the same truncated utterances, the authors noted a better than chance level of correct identification if no consonant had originally intervened between N and V, and a consistent but insignificant trend

for the number of correct vowel identifications to decrease as the number of intervening consonants increased from 0 to 2. This seems to indicate that anticipatory coarticulation effects of the vowel on the nasal aid in identification of the deleted vowel, but that as nasal and vowel move farther apart, the weakened coarticulatory effect becomes imperceptible. Thus they conclude that anticipatory coarticulation produces a strong enough cue in the nasal to facilitate identification of the upcoming vowel, and that cues present in the vowel due to carryover coarticulation with the preceding nasal aid in the correct perception of the nasal.

It remains to be seen which coarticulatory influences are perceivable and which are not, and over how long a sequence of phones coarticulatory information is usable.

CHAPTER 3

AIMS OF THE EXPERIMENT

Some major questions in the study of speech perception are: What features and cues does the listener abstract from the speech signal in attempting to identify it? Is all the acoustic information present in the signal utilizable for the perception process? Are all the fine, as well as gross, motor adjustments involved in the production of the speech signal recognized and interpreted by the listener? Research has shown that neither the acoustic value of a phoneme, nor the motor gesture that produced it, is invariant across different contexts. How much of this variation is perceivable, and to what extent does it actually provide cues for perception?

Studies on the perceptual correlates of coarticulation have begun to indicate that the listener may use some of the ever-present contextual variation as an aid in identifying speech sounds. The present experiment attempts to provide further information in this area. It asks whether coarticulation provides perceivable information, that is, whether it contains cues usable in the speech perception process. Utterances containing the sequence $-C_1C_2C_3C_4V-$ (where C_i is a consonant and V a rounded or unrounded vowel), in which coarticulated lip rounding is known to occur when V is a rounded vowel, are truncated at four points before the vowel. Edited versions thus contain different amounts of coarticulated information. By presenting these

stimuli to phonetically trained and phonetically naive native French and native English speakers, the present experiment attempts to do the following:

1. To discover whether coarticulation of lip rounding in French produces perceivable information, by asking subjects to identify a missing vowel for which coarticulation is present.

2. To discover over how many segments such information is perceivable. Coarticulation on the articulatory level is known to extend over all four consonants in the type of utterance described above.

3. To investigate the language-dependent nature of the perception of coarticulated information, by comparing results from French and English speakers; and to reveal whether perception of these cues plays a normal part in the speech perception process, or whether they may nevertheless be abstracted from speech by a suitably trained listener, by comparing results from phonetically trained and phonetically naive subjects.

CHAPTER 4

MATERIALS AND METHODS

4.1 Pilot Study

Preparation of Test Tapes

Two pilot test tapes were constructed. The items of the first test contained the consonant cluster /kstr/ followed by each of the three vowels /i/, /y/, and /u/. The sequences were derived from the three French utterances "la dextre inimitable," "la dextre universelle," and "la dextre outragée." These utterances were recorded during the course of a previous experiment [Cowan, 1973] in a non-soundproof environment. A wide-band hum due to the operation of a graphic recorder during their recording produced distracting background noise on the original tapes. However, it was decided to use these recordings because the speech wave, the duplex oscillogram, the log intensity of the speech signal, and a graphic representation of the speaker's upper lip protrusion were all available, displayed on separate channels of a Siemens Oscillomink graphic recorder. The speaker was a male native speaker of French, from Lausanne, Switzerland.

The utterances were edited at three points each, on a PDP-12 digital computer, using a set of computer programs written by L. Rice at the UCLA Phonetics Laboratory. (The editing process

will be described in Section 4.2). Three edited versions were made:

/ladɛkstr/

/ladɛkst/

/ladɛks/

The test items were recorded onto a Revox A77 tape recorder. (This procedure is also described in Section 4.2). The pilot test tape consisted of three samples of each of the three utterances truncated at each of three points, for a total of 27 items. The test was constructed so that the longest of the edited versions made up the first third of the test, the next longest the second third, and the shortest the last third, i.e.:

Group 1: 9 items of /ladɛkstr/

Group 2: 9 items of /ladɛkst/

Group 3: 9 items of /ladɛks/

However, the order of presentation with respect to the missing vowel was random within each group, with each vowel being represented 1/3 of the time.

The second pilot test tape was made in response to some subjects' comments that the first tape was noisy and distracting, and that they had felt unsure of the task required of them until at least one or two utterances had been played. It was constructed similarly, except that the utterances were recorded under soundproof conditions, using an Altec 681A LO microphone and a Scully 280 tape recorder. The

same speaker recorded the same utterances as used in the first test. These speech samples were edited with the same set of computer programs and at the same three points as described above. It was proposed that the results of the first and second tests be compared to determine whether background noise on Cowan's tapes produced a sufficiently lower score to warrant the use of new tapes recorded under soundproof conditions for the main experiment. In response to the comment that subjects were not sure of the task until at least two items had been played, the second test contained 29 items, the first two being practice items whose results were not considered in the analysis.

Subjects

Subjects were six adults (3 male, 3 female), all of whom had some knowledge of phonetics. Only one subject, who was also the speaker on the tapes, was a native speaker of French. One subject was a native speaker of German, a language which makes use of the three vowels under study. The same 6 subjects took part in both Tests I and II.

Test Procedure

Subjects were seated, one at a time, alone in a quiet room. The test items were presented over headphones at a comfortable listening level. Subjects were asked to indicate in writing whether the missing vowel was /i/, /y/, or /u/. They were first told what the original utterances had been.

Test I was given in one session and Test II in another. At the time that Test II was administered, Test I was readministered to see if familiarity with the test situation affected test results. The tests are hereafter referred to as Test Ia (first session), Test Ib (second session), and Test II (second session).

Results

Values for relative transmission (T_{rel}) of information (a measure to be discussed in Section 4.2) and % correct score were calculated. Scores were generally higher for Test II than for Test Ia or Ib. Since no significant differences were noted between Tests Ia, given in the first session, and Ib, given in the second session, it was assumed that no practice effect was contributing to the increase in T_{rel} and score from Test I to Test II. This suggests that improvement from Test I to Test II was probably due to the better listening conditions on the second tape.

A distribution of T_{rel} based on random responses to a 27-item test was calculated. This distribution is shown in Figure 2 and described in detail in Section 4.2. From the distribution, the maximum value of T_{rel} which a subject could obtain by chance 10% of the time was determined. T_{rel} values above this level were considered significant values of information transmission and the following was observed: all subjects obtained significant T_{rel} values for Group 1 items; 2 out of 6 obtained significant values for Group 2; no subject obtained a significant score for Group 3. Responses were

also analyzed to see if they tended to have a feature in common with the stimulus. The ability to perceive front/back distinctions and unrounded/rounded distinctions was examined. For all groups of items, the front/back distinction was made more often than the unrounded/rounded distinction. Both distinctions were made more often for Group 1 than for Group 2, and for Group 3, which contained the shortest edited versions, subjects were giving responses no different from random guessing.

4.2 Main Study

Speech Samples

Because scores were generally higher on Pilot Test II than on Test I (a or b), it was decided to use utterances recorded under soundproof conditions for the main study.

Three male native speakers of French recorded the utterances. Speaker #1 was born in Lausanne, Switzerland, and had been in North America for 14 years. Speaker #2 was born in Grenoble, France, and had been in North America for 4 years. Speaker #3 was born in Albi, France, and had been in North America for 9 years.

Fifteen utterances were recorded by each speaker, at least twice each. Each utterance contained one of the consonant sequences /kstr/, /rstr/, /rskr/, followed by one of the three vowels /i/, /y/, or /u/ in all possible combinations. Cowan [1973] had shown that, for the utterances decribed above, upper lip protrusion most often begins with the approach to the first consonant in the cluster, if the cluster is followed by the rounded vowel /y/ or /u/. Cowan's

findings also applied to utterances with 6-consonant clusters. Such utterances were considered for use in the experiment, but since the pilot test had shown no significant information to be available when the utterance was truncated after the second consonant of a 4-consonant cluster, utterances with 6-consonant clusters were not used.

Recordings for the present experiment were made in an IAC 1204 soundproof room using an Altec 681 A LO microphone and a Scully 280 tape recorder.

One set of 9 utterances, consisting of examples of each of the three clusters followed by each of the three vowels, was chosen from each speaker. Utterances were chosen on the subjective bases of clarity of the speaker's voice, absence of background noise, similarity of intonation patterns of utterances containing the same cluster, and presence of all phonemes in the cluster. These 9 utterances are listed in Appendix I. The remaining 6 utterances from each speaker contained additional samples of clusters which were present in the other utterances, and these samples were not used. Spectrograms, on a Kay Sona-Graph Model 7029A, and mingograms, on a Siemens Oscillomink graphic recorder, were made of all utterances, for reference in the editing process.

