INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the original text directly from the copy submitted. Thus, some dissertation copies are in typewriter face, while others may be from a computer printer.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyrighted material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is available as one exposure on a standard 35 mm slide or as a 17″ × 23″ black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. 35 mm slides or 6″ × 9″ black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
Accessing the World's Information since 1938

300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA

Order Number 8820900

An analysis of /p/, /b/ and /m/ in the speechreading signal

Scheinberg, Judith Carol Schwarz, Ph.D.

City University of New York, 1988

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __✓__.

1. Glossy photographs or pages __✓__
2. Colored illustrations, paper or print _____
3. Photographs with dark background __✓__
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages _____
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received _____
16. Other _____

_____

_____

AN ANALYSIS OF /p/, /b/ AND /m/ IN THE SPEECHREADING SIGNAL

by

JUDITH CAROL SCHWARZ SCHEINBERG

A dissertation submitted to the Graduate Faculty in
Speech and Hearing Sciences in partial fulfillment of
the requirement for the degree of Doctor of Philosophy
The City University of New York.

1988

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9-22-87
date

Harry Levitt, Ph.D.
Chairman of Examining Committee

9-22-87
date

Irving Hochberg, Ph.D.
Executive Officer

Katherine S. Harris, Ph.D.

Irving Hochberg, Ph.D.

Allen A. Montgomery, Ph.D.

The City University of New York

Abstract

AN ANALYSIS OF /p/, /b/ AND /m/ IN THE SPEECHREADING SIGNAL

by

Judith Carol Schwarz Scheinberg

Advisor: Professor Harry Levitt

Visible articulatory movements of /p/, /b/ and /m/ were examined in a VCV context with the vowels /i/, /a/ and /u/. In the first experiment, the locations for measurement were determined using an interleaving technique in which odd frames of one VCV sequence were alternated with even frames of another VCV sequence. Areas of jitter appeared in those regions of the face that were not in identical positions at the same time during the production of the two utterances. Jitter occurred primarily in the lower lip, chin and cheeks. The areas of jitter indicated the locations of interest for detailed measurement. In the second experiment, VCV productions were recorded with face markers placed at the jitter locations. Automatic tracking of the face markers was accomplished by computer. Plots of face marker movement for the mid upper lip, mid lower lip and chin were analyzed and the major findings were as follows: Both lips and the chin were displaced further upward for the production of /m/ than for /p/ and /b/. The rate of movement away from closure for the lower lip and chin was greater for /m/ than for /p/ or /b/. The rate of movement of the upper lip towards closure was faster for /p/ and /b/ than for /m/. There were only subtle durational differences measured between /p/, /b/ and /m/. In the third experiment, a forced-choice speechreading task tested the two-way discrimination between /m/ and /b/, /m/ and /p/, and /p/ and /b/ in the CV context with the vowels /i/, /a/ and /u/. The results showed that observers were

able to discriminate between consonants with the same place of production, with the best performance for the /u/ context. The results suggest that /p/, /b/ and /m/ should be grouped into two visemes: /p,b/ and /m/.

ACKNOWLEDGEMENTS

This dissertation could not have been completed without the assistance of many people. I would like to say thank you:

To Professor Harry Levitt, for his guidance, patience and kindness. His critical thinking helped me in every aspect of this dissertation. Through all the years and across all the miles, he has remained interested and supportive.

To Professor Katherine Harris for her insight and ideas. Professor Harris' advice, many years ago, encouraged me to complete this dissertation.

To Professor Irving Hochberg for his important suggestions. I always appreciated his understanding and encouragement throughout the years.

To Harvey Stromberg, for his technical assistance, humor and moral support. This dissertation could not have been completed without Harvey.

To Dr. Birin Prasada and all the people at Bell Laboratories, Holmdel, who were so generous with their time, technical assistance and their facilities.

To Jane Carp, for her friendship and support, and for all the time she spent at Bell Laboratories.

To my husband, Bob, for his patience, his friendship, his moral support, and for his continued interest in this dissertation. Without his help and his interest, it would have been impossible.

To my daughters, Erica and Emily. Their cheerful understanding and cooperation enabled me to complete this project.

To my parents, my sister, the rest of my family, and my friends, for always helping me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER I

INTRODUCTION

Speechreading is the process of deriving verbal information by seeing the facial movements associated with speech production. The speechreading signal is the visible aspect of the articulatory movements that occur during speech production.

One major difference between processing the speechreading signal and listening to speech is that the acoustic signal normally contains all the information necessary for the accurate processing of speech by the auditory system of the listener, whereas the speechreading signal is incomplete.

The primary reason for the inadequate information in the speechreading signal is the existence of homophenous sounds. Homophenous sounds are those speech sounds which are visually indistinguishable to the speechreader. That is, although there are differences in production between two particular phonemes, the visible aspects of the production may look nearly identical. Generally, phonemes that differ in place of production are not homophenous because place of production is easily visible to the speechreader; phonemes that differ in either voicing or manner may be homophenous.

Woodward and Barber's (1960) study categorized phonemes into homophenous groups. The accuracy of this study and the definition of homophenous are both important considerations since the problem of homophenous sounds leads to a controversy. There is a debate over how much importance should be attached to individual movements in the speechreading signal versus the global aspects of language.

In the analytic method the speechreader is taught to be aware of the individual movements associated with all the consonants and vowels in an utterance. In the synthetic method the speechreader learns to reconstruct the meaning of a phrase or sentence from a few recognizable words. The general meaning of the utterance takes priority over the exact decoding of the speechreading signal. The phonetic information that eludes the speechreader is compensated for by the linguistic and situational knowledge that the speechreader does have. Training focuses on increasing and utilizing this non-phonetic information.

This controversy has obscured an important aspect of speechreading. Whether the speechreader depends on a phoneme by phoneme analysis or on inference from words and phrases, speechreading must involve the perception and processing of the articulatory movements that make up the words and phrases. Thus, both the analytic and synthetic methods require that the speechreader derive information from articulatory movements. The interest in this study is to examine the visible articulatory movements of the most visible group of homophenous sounds in great detail.

Since little is known about which articulatory cues are used by a good speechreader, it has been difficult to develop systematic, objective procedures for training deaf adults and children to speechread. The results of this study will provide the kind of information necessary for an understanding of the speechreading process. This information (e.g. which sounds are potentially discriminable) can be used to develop more effective procedures for improving the speechreading skills of hearing-impaired children and adults.

It is important to understand the physical characteristics of the speechreading signal and what can be seen perceptually whether you teach the analytic or synthetic methods. The most common homophenous grouping is /p,b,m/. Therefore this grouping was chosen for study.

There were two purposes to the present study. The first was to compare the visible articulatory movements associated with three English consonants that were considered homophenous: /p/, /b/ and /m/. The movement patterns were examined in three vowel contexts (/i/, /a/ and /u/) and for two different speakers.

