

COARTICULATION AND LIPREADING

by

MARGARET KATHLEEN PICHORA-FULLER

B.A., University of Toronto, 1977

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

IN THE FACULTY OF GRADUATE STUDIES
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

February 1980

© MARGARET KATHLEEN PICHORA-FULLER, 1980

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 Audiology and Speech Sciences

The University of British Columbia
2075 Wesbrook Place
Vancouver, Canada
V6T 1W5

Date March 29, 1980

ABSTRACT

This study investigates the relationship of articulatory variation to the visual perception of phonemes. Normal hearing and hearing-impaired subjects who had demonstrated good lipreading skills on a pilot test were selected to lipread videotaped tests under visual only conditions. Eighty-one V_1CV_2 utterances where V could be /i,æ,u/ and C could be /p, t, k, tʃ, f, θ, s, ʃ, w/ were spoken by a speaker who had been selected in a pilot study as being easy to lipread. The 81 stimuli were used to construct three test tapes, one where the speaker spoke slowly, one where she spoke faster, and one in which the fast tape was reversed. Coarticulatory influences were expected to be present in these stimuli. Lipreading scores and measurements of the articulations were compared in an effort to explain some of the variability in the visual perception of phonemes which was suggested by existing literature.

Lipreading performance was nearly perfect for /p,f,w,θ,u/ on all tapes in all disyllables. Lipreading performance on /t,k,tʃ,ʃ,s,i,æ/ varied depending on phonological context, especially on the fast test tape. Variation in the identification of the less visually dominant phonemes could be directly related to coarticulatory effects revealed in the measurement of articulatory parameters (vertical and horizontal lip opening) of the visual signal. Improvement in lipreading ability throughout the task was evidenced by normal hearing subjects. The features labial, rounded, and alveolar or palatal place of articulation transmitted more information to lipreaders than did the feature continuant.

It was concluded that variability in articulatory parameters resulting from coarticulatory effects in faster speech increases lipreading difficulty, especially initially. Lipreaders are sensitive to subphonemic and subvisemic variations.

TABLE OF CONTENTS

Chapter	Page
ABSTRACT	ii
LIST OF TABLES	vi
LIST OF FIGURES	vii
LIST OF APPENDICES	ix
ACKNOWLEDGEMENT	x
 1. INTRODUCTION	 1
2. REVIEW OF THE LITERATURE	
2.1 Visual Identification of Phonemes without Consideration of Coarticulation	3
2.2 Phoneme Identification with Consideration of Coarticulation	10
2.3 Measurement of Articulatory Gestures	11
2.4 Phonetic Literature on Coarticulation	15
2.5 General Lipreading Literature	21
3. PILOT STUDY	
3.1 Aims	26
3.2 Preparation of The Pilot Test	26
3.3 Administration of The Pilot Test to Lipreaders	28
3.4 Results	32
3.5 Comments	37

Chapter	Page
4. MAIN STUDY	
4.1 Introduction	40
4.2 Aims	41
4.3 Preparation of The Main Test	42
4.4 Administration of The Main Test to Lipreaders	46
4.5 Measurements of Articulatory Gestures	49
5. RESULTS OF THE MAIN TEST	
5.1 Overview of Viseme Identification	51
5.2 Context Effects on Viseme Identification	60
5.3 Sequential Effects	67
5.4 Reverse Test Results to Investigate the Importance of Syllable Position of Vowels	71
5.5 Measurement of Articulations	75
6. CONCLUSIONS	84
BIBLIOGRAPHY	87
APPENDIX I - Instructions for Pilot Test	91
APPENDIX II - Instructions for Main Test	92
APPENDIX III - Measurement of Transmission of Information	94

LIST OF TABLES

Table	Legend	Page
I	Lipreader and Testing Situation Information	29
II	Lipreaders - Background Information	31
III	Frequency of Common Consonant Confusion Errors	58
IV	Frequency of Common Feature-Related Errors	59
V	Duration Ratios for V_1 and V_2	78
VI	Duration Ratios of Vowels for Fast Compared to Slow Test Tape	80
VII	Duration Ratios of Consonants for Fast Compared to Slow Test Tape	80

LIST OF FIGURES

Figure	Legend	Page
1.	Proposed Visemes	5
2.	Confusion Matrix Showing Vowel Confusions of Twenty Lipreaders	36
3.	Equipment Set-Up for Recording Session	44
4.	Display of Markers and Signal Information Used to Match Visual Events, Acoustic Events, and Measurements of Articulations	50
5.	Group Trends for Lipreaders Revealed by Results of the Main Test	52
6.	Identification of Visemes	55
7.	Transmission of Information by Articulatory Features	57
8.	Context Effects on Initial Vowel Errors	61
9.	Consonant Errors in Initial Vowel Context	62

Figure	Legend	Page
10.	Consonant Errors in Final Vowel Context	63
11.	Context Effects on Final Vowel Errors	64
12.	Sequential Effects on Raw Score and Transmission Score of Consonants	68
13.	Sequential Effects on the Transmission of Information by Articulatory Features	70
14.	Reverse Test - Pattern of Errors	73
15.	Measurements of Articulations on the Fast and Slow Test Tapes	76
16.	Measurements of Visual Parameters for Selected Stimuli Sampled Every Five Fields	77
17.	Relationship of Acoustic and Visual Parameters	79
18.	Evidence of Coarticulation of Rounding	83

LIST OF APPENDICES

Number	Title	Page
I	Instructions For Pilot Test	91
II	Instructions for Main Test	92
III	Measurement of Transmission of Information	94

ACKNOWLEDGEMENT

I would like to recognize with thanks those who have contributed to this work:

- Dr. André-Pierre Benguerel for the considerate assistance and for the fine model he has provided as a teacher.
- Noelle Lamb for her assistance.
- My subjects for their co-operation.
- The audiologists and teachers involved for their interest and for their assistance in recruiting subjects.
- The audio-visual technicians for sharing their expertise and equipment.
- My classmates for their friendship.
- My family, Keith and Mihai for their support.

CHAPTER 1

INTRODUCTION

The term lipreading, or speechreading, refers to the use of visually available articulatory information in speech and language perception. The two terms are often used synonymously. There is a tendency for lipreading to be used when the message can be described in linguistic units, especially phonological units. The lips provide the information used by the lipreader. More recently, the term speechreading has been used when paralinguistic and extralinguistic information is also considered. The information obtained in speechreading is recognized to include kinesic, pragmatic, sociologic, and other information.

Auditory perception of speech involves the correct abstraction, identification, and use of acoustic cues which are present in the speech signal. The speech signal is produced by complex articulatory gestures. Observation of these articulatory gestures, with or without accompanying auditory perception of speech, provides visual cues which can be abstracted, identified, and used by a lipreader in the decoding of speech.

The speech signal does not consist of discrete acoustical units. Articulatory gestures and, in turn, the resulting acoustic cues associated with any particular phoneme are subject to change depending on context. The term coarticulation refers to the altering of the set of articulatory movements made in the production of one phoneme by the set of articulatory movements made in the production of an adjacent or

nearby phoneme in the utterance. During the production of a bilabial consonant, the tongue could begin to move into position for a following vowel. Despite the constant changes in the articulatory movements, and despite coarticulatory effects, the listener is able to decode the linguistic message. Visual cues which can be used to decode the linguistic message are obtained in the observation of some of the same articulatory movements which produce the acoustic signal for auditory speech perception. The visual cues derived from articulatory gestures must also be subject to the effects of coarticulation. The lipreader is apparently able to use varying visual information to decode speech.

This study investigates the ability of lipreaders to use visual information alone to identify phonemes in varying contexts. The phonemes and contexts represented in the study were chosen to allow careful examination of nearby as opposed to remote coarticulatory effects. The accuracy of identification of phonemes and the nature of the errors in the identification of phonemes in varying contexts will be considered with reference to articulatory movements and coarticulatory effects.

CHAPTER 2

REVIEW OF THE LITERATURE

2.1 Visual Identification of Phonemes Without Consideration of Coarticulation

Alexander Graham Bell is credited with first hypothesizing that all phonemes are not visually distinguishable (Deland, 1931, cited in Fisher, 1968). Classes of visually distinguishable phonemes have been called visemes (Fisher, 1968). Visemes in the visual mode of speech perception would be analagous to phonemes in the auditory mode. Phonemes which cannot be distinguished visually would be classed as one viseme in much the same way that allophones are classed as one phoneme in the auditory mode. Utterances consisting of the same visemes are called "homophenous"; for instance, "mope" and "pope" are visually indistinguishable and /m/ and /p/ are classed as one viseme.

While some of the features used to distinguish visemes are similar to those used in the description of phonemes, the importance of other visually pertinent features must be considered. 'Speechreading movements' have been hypothesized (Jeffers and Barley, 1971) as the visual cues to identification of visemes. There is no good evidence that visemes can be contrasted through voicing. On the other hand, rounding may be a basis for distinguishing both phonemes and visemes. The source of the contrast would be found in the acoustic signal in the former case and in visually observed gestures in the latter case. The nature of visual cues pertinent to viseme identification has not been isolated. Features used to describe phonemes may be inappropriate

to describe visemes. Allophones of one phoneme may differ in regard to available visual features. The existing notion of viseme overlooks this possibility. Supposedly, categories do exist in the visual perception of speech, however, the nature of these categories has not been precisely defined.

Consonants

Visemes described as a set of visually non-contrastive phonemes have been suggested by many studies (Berger, 1972; Heider and Heider, 1940; Jeffers and Barley, 1971; Woodward and Barber, 1960; Fisher, 1968; Erber, 1972; Binnie et al., 1974 a, b; Walden et al., 1974, 1975). The number of visemes reported ranged from three to twelve. Non-contrastive sets of phonemes are typically delineated in terms of place and manner of articulation. The amount of visual information contributed by place and manner of articulation in these studies varies considerably. Figure 1 illustrates visemic contrasts encountered in the literature. Explanations provided for the variations illustrated in Figure 1 have included: test conditions, the nature of the experimental task, precision of the speaker in articulating, training of the lipreader, control of phonological environment, and linguistic redundancy.

Several studies examined the confusion errors made for consonants by using controlled phonological environments and a large set of response choices.

Erber (1972) presented nonsense syllables to hearing-impaired, profoundly deaf, and normal hearing children under auditory, visual, and combined auditory-visual conditions. Stimuli consisted of /b,d,g,k,m,n,p,t/ in the environment /a/ _ /a/. Three places of articulation, voicing, and nasality were investigated. Consonants were distinguished by place of

FIGURE 1: PROPOSED VISEMES

STUDY	PHONEMES																												CONTEXT					
	d p b m f v ɱ w ɹ k g ŋ h j m θ ð l ɹ ɪ d n k g j h s z ʃ ʒ ɪʃ dʒ																																	
1	p	b	m	f	v		w	ɹ	k	g						θ	ð	l		ɪ	d	n		j	h	s	z	ʃ	ʒ	ɪʃ	dʒ	#_[a]		
2	p	b	m				w	ɹ	k	g	ŋ					θ	ð	l		ɪ	d	n		j	h	s	z	ʃ	ʒ	ɪʃ	dʒ	V_#, #_V (3V's)		
3	p	b	m						k	g											ɪ	d	n										[a] _ [a]	
4	p	b	m	f	v											θ	ð			ɪ	d	n	k	g			s	z	ʃ	ʒ			#_[a]	
5	p	b	m	f	v		w									θ	ð										s	z	ʃ	ʒ			#_[a] (PRE-TRAINING)	
6	p	b	m	f	v		w	ɹ								θ	ð	l		ɪ	d	n	k	g	j		s	z	ʃ	ʒ			#_[a] (POST-TRAINING)	
7	d	p	b	m	f	v	ɱ	w	ɹ	k	g								l		ɪ	n			j	h	s	z	ʃ		dʒ		# _ (WORDS)	
8	p	b			f	v				k	g	ŋ			m	θ	ð	l	ɹ	ɪ	d	n					s	z	ʃ	ʒ	ɪʃ	dʒ	# _ (WORDS)	
9	p	b	m	f	v		w	ɹ	k	g	ŋ										ɪ	d	n					s	z	ʃ	ʒ	ɪʃ	dʒ	NON-EXPERIMENTAL
10	p	b	m	f	v		w		k	g	ŋ	h	j			θ	ð	l	ɹ	ɪ	d	n					s	z	ʃ	ʒ	ɪʃ	dʒ		NON-EXPERIMENTAL
11	p	b	m	f	v	ɱ		ɹ	k	g						θ	ð	l		ɪ	d	n			j	h	s		ʃ		ʒ	ɪʃ	dʒ	NON-EXPERIMENTAL

Phonemes grouped as one viseme are contained with solid vertical bars. Dotted vertical bars occur if there is no counter evidence to a grouping suggested by another study. Some phonemes appear twice in the figure for convenience in illustrating groups.

Studies: 1. Woodward and Barber (1960); 2. Berger (1972); 3. Erber (1972); 4. Binnie et al. (1974 a, b); 5. Walden et al. (1977); 6. Walden et al. (1977); 7. Fisher (1968); 9. Alexander G. Bell (cited in Jeffers and Barley (1971)); 10. Jena Method (cited in Jeffers and Barley (1971)); 11. Jeffers and Barley (1971).

articulation by all subjects under visual only conditions. However, confusions between homorganic phonemes occurred.

Perceptual confusions of sixteen CV monosyllables were studied by Binnie, Montgomery and Jackson (1974_{a,b}). Conditions of auditory-visual presentation at various signal-to-noise ratios as well as auditory only and visual only conditions were investigated. The analysis included intelligibility scores as well as a measure of relative transmission of information by articulatory features. These features were the same as those used by Miller and Nicely (1955); that is, voicing, nasality, affrication, duration, place of articulation. A sixth feature, "stop", was added. This type of analysis affords a more thorough consideration of the pattern of errors made in the visual perception of phonemes. Consonants used were /p,t,k,f,θ,s,ʃ,b,d,g,v,ɣ,z,ʒ,m,n,l/. Relative transmission for place of articulation information was found to be 98.2%. Nasality (11.3%) and voicing (11.0%) were found to convey little information visually. Affrication (63.13%) was relatively efficiently transmitted and was a basis for distinguishing /p,t,k,b,d,g,m,n,/ from /f,θ,ʃ,s,v,ɣ,z,ʒ/. Duration provided 58.1% and stop 27.7% relative transmission of information. For affrication, duration, and stops, identifications arose from subtests of phonemes distinguishable by place of articulation; for example, /p,b,m/ contrasted with /f,v,θ,ɣ/. Five visemes were found (see Figure 1). Place of articulation may contribute to many of the distinctions indicated as resulting from affrication, duration and stop features.

Woodward and Barber (1960), Fisher (1968), and Binnie et al. (1974_{a,b}) agree in that /s,z,t,d,n/ are never categorized separately on the basis of visual information alone. Differences in their results, such as a separate

category for /k,g,/ being found by Fisher but not by Binnie et al., or /ʃ,z/ being described as a category by Binnie et al. but not by Woodward and Barber, are suggested to result from test and lighting conditions.

Walden et al. (1974) set out to further investigate the nature of visual recognition of consonants by hearing-impaired subjects. CV sequences where the vowel is always /a/ and the consonant is one of /p,t,k,f,θ,s,ʃ,b,d,g,v,ʒ,z,ʒ,m,n,w,r,l/ were used. The mean correct identification score was 36.5% (one S.D.=5.4%) for thirty subjects on the visual only condition. The mean identification score for four visemes (Woodward and Barber, 1960) was 95.3%. The median score was 97%. It was concluded that subjects rarely confuse visually contrastive consonants, and that the same information is transmitted visually to all subjects.

Walden et al. (1975) presented CV's similar to the ones used in their 1974 study. Confusion matrices were analyzed according to the features voicing, nasality, liquid glide, frication, duration and place of articulation. Transmission of these features in auditory only and auditory-visual conditions were compared. Subjects with the poorest auditory only scores improved the most when visual information was also available. It was found that the transmission of the features duration, place of articulation, frication, and nasality information increases substantially when visual cues are provided. Some explanation for the improved transmission of features is directly attributable to place of articulation information; for example, auditorily, a nasality error may arise in the confusion of /m/ and /l/, however, since these phonemes differ by place of articulation, visual resolution of nasality is actually secondary. These results show the possible inappropriateness of existing

phonetic features to describe visual perceptions.

In another study by Walden et al. (1977), the same stimuli were used in an experiment which involved the comparison of pre-training to post-training confusions. Responses were analysed from the confusion matrices. Phonemes were classed in one viseme whenever within-class responses constituted 75% or more of the responses. Five visemes were determined in pre-training results and nine visemes were determined in post-training results (see Figure 1). Increases in the consistency within visemes were noted.

In the present study, coarticulation will be examined as a factor contributing to the variations in the visual perception of phonemes which have been suggested in these studies which report a diversity of identifiable visemes.

Vowels

Each vowel in English can be considered to have a unique place of articulation. It is, therefore, possible to hypothesize that each vowel could be a viseme.

Berger (1972) mentions two of his studies on vowel confusions. In the first study, vowels were presented in initial and final syllable position where the consonant of the monosyllable could take one of three values. In the second study, word pairs differing only in the vowel were presented to subjects. The greatest number of errors was made on vowels which are adjacent in the vowel triangle. Few errors involving confusions of front and back vowels or high and low vowels were recorded.

Erber (1971) tested for the visual intelligibility of the vowels /i, I, e, æ, ɔ, ʊ, u, ʌ, ɒ, ɜ/ in the context /b/_/b/. The vowels were

found to fall into two groups. The first group was easily distinguished. It consisted of /u, ʊ, i, a, ɔ/ which represent extremes in lip position and tongue articulations and are mainly "tense" vowels. The second group, which was less easily distinguished, consisted of /æ, ʌ, ɛ, ʊ, ɪ/. It represents less extreme articulatory positions and consists of mainly "lax" vowels. Although Erber does not mention it, duration may also have played a rôle.

Lowell (1974) examined the perceptibility of vocalic nuclei in CVC English words. Four visemes were hypothesized. These were based on the articulatory characteristics of lipspreading, liprounding, change in lip position from spread to rounded (/aw/), and change from rounded to spread lips (/ɔj/). The hypothesized visemes were born out for the most part. Within-viseme confusions did not follow a neat pattern; there were a significant number of confusions of central spread vowel responses for back rounded vowel stimuli. Internal confusions within groups were considerable. The scores varied somewhat depending on the phonological environment as will be discussed below. The task required identification of the word spoken. Difficulty of this task was suggested as a possible explanation for the low scores obtained for some visemes. The rounded viseme was 74% correct, the spread 42%, the spread to rounded 97%, and the rounded to spread 54%.

In summary, the results of these studies show that there are about four vowel visemes that are easily distinguishable. These are based on rounding, height, and diphthongal changes in the vowel nuclei. Within these visemes there is considerable confusion of the stimuli. The articulatory movements contributing to vowel productions vary more gradually than do those contributing to the more closed consonant articulations. Further consideration of vowels which takes coarticulation

into account could be important in understanding visual consonant confusions. In the consonant studies mentioned above, the vowel(s) was given.

2.2 Phoneme Identification with Consideration of Coarticulation

In the studies outlined in section 2.1, few phonological environments were sampled and varied phonological environments were seldom sampled. When phonological environment was varied, the lip-reader was typically informed of the variation that would occur. It was found, however, that some consonants and vowels were more difficult to lipread than others. This range of difficulty, as well as what is known about coarticulation in speech production and perception, suggest that it would be important to consider coarticulation in lipreading.

Initial consonant clusters preceding /ʌ/ were studied in a lipreading experiment by Franks and Kimble (1972). The purpose of the study was not to directly investigate coarticulatory effects, however, the results suggest that the perception of consonants is influenced by phonological environment. Groups of consonant clusters which were easily distinguishable seemed to include visually dominant consonants /b, m, p, r, w, ʃ, θ/.

Lowell (1974), in his investigation of vowel confusions, used four consonant environments. All vowels were presented in the environment after /b/ and before a non-labial consonant. An incomplete set of vowels was presented in three additional consonant environments: 1. after a labiodental, before a non-labial; 2. after an alveolar, before a bilabial; 3. after a glottal, before an alveolar. All stimuli were English words. The range of scores of correct identification of vowels varied considerably in the four contexts. The perception of phonemes entails

more than identification of articulatory gestures, but it is difficult to evaluate other contributing factors since three of the environments were only partially represented. Word frequency and expectancy were suggested by Lowell as being possible contributing factors. Frequency of occurrence of phonemes was not found to be important. The work of Franks and Oyer (1967) would also indicate that the number of homophenous words, as well as familiarity, would be important. Lowell suggested that errors in a preceding consonant could contribute to the incorrect identification of the vowel.