Editing of Speech Samples and Preparation of Test Tapes

Editing of utterances was carried out using a set of computer programs written by Lloyd Rice for a PDP-12 digital computer. This set of programs digitizes the speech signal and displays it on the computer oscilloscope screen, and allows the speech waveform data to be manipulated in various ways. The speech signal was first low pass filtered at 6 kHz to prevent aliasing of the input signal. It was intended to digitize the speech wave at 12 kHz; however, limitations

of the equipment meant that the computer could not keep up with such a fast transfer rate for the length of time it took to sample the utterance. The computer was therefore skipping some samples, once the core buffer had been filled, and notable distortion resulted. To overcome this problem, each utterance was played at half speed and digitized with a 10 bit analog-to-digital converter at 6 kHz sample frequency, for an equivalent of 12,000 samples per second. The digitized speech wave thus produced was stored on digital tape and could be displayed on the computer screen. A knob controlled the velocity of the speech waveform data as it moved backward or forward across the screen. The waveform could also be made stationary on the screen. In this way, the speech wave could be visually examined as the operator saw fit. The speech wave was then edited as follows: the speech wave of the whole utterance was displayed on the screen, and the operator marked the desired initial point of the truncated utterance by a command on the teletype. In all cases, this point was marked just before the onset of phonation at the beginning of the utterance. The waveform was then moved slowly across the screen until the desired endpoint was visible. This point was also entered by a teletype command. The entire edited segment was then stored elsewhere on the digital tape. In this way, an edited utterance could be obtained, leaving the original utterance intact and available for making further editions. Each utterance was truncated at four different points, producing the four groups of stimuli shown in Table I. Since results of the pilot study showed that no significant information is available when truncation takes place after the second

TABLE I

Groups of Edited Stimuli with Consonant Clusters /kstr/, /rstr/, and /rskr/. Each sample as Described Above has 3 Versions, One of Which Originally had the Following Vowel /i/, One /y/, and One /u/. Original Utterances From Which the Edited Stimuli Were Derived Are Listed in Appendix I

GROUP I	
Truncation immediately	/ladɛkstr/
after the final	/lavɛrstr/
consonant of the cluster	/lamɔrskr/
GROUP II	
Truncation in the	/ladɛkstɣ/
middle of the final	/lavɛrstɣ/
consonant	/lamɔrskɣ/
GROUP III	
Truncation immediately	/ladɛkst^{h}/
after aspiration of	/lavɛrst^{h}/
the third consonant	/lamɔrsk^{h}/
GROUP IV	
Truncation immediately	/ladɛkst/
after release of	/lavɛrst/
the third consonant,	/lamɔrsk/
before aspiration	

consonant of the cluster, the shortest group of stimuli for the main experiment were truncated after release of the third consonant (either a /t/ or a /k/) of the cluster. With three speakers, 9 utterances per speaker, and 4 truncation points per utterance, a total of 108 test items was available.

Truncation points were identified primarily by visual examination of the speech wave on the computer oscilloscope screen. Spectrograms and mingograms were examined for additional cues when necessary. Figure 1 shows a minogram of one of the utterances, and the four points of truncation. Identification of truncation points proved difficult for only one case: the identification of the end of aspiration of the third consonant. As displayed on the computer screen and the mingograph, aspiration was not easily separated from the following final consonant, /r/. Spectrograms were heavily relied upon for this information. Each edited utterance was checked by two listeners for auditory confirmation of the point of truncation.

Truncated utterances were played back from the computer through a digital-to-analog converter, low pass filtered at 6 kHz to remove high frequency digital noise generated by the computer, and recorded onto both channels of a two-channel Scully 280 tape recorder. The computer program also controlled the operation of the tape recorder; it was set so that 3.25 seconds of silence was recorded before and after each utterance, for a total of 6.5 seconds of silence between each item.

The order of taping items was randomized with respect to speaker, cluster, and vowel. Three practice items, picked at random

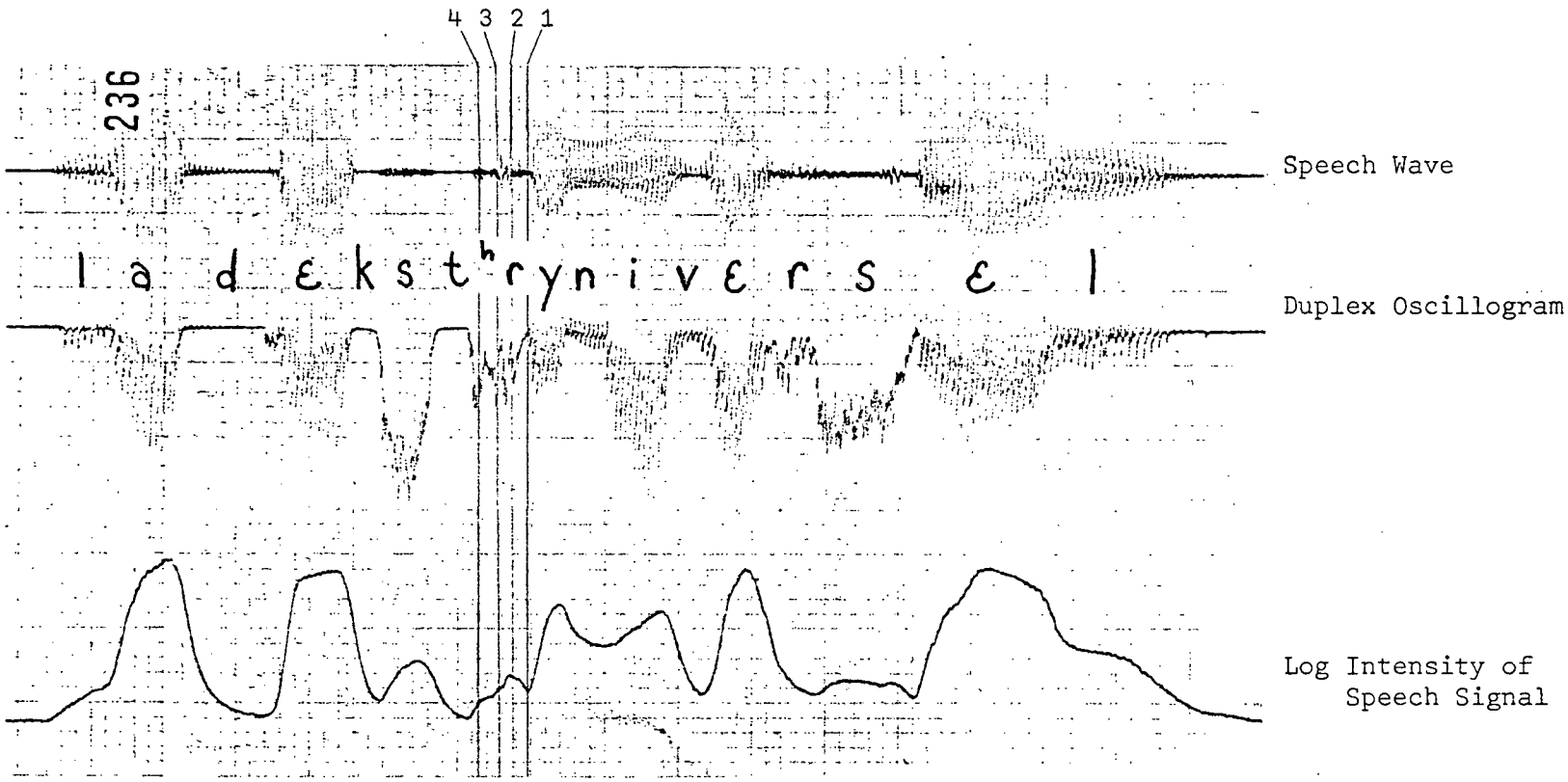


Figure 1. Mingogram of one of the test utterances, "la dextre universelle", showing the 4 points of truncation:

1. after the final consonant of the cluster (/r/)
2. in the middle of the final consonant
3. after aspiration of the third consonant of the cluster (/t^h/)
4. after release of the third consonant, before aspiration

from among the 108 test items, were recorded at the beginning of the tape. Two buffer items, also chosen at random from among the test items, were also recorded, one before the test items and one after the test items. Thus the tape contained three practice items, followed by utterances #1 to 110, of which #2 to 109 were the test items, and #1 and #110 were buffer items whose results were not considered in the analysis.

Two tapes were made from the original tape, using two Revox A77 tape recorders. Tape A contained items in the original random order. Tape B contained the same practice items, but the two halves of the test (i.e. #1 to 55, and #56 to 110) were interchanged. Thus two test tapes were available. The test was recorded on both tracks I and II of each tape.

Numbers were recorded before each test item. A non-native speaker of French recorded French numbers on channel I of each test tape, and an English speaker recorded English numbers on channel II.

Subjects

A group of 10 native French speaking adults and a group of 10 native English speaking adults participated in the experiment.