The second purpose of the study was to measure the ability of subjects to visually discriminate these so-called homophenous sounds using a forced choice discrimination task.

CHAPTER II

REVIEW OF RELATED LITERATURE

In order to study the homophenous grouping /p,b,m/ in terms of physical and perceptual measurements it is important to understand the following aspects of speechreading:

1) consonant perception in speechreading
2) speechreading training
3) speech production differences for bilabial consonants
4) articulatory displacement measurements
5) articulatory measurement of the speechreading signal
6) synthesis of the speechreading signal

This chapter will present a review of studies concerning these issues. The chapter will be divided into six parts corresponding to these six aspects of speechreading.

Consonant Perception in Speechreading

The speechreading signal has been studied by looking at subjects' phonemic confusions in identification or same-different speechreading tasks. When the confusion rate is high for two phonemes those phonemes are said to be homophenous, i.e., visually indistinctive. These studies are important since speechreading training programs and expectations for speechreading performance have been based on the results of phoneme confusion studies.

The earliest speechreading studies used live stimuli (i.e. face-to-face testing). Heider and Heider (1940) measured phoneme identification using CV (consonant-vowel) nonsense syllables. Each consonant was presented with two different vowels. Thirty-nine children at a school for the deaf participated in the study.

An analysis of the consonant confusions showed three contrastive groups: /m,b,p/, /ʃ,dʒ,tʃ/ and /l,n,t,d,g,j,s,z,k,h,v,f/. Consonants within a group were confused with each other and consonants in different groups were contrastive. These contrastive groups were subsequently labelled visemes by Fisher (1968).

The phoneme confusions within a group were not equally distributed. For example, in the /pbm/ group the percentage of correct responses for /b/ was 27% while for /m/ it was as high as 51%.

Not all the confusions within a group were reciprocal. For example, the consonant /j/ was identified as either /s/ or /t/ as often as it was correctly identified, but /j/ was not given as a response for /t/ and /s/ stimuli. The same pattern for /j/ was seen in the data reported by Binnie et al. (1976).

Woodward and Barber (1960) used a filmed speechreading test and thereby eliminated the subject and trial by trial variability inherent with live presentation. This was a very important advancement in the study of speechreading. The Woodward and Barber study is therefore widely cited and has been used as a basis for data interpretation and as a guide in planning speechreading training methods. For example, Walden, Prosek and Worthington (1974) studied the relationship between auditory and audiovisual perception and analyzed their visual data according to the contrastive groups described by Woodward and Barber. That is, confusions within a Woodward and Barber contrastive group were scored as correct responses.

Woodward and Barber presented a filmed lipreading test to 185 normal-hearing subjects. The test consisted of 102 different CV nonsense syllable pairs such as /pa/-

/ma/, 102 corresponding reciprocal pairs (e.g. /ma/-/pa/) and 25 pairs of identical syllables such as /pa/-/pa/. The 102 different pairs were a subset of the possible 2-way combinations of all phonemes in English. Each pair and its reciprocal were presented only once during the test. The subjects were required to state whether the pairs were the same or different.

The method of analysis used by Woodward and Barber was based on a pilot study of 38 subjects (Woodward, 1957). She used a ranking system to summarize the data. A score was computed for each pair of syllables as follows: The percentage of subjects calling the pair "alike" was subtracted from the percentage of subjects calling the pair "different". The scores of reciprocal syllables were added in order to get a combined score for the particular phoneme contrast. This is shown below:

P = percentage score for pair

P_r = percentage score for reciprocal

$X = P_{different} - P_{alike}$

$X_r = P_{r\ different} - P_{r\ alike}$

$X_t = X + X_r$ = total of pair and reciprocal

Note that a larger X_t meant that more subjects called the two contrasting phonemes "different". The visual difference rating for p-b was .35, for b-m .41, for p-m .66.

The X_t scores were ranked and the list of scores was divided into three sections. The high ranking group was called contrasting phonemes, the middle group was called similar phonemes and the lowest group was called equivalent phonemes. The pairs in the equivalent group and the 22 lowest ranked pairs in the similar group were labelled non-contrastive.

Woodward and Barber then grouped the phonemes so that the phonemes they called contrastive were in different groups and the non-contrastive phonemes were in

the same group. This resulted in four groups: /pbm/, /Wwr/, /fv/ and /tdnløðsztʃdʒʃʒ jkgh/. Phonemes within a non-contrastive group were not necessarily found to be non-contrastive with all the phonemes within the group. For example, /n/ had a high difference rating with /t/, /d/ and /l/ only.

Several limitations of the Woodward and Barber study should be noted:

1) A same-different task was used however only 25/229 of the items were actually the same. The results for those items were not reported.

2) Some phoneme pairs within a contrastive unit had a higher visual difference rating than did other phoneme pairs assigned to different units. For example, the /f/-/r/ contrast was rated more similar than /tʃ/-/j/ yet /f/ and /r/ were assigned to different units and /tʃ/ and /j/ were assigned to the same unit.

3) The reported data did not indicate whether particular subjects were consistently able to see differences among the so-called non-contrastive phonemes or whether all subjects showed the same response pattern.

4) There were only two samples of each phoneme contrast. The fact that the results for a contrast and its reciprocal were sometimes dissimilar indicates the need for many samples of each stimulus type.

Other researchers followed Woodward and Barber in trying to define homophenous groups of phonemes. Table 1 shows that each study resulted in a different number of homophenous groups. A review of these studies indicates that there were sufficient differences in procedures to account for these different groupings. In addition, some disparities occurred because the set of consonants differed from study to study. All the studies measured consonant perception using either an identification or a closed set response task with no more than approximately 20 presentations of each stimulus type.

TABLE 1. Summary of Consonant Perception Studies

Authors	Visemes for Initial Consonants
Heider and Heider (1940)	/p,b,m/ /ʃ,dʒ,tʃ/ /l,n,t,d,g,j,s,z,k,h,v,f/
Woodward and Barber (1960)	/p,b,m/ /W,w,r/ /f,v/
	/t,d,n,l,ø,ð,s,z,tʃ,dʒ,ʃ,ʒ,j,k,g,h/
Fisher (1968)	/p,b,m,d/ /f,v/ /k,g/ /W,w,r/
	/ʃ,t,n,l,s,z,dʒ,j,h/
Binnie, Montgomery and Jackson (1974)	/p,b,m/ /f,v/ /ø,ð/ /ʃ,ʒ/ /t,d,n,s,k,d/
Binnie, Jackson and Montgomery (1976)	/p,b,m/ /f,v/ /w/ /r/ /ø,ð/ /ʃ,ʒ/ /t,d,s,z/
	/l,n/ /k,g/
Walden et al. (1977)	/p,b,m/ /f,v/ /w/ /r/ /ø,ð, / /ʃ,ʒ/ /s,z/
	/t,d,n,k,g,j/ /l/
	Visemes for Initial Consonant Clusters
Franks and Kimble (1972)	/p,b,m/ clusters; /r,w/ clusters;
	/f/ clusters; /ʃ,d/ clusters;
	alveolar,velar and labio-dental clusters
	Visemes for Medial Consonants
Erber (1972)	/p,b,m/ /t,d,n/ /k,g/
Owens and Blazek (1985)	/p,b,m/ /f,v/ /ø,ð, / /w,r/ /tʃ,dʒ,ʃ,ʒ/
	/t,d,s,k,n,g,l/
	Visemes for Final Consonants
Fisher (1968)	/p,b/ /f,v/ /k,g,ŋ,m/ /ʃ,ʒ,d ,tʃ/
	/t,d,n,ø,ð,z,s,r,l/