Erber (1971) reported the finding by Pesonen (1969) that vowel context influences visual recognition of consonants. Erber selected the environments /i/_/i/, /a/_/a/, /u/_/u/ used by Pesonen to test for the identification of the consonants /b,d,g,h,l,r,x,v,z,ʒ/. Consonants were presented in only one vowel environment at a time. The lipreader was informed of the set of consonants and of the environments. Consonant intelligibility was found to be the best in the environment of the open vowel /a/. Scores were found to be considerably lower for alveolar consonants in the environment /i/_/i/ and /u/_/u/.

These three studies differ in the variety of stimuli used. Franks and Kimble (1972) used a large set of consonant clusters. Lowell (1974) used English words representing four contexts. Erber (1971) used non-sense syllables varying in context, although the lipreader was informed of the context. It is clear in all of these studies that identification of phonemes varies considerably depending on context. Explanations based on interpretation by the lipreader of visual coarticulatory effects have not been thoroughly explored.

2.3 Measurements of Articulatory Gestures

Recently, lipreading studies have begun to combine physical

measurements of articulators with investigations of visual perception. Existing phonetic literature involving physical measurements promises many explanations of lipreading phenomena. Demonstration of these explanations has just begun.

In a phonetic study, Fromkin (1964) used frontal and lateral photographs, lateral x-rays, and plaster casts of speakers' lips in order to determine measurements of variables by which lip positions could be specified. She measured mouth width, midline height of mouth opening, area of frontal view, protrusion, and retraction of the corners of the mouth. Five subjects were measured uttering the words "heed", "hid", "hay", "head", "had", "hod", "hud", "herd", "hawed", "hoed", "hood", "who'd" with vowels /i, I, e, ɛ, æ, ə, ʌ, ɜ, ɔ, ʊ, u/ respectively. Findings for several utterances of one speaker include the following:

1. /u/ was distinguished from other vowels by having consistently the lowest values of lip height and width;
2. /ɔ, o, ʊ/ were not statistically distinguishable from each other by measurements of height or width. They were distinguishable from other vowels with respect to width.
3. /i, I, e, ɛ, æ, ə, ʌ/ were not distinguishable from each other, although a hierarchy in terms of height existed. /i/ had the minimum value and /æ/ or /ɛ/ had the maximum.

For /u, ʊ, ɔ, o, a/, width of opening decreased with height. For front vowels, width was relatively constant as height increased. Although lip distances were equal for /ɔ, I, o, i/, the distance between the teeth was much greater for the back vowels. The action of the lip muscles is involved in closing the lips in back vowels. Protrusion was found to be greater for the lower lip. An imperfect correlation was found between protrusion and lip height and width. Four muscle actions were proposed: rounding by the obicularis

oris muscle; pulling back of the lips; protrusion of the lower lip (without much rounding, e.g. [ɔ]); jaw raising and lowering. There were variations in absolute measurements from one subject to the next, however, relative measurements and range within each subject were constant. These may be the articulatory gestures perceived by lip-readers.

Sahlstrom and Oyer (1967) are reported by Berger (1972) to have made objective measurements of certain facial movements during the production of homophenous words. They found significant differences in the rate and intensity of movements occurring on the surface of the face as measured by strain gauge. The use of movements as measured for lipreading purposes was not determined.

Jackson, Montgomery and Binnie (1976) conducted an experiment in which they tried to correlate perceptual dimensions underlying vowel lipreading with physical lip measurements. Previous investigations had identified vowel features based on knowledge of articulation and on visual perceptual confusions. Judgements of similarity between two stimuli were made by lipreaders. The underlying assumption was that the lipreader establishes a perceptual space where stimuli are separated by distances determined by their similarities. Patterns of similarity would be expected to relate to perceptual features that are suggested by hypothesis or by actual measurements. Stimuli consisted of fifteen monosyllabic nonsense words, /hɪg, hɪg, hɛg, hæg, hag, hɔg, hʊg, hug, hʌg, hɜg, heɪg, haɪg, hɔɪg, haʊg, hoʊg/ spoken by six speakers. Five dimensions were identified by multidimensional scaling. The first correlated strongly with measurements of width of mouth opening. The second dimension correlated less strongly with vertical mouth opening

measurements. Dimension three correlated to some degree with all measurements. Dimensions four and five showed a relationship to the second vowel in diphthongs. Duration and rate were not measured.

Erber et al. (1978) conducted a study which involved physical dimensions and visual perception. A new technique to study lipreading was used in the study which involved the visual synthesis of isolated physical features in order to examine the use of only those features in perception. Mouth opening was produced on an oscilloscope while voicing was simulated by a voice-like buzz which was presented tactually. Various voice onset times relative to mouth opening were presented to subjects for identification as /pa, ba, ma/. Categorical perception was evidenced to some extent for trained subjects. Categories were learned to some extent by one subject who was given training.

Erber et al. (1979) continued to look for explanations of lipreading performance in physical measurements. Five tokens of each of /i, I, ε, æ, ə, ɔ, ʊ, u, ʌ, ɜ, a^ɪ, e^ɪ, ɔ^ɪ, a^ʊ, o^ʊ, jʊ/ in the environment /b/ _ /b/ were recorded. Measurements of vertical and horizontal mouth opening, lip thickness, teeth visibility, and jaw displacement were taken. Highly intelligible tokens were found to be characterized by extreme articulatory changes and quasi steady-state segments.

Scheinberg (1979) reports examination of articulatory cues used in lipreading as found by an inter-leaving technique. VCV syllables differing in the consonant which could be /p, b, m/ were videotaped. Alternate frames of the different consonants and of different tokens of the same consonant were interleaved. Areas of jitter appeared in the former case but not in the latter. Areas of most jitter (lower lips and cheeks) were suggested as areas requiring

detailed measurement. Perceptually, good speechreaders could discriminate homophenous pairs in a forced-choice experiment.

These approaches involving the integration of information obtained by physical measurements and by perceptual studies may be used to isolate visual features contributing to the identification of visemes. The need for quantitative data on facial dynamics and the use of this information in lipreading is the object of presently ongoing studies (Brooke, 1978; Summerfield and Spencer, 1978).

2.4 Phonetic Literature on Coarticulation

Studies in acoustic, articulatory, and perceptual phonetics have provided evidence of coarticulation. An outline of findings regarding the domain and nature of coarticulatory effects will be included in this section.

Domain

Phonetic studies outlined below illustrate the domain over which coarticulatory effects have been discovered.

Measurements of the first (F_1), second (F_2) and third (F_3) formants of vowels in a fixed environment (/h_d/) were determined by Peterson and Barney (1952). Stevens and House (1963) measured F_1 and F_2 of vowels in consonantal environments of the form /h C_1 — C_2 / where $C_1=C_2$. Values of F_1 and F_2 for vowels were found to diverge from those found by Peterson and Barney depending on the consonantal context. The extent of the influence of the consonantal environment on the vowel formant frequencies varied depending on the vowel. The influence of the consonant on the vowel varied systematically according to place of articulation, manner of articulation, and voicing characteristics of the consonant. Explanations for these effects were given in articulatory terms.

The domain of coarticulation studied was extended by Öhman (1966). House and Stevens were unable to distinguish the influence of the initial from that of the final consonant because C_1 was equal to C_2 . By using $V_1 CV_2$ utterances, where $V_1 \neq V_2$, Öhman was able to identify the influence of both V_1 and V_2 . Previous studies of CV's (Delattre et al., 1955) had been the foundation for the "locus theory" which claimed that each consonant is characterized by a particular frequency to or from which formant transitions of adjacent vowels are directed. Öhman's observations evidence context dependent characteristics of consonants rather than fixed characteristics. The notion of an utterance consisting of a linear assembly of phoneme segments having unique descriptions was replaced by the notion of overlapping of phonological features. Overlap could be accounted for according to the involvement of the articulators in producing tokens of phonemes.

Later studies further extended the domain of coarticulation studied. Evidence of coarticulatory effects in French ranging over six phonemes before a vowel were found for the feature rounding (Benguereel and Cowan, 1974).

Accurate description of the productions of tokens of phonemes must allow for coarticulatory effects by sampling productions in a variety of phonological environments. Context dependent characteristics of productions could contribute to differences in the perceptibility of phonological features. These characteristics could be influenced by near as well as far phonemes in the utterance.

Nature of Coarticulation

The nature of coarticulation determined by acoustic, articulatory

and perceptual studies will be outlined. These will include investigations of lip rounding, jaw lowering, and place of articulation.

Rounding

Electromyographic studies of the obicularis oris conducted by Kozhevnikov and Chistovich (1965) indicated that liprounding begins on C_1 of a C_1C_2V syllable where the vowel is rounded. The particular combination of vowels and consonants was related to differing coarticulatory influences in a study by Fromkin (1966). The peak of electromyographic activity for rounded vowels /u/ and /o/ was less following initial /b/ than initial /d/. /b/ itself involves contraction of the obicularis oris. Sussman and Westbury (1979) measured electromyographic activity from the obicularis oris superior muscle, which is the agonist for liprounding, and from the risorius muscle, which is the antagonist for liprounding and the agonist for lipspreading. Measurements were taken during the vowels /i/ and /a/ when they were followed by a consonant or consonant cluster and /u/. Coarticulatory influences of /u/ on the preceding vowel were found to include earlier and greater force of obicularis oris activity when /i/ was syllable-initial as compared to when /a/ was syllable-initial. These adjustments are thought to occur since the prior vowel /i/ is biomechanically antagonistic to the anticipated rounding, whereas they do not occur when the prior vowel (/a/) is biomechanically neutral in regard to rounding. Further evidence of coarticulatory effects involving rounding was determined from lateral fluorographic films (Daniloff and Moll, 1968). By using a photoelectric measurement of upper lip protrusion, Benguerel and Cowan (1974) examined sequences of four to six consonants before a rounded vowel in French utterances. Most clusters exhibited rounding on the first consonant; in half of the clusters, rounding was also exhibited during production of the vowel preceding the consonant cluster. In a

perceptual study, Benguerel and Adelman (1976) asked subjects to predict the identity of a missing vowel upon hearing truncated utterances. Subjects could use coarticulatory information due to liprounding which was present in the consonants to identify the missing vowel which would have followed.

Jaw Lowering

Amerman et al. (1970) used cineradiographic techniques to study the coarticulatory effects of lip and jaw movements. Utterances consisted of clusters of up to four consonants followed by /ə/. Jaw lowering, and, to a lesser extent, lip retraction were found to coarticulate over two and sometimes three consonants. Articulations required in the production of /s/ seemed to inhibit jaw lowering and lip retraction.

Place of Articulation

As mentioned above, Ohman (1966) found acoustic evidence for coarticulation. Transitions to stop consonants of particular places of articulation were found to be influenced by preceding and following vowels. Carney and Moll (1971) also investigated coarticulatory effects. Their results indicate that in symmetric VCV utterances, a tongue position similar to that required for the vowel was maintained during /f,v/. The tongue tip was not subject to coarticulatory effects for /s,z/, however, tongue body and root moved nearer to vowel positions. In asymmetric VCV's, evidence was obtained to confirm Ohman's suggestion of a vowel-to-vowel movement during consonant production. Lehiste and Shockey (1972) failed to find perceptual correlates to Ohman's acoustic findings.

Rate and Duration

Daniloff and Moll (1968) comment that although fast and slow

utterance rates resulted in similar extents of lip protrusion, fast rate durations were found to be 71% to 87% of the conversational rate durations. There were minimal differences observed in protrusion for the two rates.

Stevens, House and Paul (1966) found that there were consistent differences in vowel durations of three different speakers. The speaker producing shorter vowels also produced the greatest amount of undershoot of F_2 reflecting greater coarticulatory effects on control of articulatory movements. Rate of speech is suggested to have a possible influence on the extent of coarticulatory information.

Observations of particular vowel durations also illustrate a possible source of information in the perception of coarticulation. Stevens, House and Paul (1966) note that tense vowels are longer than their lax counterparts. Lax vowels are characterized as exhibiting shorter duration, less precise approach to the target articulation, and a drift of most of the articulatory structures toward neutral articulation immediately after the release of the initial consonant in a C_1VC_1 . They comment that the diffuse vowels /i/ and /u/ are shorter than the non-diffuse /a/ and /æ/ which may be a result of longer time required to achieve greater jaw opening. Longest durations were measured for tense, open vowels in the environment of voiced continuant consonants; shortest were observed for lax, close vowels in the environment of voiceless stop consonants. MacNeilage (1963) observed that although spectrographic findings indicate that /f/ in the environment $V_\#$ is twice the duration of /f/ in the environment $VC_\#$, nevertheless, electromyographic measurements did not reveal duration differences. Some visual coarticulatory information could be influenced by duration. The visual cues used by lipreaders have yet to be identified.

Syllable Position

"Öhman (1966) found that the formant frequencies of /y,ø,a/ tended to show more central values in final position than in initial position. These vowels were somewhat neutralized. The trend was not true of /o,u/. Kozhevnikov and Chistovich (1965) and MacNeilage (1963) have claimed that CV is the basic unit of production in which coarticulatory effects are strongest. Coarticulations in VC's are also well documented.

Summary

Phonological features of different natures show different domains of coarticulation, some of which are far ranging. Domains of coarticulation also show language dependent differences. Consideration of the ability of the articulators involved to move is necessary in order to understand coarticulation.

Comments

It is well to consider some of the cautions raised by authors relative to perceptual tasks.

Benguerel and Adelman (1976) note that subphonemic judgements are required of subjects in perceptual studies of coarticulation. In lip-reading, phoneme identification can be considered to be a sub-visemic judgement. Lipreaders are, however, constantly required to make these interpretations in daily language comprehension situations. Identification tasks could result in fewer distinctions than discrimination tasks.

Language learning may decrease the ability of subjects to make subphonemic judgements. The nativeness of lipreading is open to question. Erber (1974) has suggested that the visual code may serve a different linguistic

role for subjects having different degrees of hearing loss and training.

Memory restrictions in linguistic processing may come into play. Clark and Sharf (1973) presented VC, VCV, and truncated VCV's of the form VC to subjects in a short term memory task. Truncation resulted in lower recall scores, however, error patterns were unchanged. Coarticulation may contribute to certainty of identification of phonemes. Lehiste (1972) also suggests that memory factors could intervene between peripheral recognition of subphonemic features and a perceptual phonemic decision.

2.5 General Lipreading Literature

Lipreaders

The lipreading ability of a normal-hearing, hearing-impaired, and deaf subjects has been studied. Lipreading ability for normals was found to be similar in nature to that of hearing-impaired subjects (Woodward and Barber, 1960). Metz (1965 cited in Berger, 1972), found that normal-hearing subjects performed as well as hearing-impaired subjects on monosyllabic items. Hearing-impaired may be better lipreaders when a variety of test materials is used (Brannon and Kodman, 1959 cited in Berger, 1972). Within hearing loss groups, those with less loss have been found to be better lipreaders (Costello 1957, cited in Berger, 1972). Evans (1960, cited in Berger, 1972) found differences between hearing-impaired and deaf subjects on auditory and auditory-visual conditions, but not on visual only conditions. Erber (1974) has suggested that there may be two types of lipreading use. For unpracticed, or for normal-hearing subjects, lipreading would be a supplementary language code. For profoundly deaf subjects, it may be a code in itself based on visual,

articulatory (kinesthetic) and vibratory information.

Training is a difficult variable to isolate from factors such as age of onset of hearing loss, degree of loss, type of education. Walden et al. (1977) reported improvements in the visual identification of consonants as a result of training for hearing-impaired adults exhibiting noise-induced hearing loss who had had no previous formal lipreading instruction. The improvements included an increase in the number of visemes consistently recognized and an increase in the percentage of within viseme responses. Most of these changes were reported to occur in the first few hours of the total fourteen hours of concentrated, individualized training.

Lipreading was originally taught in conjunction with articulation therapy. Most schools of lipreading later denied the appropriateness of combining lipreading and articulation. The Jena method did continue to advocate subvocalization as a successful learning mechanism (Jeffers and Barley, 1971). Mlott (1964 cited in Berger, 1972) has found some evidence that lipreading scores are improved when a response involving articulation is required. Responses were oral, written, or passively given. Subjects using kinesthetic information performed three times better than the control group. How production and reception are linked in lipreading or how this information plays a rôle in lipreading strategies is not understood.

Studies are reported by Berger (1972) in which visual acuities of 20/20 or 20/40 were found to be adequate for lipreading purposes. He concluded that for successful lipreading, the articulators must be seen, although acuity is not of great significance.

The visual perception involved in lipreading is dynamic. It is difficult to relate common tests of visual perception to lipreading skill.

Shephard et al. (1977) found a significant correlation between firing time and speechreading ability when they measured average visual electroencephalic responses and speechreading scores. Scores increased as latency of the peak decreased. Processing of visual information was felt to be measured in this study.

Berger (1972) summarizes that the relationship between speechreading performance and intelligence reported in most studies seems to be small. This is at least the case when IQ is greater than 80. Verbal IQ tests show the most correlation.

Farrimond (1959 cited in Berger, 1972) reported that lipreading ability may improve from the second to the third decade. Thereafter, ability declines such that those over sixty years old score half of that scored by the 30 to 39 year old group when hearing level was controlled for. Goetzinger (1967 cited in Berger, 1972) reported a plateau a decade earlier, with significant differences between those in their twenties and those in their thirties. Coscarelli (1968 cited in Berger, 1972) also found that those under forty scored significantly better than older subjects.

Speakers

Berger (1972) reports studies suggesting that slower rates of speech may be easier to lipread. These studies often involve altering film presentation speed. Fast and slow natural speech would not vary by simple multiplication as is accomplished by changing film speed. Sahlström and Oyer (1967 cited in Berger, 1972) found significant differences in rate and intensity of movements on the surface of the face during productions by men and women. The relevance of these findings to lipreading ability is undetermined.

Speakers familiar to the lipreader have been found to be more easily lipread, especially in daily situations. Black et al. (1963 cited in Berger, 1972) found that normal-hearing subjects improved more in lipreading the same rather than different speakers over a period of time.

Stimuli

Sumby and Pollack (1954) found that scores for a vision only lipreading task decreased as set size increased. Franks and Oyer (cited in Berger, 1972) found that scores on words were more related to the number of homophenous words than to familiarity.

Sentence difficulty increases as length increases in terms of words or syllables (Morris, 1944; Taaffe and Wong 1957, cited in Berger 1972). More complex syntactic forms were found to be more difficult (Schwartz and Black 1967, cited in Berger 1972). Familiarity and difficulty of sentences are related (Lloyd, 1964; Lloyd and Price, 1971, cited in Berger, 1972). Scores for speechreading words in sentences are better than those obtained for words in isolation (Brannon, 1961). Costello (1957 cited in Berger 1972), in comparing deaf and hearing-impaired children's lipreading ability, found a greater significant difference for words than for sentences, the deaf performing less well on both types of stimuli. Deaf may use different redundancy cues than are used by hearing-impaired, or normal-hearing subjects.

The Utley test is in common use. Scoring has been standardized for the original filmed version of the test. There is considerable discrepancy between raw scores obtained on the Utley test when it was given in various live or filmed conditions, (Jeffers and Barley, 1971). It was found that scores on live and filmed presentations were highly correlated suggesting

that the same skill is being evaluated in both cases. Relative rather than absolute scores are of primary importance given the variability in results due to differences in presentation.

Conditions Influencing Lipreading Performance

Berger (1972) summarized that studies suggest that the more of the speaker that is visible, the easier the speaker is to lipread. Greenberg and Bode (1968) used consonant stimuli in comparing a full face to a lips-only view of the speaker. Significant differences favoured the full face view.

Berger (1972) reported studies to indicate that superior lipreading scores were obtained for viewing angles of 0° to 45° (front view) as opposed to a viewing angle of 90° (sideview).

Berger (1972) considered several studies and concluded that viewing distances of up to 20' to 24' do not appear to have a significant effect on speechreading performance. He reported that regular classroom lighting was adequate for normal lipreading purposes. Jackson (1967 cited in Berger, 1972) did find that a light source low in front of the speaker contributed to improved scores for inexperienced lipreaders but not for experienced lipreaders.