Four females and six males made up the French speaking group. They had been in North America from 3 to 14 years. Seven subjects had been born in France, two in Switzerland, and one in Haïti. One of the French-born subjects had lived in several places in Europe as a child, but had always spoken French in the home. One of the Swiss subjects had grown up speaking both French and German, though French was her

mother language. The Haïtian subject had grown up speaking both French and Spanish. All subjects had at least a working knowledge of English. Four subjects within the French speaking group had no knowledge of phonetics, while three had had formal phonetic training and three were teachers of the French language with some informal phonetic background. Three of the subjects had served as the speakers on the test.

Six females and four males made up the English speaking group. All subjects had had approximately 4 years of high school French in Canada, while two had had additional French courses in university, also in Canada, and had each spent several months in France. None of the subjects considered himself fluent in French. Six subjects had no knowledge of phonetics while the other four had some degree of phonetic training.

All 20 subjects passed a pure tone hearing screening test at 15 dB HL for the frequencies 500, 1000, 2000, 4000, and 6000 Hz.

Test Procedure

The subjects were seated, one at a time, in a soundproof room with the experimenter. The test tape was played on a Scully 280 tape recorder and presented over TDH-39 Maico headphones at a level of 60-70 dB SPL as measured on a Brüel and Kjaer 2203 precision sound level meter with a Brüel and Kjaer 6 cm^3 4152 artificial ear. The experimenter monitored the test over headphones and controlled movement of the tape in the soundproof room by a remote control unit.

Subjects were instructed in writing to listen to each utterance and to mark the missing vowel on an answer sheet. The

missing vowels were described as "i" as in "dites," "u" as in "une," and "ou" as in "bout." (See Appendix II for complete instructions) The vowels were phonetically transcribed as /i/, /y/, and /u/ for those who had a knowledge of phonetics. English subjects were first asked whether they were familiar with the vowels as represented in French orthography. The experimenter then pronounced each vowel in isolation for the English subjects.

Included in the instructions were the nine whole utterances from which the edited versions had been taken. Inclusion of this list was meant to show subjects that each truncated utterance could in fact be followed by each of the three vowels. Subjects were told that the vowels were represented in approximately equal proportion on the test (that is, that each vowel appeared approximately 1/3 of the time). Guessing was strongly encouraged. Subjects were asked to mark an indication of the confidence they had in their answers by marking their response with a 1 (for most confident), 2 or 3, only if they felt they had time to make this judgment.

The tape track containing French numbers was played for all but three subjects. It was one of these subjects, a French speaker who was one of the speakers on the test, who suggested that numbering be done in French instead of the original English. Subsequently all French subjects heard French numbers. Each English subject was asked whether he preferred to hear the numbers in French or English, and each chose French.

The tape was stopped after the three practice items and subjects were given the opportunity to hear these items again.

Measures of Perceivability

The two measures described below, relative transmission (T_{rel}) and correct score (S), were used in analyzing the results of both the pilot and the main experiment.

The relative transmission is a measure of covariance between input (the stimulus), and output (the subject's response)[Miller and Nicely, 1955]. This measure was used to describe the amount of transmissible information available in the truncated utterances, and is given by

$$T_{rel}(x;y) = - \sum_{i,j} p_{ij} \log_2 \frac{p_i p_j}{p_{ij}} \Big/ - \sum_{i} p_i \log_2 p_i$$

where the input variable is x, with any one input x_i having the probability p_i, and the output variable is y, with any one output y_j having the probability p_j. The symbol p_{ij} represents the probability that a particular input x_i will elicit the particular response y_j. The more consistently a response can be predicted from the stimulus, that is, the better the transmission of information, then the closer T_{rel} is to a value of 1. If the transmission of information is poor, then stimulus and response are unrelated, and T_{rel} has a value near 0.

Values of relative transmission for a series of computer-generated random responses were calculated. Figure 2 shows the distribution of T_{rel}, based on random responses (with equal probabil-

ities of 0.333 each) to 1000 27-item tests. In this graph, the bins for T_{rel} values from 0 to 50% represent intervals of 1%, the n^{th} bin representing the number of cases where T_{rel} has a value between $0.01 \times n$ and $0.01 \times (n+1)$. Each asterisk represents 2 cases. A bin with less than 2 cases shows one asterisk.

The correct score, either in % or absolute value, was also calculated for each subject. Figure 3 shows the distribution of correct scores, based on computer-generated random responses (with equal probabilities of 0.333 each) to 1000 27-item tests. Each asterisk represents 2 cases, a bin with less than 2 cases showing one asterisk.

The above distributions are based on tests of 27 items for comparison with each group of edited utterances, as there were 27 items of Group I utterances, 27 items of Group II utterances, and so on.

Figure 2. Distribution of T_{rel} based on random responses to a 27-item test.

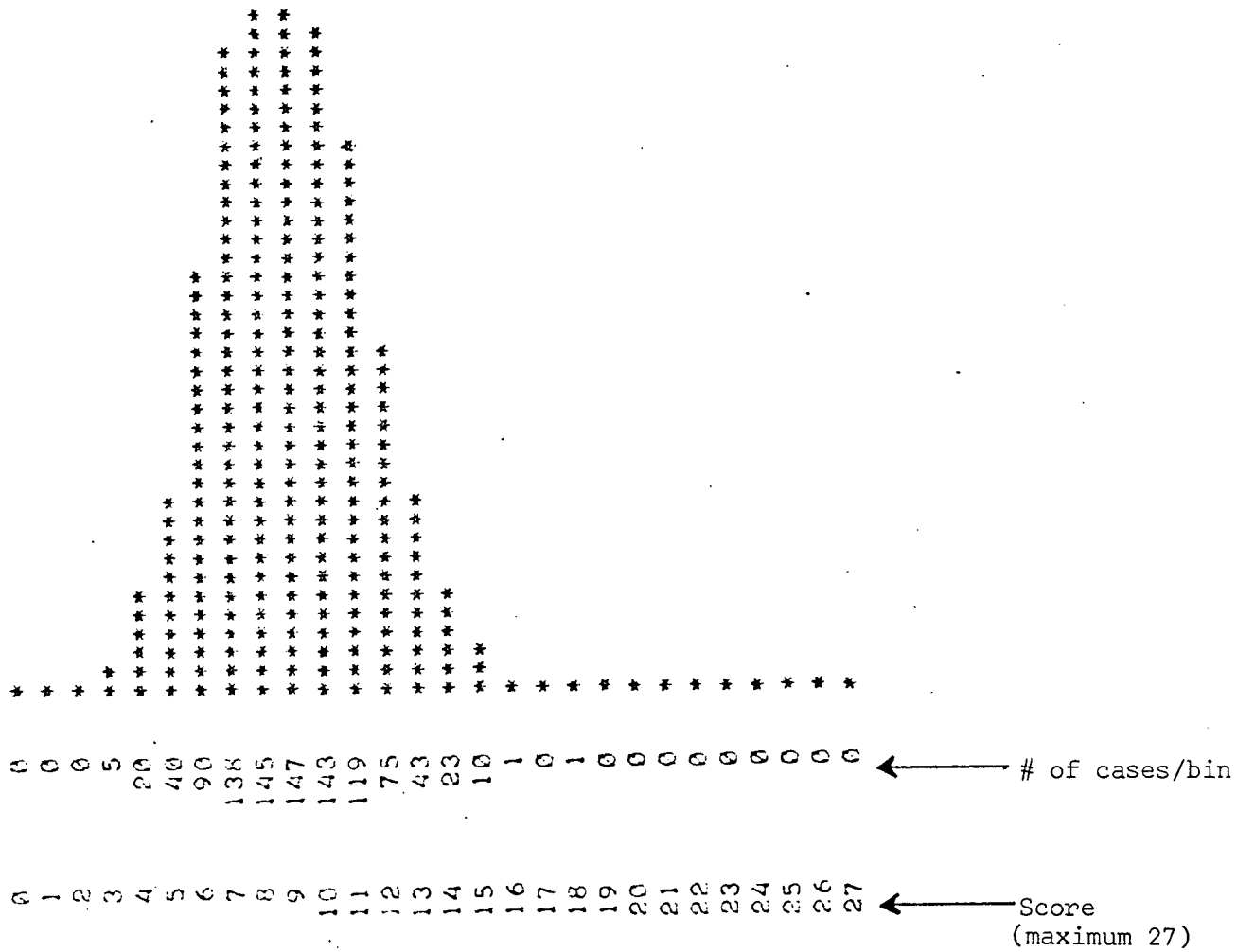


Figure 3. Distribution of Scores based on random responses to a 27-item test.