Fisher (1968), using 18 normal hearing subjects, presented a filmed test of words in phrases rather than nonsense syllables to better approximate conversational English. Five words were randomly grouped together resulting in a non-sensical phrase. The initial consonant of the entire phrase was the test phoneme for 23 of the phrases and the very final consonant was the test phoneme for 20 phrases.

The important distinction of this study is that the response sheet contained only incorrect foils for each test item, however the subjects were not told that the correct responses were omitted. The foils were words where the initial (or final) consonant had the same place or manner of production as the test phonemes.

Based on a confusion matrix of the responses Fisher defined five groups of initial consonants as contrastive and called each group a viseme: /p,b,m,d/, /f,v/, /k,g/, /W,w,r/ and /ʃ,t,n,l,s,z,dʒ,j,h/. Some of the consonants within a group were not confused in a reciprocal manner and therefore were not strictly homophenous. For example, /m/ was confused with /b/, but /b/ was not confused with /m/. Initial /m/ was chosen as a response only 3 out of 385 times.

Unlike Woodward and Barber, Fisher separated the velar and alveolar consonants. In addition, Fisher included initial /d/ in the bilabial group since it was called /p/ or /b/ in 11 out of 17 cases. Both Fisher and Woodward and Barber put /f/ and /v/ in a separate group, whereas Heider and Heider (1940) included /f/ and /v/ with the alveolar and velar phonemes.

The five groups for the final consonants were /p,b/, /f,v/, /k,g,ŋ,m/, /ʃ,ʒ,dʒ,tʃ/ and /t,d,n,ø,ð,z,s,r,l/. The phonemes within these groups were also not always reciprocally confused. Fisher did not group final /m/ with /p/ and /b/ since /p/ and /b/ were never identified as /m/ by the subjects whereas /ŋ/ was called /m/ in 7 out of 12 cases.

Franks and Kimble (1972) looked at phonemic confusions in speechreading in terms of consonant clusters rather than single consonants. They used three speakers and 275 normal hearing subjects. The subjects viewed a film in which each of 32 consonant clusters in nonsense syllables was spoken once by each of three speakers. The subjects were not told that the utterances were consonant clusters and they were not given feedback regarding their responses.

Franks and Kimble divided the consonant clusters into five contrastive groups on the basis of their confusion matrix: 1) clusters containing initial /p/, /b/ and /m/ and some medial position bilabials, 2) clusters containing /r/ and /w/, 3) clusters containing /f/, 4) clusters containing /ʃ/ and /d / and 5) clusters containing alveolars, velars and labio-dentals.

Erber (1972) compared normal hearing, severely hearing-impaired and profoundly deaf children. Previous studies looked at either hearing-impaired or normal hearing subjects. Erber presented 45 video-taped samples of each consonant rather than only a few samples as in previous studies. The stimuli were in an /a/-C-/a/ context and were presented in a closed response format. The consonants were bilabials (p,b,m), alveolars (t,d,n) and velars (k,g). Special lighting was used to provide intense illumination of the talker's oral cavity.

The results were similar for the three groups of subjects with the normal hearing subjects showing the poorest overall performance. There was confusion within a place of production; however there was a tendency not to confuse nasals and non-nasal at the same place of production. For example, in the severely hearing-impaired group /p/ was identified correctly 104 out of 225 times and was incorrectly called /m/ only 30 times. The /m/ stimuli were correctly identified 124 out of 225 times and called /p/ 43 times and /b/ 43 times. For all three groups the responses for /b/ stimuli were divided

equally among /b/, /p/ and /m/. The response /b/ was given equally often for /b/ and /p/, but was less frequent for /m/.

Normal hearing subjects did not differentiate between the nasal and non-nasals (for bilabials and alveolars) as well as the hearing-impaired groups.

Erber (1974) reported another study with profoundly deaf children in which the talker's mouth was intensely illuminated. This study differed from Erber (1972) in that consonants were presented in three different vowel contexts: /aCa/, /iCi/, /uCu/. The results show that alveolar consonants were confused with each other more frequently when the vowel was /u/. For the bilabials there was a response bias in favor of /b/ and this was especially true for /u/. Erber found that the subjects were able to discriminate between the velar, alveolar and labio-dental phonemes which Woodward and Barber had classified into one non-labial group. Subjects could not discriminate between phonemes sharing place of production.

Binnie, Montgomery and Jackson (1974) presented ten normal hearing subjects with sixteen consonants in a C/a/ context in a closed set response speechreading test. Each C/a/ was presented twenty times for a total of 320 recorded stimuli. The resulting data were divided into five groups. Phonemes that were not confused with each other were put into different groups. The five groups were /pbm/, /fv/, /øð/, /ʃʒ/, /tdnszkg/. Binnie, Montgomery and Jackson did not report the extent of the confusions within each group.

There were some differences between these results and those of other studies. Binnie, Montgomery and Jackson and Erber (1974) found /ʃ/ and /ʒ/ to be contrastive with the alveolar phonemes, however Woodward and Barber and Fisher did not. Unlike Binnie, Montgomery and Jackson, Erber (1972, 1974) and Fisher grouped /kg/ separately. Binnie, Montgomery and Jackson attributed this to differences in lighting and test procedures.

Binnie, Jackson and Montgomery (1976) repeated their earlier study with a larger number of subjects (34 speech and hearing students with normal hearing). The resulting confusion matrix was divided into nine visemes using the criterion of each viseme having a score of at least 70% correct. The authors attributed the increase in the number of visemes (compared with previous studies) to better lighting and viewing procedures.

Walden et al. (1977) demonstrated that the number of visemes can be increased as a result of speechreading training. In this study the subjects were 31 untrained hearing-impaired adults most of whom had noise induced hearing loss. The subjects were given a pre-test of 20 consonants in a C/a/ context. A control group received the pre-test and post-test and did not participate in the training program.