CHAPTER THREE

PILOT STUDY

3.1 Aims

The pilot test was designed to provide information to assist in decisions to be made regarding the construction of the main test tape and to help answer the following questions:

1. Who would be a good speaker to lipread?
2. Who would be good lipreaders? Only good lipreaders were to be chosen for the main test so that optimum rather than average ability would be measured. It was felt that the good lipreaders would be more likely to use coarticulatory cues while lipreading.
3. Which vowels would it be best to use as part of the stimuli of the test tape? If vowels which could be easily distinguished from each other in a neutral context were chosen, errors in the identification of vowels in the main test could be attributed mainly to coarticulatory effects since general difficulty of identification would not be in question.
4. What conditions of presentation would be preferable to the experimenter and to the subjects?

3.2 Preparation of the Pilot Test

Six speakers were recorded using a head and shoulders front view.

Items spoken by each speaker included the following:

1. five practice sentences;
2. ten numbers;
3. ten monosyllables;
4. ten or eleven sentences;
5. ten monosyllables

The five practice sentences were those of the Utley Test of Lipreading (Jeffers and Barley, 1971). They were presented in the same order by each speaker.

The ten numbers were different for each speaker, however, the list for each speaker was comparable to that of any other speaker. Each list consisted of an unique arrangement of a given set of digits combined to give numbers less than one hundred.

The same ten monosyllables were spoken by each speaker in randomized order. The monosyllables were "heed" /hid/, "hid" /hId/, "hay" /he/, "head", /hɛd/, "had" /hæd/, "hud" /hʌd/, "hod" /hɔd/, "hoed" /hod/, "hood" /hʊd/, "who'd" /hud/. These were chosen to sample lipreading of a variety of English vowels for which there is information in the phonetic literature (Fromkin, 1964).

The list of sentences was unique for each speaker. Sentences were selected from the Utley Test, forms A and B. A few changes were made so that sentences were more natural in the dialect of the subjects. Sentences were ordered by using a table of random numbers with the restrictions that each sentence be chosen only once, that each list contain ten or eleven sentences totalling 42 or 43 words, and that a range of sentence lengths be represented. There were three dubbings of the pilot test which differed only in the order of presentation of the speakers.

A short version of the pilot was used in the later stages of the study as a screening test for selecting lipreaders. The screening test consisted of the entire recordings for speakers 3 and 5. The two speakers were selected based on results of the administration of the complete pilot test to the first twenty lipreaders. Scores obtained by lipreaders for stimuli spoken by speaker 5 provided good information to distinguish lipreaders of different abilities. Speaker 3 was the most difficult to lipread; so that, lipreading ability was reflected to a great extent in scores obtained for stimuli spoken by her. Criteria for passing the test were established based on scores obtained by the lipreaders selected for the main test who had been administered the complete pilot.

Instrumentation, recording session format, and editing procedures were the same as those described for the main test (see Chapter 4.3). Measurement procedures were the same as those for the main test (see Chapter 4.5).

Speakers were six students of the Division of Audiology and Speech Sciences. All had training in phonetics. All had worked with hearing-impaired persons. All were females ranging in age from 22 to 27 years.

3.3 Administration of the Pilot Test to Lipreaders

The pilot tapes were presented on various TV monitors. Variation in equipment was considered to be less important, for the purposes of the pilot test, than was the opportunity of testing more subjects by going to locations in the community where lipreaders were readily available. The variety of locations is indicated in Table 1. The most comfortable viewing

Table 1: Lipreader and Testing Situation Information

Subject	Group Size	Order of Speakers	Location	Preferred Speaker	Tests Given	Hours of TV Watched at Home Each Week
1	3	1	A	1	F	3
2	3	1	A	1		5
3	3	1	A	1	F,S	2
4	2	2	B	5	F	10
5	2	2	B	5	F	0
6	3	2	C	4		5
7	3	2	C	1		0
8	3	2	C	4		1
9	3	3	C	5	F,S,R	10
10	3	3	C	5		0
11	3	3	C	5		3
12	7	3	B			?
13	7	3	B			?
14	7	3	B			?
15	7	3	B			20
16	7	3	B			25
17	7	3	B			7
18	7	3	B			36
20	2	2	C	5	F,R	8
21	2	2	C	5	F,S,R	6
22	1		C		F,S,R	0
23	2		C			2
24	2		C			0
25	2		C			0
26	2		C		F,R	0
27	4		A			16
28	4		A		F	10
29	4		A			4
30	4		A			10
31	1		C		F,S	0
32	2		C		F,F,R	2
33	2		C			?
34	4		D			2
35	4		D			10
36	4		D		F	5
37	4		D			2

situation was sought at each location.

Lipreaders were sought from two populations: normal hearing persons with phonetic training and hearing-impaired persons known to be lipreaders. Normal hearing subjects were recruited from the Division of Audiology and Speech Sciences. All but one of the subjects were females in their twenties. Hearing-impaired subjects were recruited at the recommendation of teachers of lipreading or audiologists. A variety of hearing losses was represented. There was considerable scatter in age and training of the subjects. Men and women were represented. Details are outlined in Table II.

Lipreaders were tested in groups ranging in size from two to seven. Answer sheets were distributed to aid in the explanation of the task. Lipreaders were told what type of stimuli to expect and the order of the presentation of the types of stimuli. Instructions are detailed in Appendix I.

During testing, lipreaders were assisted in identifying the number of the upcoming item if they seemed to have lost their place in the test. This problem arose because no auditory or visual cue had been recorded to indicate an upcoming stimulus. The experimenter considered the assistance given to be adequate for keeping average lipreaders at the correct place in the test.

Following presentation of all six speakers, lipreaders were asked which speaker they preferred to watch. They were then shown the speaker they preferred in sideview.

All subjects were requested to comment on all aspects of the tests and to give their opinions on lipreading. Some of these comments were elicited by questionnaire. The questionnaire also requested back-

Table II: Lipreaders - Background Information

Subject	Hearing Loss	Type of Training	Years Aided	Sex	Age
1	moderate	15 week course	10	F	36
2	mild-moderate	15 week course	1	F	33
3	normal	phonetics	--	F	31
4	profound	deaf education	8	F	27
5	moderate	preschool program	1.5 wk.	F	23
6	normal	phonetics	--	F	24
7	normal	phonetics	--	F	23
8	normal	phonetics	--	F	24
9	normal	phonetics	--	F	24
10	normal	phonetics	--	F	23
11	normal	phonetics	--	M	41
12	presbycusis	lipreading club	--	F	80
13	presbycusis	lipreading club	10	M	65
14	presbycusis	lipreading club	3	F	72
15	cros aid	lipreading club	6	F	?
16	presbycusis and trauma	lipreading club	8.5	F	66
17	onset at 30	lipreading club	20	F	80
18	presbycusis	lipreading club	13	F	83
19	normal	DID NOT COMPLETE TEST			
20	normal	phonetics	--	F	26
21	normal	phonetics	--	F	22
22	normal	phonetics	--	F	26
23	normal	phonetics	--	F	24
24	normal	phonetics	--	F	27
25	normal	phonetics	--	F	25
26	normal	phonetics	--	F	23
27	mild-moderate	15 week course	--	F	58
28	mild-moderate	15 week course	--	F	?
29	profound	15 week course	27	M	73
30	moderate	15 week course	3	M	66
31	normal	phonetics	--	F	23
32	normal	phonetics	--	F	24
33	normal	phonetics	--	F	27
34	severe-profound	deaf education	9	M	15
35	severe-profound	deaf education	8	M	16
36	severe-profound	deaf education	12	M	16
37	normal	phonetics	--	F	28

1. Subjects 1-21 were administered the complete pilot test.

Subjects 22-37 were administered the screening test.

2. Hearing loss was suggested by the history reported in the questionnaire. In some cases, the information was insufficient to pinpoint the degree of loss. Best indications of the hearing loss are given here.

ground information from the lipreaders.

3.4 Results

Speaker

The clearest objective differences between speakers as measured by errors made by lipreaders were found for sentence stimuli. Scores for whole sentences and for words showed basically similar differences. Little difference between speakers was evidenced on the number or the monosyllable stimuli.

Speaker 5 was chosen to be the speaker in the main test for the following reasons:

1. Of all the speakers, speaker 5 was most accurately speechread on words and syllables, and was second most accurately speechread on sentences and numbers.
2. Scores obtained for speaker 5 were consistently high on all types of stimuli.
3. Speaker 5 was easy to speechread for "good" lipreaders, while she was difficult to speechread for poor lipreaders.
4. Seven of the thirteen lipreaders asked indicated a preference for speaker 5. Preference and best performance were in agreement for less than 2/3 of the thirteen subjects.
5. Performance was most consistent for monosyllables spoken by speaker 5. This was considered to be an important advantage since the stimuli of the main test were to be similar; i.e., disyllables. All ten vowels spoken by speaker 5 were identified at least 33% accurately. Other speakers were speechread with less accuracy. Monosyllables containing

/i,I/, /æ/, /ʌ,u/, which were the vowels selected for the main test, were identified with at least 63% accuracy when they were spoken by speaker 5. Lower scores were obtained when monosyllables spoken by other speakers were lipread. Finally, for speaker 5, at least 70% of the time, an attempt was made by lipreaders to identify each vowel. Items spoken by speaker 1 elicited slightly more attempts; items spoken by other speakers elicited fewer attempts to identify vowels.

7. Speaker 5 seemed to exhibit features of a speaker who would be easy to lipread, as these features were described by lipreaders in the questionnaire. Good features included: facial expression only when appropriate, maintenance of eye contact (looking directly into the camera), slower pace of speech, even teeth, average lip size, good vertical and horizontal lip movement.
8. Measurements indicated that speaker 5 exhibited the greatest range of vertical lip movement and the second greatest range of horizontal lip movement of the six speakers. Speaker 5 exhibited the second greatest range in vowel durations of the six speakers. These measurements agree with speaker features claimed to make lipreading easier. The pattern and range of measurements found for speaker 5 suggest that her articulations may provide clearer visual cues to lipreaders.

Lipreaders

Lipreading performance on the pilot test was observed to differentiate three groups of lipreaders.

Group I consisted of normal hearing and hearing-impaired subjects. This group was characterized by good scores on all types of stimuli. All achieved at least 33% correct sentences; 40% correct words, 47% correct syllables, and 75% correct number scores. Their scores for /i,I/, /æ/, /ʊ, u/ were at least 58% correct. These subjects were chosen to participate in the main test.

Group II consisted of normal hearing subjects with phonetic training. They performed as well as Group I on numbers and syllables. Scores obtained for words and sentences, however, fell noticeably below those of Group I. This pattern is in agreement with reports of good performance by normal hearing subjects on monosyllables but not on longer units of language (Berger, 1972, pg. 125).

Group III consisted of elderly hearing-impaired lipreaders. Scores obtained by this group were below the levels of accuracy of Group I on all types of stimuli. In some cases, sentence and word scores for Group III lipreaders were comparable to those obtained by Group II lipreaders. Digit and syllable scores were worse than those of Group II lipreaders.

The three groups will be discussed further below. The distinctions separating the groups were still evident if a percentage of correct answers divided by attempted answers were considered rather than raw scores. This will also receive further comment below.

Stimuli

Vowels which could be easily distinguished from each other were

selected to be part of the main test stimuli. Three groups of vowels were determined by considering perceptual confusions (see Figure 2). These were /i, I, e, &/, /æ, ʌ, ɒ/, and /o, ʊ, u/. Although confusions within these groups were common, confusions between groups were not common. Representatives of each of the three groups were chosen by considering measurements of articulatory parameters, especially for speaker 5. Greatest vowel durations were measured for /i, e, æ, ɔ, u/. Measurements of vertical and horizontal lip separation during the production of /u, æ, i/ suggested that these vowels would be most easily distinguished by lipreaders. The smallest height and width of lip separation were measured for /u/. The greatest height and width of lip separation were measured for /æ/. Of the front, unrounded vowels characterized by large width measurements, /i/ was measured to have the smallest height. The findings agree with those of Fromkin, (1964).

These choices maximize contrast in terms of the features of rounding and tongue height which have been found to be the most important features in vowel identification in lipreading (Berger, 1972).

Conditions of Filming and Administration of the Test

A front rather than a sideview of the speaker was considered to be better for the following reasons:

1. Scores obtained on monosyllables spoken by speaker 5 at a front view equalled or exceeded scores obtained at a sideview for twelve of the sixteen subjects tested on both views.
2. Scores obtained for the front view showed less variation across lipreaders than did scores obtained for the sideview.
3. Most lipreaders claimed that the front view was easier to lipread.

Figure 2. Confusion Matrix Showing Vowel
Confusions of Twenty Lipreaders
(X 6 speakers X 2 vowel presentations)

STIMULUS	RESPONSE			
	i I e ε	æ ʌ ɒ	o ʊ u	No Response
i I e ε	569/960 59.3%	133/960 13.9%	18/960 1.9%	240/960 25.0%
æ ʌ ɒ	88/720 12.%	486/720 67.5%	6/720 .8%	140/720 19.4%
o ʊ u	27/720 3.8%	27/720 3.8%	536/720 74.4%	130/720 18.1%

Matrix entries show the number of responses given in four response categories (/i,I,e,ε/, /æ,ʌ,ɒ/, /o,ʊ,u/, no response) divided by the number of items in each stimulus category (/i,I,e,ε/, /æ,ʌ,ɒ/, /o,ʊ,u/) for which the responses were given. Percentage scores are also expressed.

Some subjects commented that the counter was distracting. The inter-stimulus interval seemed comfortable to the lipreaders who were to participate in the main study. Comments by lipreaders suggested that the visibility of the tongue and teeth of the speaker made lipreading easier. Many lipreaders were observed to miss some of the presentations because no cue preceded the presentation of each test item.

Variation in the contrast obtainable on the four monitors used for the pilot and screening tests, variation in the monitor size, and variation in distortion of the image were noted. Although it is difficult to evaluate the contribution of these variations, they were eliminated in the main study.

The smallest group performed well on the pilot test, whereas the largest group performed poorly. Variables other than grouping seemed to contribute to these differences. Although it is difficult to evaluate the contribution of group size to the results, it was standardized for the main test.

Comparison of scores obtained for the ten monosyllables presented before the sentences to the scores obtained for the ten monosyllables presented after the sentences by each speaker provided no evidence of improvement in score from the first to second presentation. Comparison of scores obtained for stimuli spoken by each speaker in the three dubbing orders provided no evidence that the order of the speakers on the tape influenced lipreading scores.

3.5 Comments

Stimuli and Lipreading Abilities

Three types of stimuli were used in this pilot study: mono-

syllables, numbers and sentences. Associated with each of the types of stimuli are various levels of redundancy. Members of lipreader Groups I, II, and III were determined by considering performance on the types of stimuli. The three Groups may have been using redundancy cues in different ways. Group I consisted of young adults with training in either lipreading or phonetics. They could take advantage of most available cues and scored well on all types of stimuli. Group II were young adults who had all had phonetic training. They were able to use cues for short units like syllables and numbers, but were less able to use cues for the longer, linguistically more complex sentence stimuli. Group III consisted of elderly hearing-impaired persons who had had less formal training. Group III performed as well as Group II on sentences, but not on numbers or syllables. Daily use of lipreading may have helped them in the use of redundancy cues for longer units but not for the less natural, short units.

Age as a Lipreader Variable

Lipreading ability has been claimed to decrease with age beginning at the fourth decade (Farrimond, 1959; Goetzinger, 1967; Coscarelli, 1968 cited in Berger, 1972). It was the feeling of the experimenter that the task required in this study was very difficult for the older subjects. They were often slow in writing responses because of poor fine motor abilities. As a result, many stimuli were not seen and fewer responses were attempted. There were also more misunderstandings regarding procedures involved in the test for Group III. Fatigue and disagreement were expressed only by lipreaders in this group. When questioned about their daily use of various types of communication,

many of these lipreaders replied that they only communicated at lip-reading class because they lived alone. The importance of these considerations in evaluating the test situation and daily use of lipreading should be investigated more directly than has been reported by previous studies.

Strategies

A synthesizing strategy was exhibited by some subjects who were observed to write a few of the words of a sentence immediately and to later complete the sentence by filling in the remaining words. High scores on all types of stimuli were obtained by these lipreaders.

A subvocalizing strategy was exhibited by some subjects who vocalized or subvocalized before writing. These lipreaders commented that doing this made lipreading easier. The subjects who used this strategy had had good phonetic training. They scored very well on all types of stimuli.

CHAPTER FOUR

MAIN STUDY

4.1 Introduction

Many of the questions motivating studies of lipreading which have been raised in the literature have not been resolved.

There have been efforts to establish categories of visually distinguishable phonemes by using methods of descriptive linguistics. Minimal units of visual contrast which can be used to differentiate meaning have been hypothesized. These units are sometimes referred to as "visemes" (Fisher, 1968). No agreement has been reached regarding the exact nature of visemes.

The pertinent visual cues used to identify visemes remain in question. Jeffers and Barley(1971) have proposed that certain articulatory gestures are likely to be movements identified as cues by a speechreader. Few studies (Jackson et al.,1976; Erber et al., 1979) have been undertaken to confirm proposed speechreading movements.

Vision cannot be used to obtain all the phonological or phonetic information which can be obtained through audition. The amount of such information that can be conveyed through vision is in dispute.

Studies reported in the literature (Fisher, 1968; Berger, 1972) bring into question the reliability of speechreading movements for the identification of visemes. Variability in speechreading movements acting as visual cues for viseme identification must be interpreted by the speechreader. What is the nature of this variability? How is the accuracy of the speechreader in identifying visemes influenced by such variability?

4.2 Aims

This study attempts to provide more information about the identification of visemes. Coarticulation is examined as a source of possible variability in the speechreading movements acting as visual cues. The influence of such variability on the accuracy of the speech-reader in identifying visemes is considered. Eighty-one utterances of the form V_1CV_2 were recorded on videotape. Coarticulatory effects were expected to occur so that each of the three component phonemes of an utterance could alter the production of the other phonemes to some extent. The stimuli were presented visually to subjects who were chosen on the basis of their good performance on the pilot test or the screening test. Specifically, the aims of the experiment were the following:

1. To evaluate the visual perception of phonemes presented in varying phonological environments to good lipreaders who were asked to identify the utterances which were presented to them visually.
2. To determine how much information was transmitted visually in terms of articulatory features.
3. To determine if there was apparent variability in the perception of visemes or in the transmission of information by articulatory features when phonemes were presented visually in different phonological environments.
4. If variability in the perception of visemes as a function of context was evidenced, to determine if this could be explained by variability or reliability of speechreading movements that could be physically measured from the videotape providing the visual signal.
5. To consider how a lipreader may be helped or hindered by visual coarticulatory effects.

4.3 Preparation of the Main Test

Corpus

Stimuli consisted of eighty-one ($3 \times 9 \times 3$) V_1CV_2 disyllables where V_1 and V_2 were selected from /i,æ,u/, and where C was selected from /p,t,k,tʃ,f,θ,s,ʃ,w/.

Consonants were chosen to represent a wide variety of places of articulation. Manners of articulation represented were stop, fricative, and glide. Voicing and nasality were not expected to be perceptible visually. Rounding is a feature of /w/ and usually also of /tʃ,ʃ/. /tʃ/ was included to increase the occurrence of rounded consonants in the test and to provide some information about the perception of articulation within a phoneme since /tʃ/ can be described as $t + ʃ$.

Consonants were of main interest since consonant viseme confusions have been studied previously. Vowels were selected to be easy to lipread, while performance on consonants was designed to be more difficult. The three vowels were selected based on the results of the pilot study.