CHAPTER 5

RESULTS

5.1 Correlation Between Relative Transmission (T_{rel}) and Score (S)

Because the relative transmission is a measure of information transmission, and not necessarily *correct* information transmission, a high T_{rel} value does not always reflect a high score. For example, if a subject consistently responds with the vowel /i/ when the stimulus is /y/, he is receiving predictable information from the stimulus. Although he misinterprets this information consistently, he obtains a high value of T_{rel} (assuming other responses are also highly predictable from their stimuli). If a subject responds in a random manner, using no information from the stimulus, S will be at chance level, for example in the test described here, distributed around 9 (out of a maximum of 27), as shown in Figure 3; T_{rel} will be relatively low, distributed as shown in Figure 2. The better a subject performs on the test, the better one would expect T_{rel} and S to correlate.

To determine whether this was so for performances on the present test, Pearson correlation coefficients were calculated between T_{rel} and S for each of the 4 groups of items and also for a series of random responses. Correlations were as follows:

Group I:	R = 0.91
Group II:	R = 0.73
Group III:	R = 0.76
Group IV:	R = 0.50
Random Responses:	R =-0.08

From the above, one sees that the longer the portion of the cluster in the test item, the better the correlation between the two measures used here to describe performance. It will also be seen in Section 5.2 that the longer the test item, the better the subjects' performance. Therefore, as expected, the highest correlations between T_{rel} and S occur for the items for which the subject does best. In examining performance for Group IV items, one measure is not a good indicator of the other measure.

5.2 Identification of the Missing Vowel

Responses were tabulated in 3 x 3 confusion matrices, one matrix per group of items and per subject. There were therefore 4 matrices per subject, each with 27 items.

T_{rel} and S were calculated for each matrix. Figure 4 shows mean T_{rel} and S values for each of the four groups of items, displayed separately for French and English speakers. T_{rel} and S values one standard deviation about the mean are also shown. Levels that one subject would obtain by chance 1%, 5%, and 10% of the time (obtained from Figures 2 and 3, Chapter 4) are also shown in Figure 4. As can be seen in that figure, all subjects showed a downward trend in both T_{rel} and S, from Group I to Group IV. That is, the farther from the vowel the utterance was truncated, the less able subjects were to correctly identify the vowel. An analysis of variance showed a significant treatment effect among the groups of items for both T_{rel} and S for both French and English subjects. This effect is significant

at the levels indicated in the table below.

		French	English
Treatment Effects On:	T_{rel}	$p < 0.001$	$p < 0.05$
	S	$p < 0.001$	$p < 0.005$

The Newman-Keuls test, which indicates between which groups of items significant differences exist [Winer, 1971, pp. 191-196] was also applied to the data. Significant differences were found between several pairs of groups of items, for both French and English speakers, as shown in Table II.

Results show a great deal of individual variation. Figure 5 shows T_{rel} and S values for each group of utterances, for 3 different subjects. Levels that one subject would obtain by chance 1%, 5%, and 10% of the time are indicated. Subject AS scored consistently higher than any other subject. She was a native French speaker who was a teacher of French, but had had no formal phonetic training. Subject CM was a female native English speaker who had had some phonetic training. Subject CB was a male native speaker of French and one of the speakers on the test; he had also had some phonetic training. Because of her high performance relative to other subjects, subject AS was retested. On the second run of the test she maintained her high levels, scoring slightly higher than she had on the first run.

Results as shown in Figures 4 and 5 indicate that, for several of the item groups, subjects were able to identify the missing vowel above chance levels. For example, on the average, English

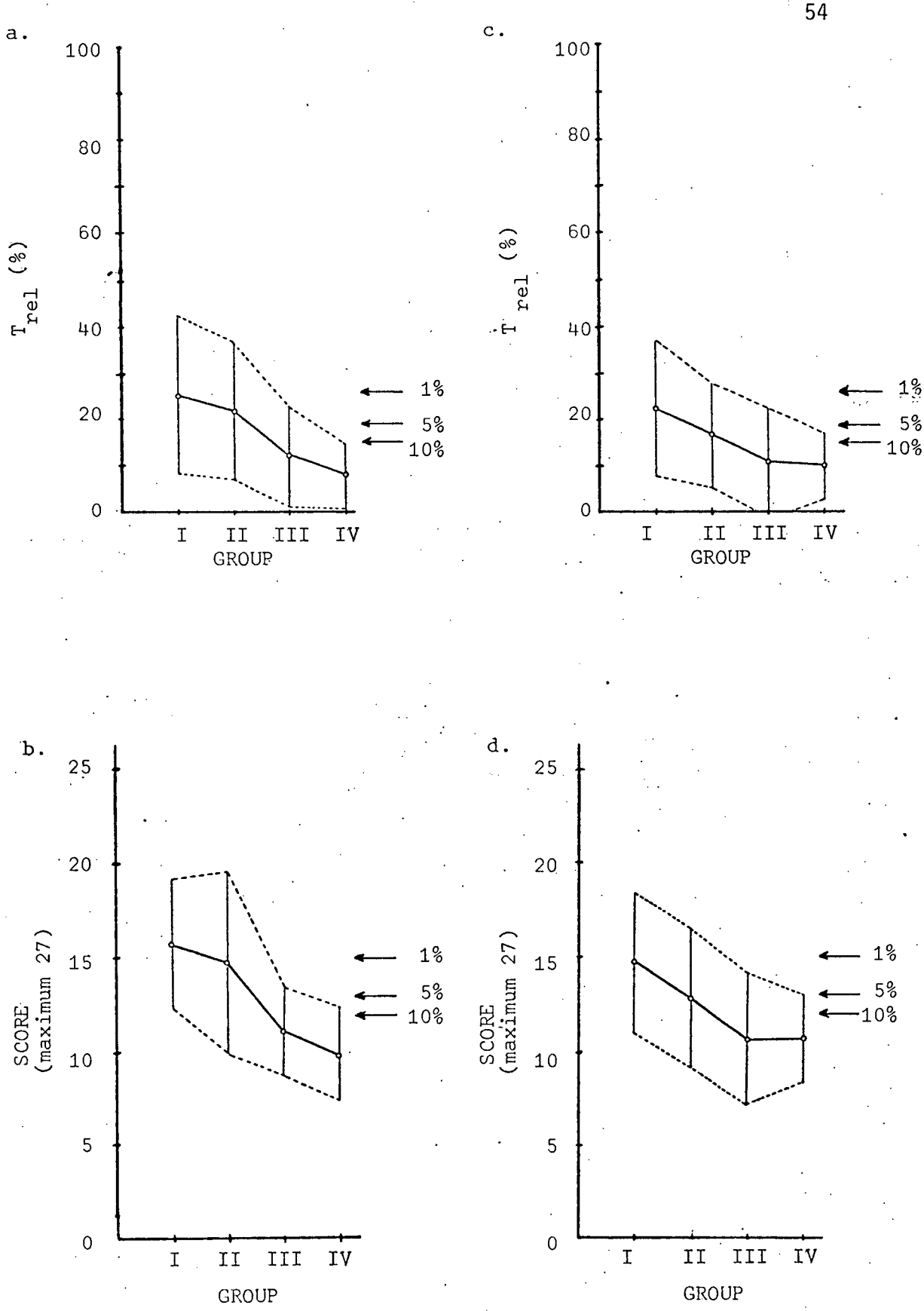


Figure 4. Mean values of T_{rel} and S, plus or minus one standard deviation, for each group of items.
a,b - French speakers
c,d - English speakers

TABLE II

Differences in T_{rel} and in S Between Each Group of Items, for French and English Subjects

FRENCH SPEAKERS

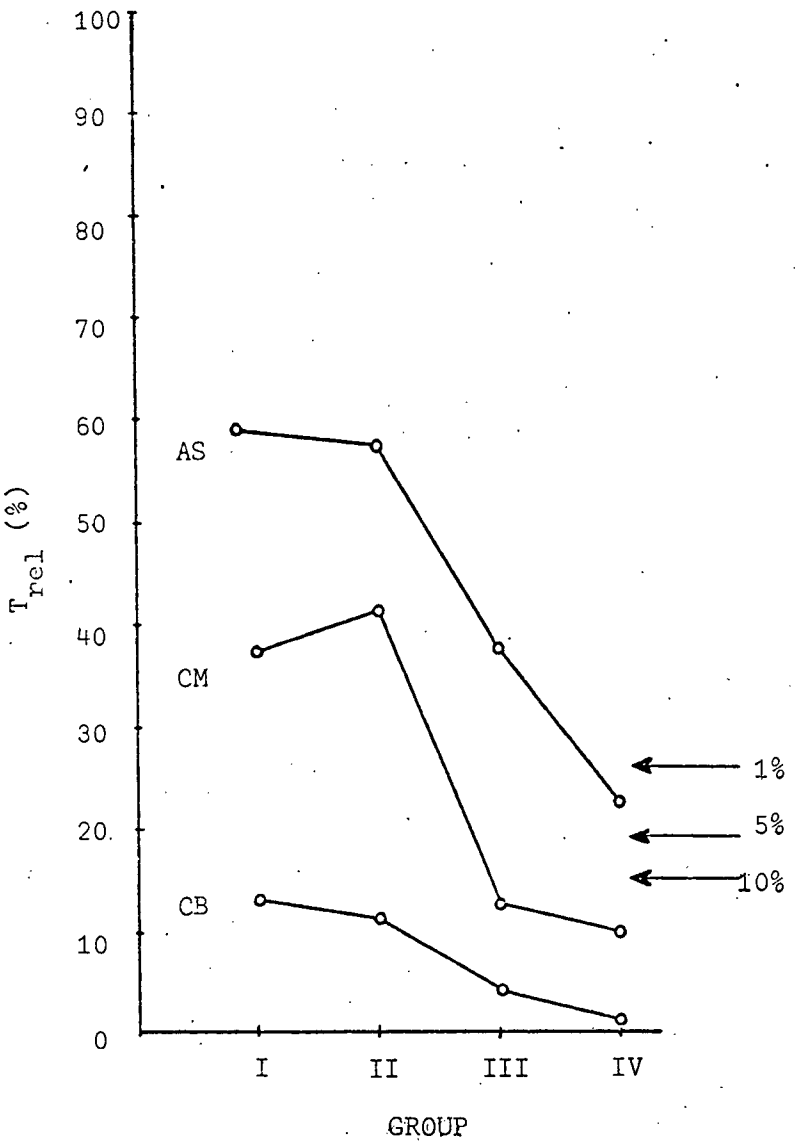
T_{rel} (%)				Score			
Group	I	II	III	Group	I	II	III
II	35.18			II	10		
III	132.03**	96.85*		III	46**	36**	
IV	172.97**	137.79**	40.92	IV	58**	48**	12