The training materials consisted of lists of four to nine syllables (identical to the ones used in the pre-test). Initially subjects were presented with CV's that were contrastive according to the Woodward and Barber (1960) classification. Then they were trained with CV's that were homophenous according to the Woodward and Barber classification. However, the subjects were never presented with voiced cognates as part of the discrimination exercises. At the conclusion of fourteen hours of individualized training the original CV test was presented again.

The data were analyzed using a hierarchical clustering technique in which similar items were put into one cluster and then similar clusters were combined in stages until all items formed a single cluster. The criterion for the number of clusters was that each cluster should contain the phonemes that account for 75% of the responses for the stimuli in that cluster.

There were six pre-training visemes and nine post-training visemes based on the 75% criterion. The control group showed no improvement. These results indicated that the training program resulted in improved discrimination.

The effect of vowel context was studied by Owens and Blazek (1985). Subjects were asked to identify the VCV that was presented. They were familiarized with the 23 consonants used in the experiment. Owens and Blazek grouped the subjects' responses into visemes using a criterion similar to Binnie et al. (1976). The criterion for including a consonant in a viseme grouping was that 75% of the responses to that consonant were consonants within that grouping. (For example, /m/ was included in the /p,b,m/ viseme since at least 75% of the responses to stimulus /m/ were either /p/, /b/ or /m/.) According to this criterion there were seven visemes for /aCa/ and six for /uCu/ and /iCi/. There were only two visemes for /uCu/ stimuli (/p,b,m/ and /f,v/). Erber (1974) found that the consonant confusion matrices differed slightly for /uCu/ compared with /iCi/ and /aCa/. Owens and Blazek attributed the more random responses for /u/ to the lack of visibility of non-labial consonants during the production of /u/.

When all the data were combined, the responses were grouped into 6 visemes: /p,b,m/; /f,v/; /ø,ð/; /w,r/; /tʃ,dʒ,ʃ,ʒ/; and /t,d,s,k,n,g,l/.

Owens and Blazek compared the confusion matrices for two groups of subjects: post-lingually hearing-impaired adults and normal hearing adults. The results for the two groups were very similar. Owens and Blazek concluded that it was possible to generalize such speechreading data from normals to hearing-impaired subjects.

In contrast, Erber (1972) found differences as a function of hearing impairment on an analytic, visual consonant recognition task. Normal hearing subjects did not perform as well as the severely or profoundly impaired subjects. Subjects with severe losses performed slightly better than those with profound losses.

In summary, most studies showed that subjects could not discriminate between more than five to nine visemes. The number of visemes obtained in a confusion experiment depended on the subject population, the type of task (recognition or discrimination), the utterance type, the recording conditions, the speaker, the statistical

method used to divide the phonemes into contrastive groups and the training of the subjects (Walden et al., 1977; Walden et al., 1981). However, there was a common trend. The smallest and most consistent groupings were /p,b,m/, /w,r/ and /f,v/. The /p,b,m/ viseme is of the most interest since it appeared in all studies, whereas the other consonants were not used as stimuli in some of the studies. It is likely that /p/, /b/, /m/ differences exist since subjects discriminated between /p/, /b/ and /m/ to some extent in some of the studies (e.g. Erber, 1972).

Speechreading Training

There are two basic approaches to speechreading training: the analytic method and the synthetic, or global method. In the analytic method the speechreader learns the movements associated with each phoneme. The global method teaches the speechreader to rely on contextual information and does not focus on the fine structure of the signal.

Binnie, Montgomery and Jackson (1974) recommended the global approach since confusion studies had shown that visual phoneme perception provided only place of articulation information and subjects were able to utilize that information without any training.

Subsequently, Walden et al. (1977) pointed out that a certain level of phonetic discrimination competency is necessary in order to speechread, even if the speechreader relies primarily on contextual information. However, they raised the question of whether an improvement in phoneme discrimination would result in improved speechreading of words and sentences.

In 1981 Walden et al. studied the effect of consonant recognition training on audiovisual sentence recognition. They found that either auditory or visual training with CV's and VC's resulted in improved performance for audiovisual sentence

recognition. Walden et al. cautioned that these results may be interpreted several ways. There may be a direct relationship between sentence recognition and improved phonemic recognition. On the other hand it may be that the phonemic training just focuses attention on visual and auditory cues. In addition, they pointed out that audiovisual sentence recognition showed an improvement immediately after training, but that long-term effects of consonant recognition training have not been studied.

The 1981 study of Walden et al. is important since there are few data to support either the analytic or global approaches to speechreading training. Note that this study did not compare analytic with global training. It showed, however, that improved performance was possible with analytic training.

Within the analytic framework, there are two approaches to phonemic training. One viewpoint is that training should consist of discrimination between viseme groups since phonemes within a viseme are thought to be indistinguishable.

The other approach is to train speechreaders to see differences between sounds within a viseme group (i.e. to increase the number of visemes). This is somewhat contradictory since the phonemes within a viseme are by definition, indistinguishable. However, since different experiments have resulted in different numbers of visemes, and since it was shown that training increases the number of visemes (Walden, 1977), it is clear that the actual number of visemes depends on both the experimental conditions and the criteria for defining a viseme.

The first approach, training between viseme groups, has been used widely. For example, Jeffers and Barley's (1971) Quick Recognition Exercises teach rapid discrimination between visemes (e.g. discrimination between "sing" and "thing"). In Jeffers and Barley's Quick Identification Exercises the speechreader is asked to give a homophenous substitution for a given word. The speechreader is taught that a particular movement may be associated with many words.

The limitation of this speechreading training method is that the visemes have been defined based on the performance of untrained speechreaders who are not necessarily good speechreaders. The level of performance in a particular study may not be a satisfactory goal for other speechreaders. That is, the use of previously defined visemes limits the speechreading student who may be able to learn finer discriminations. On the other hand, in both research and rehabilitation, time is an important limiting factor, and there may not be enough time to train awareness of subtle differences. Training discrimination between visemes may be viewed as an efficient first step. If this approach is used, it is crucial that the viseme groups be defined accurately. Therefore, demonstrating that /p/, /b/ and /m/ differences exist will have implications for speechreading training and will indicate a direction for future speechreading research.

Synthesis of the Speechreading Signal

Electronic synthesis of lip shapes is being to be used to study speechreading. Erber and De Filippo (1978) synthesized mouth shapes using Lissajous figures on an oscilloscope. The horizontal aspect of the Lissajous figures corresponded to the width of the mouth and the vertical aspect corresponded to the height of the mouth opening. The mouth opening approximated the vowel /a/. The figures were presented in combination with vibratory buzz signals at time delays representing voice onset for the syllables /pa/, /ba/ and /ma/. Erber and DeFelippo wanted to determine whether this representation of voice onset timing in relation to mouth opening is a feasible cue for identifying /pa/, /ba/ and /ma/. Performance on the task was poor initially, however one subject was then trained to label stimuli correctly on a consistent basis.