The form of the disyllable was chosen as V_1CV_2 in order to provide a good opportunity to observe a variety of coarticulatory effects. Existing lipreading literature documents confusions of consonants in a single vowel environment such as before /a/ (Binnie et al., 1974) or in the environment $V_1_V_2$ where V_1 is specified (Erber 1971). Figure 1 outlines other examples of simple environments. In running speech, vowel environments would not be predictable as they have been in previous studies. Phonetic studies of coarticulation, regardless of the particular theoretical model suggested, all indicate that coarticulatory effects can

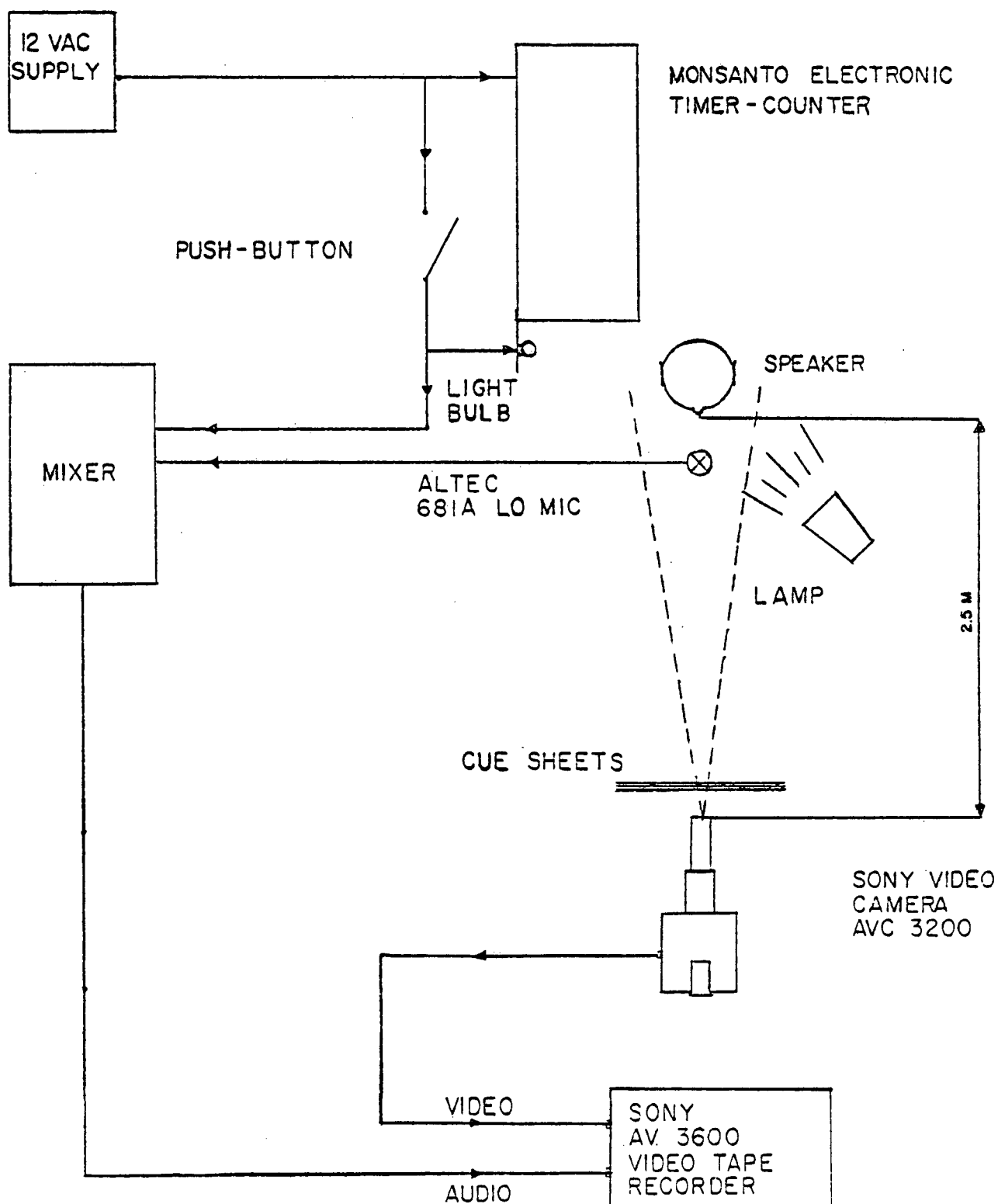
extend beyond one adjacent phoneme (Ohman, 1966; Kozhevnikov and Chistovich, 1965; Daniloff and Moll, 1968; Benguerel and Cowan, 1974). It was decided to include three phonemes so that there would be one phoneme on each side of the consonant. The findings of Ohman (1966) that coarticulatory effects are evidenced by altering preceding or following vowels in a V_1CV_2 contradicted conclusions of others (Delattre et al., 1955) that had been based on consideration of CV syllables. It is also possible that the use of V_1CV_2 's and CV's in a visual investigation could lead to divergent observations.

Ordering of the test items on the recording was as follows: six practice items, two repetitions of the pseudo-randomized sequence of eighty-one V_1CV_2 's, a rest break, a repetition of the six practice items and another two repetitions of the sequence of eighty-one V_1CV_2 's. The sequence of eighty-one V_1CV_2 's consisted of nine subtests. Each subtest contained one of each consonant. The ordering of the V_1CV_2 's was such that adjacent items never contained the same initial vowel, consonant or final vowel; for example, /æti/ could not follow or precede /æku/ because the two initial vowels are the same. Three versions of the test, one slow, one fast, one reverse, were prepared. For the fast version, the speaker was instructed to speak more quickly. For the reverse version, each item of the fast tape was recorded in reverse. Although the item was reversed so that V_1CV_2 became V_2CV_1 , the order of the eighty-one items was maintained.

Instrumentation

The equipment setup for the recording session is displayed in Figure 3.

FIGURE 3: EQUIPMENT SET - UP FOR
RECORDING SESSION



As well as recording the speaker, visual and auditory time markers were recorded onto the videotape to allow the precise identification of visual fields and to allow matching of any particular visual field to acoustic events recorded at the same moment in time. The visual markers were produced by the light bulb and the electronic counter display, the auditory markers by a 60 Hz buzz. When the speaker pressed the button, the 60 Hz buzz was activated and the light bulb was illuminated. The counter was set to increment at 1/60 second intervals throughout the entire recording session. The increments were synchronized to the 60 Hz buzz and to the change in visual fields on the videotape (there are 60 fields/second and 2 fields/frame). A grid (No. 340.1 Dietzgen Isometric Graph Paper) was filmed on the videotape so that lip measurements taken from the monitor could be scaled and compared to actual lip dimensions.

The main test differed from the pilot in that a desk lamp was used to provide additional lighting on the speakers face in an attempt to improve visibility of the articulators. A head and shoulders front view of the selected speaker was used.

Recording Session

Before filming, a practice session took place in which the speaker was familiarized with the stimuli and procedures. Each item was cued by the clicking of a stop watch. Items were timed so that all stimuli (except sentences in the pilot) were separated by ten seconds. The speaker pressed the button (described above) when she was ready to say the next item. Upon hearing the click, she released the button, spoke the item, and again pressed the button. The experimenter was

responsible for timing and for indicating to the speaker which item was next by pointing to the cue sheets. The speaker was told to look at the camera as much as possible. No particular instructions were given regarding mouth position at rest. It was recommended to the speaker that she speak in a slow, normal way.

Editing

Editing and assembly of the test tapes was performed using a Sony 8650 videotape recorder. Most errors made during recording were corrected immediately by re-recording. Some editing was done later. Most of the editing work involved assembly of the recordings of the speaker(s) in the test order.

Editing procedures used for the main test included the following additions to the procedures used for the pilot test. In order to provide a cue to the lipreader that the next item was about to be presented, the click of the stop watch was retained on the sound track of the tape. During editing, the click was copied whereas the sound of the utterance was not copied. In copying the fast version in reverse, the sound track was lost. New clicks were recorded to serve as cues on the reverse tape. The reverse tape itself was made by recording the V_1CV_2 's, thirty seconds at a time, onto a playback loop. The loop was then recorded in reverse at normal speed. This procedure was carried out by CBC technicians using CBC equipment.

4.4 Administration of the Main Test to Lipreaders

Instrumentation

The test was administered to all subjects in one location. The

location was arranged for testing by placing the monitor in front of a blank wall. A table was situated so that the lipreader would be sitting at a viewing distance of about two meters from the monitor. Movable partitions were placed along side the table to provide a more enclosed, less distracting surrounding. A cardboard frame was placed in front of the monitor so that the counter and light were blocked from view. Only the speaker's face was visible. The videotape recorder and monitor dials were also screened from view. An 18" black and white monitor was used.

Two identical sets of answer sheets were provided--one for each half of the test. The practice items were numbered as if they were test stimuli. Each page of the answer sheet was numbered for sixty items so that repetition of the eighty-one items would not be evident. Although the final item of each half of the test was number 168 (6×28), the third page was numbered to 180. The subject was not aware of the number of the final item before the test stopped. The numbering of the items was considered to conceal the pattern of the test and to provide for sampling of the lipreaders' best performances.

Subjects

Lipreaders were selected from those tested with the pilot test or screening test. All lipreaders were administered the fast version of the test. This version was expected to be more difficult than the slow version. Some of the normal-hearing subjects were also administered the slow and reverse versions. One normal-hearing subject was administered the fast version twice so that practice effect could be checked. Lipreaders were not aware of the nature of the difference between the fast and slow versions of the test. For the reverse test, three of the

six subjects were aware that the items were reversed.

Testing Session

Lipreaders were tested individually. During the instructions, the experimenter wrote the set of possible vowels and consonants on a card. The structure of the disyllables was explained. The card remained on the table in view of the lipreader for the duration of the test (see Appendix II for instructions).

A practice session was conducted before the test. First, the lipreader was asked to write sample V_1CV_2 's spoken by the experimenter. Next, the first six items spoken by the recorded speaker were shown on the monitor. The lipreader wrote answers on a spare sheet of paper during the practice session. Transcription problems were resolved. The click was adjusted until it was easily heard by the lipreader. The purpose of the click was explained to the lipreader. For subject four, who was profoundly deaf, the experimenter gave a hand tap on the table when the click occurred. The tape was restarted after the practice session. The first six responses on the answer sheets were the practice items.

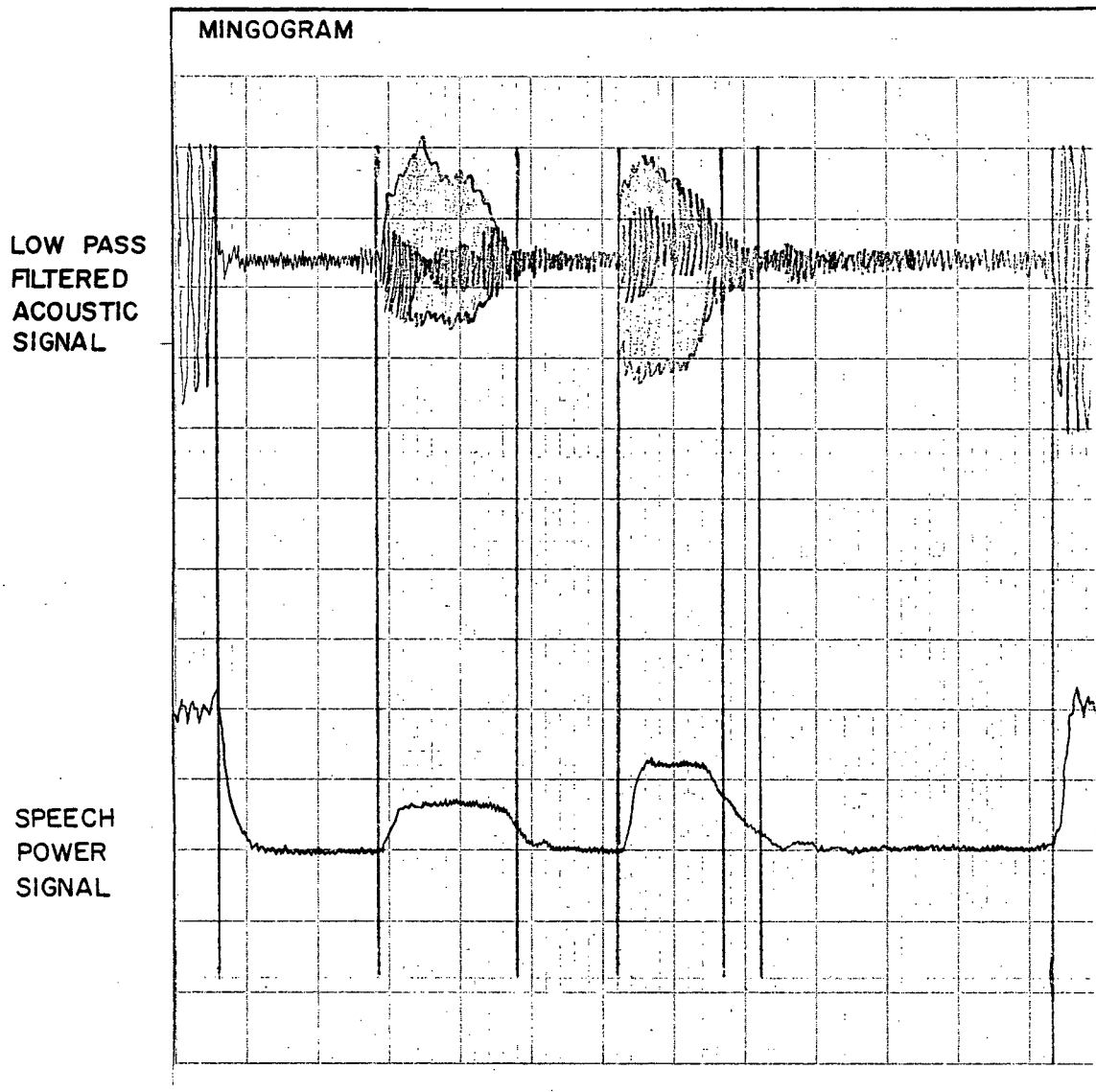
Comments were elicited from the lipreader during the break and following the test. The break was usually about five minutes long, depending on the wishes of the lipreader.

Subjects who were administered the slow version returned for a second session. The interval between sessions ranged from several hours to several days. Subjects who were administered the reverse test returned for a third or second session about two months after the initial sessions.

4.5 Measurements of Articulatory Gestures

Mingograms showing the low-pass filtered signal (including the 60 Hz buzz), and the speech power signal were made for each stimulus using a Siemens Oscillomink graphic recorder (see Figure 4). Durations of the vowels and consonants were measured. There was some difficulty in determining the termination of the vowels, especially the final vowel, so two duration measurements were taken for V_2 . The time scale on the mingograms was 10 cm/sec. Measurement error was on the order of .5 mm. or .005 seconds. In relation to the fields of the videotape, there were 60 fields/second or six fields/centimeter on the mingograms. The location of the points measured on the videotape were identified on the mingogram by using the markers described in Section 4.3, Instrumentation. Measurements of vertical and horizontal lip opening seen on the videotape were taken near the middle of all vowels and consonants of the fast tape. A selection of items of the slow tape were measured in the same way. Measurements of vertical and horizontal lip opening were taken at five field intervals for selected items of the fast and slow tapes. The filmed grid was copied onto a transparency which was used in measuring lip opening. The grid provided some correction for monitor distortion and allowed scaling to actual lip dimensions. Measurement error for lip opening was in the order of .25 grid unit or .83 mm. after scaling to actual lip dimensions was calculated. One grid unit represented $2/3$ cm.

FIGURE 4 : DISPLAY OF MARKERS AND SIGNAL INFORMATION USED TO MATCH VISUAL EVENTS, ACOUSTIC EVENTS, AND MEASUREMENTS OF ARTICULATIONS



CM	0	1	2	3	4	5	6	7	8	9	10	11	12
SECONDS	0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0	1.1	1.2
FIELDS	0		15		30		45		60				

COUNTER	862			874			883			915		
MESSAGE	BUZZ	SILENCE	∞	f	u	a	b					BUZZ
LIGHT	ON	OFF										ON
BUTTON	DOWN	RELEASED										DOWN
CLICK	↑											

VERTICAL LIP OPENING (mm)	18.8	4.95	4.95
HORIZONTAL LIP OPENING (mm)	42.9	23.1	14.85

a. termination of V_2 short
 b. termination of V_2 long

CHAPTER FIVE

RESULTS OF THE MAIN TEST

5.1 Overview of Viseme Identification

Subject Differences

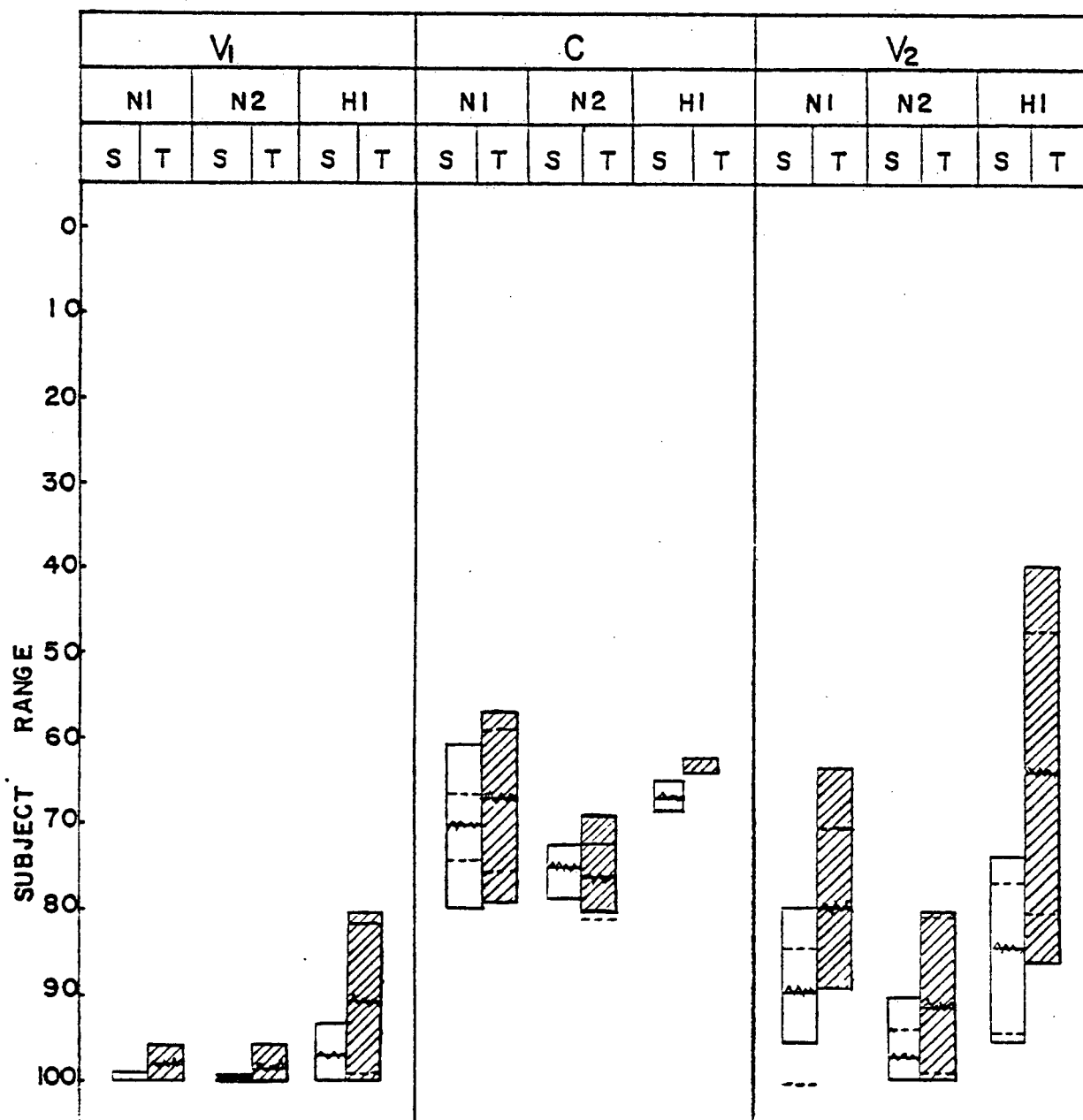
All lipreaders chosen to participate in the main test, including hearing-impaired and normal hearing persons, had performed well on the pilot or on the screening test. Main test results did, however, reveal trends for groups of these lipreaders. Data obtained for normal hearing lipreaders on the first administration of the main test (Group N1), for normal hearing lipreaders on the second administration of the main test (Group N2), and for the hearing-impaired lipreaders on the first administration of the main test (Group HI), were considered. (see Figure 5)

Raw scores and transmission scores were statistically analysed. Means obtained for members of Groups N1 and N2 were compared by using t-tests of paired differences. Group means for all three groups were also compared. Means for any two groups were considered to be significantly different if one mean was greater than two standard deviations away from the other mean. Significant differences in variances of scores for the groups were determined by using F tests where two groups were compared in each F test.

The following trends were observed:

1. Normal hearing lipreaders performed better on the second administration of the test than on the first administration. Improved performance was noted on mean raw scores obtained for consonants by all but one subject, and for final vowels (V_2) by all subjects.

FIGURE 5: GROUP TRENDS FOR LIPREADERS REVEALED BY RESULTS OF MAIN TEST



S = RAW SCORE (% CORRECT)

T = TRANSMISSION SCORE (% CORRECT)

--- MEAN

-- INDICATES VALUE ONE STANDARD DEVIATION AWAY FROM THE MEAN (I.S.D.)