ENGLISH SPEAKERS

T_{rel} (%)				Score			
Group	I	II	III	Group	I	II	III
II	60.99			II	19		
III	121.33*	60.34		III	41**	22	
IV	123.27*	62.28	1.94	IV	40**	21*	1

** 0.01 level of significance.

* 0.05 level of significance.



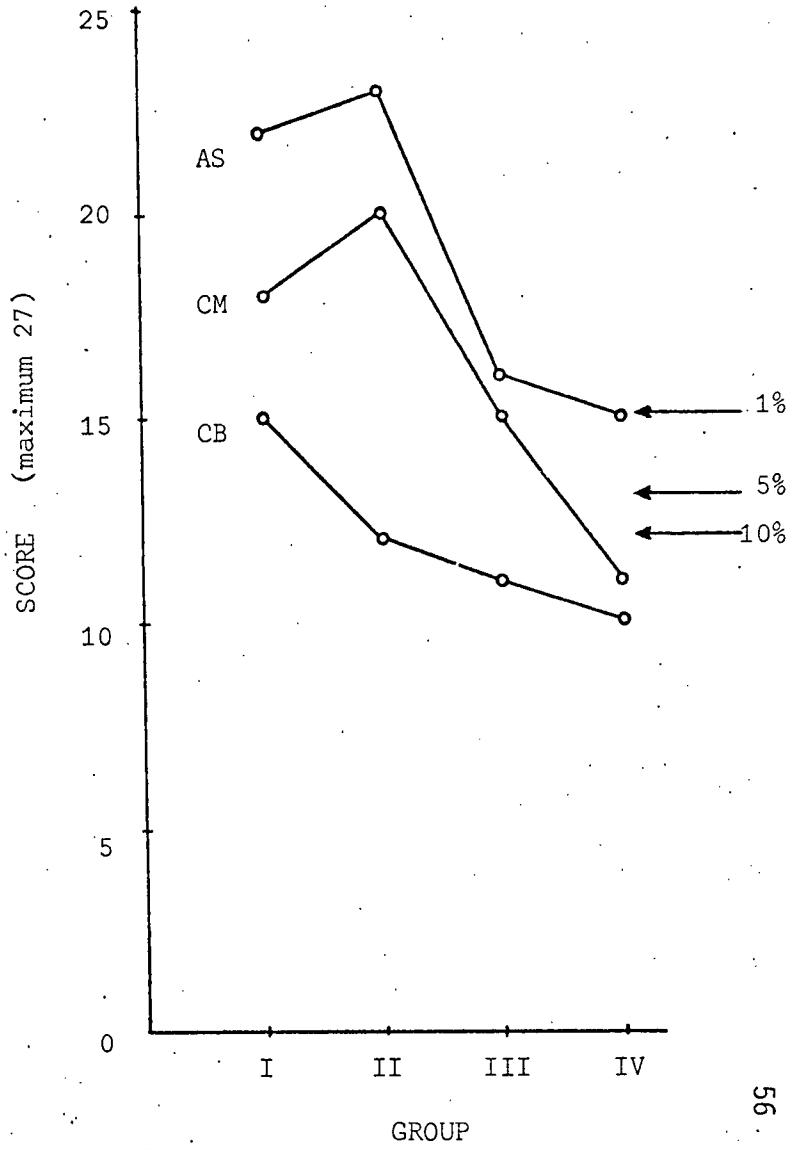


Figure 5. T_{rel} and S values for each group of items for three different subjects.

subjects obtained T_{rel} and/or S values higher than those one subject would obtain by chance 5% of the time, for items of Group I, and French subjects, on the average, obtained similarly high levels for items of Group I and II. Several individuals of both languages performed well above the 5% chance levels for Groups I and II, and 5 individuals did so for Group III. In general though, Group III and IV performances were at the level which one subject would obtain by chance 80% of the time. It is interesting to note that individual variations were so great that some subjects were able to identify the vowel for Group III and IV items better than others were able to identify vowels for Groups I and II.

In general, French subjects tended to distribute their responses evenly, responding approximately 1/3 of the time with each vowel. This tendency was somewhat weaker for the English subjects, who for Groups III and IV tended to make more /u/ and /i/ responses respectively.

Correct answers were not evenly distributed for either language group, for any group of items. For all groups except Group IV, correct /y/ responses were less frequent than correct /i/ or /u/ responses. French subjects did not show a different pattern of correct responses from English subjects. The percent of correct responses for all items of a particular stimulus, pooled for all subjects, is shown in Table III.

TABLE III

Percent of Items Answered Correctly in Each Vowel Category for Each Group of Items, All Subjects Pooled Together

GROUP	STIMULUS		
	/i/	/y/	/u/
I	64.4%	40.5%	63.9%
II	51.7	40.0	61.1
III	41.2	38.3	41.5
IV	45.0	35.6	33.4

Examination of the confusion matrices showed certain confusions to be more common than others. Confusions between /i/ and /y/, and /y/ and /u/, were more common than the /i/-/u/ confusion. This is not surprising when one considers that while /i/ and /y/ share the front feature, and /y/ and /u/ share the rounding feature, /i/ and /u/ share neither of these. Confusions between all pairs of vowels increased as the items got shorter, the only exception to this downward trend being for the /y/-/u/ confusion, which was made less often in Group IV than in Group III. Table IV shows the percent of time each confusion was made. For example, in Group I the /i/-/y/ confusion was made on 27.5% of the items for which either /i/ or /y/ was the stimulus.

Several subjects reported that they were most often undecided as to whether the missing vowel was /i/ or /y/, or /y/ or /u/, and several reported that they were never confused between /i/ and /u/. Performances seem consistent with the first observation, but not strictly so with the second.

TABLE IV

Percent of Items in Each Group for Which Various Vowel Confusions Were Made, All Subjects Pooled Together

GROUP	VOWEL CONFUSIONS		
	/i/-/y/	/y/-/u/	/i/-/u/
I	27.5%	27.8%	10.3%
II	27.5	27.2	16.1
III	30.8	37.2	21.7
IV	40.6	28.1	24.2

Further findings on feature relationships between stimulus and response are discussed in Section 5.3 below.