Erber (1979) synthesized lip shapes representing vowels. The height and width of the shapes were determined by the vowel formant frequencies. The validity of this method was tested by presenting these synthetic vowels to subjects experienced in

lipreading. Only 39% of the vowels were identified correctly, however the confusions were with neighboring vowels. This particular system did not produce the articulatory details that would enable a subject to distinguish between similar vowels, however Erber pointed out that more detailed synthesis techniques will be effective as a research tool in speechreading.

Erber, Sachs and De Felippo (1979) described another method for synthesizing mouth outlines varying in height and width. The parameters for these stimuli were obtained from a frame by frame analysis of video-recorded vowel production. Subjects were able to label the vowels represented by the outlines.

Walden, Montgomery and Prosek (1987) developed a synthetic visual stimulus that included lip thickness and teeth. The stimuli were computer-generated. Video tapes of a sequence (i.e. animated pictures) were then used in perceptual experiments. In this particular study the stimuli were modified along a continuum in order to measure categorical vs. continuous perception.

Brooke and Summerfield (1983) described a computerized system for generating synthetic visible speech. In this system it was possible to specify the shape and the location of the facial features and to modify the extent of the detail shown for a particular feature. Hair, ears, eyes, eyebrows, nose as well as lips and the face outline were included.

The features were made up of points that were specified as either fixed, independent or dependent. Fixed points did not move at all. Independent points moved independently of one another (e.g. points in the lip margins). The dependent points moved concurrently with independent points (e.g. the jaw line moved along with the extreme point of the jaw). The position and movement of the points in the synthesized speech were pre-determined by a frame by frame analysis of video recorded natural speech.

Brooke and Summerfield synthesized VCV bisyllables with /m/, /b/ and /p/ and /i/, /a/ and /u/. They were presented along with natural VCV's in an identification task. The consonants were not identified accurately for either natural or synthetic stimuli. The natural vowels were identified with a score of 98%. The scores for the synthetic vowels /u/ and /a/ were 97% and 87% respectively, however, the score for /i/ was only 22%, possibly because the bottom teeth and tongue tip were not included in the facial synthesis. In a subsequent perceptual study Brooke, McGrath and Summerfield (1984) looked at the effect of adding teeth to the synthetic model of the face. There was a significant positive effect for the vowel /i/, but little effect for other vowels.

Computerized synthetic stimuli are useful for speechreading training since it is possible to make systematic changes in the positions of the articulators and measure the perceptual effect (Erber, Sachs and De Filippo, 1979). It is likely that new training strategies will be developed both as a result of information from synthesis research, and as synthesized training materials are available. The availability of computerized synthetic stimuli is an important technological step in speechreading research and training, much as the introduction of filmed speechreading stimuli was nearly 40 years ago.

Speech Production Differences for Bilabial Consonants

The muscular and aerodynamic differences between /p/, /b/ and /m/ are of particular interest in this study. Although /p/, /b/ and /m/ are difficult to distinguish visually, there may be subtle visible differences resulting from the differences in production. It is expected that the p-b differences will be less visible than the differences between the stops (p,b) and the nasal (m). The difference between p and b occurs primarily at the level of the larynx, whereas the stop-nasal difference is in the manner of production and may affect the visible articulators to a greater extent.

There are several differences between the stops (p and b) and the nasal /m/ (Borden and Harris, 1984). The closure for stops is followed by a sudden release whereas the closure for /m/ can be prolonged. In the nasal the sound emanates from both the oral and nasal cavity. Nasal production requires less intraoral pressure so it is possible to have coarticulatory movement and still maintain the pressure requirement. In stop production, coarticulatory movements that would compromise the pressure build-up cannot occur. Thus, there is more coarticulation for nasals.

Although the primary difference between /p/ and /b/ is voice onset time, there are other differences in production. Some differences in /p/, /b/ and /m/ production have been studied using measures of muscle activity (electromyography (EMG)) and supraglottal pressure. EMG studies of labial muscles have shown small differences between /p/, /b/ and /m/ with a great deal of variability across subjects.

Harris, Lysaught and Schvey (1965), measuring from upper and lower lips, found no significant differences in peak response magnitude for lip opening. For lip closing the peak response magnitude for /p/ production was greater than for /b/ for four out of five subjects. The response for /b/ was greater than for /m/ for four out of five subjects. The relation /p/> /b/>/m / was true for only three of the five subjects.

Lubker and Parris (1970) looked at lower lip obicularis oris activity for /p/ and /b/ in eighteen subjects. Although there was greater EMG activity for /p/ than /b/ in initial, medial and final positions, the differences were statistically significant only in the final position.

Tatham and Morton (1973) measured the EMG response from the upper lip obicularis oris. They found that the EMG amplitude for /p/ was higher than for /b/, however the difference was not statistically significant except when the measurement was taken at the moment of the plosive release.

Sussman, MacNeilage and Hanson (1973) found that one upper lip muscle, the depressor anguli oris, showed a substantial difference in averaged EMG activity for /p/, /b/ and /m/ while there was little difference in the obicularis oris superior and the quadratus labii superior. The greatest activity was for /p/ with the least activity for /m/.

For the lower lip there was less activity for /m/ for the obicularis oris inferior, which agrees with Lubker and Parris (1970). However, /m/ showed greater activity than /p/ for the mentalis muscle (lower lip closure) and for the depressor labii inferior (lower lip depression). The jaw lowering muscle, the anterior belly of the digastric, also showed greatest activity for /m/ and least for /p/.

In a recent study, Tatham, Daniloff and Hoffman (1985) recorded from the upper lips of four subjects during lip closure. As in previous studies, there was variability across subjects, however, when differences for the upper lip existed, /p/ showed greater peak EMG amplitude than /b/.

In summary, for the majority of subjects, when differences exist, the peak height of the EMG response was greatest for /p/ and smallest for /m/. The exception to this was the lower lip data in which the greatest activity was for /m/.

Lisker (1970) evaluated four different measures of supraglottal pressure: peak pressure, onset time, decay time and pressure pulse. The data were based on one subject. For /p/ and /b/ in the final position it was possible to differentiate between them (/p/>/b/) using any one measure of pressure. Peak pressure showed a difference only for medial consonants in the post-stress position. The pressure pulse differentiated the consonants except in the initial position where only the decay time measure showed /p/-/b/ differences. In summary, the supraglottal differences depended on stress and consonantal position.

Lubker and Parris (1970) found that for eighteen subjects supraglottal differences between /p/ and /b/ were significant for initial, medial (before stressed and

unstressed syllables) and final positions. The pressure was measured in peak amplitude and was greater for /p/ than /b/.