2. Improvement in transmission of information (see Appendix III) from the first to the second administration of the test was more pronounced than was the improvement evidenced by mean raw scores. Transmission of information for consonants and final vowels improved significantly ($p = .1$ and $p = .02$ respectively) for Group N2 compared to Group N1. Improvement was noted for all subjects.
3. The variance of raw scores for V_2 obtained by the normal hearing subjects decreased significantly ($p = .02$) from the first to the second test administration.

These three findings provide evidence that normal hearing subjects were able to increase lipreading accuracy from the first to the second test administration.

4. Hearing-impaired subjects performed generally less well than did normal hearing subjects. The mean raw scores obtained by Group HI on V_1 , C, and V_2 were not significantly different from those obtained by Group N1. Mean raw scores on C and V_2 obtained by Group HI were significantly (greater than two standard deviations) less than those obtained by Group N2.
5. Transmission scores for C and V_2 were significantly (greater than two standard deviations) better for Group N2 than for Group HI.
6. Hearing-impaired subjects evidenced significantly ($p = .02$) less variance in raw scores and transmission scores for C than did Group N1. Group HI evidenced significantly ($p = .02$) greater variance in raw scores and transmission scores than did

Groups N1 or N2 for V_1 , and than Group N2 did for V_2 .

On the same test, normal hearing and hearing-impaired subjects differed in terms of variance of raw scores and transmission scores, but not in terms of mean raw scores or mean transmission scores. Compared to normal hearing subjects taking the test a second time, hearing-impaired subjects tested only once scored significantly less in terms of mean raw scores as well as exhibiting significantly different variances of raw scores and transmission scores for parts of the stimuli.

These results suggest that either learning and/or differences between the first administration (fast spoken stimuli taped) and the second administration (slow spoken stimuli taped) could contribute significantly to lipreading accuracy. Further examination of these possible factors will be undertaken (see Sections 5.3 and 5.5).

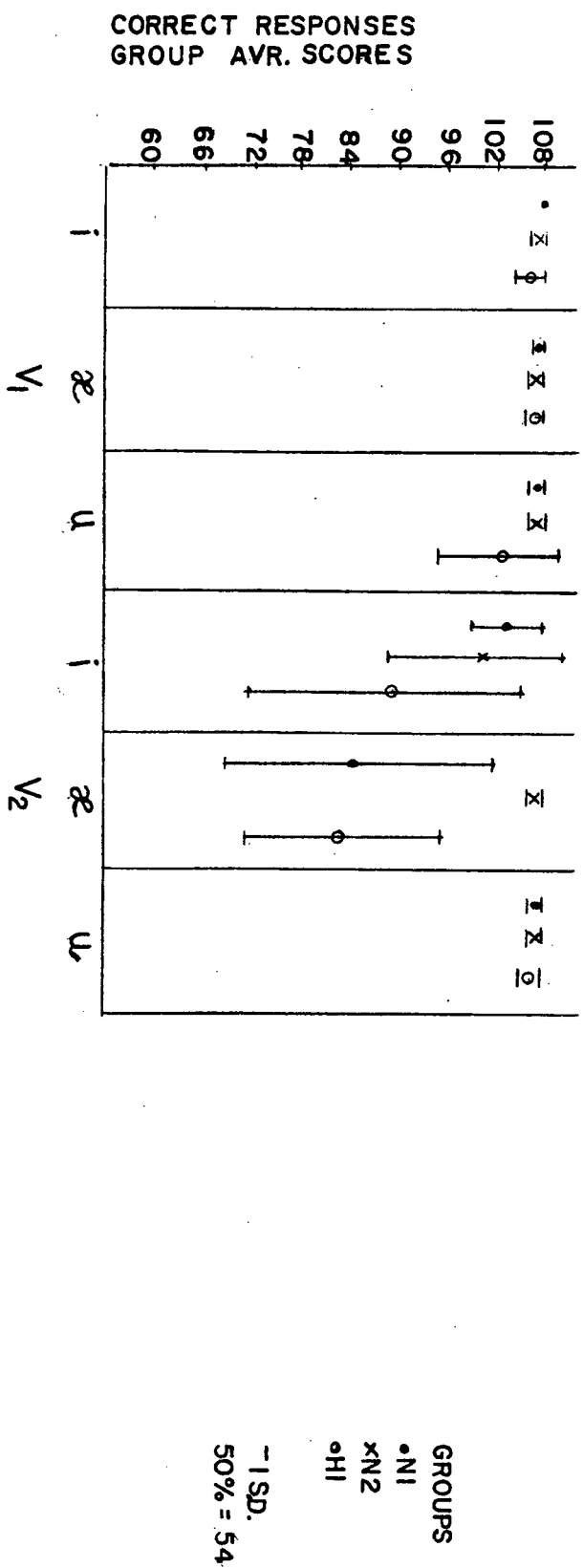
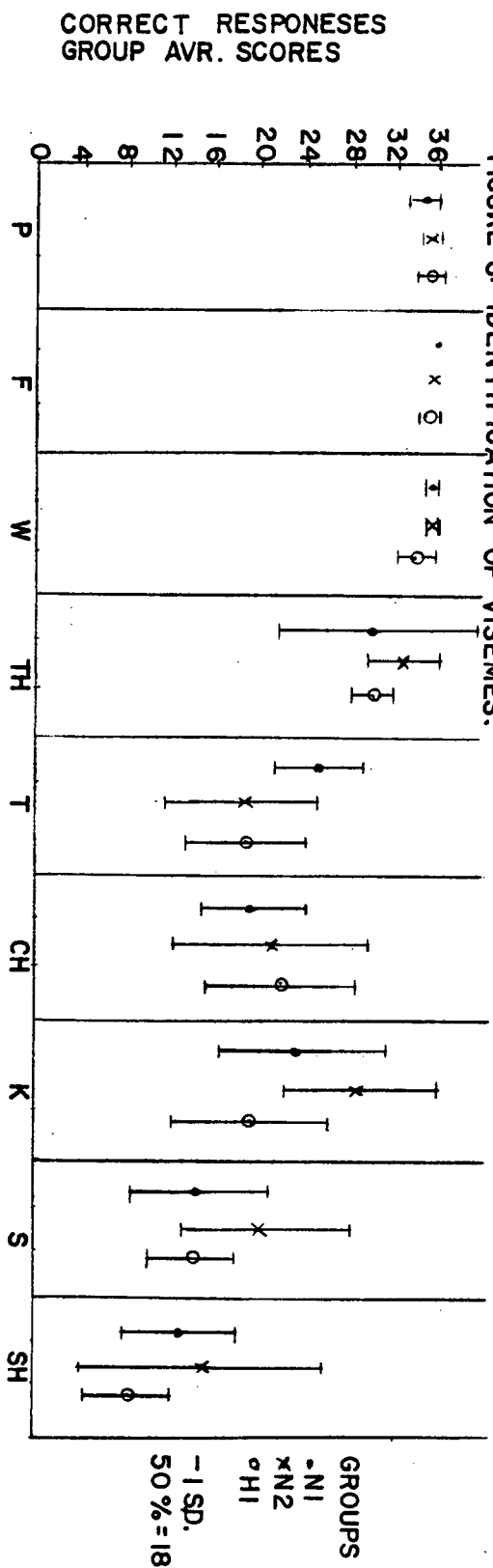
Identification of Visemes

Raw scores obtained for each consonant by each of the three groups of subjects were above chance, with one exception (/ʃ/ for Group HI). Labial articulations /p,f,w/ were lipread nearly perfectly. Very good scores were achieved for /θ/. Other consonants /t, tʃ, s, k/ were less accurately lipread, with scores ranging around 50% (see Figure 6).

Raw scores obtained for V_1 were all nearly perfect. The hearing-impaired group made more errors than Groups N1 or N2, half of the group making errors on initial /u/.

In final position, /u/ was nearly perfectly lipread by all groups. Raw scores obtained for V_2 were lower for /i,æ/ than for /u/. Considerable improvement was noted for /æ/ when Group N2

FIGURE 6: IDENTIFICATION OF VISEMES.



was compared to Group N1.

Articulatory Features

In order to evaluate the nature of overall errors, transmission of information by articulatory features was evaluated. Each consonant can be uniquely specified by the following selection of gross articulatory features:

	p	t	k	tʃ	θ	f	s	ʃ	w
labial	+	-	-	-	-	+	-	-	+
round	-	-	-	+	-	-	-	+	+
alveolar or palatal	-	+	-	+	-	-	+	+	-
continuant	-	-	-	-	+	+	+	+	+

For consonants, information derived from articulations that were labial, or alveolar or palatal, or rounded was transmitted relatively well (see Figure 7). Place of articulation and rounding seemed to be well cued to the lipreader. Manner of articulation, involving judgement of stop-vs.-continuant articulation, was less well transmitted. Improvements achieved by normal hearing lipreaders from the first to second administration of the test appeared to be contributed by improvements in the interpretation of labial, alveolar or palatal, and rounding cues; whereas, there was no improvement in the interpretation of manner of articulation cues.

For vowels, the feature rounded was better transmitted than was the feature height. Considerable improvement in the use of height information was seen in lipreading by Group N2 relative to Group N1.

Transmission of information by articulatory features for vowels and consonants was less efficient for hearing-impaired subjects than for normal hearing subjects, however, variance of feature transmission was often less for hearing-impaired subjects.

Errors

Closer inspection of the errors provides examples of the confusions resulting from poor transmission of continuant information. The four most frequent errors listed in Table III represent this type of error. The next most frequent type of error results from a confusion of an alveolar with a consonant of a more back place of articulation.

Table III: Frequency of Common Consonant Confusion Errors

	Stimulus Presented	Response	Number of Errors	Percentage of Total Responses For Stimulus	Number of 20 S's Making Errors
1.	ʃ	tʃ	272	38%	19
2.	tʃ	ʃ	201	28%	18
3.	s	t	179	25%	19
4.	t	ʃ	166	23%	18
5.	k	t	119	17%	18

There were 38 other types of confusions of consonant pairs, none of which represented more than 10% of the responses to a particular consonant presented. These 38 types of confusions can be described in terms of feature(s) in error. (See Table IV)

Table IV: Frequency of Common Feature-Related Errors

Feature of Stimuli	error involving one feature		error possibly involving more than one feature	
	number of errors	% of responses for stimuli	number of errors	% of responses for stimuli
labial	32	4%	56	8%
round	126	18%	239	33%
alveolar or palatal	16	2%	211	29%
continuant	77	11%	332	46%

While the feature continuant contributed second most of any single feature error type to the 38 additional confusions, it was involved in the greatest number of confusions where the confused consonants differed by more than one feature. Labial articulations were seldom in error.

Voicing

All consonant stimuli, except /w/, were voiceless, however, voiced consonants were included in the response set available to lipreaders in order to increase the difficulty of the experimental task. The voiced alternative was chosen within the limits of chance in 32 of 40 cases by Group N1 (8 C's x 5 S's), in 35 of 40 cases by Group N2, and in 17 of 25 cases for Group HI (5 C's x 5 S's). After the test, many subjects commented that they had not been able to differentiate voiced from voiceless consonants, although both types of answers had been guessed. Some subjects exhibited

trends to prefer voiced or voiceless responses. For the 8 of 25 cases for Group III in which the voiced alternative was not chosen within the limits of chance, the voiced alternative was chosen infrequently. Perhaps these subjects simply ignored the voiced alternative. /ʒ/ was identified as voiceless more often than as voiced by 3 of 5 members of Group N1 and by 4 of 5 members of Group N2. (/ʒ/ was not an alternative for Group HI). The relatively infrequent occurrence of /ʒ/ in English may account for this preference. The number of subjects identifying particular consonants as voiced or voiceless showed no consistent response trends that could otherwise be related to the stimuli. Differences seemed more likely to have resulted from individual response strategies.

5.2 Context Effects on Viseme Identification

One of the main purposes of this study was to examine the pattern of errors in the perception of visemes occurring in varying phonological environments.

Few errors were made on initial vowels. The perception of the initial vowel (V_1) was nearly perfect in all contexts. Hearing-impaired subjects did make more errors on /u/ when it preceded /w/ (see Figure 8).

Consonant errors were found to occur in about equal numbers after each of the three initial vowels. For Groups N1 and HI, /u/ as an initial vowel was associated with more errors on following C's than were /æ/ and /i/. Group N2 showed improvement on lipreading of C's following /u/ such that performance