5.3 Identification of Individual Features of the Missing Vowel

To examine the perceivability of a particular feature, responses were grouped in the following ways: /i/ and /y/ vs /u/ (front-vs-back), and /i/ vs /y/ and /u/ (unrounded-vs-rounded). That is, if the stimulus was a front vowel, and the response either the same or the other front vowel, the response was considered correct in the front/back analysis. A similar procedure was employed for the unrounded/rounded analysis. When a 27-item 3 x 3 confusion matrix, for which each row has a total of 9 entries, is collapsed in the manner described above, a 2 x 2 matrix results in which one row has 9 entries, and the other row 18 entries. In such a matrix, the feature for which the data are grouped forms 2/3 of the total data. Therefore an error among the

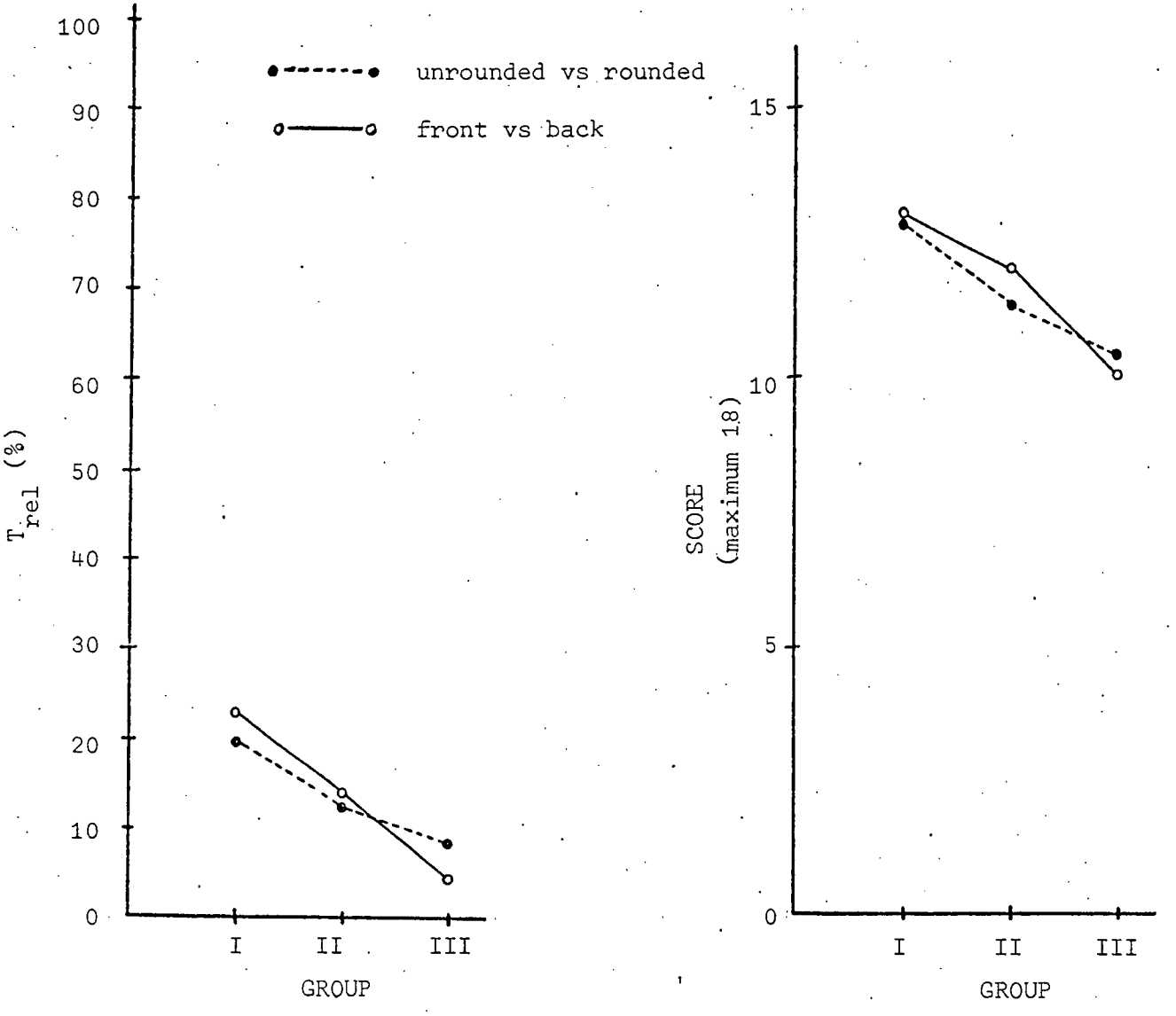


Figure 6. Mean values of T_{rel} and S for front-vs-back distinctions, and unrounded-vs-rounded distinctions, shown for three groups of items.

grouped data affects the score more than does an error among the ungrouped data. To overcome this imbalance, values in the row containing 18 entries were halved before T_{rel} and S were calculated.

Figure 6 shows mean values of T_{rel} and S for perception of the front/back distinction and for the unrounded/rounded distinction. Values shown are mean values for all 20 subjects. Only responses for Groups I to III are shown, as Group III responses are already at chance levels.

An analysis of variance showed that, for items of Group I, subjects did not make front/back distinctions significantly better than they made rounded/unrounded distinctions. This was the case when either T_{rel} or S was taken as an indication of performance. Because differences were greatest for Group I but yet were not significant, analysis of variance between feature distinctions in the other groups was not carried out.

Much individual variation was seen, both in ability to make a feature distinction, and in which distinction, either front/back or unrounded/rounded, was more easily made. The table below shows the wide range of T_{rel} values for feature distinctions for 4 French subjects. It also shows that some subjects made the front/back distinction more often, some made the unrounded/rounded distinction more often, and some made both distinctions equally. Similar individual differences were observed for English subjects.

Subject	T_{rel} for Front/Back Distinction	T_{rel} Unrounded/ Rounded Distinction
DN	0.30%	2.19%
EA	8.17	25.33
CB	10.52	11.24
PC	65.49	29.07

5.4 Differences Between Subject Groups

Tables V and VI compare mean T_{rel} values and standard deviations for the 10 French and 10 English subjects, and for the 7 phonetically trained and 13 phonetically naive subjects, respectively. Treatment-by-levels analyses of variance showed no significant differences between French and English subjects, for either T_{rel} or for S, and no significant differences between phonetically trained and phonetically naive subjects, for T_{rel} or for S.

5.5 Speaker Differences

Responses were examined to see if subjects performed better on items spoken by one of the three speakers than by the others. No significant differences were found between results for items by each speaker. However, again some individual variations were observed. Several subjects stated that items spoken by one or another of the

TABLE V

Mean T_{rel} Values ($\bar{x}$) and Sample Standard Deviations (s) for Each Group of Items, Shown for French and English Subjects. Differences Between the Two Language Groups are not Significant. N = Number of Subjects in Each Group

SUBJECT	N	GROUP			
		I	II	III	IV
French	10	$\bar{x}$ 25.4	21.9	12.2	8.1
		s 17.3	14.9	10.8	7.1
English	10	$\bar{x}$ 22.7	16.6	10.5	10.3
		s 14.8	11.2	11.7	7.2

TABLE VI

Mean T_{rel} Values ($\bar{x}$) and Sample Standard Deviations (s) For Each Group of Items, Shown for Phonetically Trained and Phonetically Naive Subjects. Differences Between the Two Groups With Different Phonetic Backgrounds are not Significant. N = Number of Subjects in Each Group

SUBJECT	N	GROUP			
		I	II	III	IV
Phonetically Trained	7	$\bar{x}$ 26.2	22.4	13.2	11.1
		s 6.2	7.4	7.5	4.9
Phonetically Naive	13	$\bar{x}$ 22.8	17.5	10.4	8.2
		s 14.3	10.6	7.9	4.9

speakers were easiest to answer. Such remarks were usually consistent with the subjects' better performance for that particular speaker, but there was no consistent trend as to who the "best" speaker was.

All of the three speakers served as subjects in the test. They did not perform consistently differently from the other subjects. Nor did they perform best on the items for which they themselves were speaking. In fact, two of the speakers performed somewhat worse on items for which they were the speakers, than for items uttered by another speaker.

CHAPTER 6

DISCUSSION

6.1 Identification of the Missing Vowel

French utterances containing the sequence $-C_1C_2C_3C_4V-$ were truncated at four points before the vowel, as shown in Table I (Chapter 4). Cowan [1973] has shown that for production of these utterances by native French speakers, upper lip protrusion most often begins with the first consonant of the cluster or earlier, when the vowel following the cluster is a rounded one. Items of each of the four groups prepared for the present experiment therefore contained different amounts of information as to the nature of the following vowel.

Subjects were able to predict the upcoming vowel above chance levels for items in Groups I and II. These items had been truncated after C_4, and in the middle of C_4 respectively. In all cases, C_4 was the phoneme /r/. In general, subjects were unable to predict the upcoming vowel for items in Groups III and IV. These items had been truncated after aspiration of C_3 (C_3 being /k/ or /t/), and after release of C_3 but before aspiration, respectively.

One sees from these results, represented graphically in Figures 4 and 5 (Chapter 5), that the segments up to and including C_4 contain information about the following vowel that is utilizable in the perception process. This information is not restricted to the C_4V juncture, since the vowel can be correctly predicted when segments up to only the middle of C_4 are heard. Though on the articu-

latory level the influence of the vowel is apparent as far back as the first consonant of the cluster or earlier, the information present in C_3 of the cluster and before is not by itself utilizable by the listener as an aid in identifying the upcoming vowel. It is not apparent whether the perceivable information is restricted to C_4 or whether it is the cumulative information present in the whole cluster up to and including at least half of C_4 which is used in perception. However, because coarticulation due to the vowel may begin by the first consonant of the cluster, it seems likely that several segments, and not just C_4, contain information which, when it is all available to the listener, can be used in the perception process, but when only early segments are available, is not perceptually useful. However, there are great individual differences, and one subject at least was able to consistently predict the upcoming vowel above the level she would obtain by chance 5% of the time even for items of Group IV.