There are differences in supraglottal pressure when initial and non-initial consonants are compared (eg. CV vs. VCV utterances). Several studies (Arkebauer, Hixon and Hardy, 1967; Lisker, 1970; Malecot, 1955) have measured greater supraglottal pressure for medial /p/ and /b/ than for initial /p/ and /b/. However, a recent study (Flege, 1983) found that the pressure for medial /p/ was greater than for initial /p/, whereas the pressure for initial /b/ was greater than pressure for medial /b/. As a result, the /p/-/b/ supraglottal pressure differences were smaller in the initial position than in the medial position.

Lubker and Parris (1970) showed that when measures of intraoral pressure, labial pressure and electromyographic activity were combined, there was error-free discrimination between /p/ and /b/.

Articulatory Displacement Measurements

Several studies have measured lip and jaw movement for /p/, /b/ and /m/. Fujimura (1961) measured the vertical separation and horizontal width of the lips for /p/, /b/ and /m/ in various contexts using an optical method of measurement. The utterances were video-recorded at the rate of 240 frames per second. There was one production of each utterance type by one speaker.

Fujimura's results suggested that there were measurable differences in lip position for /p/, /b/ and /m/. He found differences of 1 or 2 mm within 5 msec of the consonant explosion. At the maximum vowel opening these differences were again 1 or 2 mm. For initial consonants /p/ showed the greatest vertical separation between the lips 5 msec after consonant explosion. The differences were vowel dependent when the consonant followed the neutral vowel.

Fujimura suggested that one difference between the plosives and /m/ was that for plosives there is an upper lip deformation immediately before and after plosion. This is due to an overpressure prior to plosion which then results in a highly damped oscillation of the lips. The oscillation was not seen for /m/ presumably due to the lack of overpressure.

Sussman, MacNeilage and Hanson (1973) measured lip and jaw displacement for /p/, /b/ and /m/ in VCV contexts using a strain-gage transducer. For both the jaw and lower lip /m/ showed the greatest downward displacement and /p/ the least. However these were average values and in some cases /p/ and /b/ displacement was greater than /m/. For the upper lip closing gesture /p/ and /b/ were similar and /m/ showed slightly less displacement downward than /p/ and /b/.

Sussman, MacNeilage and Hanson also compared the velocity of movement for lips and jaw for /p/, /b/ and /m/. For the upper lip the closing and opening velocities were greater for /p/ than for /b/ and were greater for /b/ than for /m/ (i.e. they were in the order /p/>/b/>/m/).

The closing velocity for the jaw was fastest for /p/ and slowest for /m/. In contrast, lower lip closure showed a slower velocity for /p/ with /m/ and /b/ nearly equal.

Jaw lowering occurred at the slowest rate for /p/, with /m/ and /b/ at similar rates. The lower lip opening gesture was fastest for /m/ and slowest for /p/.

Articulatory Measurement of the Speechreading Signal

Measurement of the speechreading signal (i.e. the visible articulatory signal) requires an optical system of measurement that will detect movement that is detectable by the human visual system. For example, the articulatory displacement measures reported by Sussman, MacNeilage and Hanson (1973) describe the movement of visible articulators, however the movement detected by the strain gage transducer is not

restricted to visually detectable movement. Similarly, consonant differences measured within as short a time frame as 5 msec (Fujimura,1961) would not be visually perceptible. That is, when articulatory movement is analyzed for the purpose of understanding speechreading, it is necessary to account for visual perception.

A speechreading study of visible articulation was done to determine whether there are physical differences in lip movement for words that appear visually identical. Joergenson (1962) investigated whether there are physical differences in lip movement for words that appear visually identical. Motion pictures of 48 homophenous words spoken by four speakers were analyzed frame by frame. Words were chosen on the basis of a study by Roback (1961) in which subjects discriminated between homophenous words at a level better than expected by chance. Joergenson used the words that had the highest frequency of correct identification in Roback's study.

Joergenson measured mouth opening, mouth width, and teeth visibility of the words. Measurements were taken at intervals of 0.25 seconds. There were no significant differences between the words for any of the measurements with the exception of mouth opening. However, the differences in mouth opening were not systematic and could not be attributed to phonemic differences.

Joergenson's study was very limited since there were only five measurements per second and measurements were to the nearest 32nd of an inch. In addition, the measurement points on the face were chosen arbitrarily. Joergenson suggested that more sensitive measurements were needed in order to find differences.

Jackson, Montgomery and Binnie (1976) measured lip positions in order to study the relationship between physical measurements of the lips and perceptual dimensions of vowel speechreading. Physical measurements of lip movement were taken of vowel samples at several stages of vowel production: immediately prior to production, during production at the extreme position and during the extreme position

for the second vowel in diphthongs. Three measurements were made for each of these lip positions: from corner to corner, vertical lip separation and from the midpoint between corners to the lower lip. These measurements were then compared with the results of the perceptual dissimilarity scaling for the vowels. Five perceptual dimensions were correlated with five physical aspects: lip extension or rounding, vertical lip separation, size of opening, vertical movement from first to second nucleus of diphthong, size of second nucleus opening.

The speechreading signal has been analyzed on a frame by frame basis in speechreading synthesis studies (Erber, Sachs and De Filippo, 1979; Brooke and Summerfield, 1983). Erber, Sachs and De Filippo (1979) analyzed vowel productions using a frame rate of 54 frames per second. Measurements were taken of horizontal and vertical upper and lower lip displacement, lip thicknesses, teeth and jaw locations and tongue width and height. The measurements were taken using a superimposed horizontal and vertical grid. Mouth outlines of vowel production were synthesized from these data.

Brooke and Summerfield (1983) did a frame by frame analysis of video recorded natural speech. Articulatory trajectories for one speaker were obtained for eleven marked points on the lips and one point on the jaw. The points were chosen to represent lip movements for spreading, rounding and protrusion. There were no points on the cheeks. The point on the jaw was just under the chin.

The video recordings were stored on a video disc and could be displayed on a frame by frame basis along with an x,y coordinate indicator. The experimenter was able to move the indicator to the exact location of a point and press a key, thus digitizing the coordinate values of each point from frame to frame.

The trajectories of the points that were presented in the study were selected at random. Vowel differences were evident, however, differences between /p/, /b/ and /m/ were no larger than differences between tokens of the same consonant.

In summary, Joergenson (1976) and Brooke and Summerfield (1983) were the only researchers to analyze visible articulatory movement during consonant production. Joergenson's data was imprecise compared with present day computer capabilities. Brooke and Summerfield did not find differences between /p/, /b/ and /m/ for the particular points reported in their study.

Conclusions

Homophenous groupings have been described in many studies. That is, certain phonemes are consistently confused with each other in speechreading. However, the structure of these groupings is not definite. There is variability from study to study as to what the groupings are. The /p,b,m/ grouping is particularly important since it is the most visible to the speechreader and was defined as a viseme in every study.

New technology in speechreading is beginning to emerge. There is electronic synthesis of lip shapes. This technology requires analytic information about the speechreading signal.