FIGURE 8: CONTEXT EFFECT ON INITIAL VOWEL ERRORS

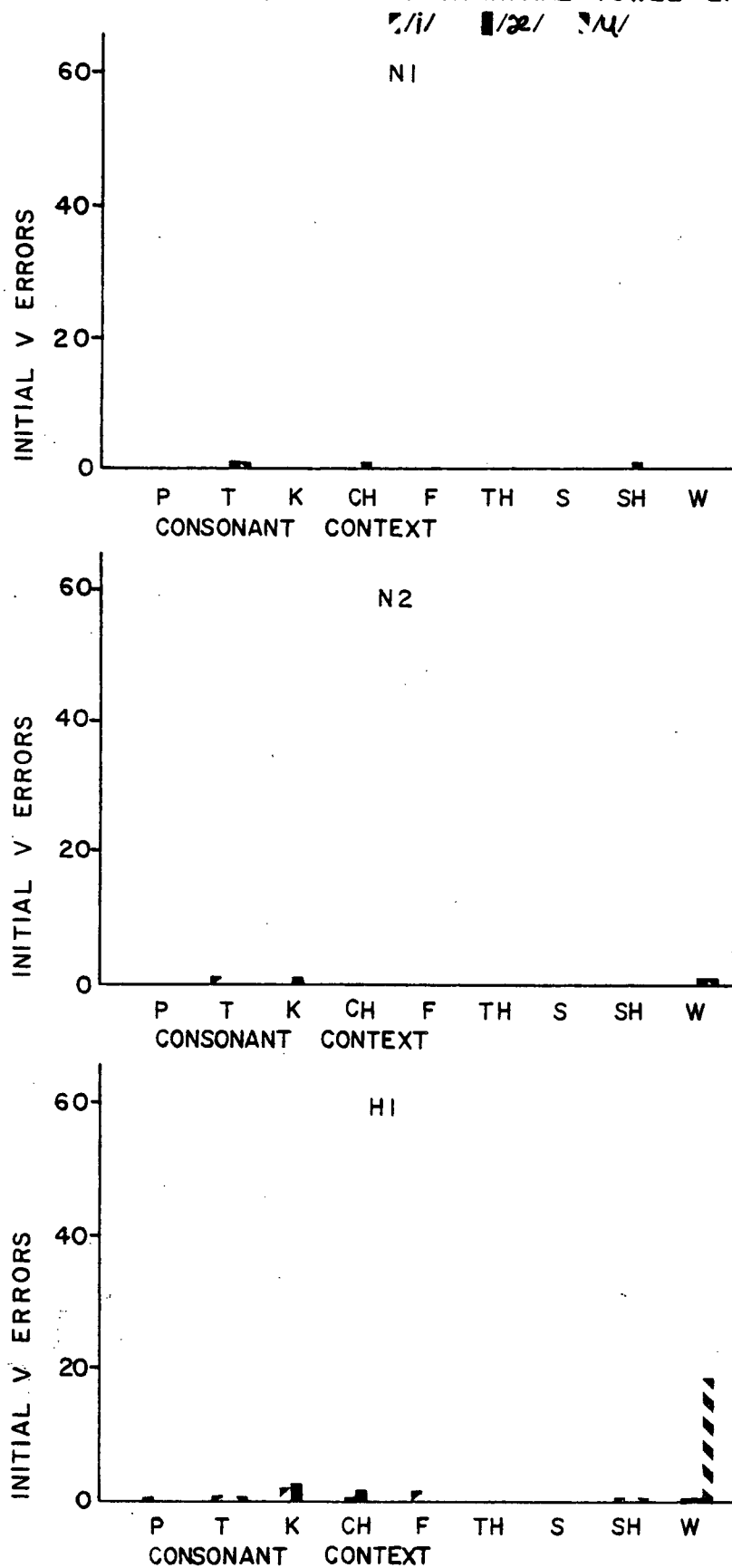


FIGURE 9: CONSONANT ERRORS IN INITIAL VOWEL CONTEXT

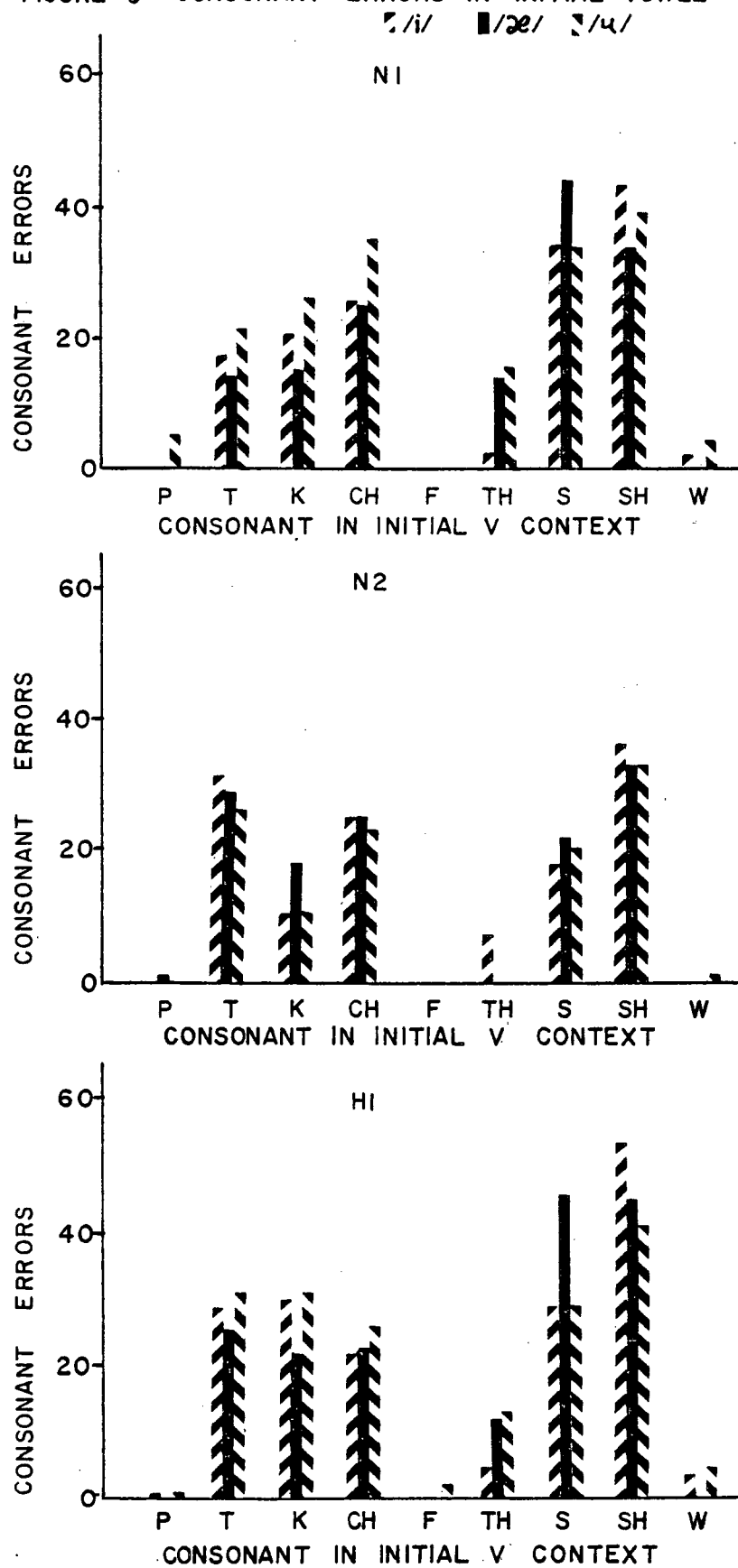


FIGURE 10: CONSONANT ERRORS IN FINAL VOWEL CONTEXT

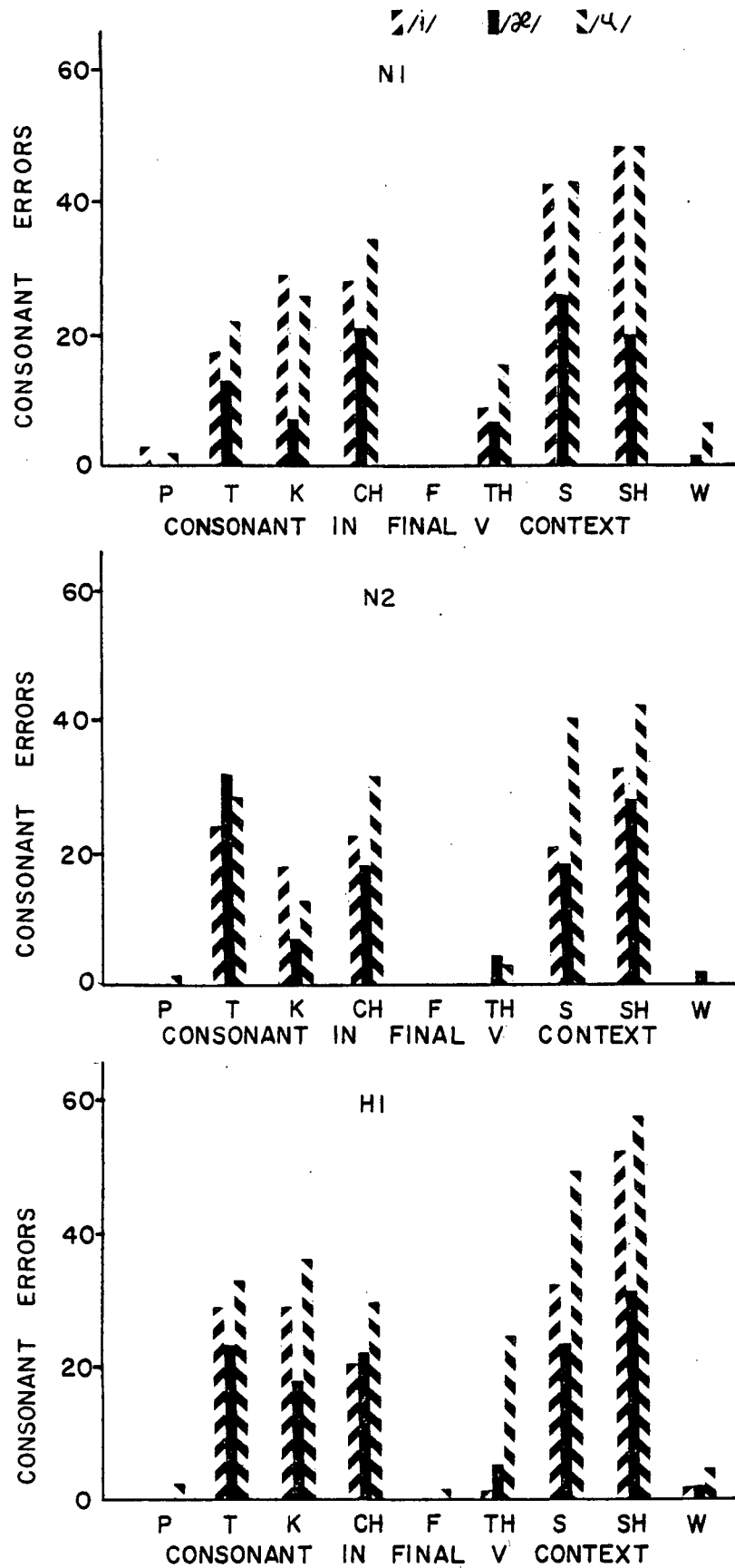
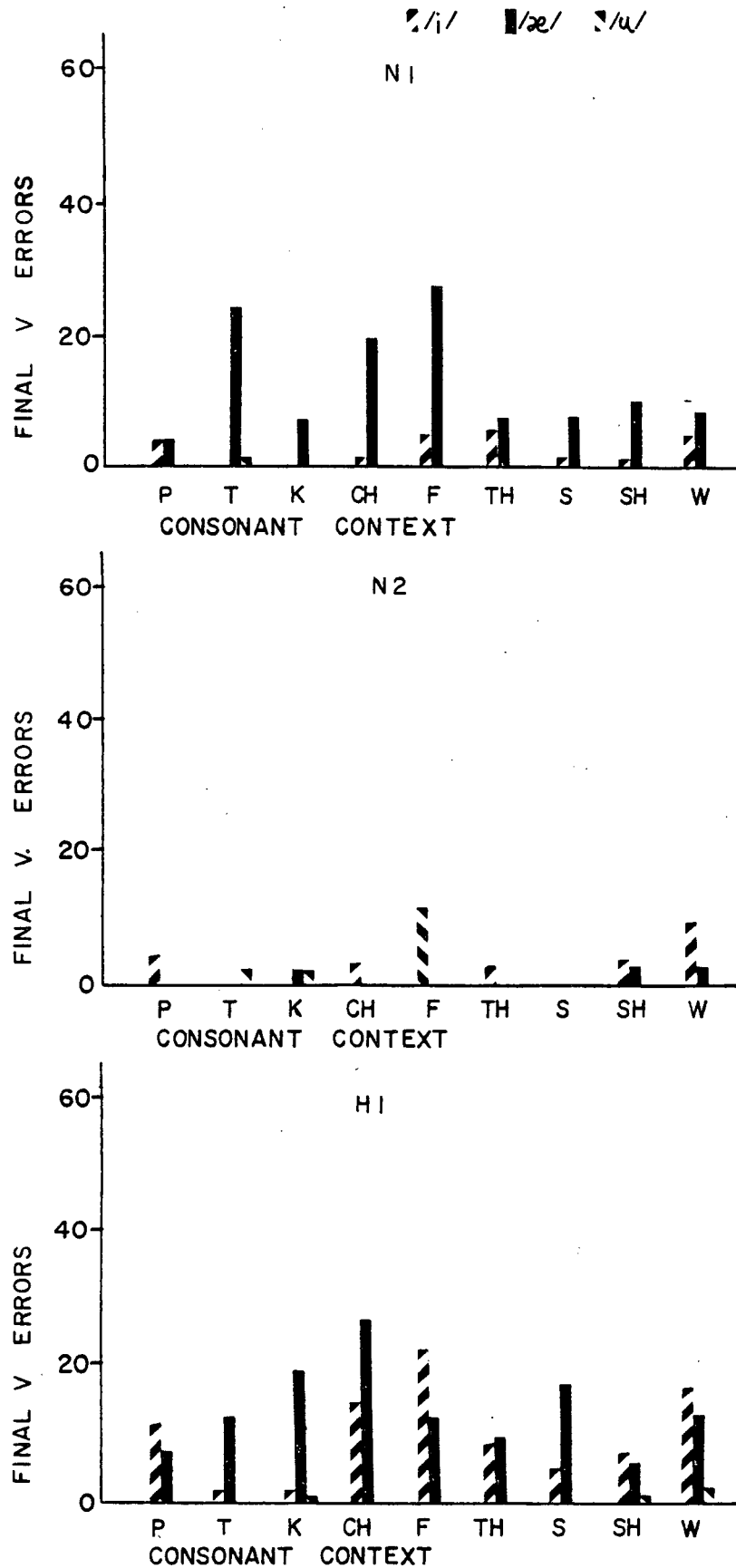


FIGURE II: CONTEXT EFFECTS ON FINAL VOWEL ERRORS



was about the same for consonants following the three initial vowels (see Figure 9).

Coarticulatory effects on lipreading of consonants is suggested when the final vowel (V_2) environment is considered. For all groups, /i/ and /u/ tended to be associated with more errors on the preceding C than was /æ/. For hearing-impaired subjects, /u/ was always associated with more consonant errors than was /i/. For Group N1, /u/ and /i/ were associated with about equal numbers of consonant errors. The greatest reduction in consonant errors associated with a following vowel for Group N2 relative to Group N1 occurred for final /i/. Difficulty of consonant perception appears to be ordered from greatest to least for the environment preceding /u/, then /i/, then /æ/ (see Figure 10).

More errors were made on final vowels than were made on initial vowels. Difficulty of perception of final vowels by Group N1 was always greatest for /æ/, then /i/, while only one error was made on final /u/. Group HI exhibited a similar distribution of final vowel errors. Group N2 continued to make errors on /i/, whereas few errors were made on final /æ/ or /u/.

Coarticulatory effects are suggested when the consonant environment preceding the final vowel is considered. Errors on /i/ following /p,f,w,θ/ definitely exceeded the number of errors occurring after other consonants for all groups. Errors on /æ/ for Group N1 occurred more frequently after /t, tʃ, ʃ, f/. For Group HI, errors on /æ/ occurred more frequently after /tʃ, k, s/ and, to a lesser extent, after /t, f, w/ (see Figure 11).

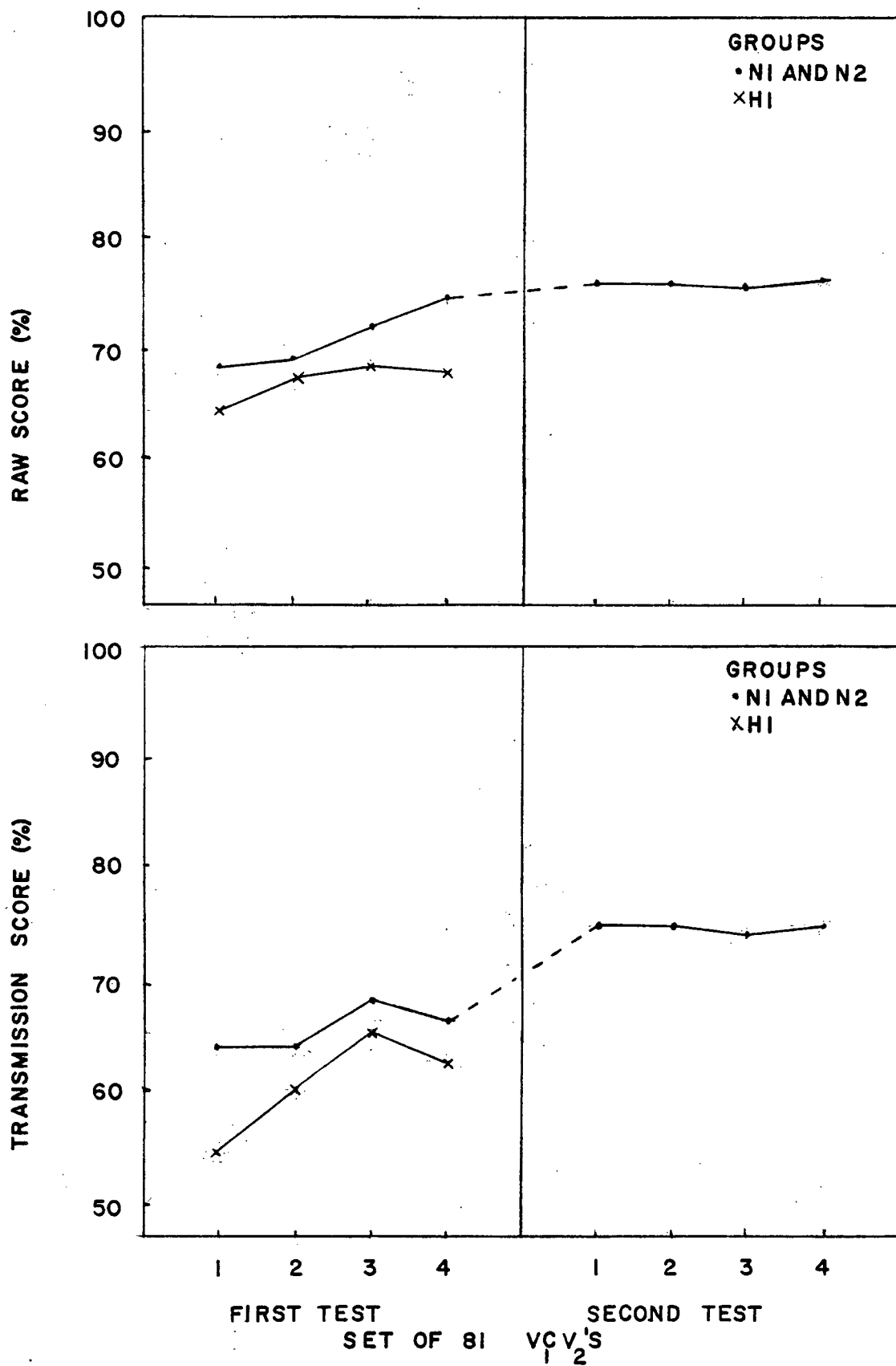
Coarticulatory effects have been claimed to be stronger in the unit CV (Kozhevnikov and Chistovich, 1965; MacNeilage, 1963; cited in Benguerel and Adelman, 1976). The results of this study suggest that coarticulatory effects are present in CV₂. The final vowel /æ/, which is itself most often in error, is associated with the least errors on the preceding consonants. The final vowel /u/, which is least often in error, is associated with the most errors on the preceding consonants. Likewise, the consonants /p,f,w,θ/, which are least often in error, are associated with the most errors on the final vowel /i/, and with a good number of errors on the final vowel /æ/. Aspects of the visual articulatory cues which could contribute to this pattern will be considered (see Section 5.5.). An hypothesis could be that there are visually dominant consonants and vowels (/p,f,w,θ,u/) which may override less visually dominant consonants and vowels (/k,t,ʃ,tʃ,s,i,æ/) in lipreading where coarticulatory effects are present. Less dominant consonants and vowels are likely to be confused. Within the less visually dominant group, patterns of confusion based on weaker coarticulatory effects may also be found; e.g. final /i/, which is a more closed articulation, seems to interfere with preceding C's more than the more open final /æ/; the more open final /æ/ seems to be more resistant to confusion than /i/, especially following labial and rounded C's where the lips would be involved in articulations that would contrast to the articulatory position of the lips for /æ/.

5.3 Sequential Effects

Better performance by Group N2 relative to Groups N1 and HI has been noted. Performance throughout the first and second administration of the test was examined. Recall that each test consisted of four sequences of 81 V_1CV_2 's. Scores for each set of 81 disyllables were analysed separately. Improvement was evidenced from the first to the second (or third) set for Group N1. Hearing-impaired subjects also improved after the first and second set during the test. Performance on the second administration of the test given to normal hearing subjects was slightly better than performance on the third and fourth sets of the first test. This was true for the five subjects who saw the slow test tape during the second test administration and for the one subject (not counted as a member of Group N2) who saw the fast test tape a second time. These findings are more marked if relative transmission of information, rather than raw score, is considered (see Figure 12).

Relative transmission of information by articulatory features was also analysed for each of the sets of the first and second test. Little improvement in transmission of information about continuant manner of articulation was observed for normal hearing or for hearing-impaired subjects. Improvement in the transmission of information by alveolar or palatal versus other places of articulation was the most pronounced and continued throughout both test administrations. Improvement

FIGURE 12: SEQUENTIAL EFFECTS ON RAW SCORE AND TRANSMISSION SCORE OF CONSONANTS



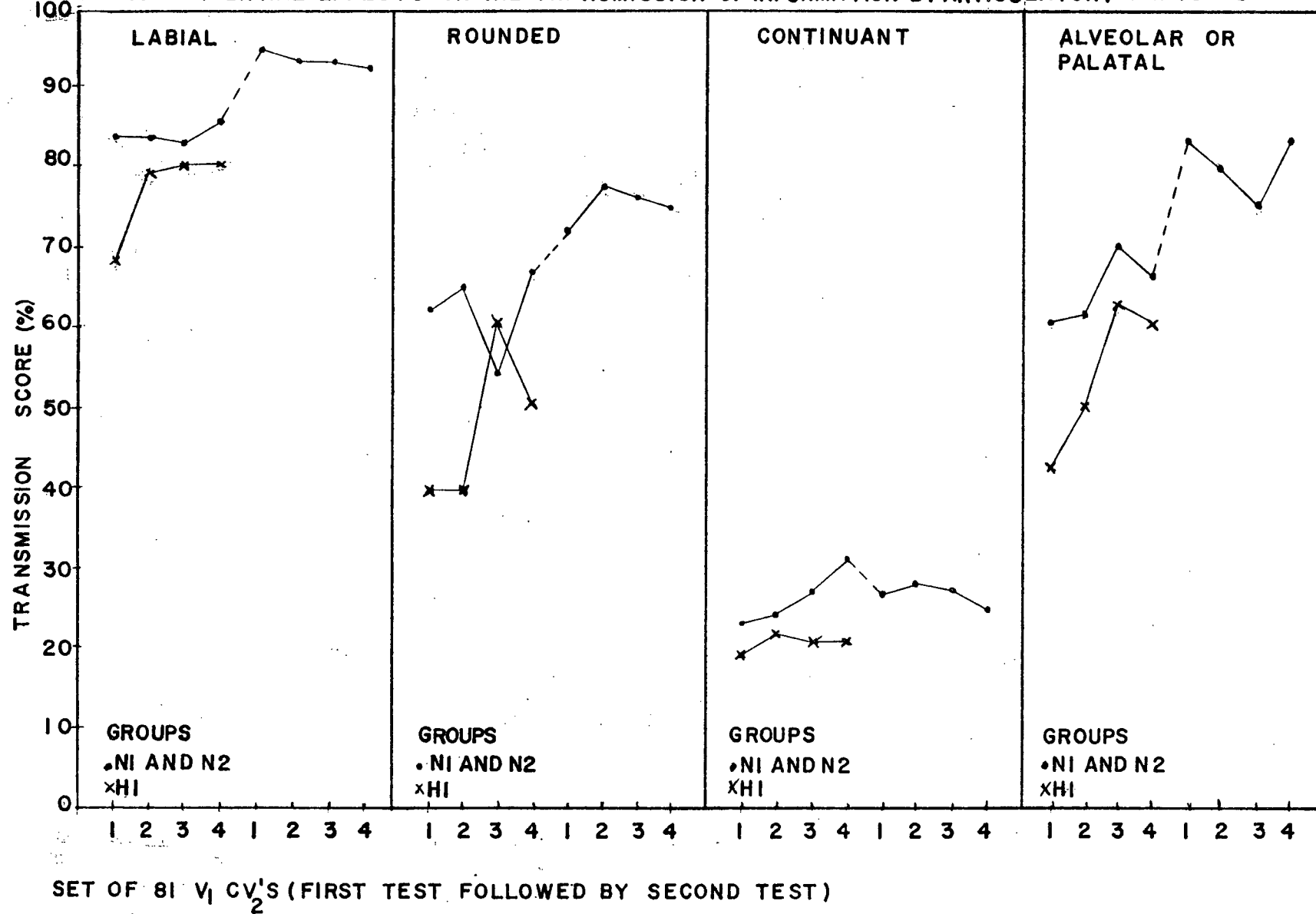
in the transmission of rounding information was the second most pronounced, followed by transmission of information by the feature labial. The hearing-impaired group showed greater gains than did normal hearing subjects on each of the features following the first or second set of the first test administered (see Figure 13).

Examination of particular changes in errors from the first to the second test administration revealed the following findings:

1. Sixteen confusions which occurred infrequently on the first test administered did not occur on the second test administered. The greatest improvement (decreased by ten errors) was for /θ/ which was no longer identified as /t/. Other improvements involved pairs which would be expected to be easily distinguished; e.g., /p/-/f/, /s/-/p/, /w/-/p/.
2. Confusions were reduced for sixteen pairs of consonants, although some errors continued to be made. These improvements occurred for less visually obvious contrasts; e.g., /k/-/t/, /k/-/s/, /k/-/w/. Context was an important contributor to some of these errors, e.g., /k/-/w/ before /u/, where the feature rounded was misidentified.
3. There were increases of more than one error on three pairs: /t/-/s/, /t/-/ʃ/, /ʃ/-/tʃ/ which were among the most confused pairs.

In summary, subjects seem to be able to learn in the task used in this lipreading test. Improvement occurs until performance reaches a level and then little further improvement is noted. Some

FIGURE 13: SEQUENTIAL EFFECTS ON THE TRANSMISSION OF INFORMATION BY ARTICULATORY FEATURES



consonant pairs which are fairly easy to distinguish may be confused at the beginning of the test, however, these pairs are not confused later in the test. Use of articulatory features improves; e.g., rounding for /k/-/w/ and /w/-/k/.

5.4 Reverse Test Results to Investigate the Importance of Syllable Positions of Vowels

Examination of the results obtained for the fast test revealed that lipreaders had difficulty, especially early in the test, in identifying the final vowel of the V_1CV_2 's. The two most likely explanations for difficulty in identifying V_2 were: 1. that V_2 was more difficult to lipread because it was in final position; or 2. that the visual cues available for V_2 were more difficult to lipread than those available for V_1 . V_1CV_2 stimuli were re-recorded in reverse to become $V_2'CV_1'$. The reversed re-recording of the fast tape was administered to six subjects, three of whom were aware of the reverse nature of the test, and three of whom were unaware of it. Test administration was carried out as it had been for the forward tests.