Lehiste and Shockey [1972] have determined that coarticulatory effects in VCV utterances are not perceivable, whereas Sharf & Ostreicher [1973] cite evidence that these effects are perceivable in CVNV utterances.

It is not clear why, for some utterances, coarticulated information is perceivable, while for others it is not. The extent of coarticulation may depend on several factors: on the articulators involved [for example, Carney and Moll, 1971], the place, manner, and voicing characteristics of the neighbouring phonemes [Stevens and House, 1963], and the language being spoken [Öhman, 1966]. Depending on factors such as these, coarticulation at the articulatory level may be of an extent to produce more or less perceivable effects, or none at all. It may be, for instance, that coarticulatory influences of a vowel on a preceding nasal (as in Sharf and Ostreicher's

study) are perceivable, whereas coarticulatory influences of a vowel on a preceding stop consonant (as in Lehiste and Shockey's study) are less so. Further studies comparing perceivability of coarticulatory influences on fricatives, nasals, stops, and glides, voiced and unvoiced, would yield results relevent to this matter. In comparing such studies to the present one, it should be noted that the /r/ used here is the uvular fricative, as opposed to the English retroflexed sonorant.

Table II (Chapter 5) shows between which groups of items performance differed significantly. One sees that for both French and English subjects, and for both T_{rel} and S, no significant differences were found between Groups I and II, although the trend was a slight decrease from I to II. Subjects were able to identify the vowel when segments only up to the middle of C_4 were present almost as well as they could identify it when they heard all segments including the entire consonant. Similarly, no significant differences were noted between Groups III and IV, indicating that hearing all of C_3 did not increase a listener's performance over hearing only part of that consonant. It seems that without the information present in C_4, the amount of other preceding information present makes no difference to a listener's ability to identify the following vowel.

6.2 Identification of Individual Features of the Missing Vowel

Ali et al. [1971] found that coarticulated nasality was perceivable in CVN and CVVN utterances from which the final nasal was deleted. The present study found that coarticulation of two other

features, front/back and unrounded/rounded, also have perceivable effects.

As shown in Tables V and VI (Chapter 5), /i/-/u/ confusions were less frequent than /i/-/y/ or /y/-/u/ confusions, and the vowel /y/ was correctly identified less of the time than the other vowels. These results are probably due to the fact that, while /i/ and /y/ share the feature value front, and /y/ and /u/ the feature value rounded, /i/ and /u/ share neither of these. Thus on hearing an item containing information for a /y/, a subject may misinterpret it as either an /u/ or an /i/, based on his perception of the shared features discussed above. Similarly, he may misinterpret an /i/ as a /y/, but is less likely to misinterpret it as an /u/; he may misinterpret an /u/ as a /y/, but is less likely to misinterpret it as an /i/. Because /y/ shares features with both other vowels, misinterpretations of the kind described here are more likely to occur for the vowel /y/ than for the other vowels.

Figure 6 (Chapter 5) compares perceivability of the front/back and unrounded/rounded distinctions. Individual features are known to differ significantly in intelligibility, some being more readily perceivable than others [Wang and Bilger, 1973]. However, the features in question here are equally perceivable, though several individuals were better able to make one distinction than the other.

As expected, perception of either feature decreased as the test item grew shorter, performances for Groups III and IV being at the chance levels indicated at the right of each graph in Figure 6. One sees that, just as segments preceding C_4 provide no usable infor-

mation regarding the vowel on their own, they also provide no usable information regarding a feature of the vowel.

Comparison of Figures 4 and 6 shows that scores (converted to %) were considerably higher for feature identification than for vowel identification, in part a consequence of collapsing the matrix and including entries off the diagonal of the 3 x 3 matrix. However T_{rel} values for feature identification were slightly lower than for vowel identification. One expects feature identification to be better than vowel identification, the subject being presented with a two-way discrimination task in the first case, and a three-way task in the second. The slight decrease in T_{rel} from the vowel to the feature condition shows that, when both feature distinctions are considered together, as is necessary for correct vowel identification, slightly more information is abstracted from the stimulus than when either feature is considered on its own.

The distinctions front/back and unrounded/rounded are the manifestations of specific articulatory gestures, necessary for the production of the vowels described above. Coarticulation in the utterances described here causes the articulators to initiate these gestures in anticipation of the upcoming vowel. To an extent, this effect is perceivable. In the case of rounding, such anticipatory coarticulation is known to occur as early as the first consonant of the $C_1..C_4V$ sequence, yet in general, is perceivable only if segments up to and including at least half of C_4 are present, and not if less than this amount of information is available. It is not known how extensively fronting is coarticulated in these utterances, but similar

to rounding, it is perceivable only if all segments including at least half of C_4 are present.

6.3 Differences Between Subject Groups

Results show that the perceivability of coarticulated information does not seem to be related to the listener's native language, even though one of the vowels employed in the study (/y/) is not an English phoneme. Such findings are in agreement with the findings of Stevens et al. [1969], that the listener's linguistic background, be it English or Swedish, did not affect his ability to make subphonemic distinctions, even among vowels that were not present in his language. Though the pattern of coarticulation may be language-dependent, its perception does not seem to be.

Phonetic training did not affect the test results. The subjects with phonetic background in general were not better able to identify the missing vowel than the phonetically naive subjects. This suggests that the ability to make use of coarticulatory information does not depend on specific training. However, large individual variations in test performance indicate that not all listeners make the subphonemic distinctions necessary to predict the missing vowel. Whether they are completely unable to do so, or whether several subjects were not sufficiently motivated or did not completely understand the task, is not clear. Other researchers of speech perception abilities have also noted considerable individual differences [Ali et al., 1971; Liberman et al., 1957; Stevens et al., 1969]. It is likely that

some subjects in these tasks are more motivated than others; however, it also seems possible that some individuals possess keener powers of discrimination than others. Several of the poorer performers in the present study had in fact shown keen interest and motivation in the task.

6.4 Subjects' Comments

Without exception, all subjects reported that they found the test difficult. Most felt certain they had performed badly (though they may or may not have), and that they had guessed a large proportion of the time. The fact that subjects thought the test was a difficult one and that they had "only guessed" does not necessarily mean that the perceptual mechanism, to a large extent working subconsciously, could not handle the task. However, only one of the twenty subjects consistently gave an indication of his confidence in each of his responses, as was suggested in the instructions. This seems to indicate that the task of identifying the vowel was sufficiently difficult to impede subjects from making the further decision of how confident they were in each response. This may mean, that though the vowel was perceivable to an extent, use of coarticulatory information is not a process used in everyday speech perception.

As noted in Section 5.5, some subjects tended to do better on items spoken by one speaker than the others, though there was no general trend for all subjects to perform best for one particular speaker. Subjects were usually correct when they stated they had

performed best for one of the speakers. Familiarity with one or more of the speakers did not affect a subject's performance, and the three speakers, who also served as subjects, did not perform best on their own utterances.

Most subjects could not describe the strategy they had used in responding. However, several subjects were seen to repeat the test item subvocally two or three times before choosing their response.

Another comment some subjects made was that their choice was sometimes influenced by a vowel heard in the test item. Specifically, for an item containing the sequence /lamɔrsk/, they would tend to choose the vowel /u/ as the missing one, because of the back vowel /ɔ/ in the test item. Other subjects reported that they tended to choose /u/ for items of the speaker with the lowest voice, and one phonetically trained subject said she often chose /i/ and /y/ for a speaker who she judged to have "more fronted speech." Subjects were sometimes, but not always, accurate in their descriptions of their response tendencies. Thus it seems several factors may have influenced a subject's response, perhaps sometimes masking out the perceivable effect due to coarticulation. However, none of the factors described above was looked at specifically in the analysis.

6.5 Conclusions

Ali et al. [1971] hypothesize that if the effects of coarticulation are perceivable, then speech perception can be said to follow speech production and make use of its idiosyncracies. This relationship is predicted by the motor theory of speech perception

[Liberman et al., 1967]. Results of the present study suggest that such a relationship between production and perception exists to an extent. The perception process can make use of some of the idiosyncracies of production; coarticulated information is only sometimes perceivable in -C..CV utterances, notably when all segments up to and including C_4 are present. The present results seem to be predicted by Wickelgren's [1969] model of context-sensitive coding, in which each unit specifies its right- and left-hand neighbours. The final consonant of the cluster contains information which specifies the immediately following vowel, but segments preceding the final consonant seem to contain no perceivable information regarding the vowel. However, as discussed previously, it is likely that it is the cumulative information present in all preceding segments that is used perceptually. Also, there is no reason to assume that coarticulatory influences of a vowel could never be strong enough to produce a completely perceivable effect on a phoneme more than one removed from the vowel. Some subjects were able to identify the vowel when hearing utterances truncated after C_3 of the cluster, suggesting that, for them at least, context sensitivity is not limited to the immediately neighbouring phoneme. In addition, C_3 in the utterances used here was always a voiceless stop. It is not known what the coarticulatory influence of the vowel on a nasal or fricative in that position may be.