There are two basic approaches to speechreading training: analytic and global. This particular study is interested in the analytic approach. In the analytic method, speechreaders are trained to discriminate between homophenous groupings. This places an emphasis on the accuracy of the homophenous groupings and raises two questions: How accurate is the /p,b,m/ grouping? How can advanced technology help to answer this question?

CHAPTER III

EXPERIMENTAL PROCEDURES

The study consisted of three experiments: In the first experiment the points of measurement were determined using a technique called interleaving. In the second experiment these points were used to measure facial movements. The third experiment of the study consisted of a speechreading task. A set of homophenous consonants was used as test material in all three parts of the study.

Experiment 1: Interleaving

Speakers

Two female graduate students age 26 and 27 served as speakers for the video recordings. (One of the speakers was the author.) The speakers had non-deviant speech patterns. Both speakers spoke a dialect common to the New York City area.

Speech Materials

The bilabial consonants /p/, /b/ and /m/ were chosen for study because of their high visibility and because they were considered to be homophenous. The consonants were combined with vowels to form VCV disyllables in which the same vowel was in the ini-tial and final position. The vowels /i/, /a/ and /u/ were chosen because they represent the extreme positions on the vowel triangle. They also involve extreme positions of the lips.

The VCV's were spoken in groups of three such that each group of three constituted a breath group. Each of these groups of three VCV's is referred to as a triplet.

There were two types of triplets. In one type the vowel was the same in each of the VCV's and there were three different consonants. An example of this is /aba/ /apa/ /ama/. In the second type the consonant was the same in each of the VCV's and there were three different vowels. An example of this is /aba/ /ibi/ /ubu/. These two types of utterances made it possible to compare movements that differed either due to a change in vowel or due to a change in consonant.

Table 2 shows the utterances recorded for the interleaving experiment. Four replications of each triplet were recorded by Speaker 1 and two replications of each listed triplet were recorded by Speaker 2.

Constraints in the computer time available made it impossible to record a complete set of utterances for Speaker 1, however, each consonant appeared with each vowel at least once. The nasal version of the triplet with a changing vowel was omitted because of the time constraint. For one triplet (/ibi/ /ipi/ /imi/) a problem with the recording was suspected so the entire recording was repeated. Both recordings turned out to be useful.

When the recordings for Speaker 2 were made a more complete design was possible since the number of replications was halved and more computer time was available. For all three vowels each consonant appeared in each position within the triplet. Only one triplet with a changing vowel was recorded.

Use of Video Recordings

Video recordings of utterances were used in order to have identical, repeatable stimuli as required for the interleaving procedure, the detailed analysis of facial movements, and the related perceptual study.

TABLE 2. Utterances Recorded for the Interleaving Experiment

SPEAKER 1	SPEAKER 2
/ibi//ipi//imi/	/ibi//ipi//imi/
/aba//apa//ama/	/imi//ibi//ipi/
/ubu//upu//umu/	/ipi//imi//ibi/
/ipi//apa//upu/	/aba//apa//ama/
/ibi//aba//ubu/	/ama//aba//apa/
	/apa//ama//aba/
	/ubu//upu//umu/
	/umu//ubu//upu/
	/upu//umu//ubu/
	/ipi//apa//upu/

Apparatus

A computer system designed for video processing was used. The system consisted of a Digital Equipment Corporation PDP-11/45 computer (core memory of 48k 16 bit words), a General Instrument Corporation rotating magnetic drum memory for bulk data storage, and a monochrome T.V. camera (R.C.A. camera head PK701 and module (115V, 50/60 Hz)) with a 55mm lens (Nikon). An A/D converter sampling at a rate of about 2.3 MHz digitized the video signals. The digitized video signals were stored on the drum and then could be transferred to the disk for processing or viewed on a video monitor (Miratel) after being converted back to the analog signal by the D/A converter.

A palindromic display format was used in which the frames were displayed first in a forward sequence and then in the reverse sequence. The use of a palindromic display eliminated discontinuities in the video display. For example, during speech, if the lips were closed in the first frame and open in the last frame there would be a jump in the picture if the sequence of frames were repeated cyclically in the same order; i.e. the last frame with the open lips would be followed by the first frame of a new cycle showing closed lips. In a palindromic display the order of the frames is reversed midway during the display and the lips appear to be opening and closing without discontinuities.

As is standard practice in video transmission systems, the video pictures were made up of a series of horizontal lines. A total of 256 lines constitutes a frame. A frame is scanned in two parts, known as fields. In the first field all even numbered lines are scanned and in the second field all odd numbered lines are scanned. The system employed 30 frames per second and since there are two fields per frame, there were 60 fields per second.

Each line consisted of 256 picture points known as pels (for picture elements). Each pel had 256 levels of gray.

Each pel required 8 bits of storage in the computer; i.e. 8 binary digits have 2^8 (=256) possible combinations. Since there are 256 pels/line and 256 lines/frame, a total of 8 x 256 x 256 bits of computer storage is required for each frame of video signal. The computer could store a total of 256 frames of uninterrupted video signal. At 30 frames per second this allowed just over eight seconds of continuous video information to be stored in the computer without interruption. Each sequence of eight seconds of continuous video recording is referred to as a pass. When recordings were made of the stimuli it was found that four triplets could be recorded on a single pass.

Method

Interleaving refers to the combination of two video recorded sequences into one sequence. Odd frames of one sequence are alternated with even frames of the other sequence to make a new video recording. In the resulting recording half the frames are from sequence 1 and half are from sequence 2. The number of frames in the new sequence equals the number of frames in either original sequence since half the frames in each original sequence are discarded. If the number of frames in the two original sequences is not equal, then the new sequence has the same number of frames as the shorter of the two original sequences. The number of frames ranged from six to twelve for most of the sequences.

In this study, sequence 1 and sequence 2 were consonant-vowel (CV) portions of previously video-recorded VCV utterances. For example, /pi/ was interleaved with /mi/. Note that sequence 1 and sequence 2 were from two different VCV utterances. In order to minimize the effects of differences in head position and head movement (as opposed to articulatory movement), the CV sequences were taken from VCV's

produced in a single eight second sitting. In most cases the two sequences were produced within four seconds of each other.

In order to interleave two VCV's, the sequence of frames making up each VCV must be synchronized in time. Lip positions were used as time markers in the utterances. These positions were determined by viewing each frame and labelling the frame in terms of the lip position with respect to consonant closure. The lips were described as either at the first point of closure, closed, protrusion closed, almost closed, almost open or open. The first point of closure was the first point at which the lips came together. This was defined as the time origin; i.e. time=0 at this point. Closed refers to the frames following the first point of closure when there continued to be no space between the upper and lower lips. Protrusion closed identified those frames when the lips protruded during the period of closure. The almost closed and almost open designations were necessary for frames showing the rapid transitional movement between open and closed. The open lip position corresponded to the vowel portions of the VCV.