Overall raw scores and transmission scores for V_1 remained nearly perfect when it appeared as V_1' in final position in the reversed stimuli. Scores for C were also about the same for the reversed stimuli as they had been for the original stimuli. Two of the subjects who were uninformed of the reverse nature of the test obtained lower raw scores and transmission scores relative to the other subjects. Raw scores for V_2' (in initial position) in the reverse test were lower than scores for V_1' ; transmission of information was even less. This finding is

similar to that found for V_2 relative to V_1 in the forward test. The results support the second explanation that visual cues derived from the articulation of V_2 must differ from those derived from the articulation of V_1 in such a way that V_2 is more difficult to lipread.

None of the three subjects who were uninformed of the reverse nature of the test could determine that the stimuli had been reversed, however, all three did comment on the bizarre appearance of the articulations. They were able to complete the test despite their observations of there being something unusual about the articulations.

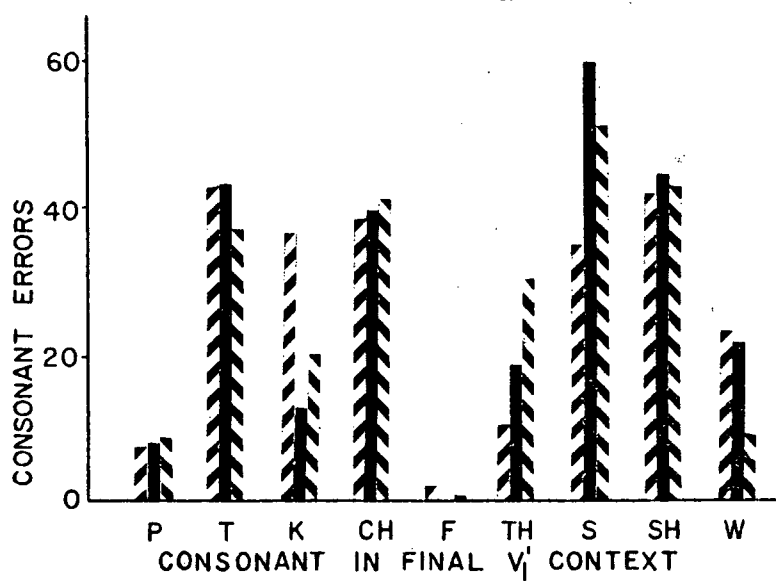
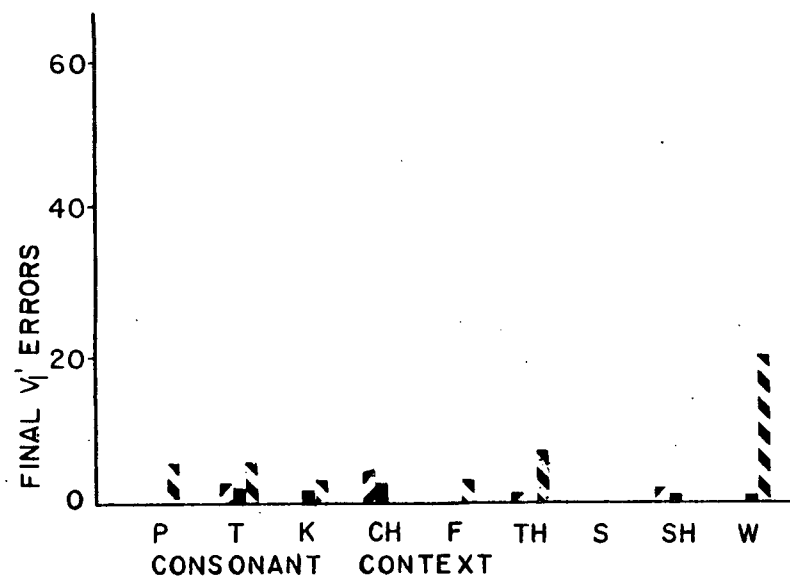
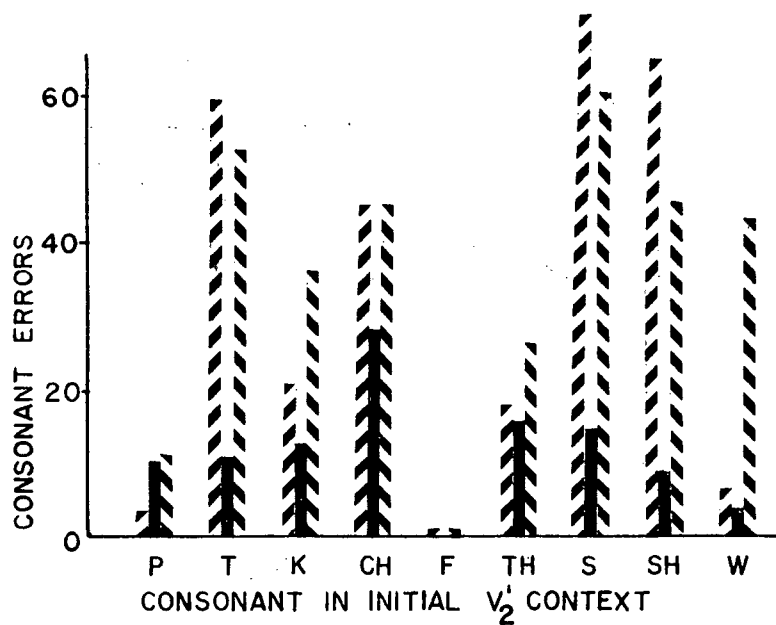
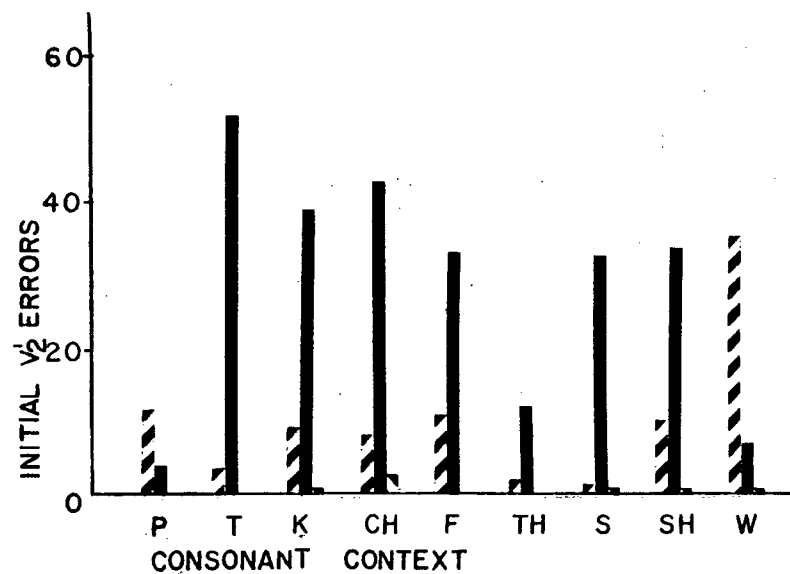
The transmission of articulatory features for V_1' and C in the reverse test were similar to those for V_1 and C in the forward test. Better transmission of information was demonstrated for subjects who were informed of the reverse nature of the stimuli, and for one of the uninformed subjects who had scored nearly perfectly on the previously administered lipreading tests. Transmission of height information for V_2' was poor (33%) although transmission of rounding was only slightly less for V_2' (79%) than for V_1' (84%). Lipreading of V_2' seems to take advantage of the same information transmitted by articulatory features as was found for V_2 .

Consideration of the context in which errors on V_2' , C and V_1' occurred in the reverse test revealed very similar results to those found for the forward test (see Figure 14).

Errors on V_1' were few, however, errors on /u/ after /w/ outnumbered other V_1' errors.

FIGURE 14: REVERSE TEST-PATTERN OF ERRORS

▨ /i/ ■ /æ/ ▨ /u/



Consonant errors in the environment before V_1' were equally distributed after /i/, /æ/, and /u/.

V_2' exhibited a closer relationship to C. Errors on C were fewer when V_2' was /æ/ than when it was a high vowel /i,u/. On the other hand, V_2' was in error most often when it was /æ/. V_2' was nearly perfectly identified when V_2' was /u/. Few errors were made for V_2' when it was /i/.

Sequential effects for the reverse test were very similar to those observed for the forward presentation of the stimuli. Generally, improvement was most marked after the first presentation of the 81 $V_2'CV_1'$'s. Some subjects also showed improvement after the second set of 81 disyllables. Performance reached a plateau on the third and fourth sets. Feature transmission for all four features was similar in the four sequential presentations of the 81 disyllables in the reverse test as it had been in the forward test. Improvements occurred in the transmission of the features labial, alveolar, or palatal, rounded, and continuant (ordered from most to least improvement).

Lipreaders were able to adjust, to some extent, to conditions of test presentation. That the same patterns of results were found for V_2' as for V_2 and for V_1' as for V_1 , and for C_1 despite test syllable position, suggests that lipreaders are influenced by articulatory information in visual cues. The exact nature of the differences between the articulations of V_1 and V_2 must be considered in order to appreciate the coarticulatory effects evidenced in the error patterns, depending on context for V_2 or V_2' .

5.5 Measurement of Articulations

All stimuli of the fast and slow test tapes were sampled near the midpoint of V_1, C_1 and V_2 . At these points, vertical and horizontal lip openings were measured. Measurements of V_1 in the fast test vary little regardless of the following CV_2 context. Measurements taken for V_2 , on the other hand, vary considerably.

In initial position, /i/ and /æ/ differ from each other consistently but slightly. In final position, these slight differences are not as consistently maintained. Co-articulatory effects by the preceding rounded consonant are suggested by a reduction in the horizontal opening and an increase in the vertical opening for /i,æ/ in this context. Measurements for /u/ in both positions are smaller horizontally and vertically than measurements for /i,æ/. In final position, /u/ is characterized by greater horizontal and vertical opening than in initial position; i.e., rounding is less pronounced. In the slow test tape, measurements for V_2 are less variable and more characteristic of the vowels than in the fast test tape. In final position in the slow test, /i/ and /æ/ are more distinct (see Figure 15).

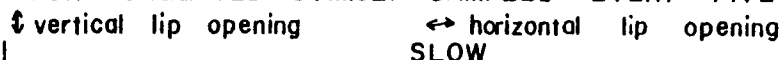
Some stimuli were also sampled every five fields, or every 1/12 seconds. Figure 16 shows a sketch of horizontal and vertical lip opening throughout initial and final /i,æ,u/ where $V_1=V_2$ and C is /w,p,t,/. It can be seen that /i/ and /æ/ are represented by similar articulations, however, /æ/ is consistently represented by slightly greater vertical and horizontal opening of the lips.

FIGURE 15: MEASUREMENTS OF ARTICULATIONS ON THE FAST AND SLOW TEST TAPES

x—horizontal lip opening o—vertical lip opening

FAST				SLOW			
#_Ci		iC_#		#_Ci		iC_#	
46.7	x	46.7	x	46.7	x	46.7	x
400	x	400	x	400	x	400	x
33.3	x	33.3	x	33.3	x	33.3	x
26.7	x	26.7	x	26.7	x	26.7	x
200	x	200	x	200	x	200	x
13.3	x	13.3	x	13.3	x	13.3	x
6.7	x	6.7	x	6.7	x	6.7	x
P T K CH F TH S SH		P T K CH F TH S SH		P T K CH F TH S SH		P T K CH F TH S SH	
#_Cæ		æC_#		#_Cæ		æC_#	
46.7	x	46.7	x	46.7	x	46.7	x
400	x	400	x	400	x	400	x
33.3	x	33.3	x	33.3	x	33.3	x
26.7	x	26.7	x	26.7	x	26.7	x
200	x	200	x	200	x	200	x
13.3	x	13.3	x	13.3	x	13.3	x
6.7	x	6.7	x	6.7	x	6.7	x
P T K CH F TH S SH		P T K CH F TH S SH		P T K CH F TH S SH		P T K CH F TH S SH	
#_Cu		uC_#		#_Cu		uC_#	
46.7	x	46.7	x	46.7	x	46.7	x
400	x	400	x	400	x	400	x
33.3	x	33.3	x	33.3	x	33.3	x
26.7	x	26.7	x	26.7	x	26.7	x
200	x	200	x	200	x	200	x
13.3	x	13.3	x	13.3	x	13.3	x
6.7	x	6.7	x	6.7	x	6.7	x
P T K CH F TH S SH		P T K CH F TH S SH		P T K CH F TH S SH		P T K CH F TH S SH	

— V=/i/ -- V=/æ/ V=/u/
FAST



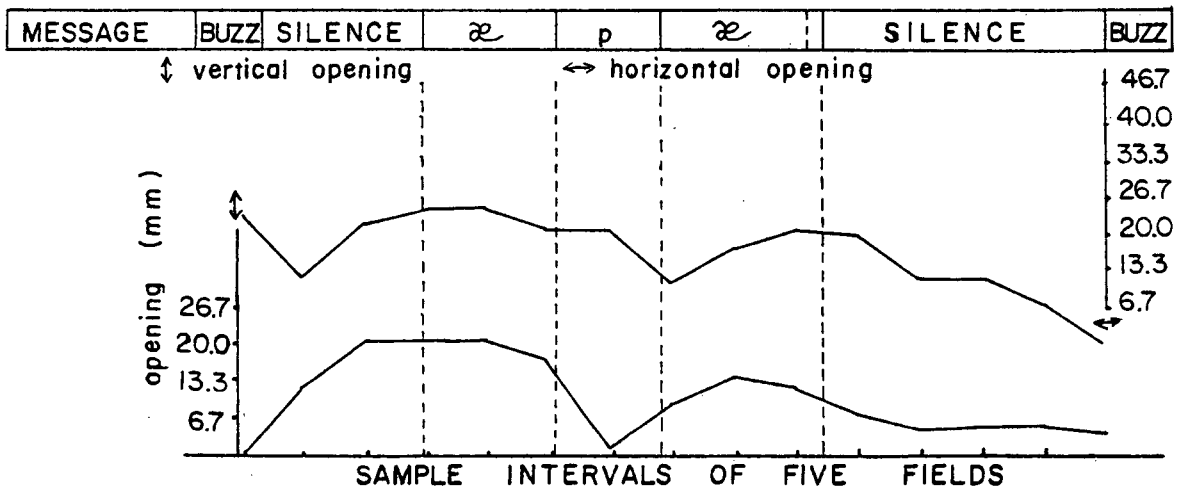
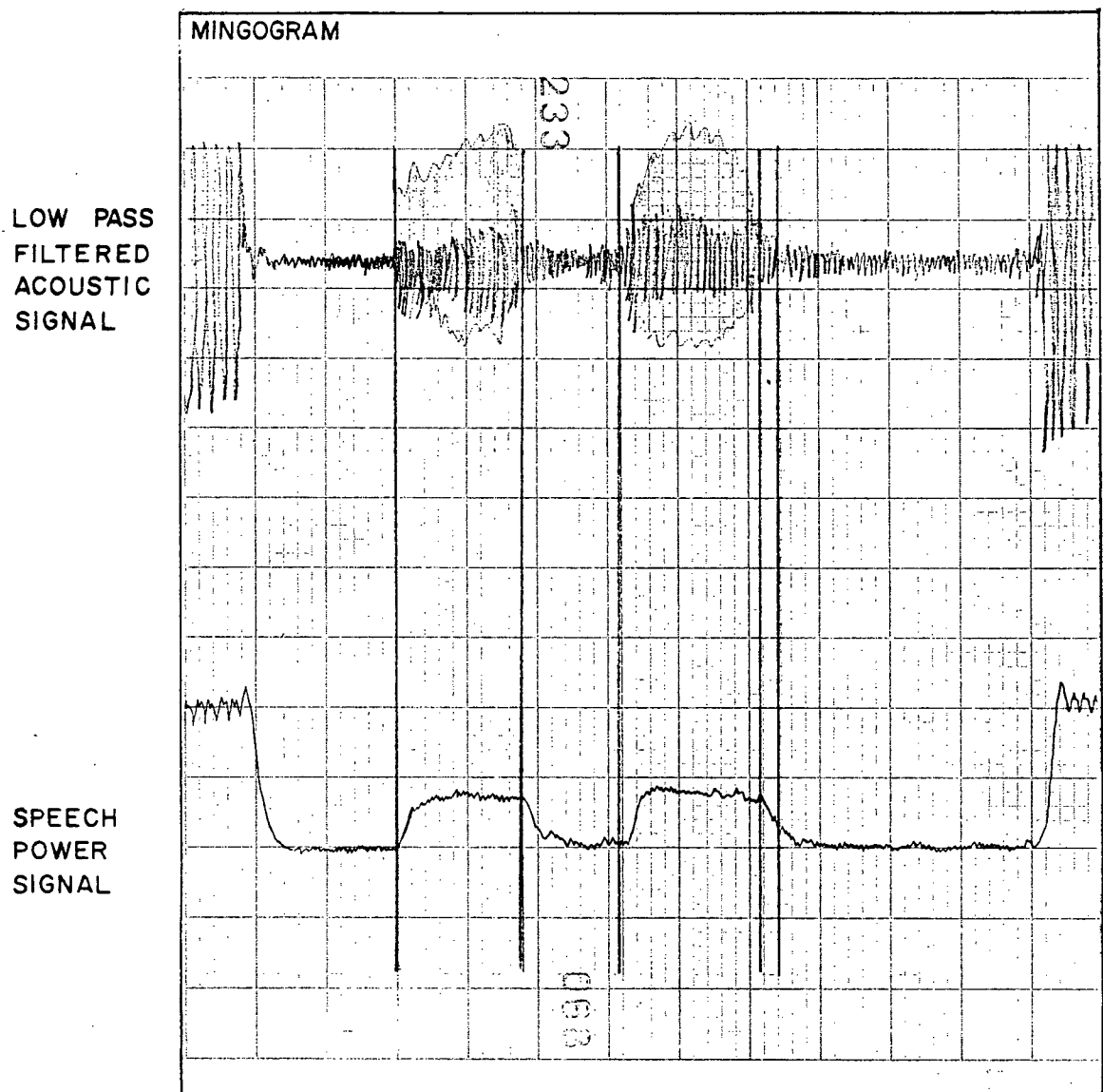
In the slow test, V_1 and V_2 articulations are nearly identical. In the fast test, V_2 vertical and horizontal openings are less for /æ/ and /i/ and more for /u/ than for V_1 ; i.e., distinctiveness of the vowels is reduced. At the onset of most of the stimuli, the lips are set in a position of reduced horizontal and increased vertical opening. The offset of the stimuli is, in most cases, more gradual and less clear from visual cues than from acoustic cues. It often appears visually that the utterance is continuing when, in fact, sound is no longer being produced. In the visual signal, the offset of /i/ and /æ/ are often the same as the lips close to return to a neutral position. Figure 17 illustrates the relationship of visual parameters to acoustic parameters measured for a given stimulus.

Duration of V_1 , C, and V_2 in the fast and slow test were measured. V_2 was assigned both a short and a long duration measurement because of difficulty in interpreting the mingogram tracing. Relative to both short and long durations of V_2 , V_1 was found to be shorter. This relative difference in duration was more pronounced in the fast test tape than in the slow test tape (see Table V).

Table V: Duration Ratios for V_1 and V_2

	Duration Ratios			
	Fast Tape		Slow Tape	
	$V_1: V_2$ short	$V_1: V_2$ long	$V_1: V_2$ short	$V_1: V_2$ long
i	.76	.63	.93	.75
æ	.79	.66	1.03	.84
u	.78	.66	.90	.75

FIGURE 17: RELATIONSHIP OF ACOUSTIC AND VISUAL PARAMETERS



The slow test was characterized by the lengthening of V_1 to almost twice the duration of V_1 in the fast test. V_2 and C were lengthened by about one third (see Table VI).

Table VI: Duration Ratios of Vowels for Fast Compared to Slow Test Tape

Duration Ratios - Fast:Slow			
	i	æ	u
V_1	.54	.58	.6
V_2 short	.67	.76	.69
V_2 long	.65	.75	.68

Table VII: Duration Ratios of Consonants for Fast Compared to Slow Test Tape

Duration Ratios - Fast:Slow									
C	p	t	k	tʃ	f	θ	s	ʃ	w
Ratio	.64	.64	.74	.81	.7	.68	.7	.65	.71

Comparing measurements of opening for the slow test to those for the fast test, it was found that the degree of opening for V_1 was, for the most part, unchanged, however, the steady-state part of the

vowel was doubled in duration. Rate of transitions decreased as much as 25% for vertical and horizontal movements. The degree of opening for consonants was minimally more pronounced. The degree of opening for V_2 was more like that for V_1 on the slow test tape than it was on the fast test tape; i.e., it was more distinctive for the slow test. The steady-state portion of the final vowel was also about doubled in duration. The rate of transition from C to V_2 was reduced 10% to 50% on the slow tape relative to the fast tape on vertical and horizontal opening measurements. Greater extent of opening and longer steady-state portions of V_2 may be related to easier lip-reading of V_2 on the slow test tape (see Figure 16).