The fact that subjects can use subphonemic coarticulatory information to identify an upcoming vowel does not mean that the perception process necessarily incorporates this ability. There is evidence that subphonemic distinctions are not as well perceived as

phonemic ones [Liberman et al., 1957; Stevens et al., 1969], and speech perception seems to be primarily a categorical process. But it is possible that in unfavorable conditions, such as a noisy environment or a large amount of information having to be processed quickly, coarticulatory effects are used as cues by the perceptual mechanism. Use of such redundant cues would facilitate correct identification of any one speech sound. It is clear that some coarticulatory effects provide significantly perceivable information to the listener.

BIBLIOGRAPHY

ALI, L., GALLAGHER, T., GOLDSTEIN, J., and DANILOFF, R. (1971). "Perception of Coarticulated Nasality," J. Acoust. Soc. Amer. 49: 538-540.

AMERMAN, J.D., DANILOFF, R., and MOLL, K.L. (1970). "Lip and Jaw Coarticulation for the Phoneme /æ/," J. Speech Hearing Res. 13: 174-161.

CARNEY, P.J., and MOLL, K.L. (1971). "A Cinefluorographic Investigation of Fricative Consonant-Vowel Coarticulation," Phonetica 23: 193-202.

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.

COWAN, H.A. (1973). "A Study of Upper Lip Protrusion in French," Master's Thesis, University of British Columbia.

DANILOFF, R., and MOLL, K.L. (1968). "Coarticulation of Lip Rounding," J. Speech Hearing Res. 11: 707-721.

DELATTRE, P.C., LIBERMAN, A.M., and COOPER, F.S. (1955). "Acoustic Loci and Transitional Cues for Consonants," J. Acoust. Soc. Amer. 27: 769-773.

FROMKIN, V.A. (1966). "Neuro-Muscular Specifications of Linguistic Units," Language and Speech 9: 170-199.

FRY, D.B. (1964). "Experimental Evidence for the Phoneme," in In Honour of Daniel Jones, David Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, and J.L. Trim, Eds. (Longmans, London), 59-72.

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," J. Acoust. Soc. Amer. 25: 105-113.

KELSEY, C.A., WOODHOUSE, R.J., and MINIFIE, F.D. (1969). "Ultrasonic Observations of Coarticulation in the Pharynx," J. Acoust. Soc. Amer. 46: 1016-1018.

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).

KUEHN, D. (1970). "Perceptual Effects of Forward Coarticulation," M.A. Thesis, University of Iowa.

LADEFOGED, P., and BROADBENT, D.E. (1960). "Perception of Sequence in Auditory Events," Quart. J. Exp. Psych. 12: 162-170.

LEHISTE, I. (1972). "The Units of Speech Perception," Working Papers in Linguistics No. 12, The Ohio State University, 1-32.

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.

LIBERMAN, A.M., COOPER, F.S., SHANKWEILER, D.P., and STUDDERT-KENNEDY, M. (1967). "Perception of the Speech Code," Psych. Review 74: 431-461.

LIBERMAN, A.M., HARRIS, K.S., HOFFMAN, H.S., and GRIFFITH, B.C. (1957). "The Discrimination of Speech Sounds within and across Phoneme Boundaries," J. Exper. Psych. 54: 358-368.

LOTZ, J., ABRAMSON, A., GERSTMAN, L., INGEMANN, F., and NEMSER, W.J. (1960). "The Perception of English Stops by Speakers of English, Spanish, Hungarian, and Thai," Language and Speech 3: 71-77.

MACNEILAGE, P.F. (1963). "Electromyographic and Acoustic Study of the Production of Certain Final Clusters," J. Acoust. Soc. Amer. 35: 461-463.

MACNEILAGE, P.F. (1972). "Speech Physiology," in Speech and Cortical Functioning, John H. Gilbert, Ed., (Academic Press, New York & London), 1-72.

MACNEILAGE, P.F., and DECLERK, J.L. (1969). "On the Motor Control of Coarticulation in CVC Monosyllables," J. Acoust. Soc. Amer. 45: 1217-1233.

MILLER, G.A., and NICELY, P.E. (1955). "An Analysis of Perceptual Confusions Among Some English Consonants," J. Acoust. Soc. Amer. 27: 338-352.

MOLL, K.L., and DANILOFF, R. (1971). "Investigation of the Timing of Velar Movements During Speech," J. Acoust. Soc. Amer. 50: 678-684.

ÖHMAN, S.E.G. (1966). "Coarticulation in VCV Utterances: Spectrographic Measurements," J. Acoust. Soc. Amer. 39: 151-168.

ÖHMAN, S.E.G. (1967). "Numerical Model of Coarticulation," J. Acoust. Soc. Amer. 41: 310-320.

PERKELL, J.S. (1969). Physiology of Speech Production: Results and Implications of a Quantitative Cineradiographic Study (M.I.T. Press, Cambridge).

PETERSON, G.E., and BARNEY, H.L. (1952). "Control Methods Used In a Study of the Vowels," J. Acoust. Soc. Amer. 24: 175-184.

SAVIN, H.B., and BEVER, T.G. (1970). "The Nonperceptual Reality of the Phoneme," J. Verb. Learning Verb. Behavior 9: 295-302.

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.

STEVENS, K.N., and HOUSE, A.S. (1963). "Perturbation of Vowel Articulations by Consonantal Context: An Acoustical Study," J. Speech Hearing Res. 6: 111-128.

STEVENS, K.N., HOUSE, A.S., and PAUL, A.P. (1966). "Acoustical Description of Syllabic Nuclei: An Interpretation in Terms of a Dynamic Model of Articulation," J. Acoust. Soc. Amer. 40: 123-132.

STEVENS, K.N., LIBERMAN, A.M., STUDDERT-KENNEDY, M., and ÖHMAN, S.E.G. (1969). "Cross Language Study of Vowel Perception," Language and Speech 12: 1-23.

WANG, M.D., and BILGER, R.C. (1973). "Consonant Confusions in Noise: A Study of Perceptual Features," J. Acoust. Soc. Amer. 54: 1248-1266.

WICKELGREN, W.A. (1969). "Context-Sensitive Coding in Speech Recognition, Articulation, and Development," in Information Processing in the Nervous System, K.N. Leibovic, Ed., (New York-Heidelberg-Berlin: Springer), 85-95.

WINER, B.J. (1971). Statistical Principles in Experimental Design, (McGraw-Hill Book Company, New York).

APPENDIX I

Utterances Used in the Experiment

la dextre inimitable	/ladɛkstrinimitabl/
la dextre universelle	/ladɛkstryniversɛl/
la dextre outragée	/ladɛkstrutraʒe/
l'averse tribale	/lavɛrstribal/
l'averse truquée	/lavɛrstryke/
l'averse troublée	/lavɛrstruble/
l'amorce criptique	/lamɔrskriptik/
l'amorce cruciforme	/lamɔrskrysifɔrm/
l'amorce croupissante	/lamɔrskrupisãnt/

APPENDIX II

Instructions

You will be hearing a tape of a series of short French utterances. The end of each utterance has been deleted. Listen carefully and decide what vowel will follow the truncated utterance. The possible answers are the French vowels "i" as in "dites," "u" as in "une," and "ou" as in "bout" (that is, the phonetic symbols /i/, /y/, /u/).

For example, the utterance may be:

la dextre inimitable
or la dextre universelle
or la dextre outragée

However, you will hear the phrase cut off before the vowel:

la dextr(e)---

In all cases, your task is to decide if the missing vowel is "i," "u,"or"ou," *Choose your answer on the basis of what you hear, and what vowel sounds as if it is coming up.* Do not be concerned with the meaning of the utterance.

The next sheet contains a list of all the utterances. Remember, you will not be hearing the whole utterance, only a shortened form. The list is meant to familiarize you with all the possible answers. Your task is to identify only the missing vowel.

Mark your answer in the appropriate column on the answer sheet. If you feel you do not know the answer, it is important that you guess. Approximately 1/3 of the answers are "i," 1/3 "u," and 1/3 "ou." These numbers are only approximate, so listen carefully and mark your answer as the vowel you feel most sure is the missing one.

If you like, you can mark an indication of the confidence you have in your choice. If you are reasonably sure you have answered correctly, mark a '1' beside your answer. If you are not too sure of the answer you have put down, or if you have no confidence at all in your response, mark a '2' or a '3' respectively beside the answer. You need not make this judgment for each response if you feel you do not have the time.

There are 110 items on the test. It will last approximately 20 minutes. You will first be hearing three practice items, after which the tape will be stopped in case you have any questions. You may ask to stop the tape any time during the test if you feel you need a break, but no item will be repeated.

Choose your answer on the basis of what you <u>hear</u>, and what vowel <u>sounds</u> as if it is coming up.