Measurements of vertical and horizontal lip opening revealed coarticulatory effects present in the visual cues available to the lipreader.

1. Horizontal lip opening during /t/ was reduced in the environment /u/___/u/ in the slow test tape, probably as a result of coarticulated liprounding. In the fast test tape, vertical opening for /t/ in the environment /u/___/u/ was also reduced compared to measurements found for /t/ in other vowel environments (see Figure 16).
2. Horizontal lip opening was greater for /w/ in the environment /æ/___/ə/ than in other vowel environments in the slow test tape. This was more pronounced in the fast test tape.
3. V_1 and V_2 did not seem to vary in the environment of /p,t,k/ in the slow test tape.

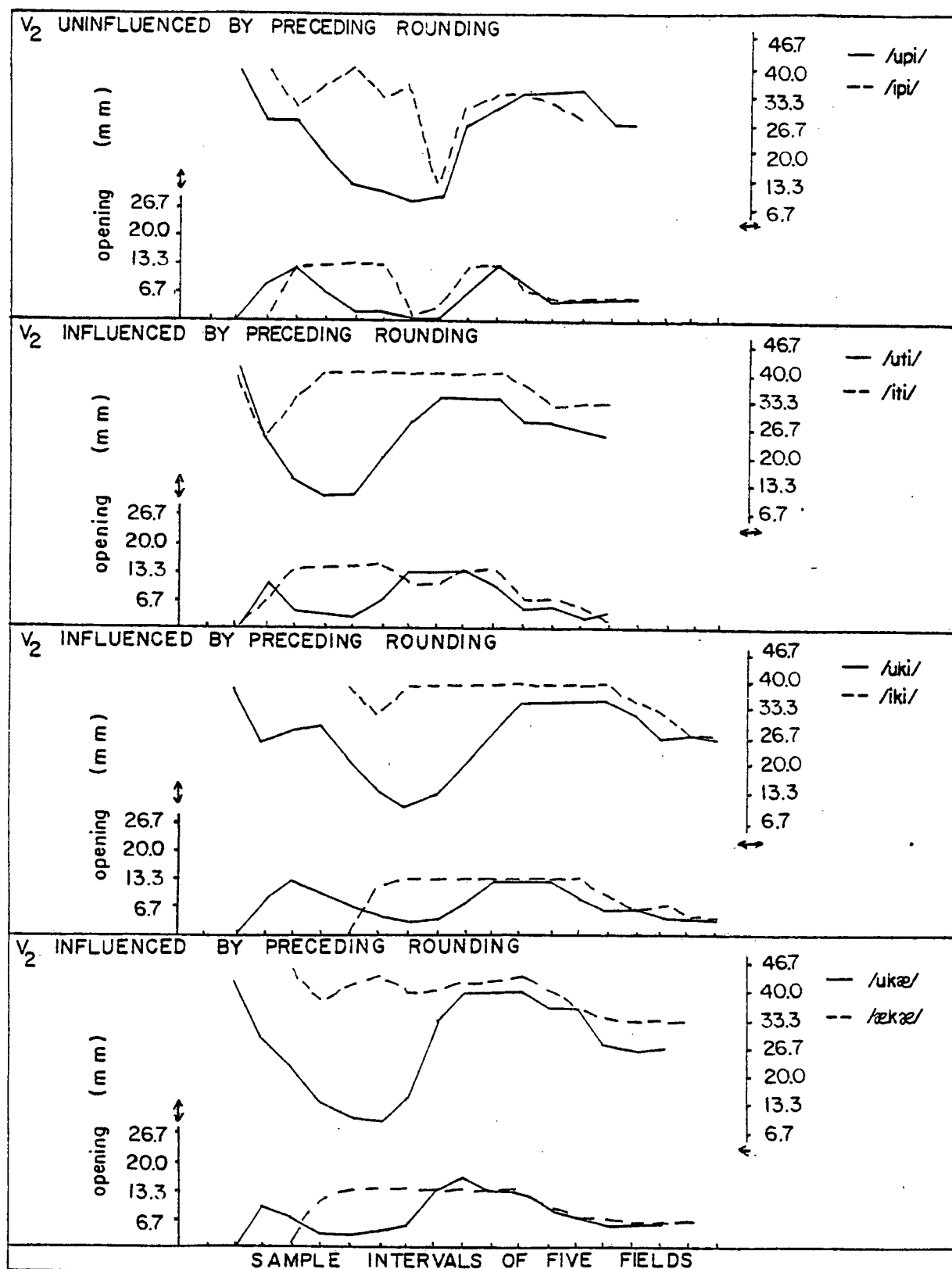
4. V_1 did not vary in the environment of /p,t,k/ in the fast test tape.
5. In the fast test tape, final /i/ and /æ/ were characterized by reduced horizontal opening in the environment of a rounded consonant; e.g. /w/ (see Figure 16).
6. On the fast test tape, final /i/ after /ut/ or /uk/ showed reduced horizontal opening. This was not true after /up/, where the lips closed before V_2 . Final /æ/ after /uk/ showed a lesser reduction of horizontal opening than did final /i/ (see Figure 18).
7. Final /u/ after /p,t,k/ did not vary in these environments in the fast test.

Coarticulatory effects do provide visual cues which influence lipreading performance.

FIGURE 18: EVIDENCE OF COARTICULATION OF ROUNDING

↑ vertical lip opening

↔ horizontal lip opening



CHAPTER 6

CONCLUSIONS

Lipreading ability, as measured in this study, was influenced by three factors: practice, rate at which stimuli were spoken, and phonetic context.

Learning was observed during the test, especially for normal hearing subjects. These subjects were able to quickly improve their skill on the experimental task. Hearing-impaired subjects, who would usually rely more on lipreading than normal hearing subjects, improved less during the test. In daily life, the use of linguistic redundancy which hearing-impaired lipreaders develop is probably different from that evaluated in this task. Phonetic information alone is examined in this task. For this reason, scores obtained may be slightly lower for hearing-impaired subjects who are accustomed to using many levels of redundancy.

The second and third factors which were found to influence lipreading could be directly related to measurements taken from the visual signal providing articulatory information.

Lipreading scores were better on the second administration of the test in which the taped stimuli were spoken at a slower rate. Measurements of the visual signal demonstrated that, in fact, in the slower recording, extent of articulations (vertical and horizontal lip opening) were more extreme, and duration of steady-state segments of the utterance were increased as compared to the fast test tape. Erber et al. (1979) found that highly intelligible tokens were found to be

characterized by extreme articulatory changes and quasi steady-state segments. Coarticulatory influences were not readily measurable for stimuli of the slow test tape by the technique used in this study.

The fast test tape was more difficult to lipread. Stevens, House and Paul (1966) found that greater coarticulatory effects were associated with shorter vowels. Coarticulatory measurements were readily noticeable for the fast tape articulations, especially for V_2 and C. This was true for forward V_1CV_2 and reversed $V_2'CV_1'$ presentations of the stimuli. Ohman (1966) also noted a tendency for final vowels to neutralize more than initial vowels. Kozhevnikov and Chistovich (1966) and MacNeilage (1963) (cited in Benguerel and Adelman, 1976) claimed that the coarticulatory effects were strongest in CV sequences. Variations in articulation appear in faster rates of speech on the videotaped recordings used here. Such variation due to coarticulatory effects made lipreading more difficult, especially at the beginning of this experimental task.

Phonological environment had a more pronounced effect on the visual perceptibility of phonemes in the fast test tape than in the slow test tape. Coarticulatory effects were measured from the visual signal; e.g., rounding of consonants /t,k/ but not /p/ when followed by /u/; rounding of vowels /æ,i/ when following /uk, ut/ but not /up/, or when following /w, tʃ, ʃ/. No coarticulatory effects were seen on measurements of /p,f,w,θ,u/. Lipreading scores prompt the hypothesis that there are visually dominant visemes /p,f,w,θ,u/ for which nearly perfect scores were obtained under all test conditions, whereas there are less dominant visemes /t,k,s,ʃ,tʃ,i,æ/ which were often in error if they occurred in the environment of a dominant viseme. Such a hypothesis could

be explained by variations measured from the visual signal. The results of Franks and Kimble (1972), in their investigation of consonant clusters, also suggested that there are dominant visemes. Articulatory findings from phonetic studies support these findings. Fromkin (1966) found that coarticulation varied so that there was less EMG activity for rounded vowels after /b/ which itself involves contraction of the obicularis oris muscle. The visual dominance suggested to characterize visemes involves labial or dental articulations. The articulatory features which are best transmitted are labial and rounding. Continuant and height information are less well transmitted. The poorly transmitted features are less visible. Two labial articulations do not interfere with the perception of each other. In this case, the articulation involved is shaped for each viseme.

In general, the lipreader has more difficulty interpreting stimuli which exhibit greater coarticulatory effects. Lipreaders do seem to be sensitive to articulatory changes which are subphonemic and subvisemic. Perception of subphonemic and subvisemic differences are suggested by Erber et al. (1979) and Scheinberg (1979). Some ability to improve in the use of this information with practice is suggested, especially if feature transmission as opposed to score is considered. Discrepancies in the literature about the visual perceptibility of phonemes may be explained to a great extent by the type of variation examined in this study.

Efforts to relate articulatory dynamics to lipreading ability promise to reveal valuable information about the process of lipreading, about the development of lipreading skill, and about speech and language perception in general.

BIBLIOGRAPHY

- 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.
- BENGUEREL, A.-P., and ADELMAN, S. (1976). "Perception of Coarticulated Lip Rounding," *Phonetica* 33: 113-126.
- BENGUEREL, A.-P., and COWAN, H.A. (1974). "Coarticulation of Upper Lip Protrusion in French," *Phonetica* 30: 41-55.
- BERGER, K.W. (1972). Speechreading. (National Educational Press, Baltimore).
- BINNIE, C.A., MONTGOMERY, A.A., JACKSON, P.L. (1974a). "Auditory and Visual Contributions to the Perception of Selected English Consonants," J. Speech Hearing Res. 17: 619-630.
- BINNIE, C.A., MONTGOMERY, A.A., JACKSON, P.L. (1974b). "Auditory and Visual Contributions to the Perception of Selected English Consonants for Normally Hearing and Hearing-Impaired Listeners," *Scandinavian Audiology Supplement* 4: 182-209.
- BRANNON, J.B. (1961). "Speechreading of Various Speech Materials," J. Speech Hearing Disorders 26: 348-359.
- BROOKE, M. (1978). "Steps Towards a Video Speech Synthesizer," Medical Research Council Institute of Hearing Research Annual Report 1-2: 20-21.
- 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 and the Short-Term Recall of Pre-consonantal Vowels," *Language and Speech* 16: 67-76.
- 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.
- ERBER, N.P. (1971). "Effects of Distance on the Visual Reception of Speech," J. Speech Hearing Res. 14: 848-857.
- ERBER, N.P. (1972). "Auditory, Visual and Auditory-Visual Recognition of Consonants by Children with Normal and Impaired Hearing," J. Speech Hearing Res. 15: 413-422.

- ERBER, N.P. (1974). "Effects of Angle, Distance and Illumination on Visual Reception of Speech by Profoundly Deaf Children," J. Speech Hearing Res. 17: 99-112.
- ERBER, N.P., DeFILIPPO, C.L. (1978). "Voice Mouth Synthesis and Tactual/Visual Perception of /pa, ba, ma/," J. Acoust. Soc. Amer. 64: 1015-1019.
- ERBER, N.P., SACHS, R.M., and DeFILIPPO, C.L. (1979). "Labimetrics 1: Analysis of Articulatory Dynamics in Relation to Perception of Vowels through Lipreading," J. Acoust. Soc. Amer. 65: Suppl. 1: S136.
- FISCHER, C.G. (1968). "Confusions Among Visually Perceived Consonants," J. Speech Hearing Res. 11: 796-804.
- FRANKS, J.R., and KIMBLE, J. (1972). "The Confusion of English Consonant Clusters in Lipreading," J. Speech Hearing Res. 15: 474-482.
- FRANKS, J.R., and OYER, H.J. (1967). "Factors Influencing the Identification of English Sounds in Lipreading," J. Speech Hearing Res. 10: 757-764.
- FROMKIN, V. (1964). "Lip Positions in American English Vowels," Language and Speech 7: 215-225.
- FROMKIN, V.A. (1966). "Neuro-muscular Specifications of Linguistic Units," Language and Speech 9: 170-199.
- GREENBERG, H.J., and BODE, D.L. (1968). "Visual Discrimination of Consonants," J. Speech Hearing Res. 11: 869-874.
- HEIDER, F., and HEIDER, G. (1940). "An Experimental Investigation of Lipreading," Psych. Monographs 52: 124-153.
- JACKSON, D.L., MONTGOMERY, A.A., and BINNIE, C.A. (1976). "Perceptual Dimensions Underlying Vowel Lipreading Performance," J. Speech Hearing Res. 19: 796-811.
- JEFFERS, J., and BARLEY, M. (1971). Speechreading. (Charles C. Thomas Publisher, Springfield, Illinois).
- 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).
- LASS, N.J. ed. (1976). Contemporary Issues in Experimental Phonetics, (Academic Press, New York).
- 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.

- LOWELL, E.L. (1974). "Perceptibility of Vocalic Nuclei," *Scandinavian Audiology Supplement* 4: 136-152.
- MacNEILAGE, P.F. (1963). "Electromyographic and Acoustic Study of the Production of Certain Final Clusters," *J. Acoust. Soc. Amer.* 35: 461-463.
- MILLER, G.A. and NICELY, P.E. (1955). "An Analysis of Perceptual Confusions Among Some English Consonants," *J. Acoust. Soc. Amer.* 27: 338-352.
- OHMAN, S.E.G. (1966). "Coarticulation in VCV Utterances: Spectrographic Measurements," *J. Acoust. Soc. Amer.* 39: 151-168.
- PESONEN, J. (1969). "Phoneme Communication of The Deaf," *Teacher of the Deaf* 67: 130-131.
- PETERSON, G.E. and BARNEY, H.L. (1952). "Control Methods Used in a Study of The Vowels," *J. Acoust. Soc. Amer.* 24: 175-184.
- SCHEINBERG, J.S. (1979). "Analysis of Speechreading Cues Using an Interleaving Technique," *J. Acoust. Soc. Amer.* 65: Suppl. 1: S136.
- SHEPHARD, D.C., DeLAVERGNE, R.W., FRUEH, F.X., and CLOBRIDGE, C. (1977). "Visual-Neural Correlates of Speechreading Ability in Normal Hearing Adults," *J. Speech Hearing Disorders* 20: 752-765.
- STEVENS, N.K. and HOUSE, A.S. (1963). "Perturbation of Vowel Articulations by Consonantal Context: An Acoustical Study," *J. Speech Hearing Res.* 6: 111-128.
- STEVENS, N.K., 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.
- SUMBY, W.H., and POLLACK, I. (1954). "Visual Contributions to Speech Intelligibility in Noise," *J. Acoust. Soc. Amer.* 26: 212-215.
- SUMMERFIELD, Q., and SPENCER, R. (1978). "The Perceptual Bases of Lip Reading," *Medical Research Council Institute of Hearing Research Annual Report* 1-2: 21-22.
- SUSSMAN, H.M., and WESTBURY, J.R. (1979). "The Effects of Antagonistic Gestures on Temporal and Amplitude Parameters of Anticipatory Labial Coarticulation," to be published in *J. Speech Hearing Res.*
- WALDEN, B.E., PROSEK, R.A., and WORTHINGTON, D.W. (1974). "Predicting Audio-Visual Consonant Recognition Performance of Hearing-Impaired Adults," *J. Speech Hearing Res.* 17: 270-278.
- WALDEN, B.E., PROSEK, R.A., and WORTHINGTON, D.W. (1975). "Auditory and Audio-Visual Feature Transmission in Hearing-Impaired Adults," *J. Speech Hearing Res.* 18: 272-280.

WALDEN, B.E., PROSEK, R.A., MONTGOMERY, A.A., SCHERR, C.K. and JONES, C.J. (1977). "Effects of Training on the Visual Recognition of Consonants," J. Speech Hearing Res. 20: 130-145.

WOODWARD, M.F., and BARBER, G.G. (1960). "Phoneme Perception in Lipreading," J. Speech Hearing Res. 3: 212-222.

APPENDIX I

Instructions for Pilot Test

You will see six different people on the TV. There will be no sound. They each say the same kind of thing (E points to answer sheet). First, you'll see five practice sentences which are listed here (E reads sentences). Then, you'll see ten numbers that can be between one and a hundred. Then, you'll see ten syllables (E reads syllables). Then you'll see ten or eleven sentences. I'll tell you how many there will be during the test. Here is a list of the possible sentences (E distributes list). I'd like you to read these now (S reads sentences and returns list). Turn the page of the answer sheet. Later you'll see ten syllables again. After the first speaker is finished, the second one will start all over again at the practice sentences. Any questions? During the test I'll remind you which item we're on from time to time. There is about equal time between the items of the test. At the end I'll be asking you about the test and about lipreading so you can save your comments until then. The test is not easy. I would expect you to get about one quarter correct. If it seems hard, don't worry and just do your best.

APPENDIX II

Instructions for Main Test

Instructions as given to hearing-impaired subjects:

"Here is an answer sheet (E shows first answer sheet). You will see one of the girls on the TV who you saw last time. This time she will say only one kind of thing. Everytime she says something you are to write down what you think she said. I'll tell you about the kind of thing she says. They are like small words. Each has one vowel, one consonant, and another vowel. They seem more like Japanese words than English words. The vowel can be [i] (E writes "e"), or [æ] (E writes "a"), or [u] (E writes "u"). The consonant can be [p] or [b] (E writes "p", "b"), or [t] or [d] (E writes "t", "d"), or [k] or [g] (E writes "k", "g"), or [tʃ] (E writes "ch"), or [f] or [v] (E writes "f", "v"), or [θ] (E writes "th"), or [s] or [z] (E writes "s", "z"), or [ʃ] (E writes "sh"), or [w] (E writes "w"). So, some examples would be [itæ, uku, æʃi] (S writes response; advice and corrections are provided by E as appropriate). Try some with no voice (E repeats without phonating; S writes response). Now we'll see what it looks like on TV. Before each time she says something, you'll hear a click. I want to make sure you can hear the click (E presents first few items and adjusts click). Any questions? We'll start at the beginning (E rewinds tape). If you get into any trouble, I'll be sitting at the back of the room doing some work, so just tell me. The first part lasts about a half hour. Then you'll have a chance to rest. The second part is the same type of test and will take about a half an hour. Ready?"

Instructions were modified for phonetically trained subjects who were asked to respond in phonetic transcription. They were told that the vowels would be /i,æ,u/ and the consonants /p,b;t,d;k,g;f,v;θ,x;t,d;tʃ,dʒ;w/.

For these subjects, a less simplified explanation of the task was employed; for example, terms such as "nonsense VCV" could be used.

APPENDIX III

Measurement of Transmission of Information

A measure of relative transmission of information has been used to evaluate lipreading performance on each of V_1 , C_1 and V_2 segments of the test stimuli, as well as to evaluate the contribution of four articulatory features to lipreading performance.

The relative transmission is a measure of the covariance between the set of inputs (stimuli), and the set of outputs (responses), (Miller and Nicely, 1955). It is calculated as follows:

$$T_{rel}(x;y) = \frac{-\sum_{ij} p_{ij} \log_2 \frac{p_i p_j}{p_{ij}}}{-\sum_i p_i \log_2 p_i}$$

where x is the input variable ensemble;

x_i is the i th input;

p_i is the probability of x_i ;

y is the output variable ensemble;

y_j is the j th output;

p_j is the probability of y_j ;

p_{ij} is the joint probability of x_i and y_j .

The best possible transmission of information is represented by $T_{rel} = 1$. If transmission is poor, then the stimulus and response are unrelated and T_{rel} is near 0. High transmission can be found if responses are related to the stimulus, even if the responses are not necessarily correct.