The interleaved recording showed articulatory movements that appeared quite natural except for areas of jitter in those regions of the face that were not in identical positions at the same time during the production of the two utterances. The areas of jitter indicated the locations of interest for detailed measurement.

Figures 1 through 7 show an example of interleaving. Odd frames from the utterance /imi/ were alternated with even frames from the utterance /ipi/. The first frame shown (figure 1) is from /imi/. The jitter in going from odd to even frames and vice versa was most apparent in the lower lip and the cheek regions.

Every type of CV was interleaved with every other type of CV whenever possible. Some triplets and some parts of triplets were inadequate for interleaving due to excessive head movement during the eight seconds of recording. Limitations on available computer time also reduced the number of interleaved samples. Thus, not all

Fig. 1. The first frame of the sequence interleaving /imi/ and /ipi/. This frame is from /imi/ and is the first point of closure.

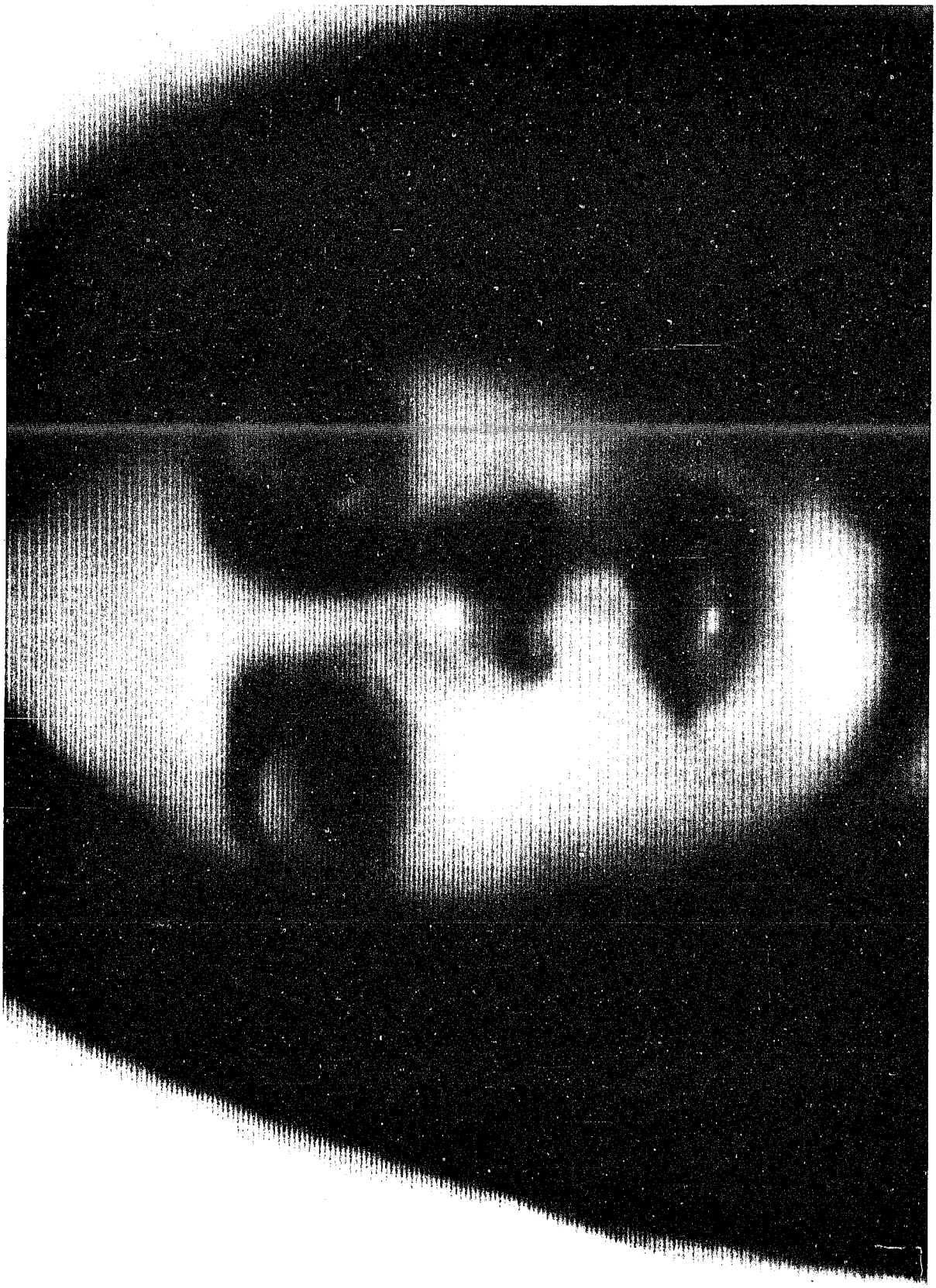


Fig. 2. The second frame of the sequence interleaving /imi/ and /ipi/. This frame is from /ipi/.

Fig. 3. The third frame of the sequence interleaving /imi/ and /ipi/. This frame is from /imi/.

Fig. 4. The fourth frame of the sequence interleaving /imi/ and /ipi/. This frame is from /ipi/.

Fig. 5. The fifth frame of the sequence interleaving /imi/ and /ipi/. This frame is from /imi/.

Fig. 6. The sixth frame of the sequence interleaving /imi/ and /ipi/. This frame is from /ipi/.

Fig. 7. The seventh frame of the sequence interleaving /imi/ and /ipi/. This frame is from /imi/.

CV-VC combinations were interleaved from each triplet. For Speaker 1, some replications of triplets could not be used. For Speaker 2, no recordings of the following triplets could be used: /imi/ /ibi/ /ipi/, /upu/ /umu/ /ubu/, /ubu/ /upu/ /umu/.

Phonemes in the initial and final position were interleaved with each other and with the phonemes in the medial position. Some of the interleaved samples consisted of two different productions of the same CV (e.g. /pa/-/pa/). This was done to test whether there was any observable jitter for identical phonemes.

Table 3 shows that there were twenty-four interleaved samples for Speaker 1 and 35 samples for Speaker 2. The information obtained from viewing these interleaved samples was used in Experiment 2.

Experiment 2: Measurement of Face Markers

The purpose of this experiment was to track the movement of points on the face for /p/, /b/ and /m/ utterances.

The speakers for experiment 2 were the same as for the interleaving experiment. The same video recording apparatus was used as well.

Location of Face Markers

The points of measurement were chosen so as to sample those regions of the face that showed jitter in the interleaving experiment. These points were located in the regions in which noticeable jitter was observed; i.e. the regions in which differences in articulatory movement between phonemes were observed.

There were fifteen points of measurement for Speaker 1 and twenty-one points of measurement for Speaker 2. For both speakers there were three additional reference points which formed a calibration triangle. These reference points were not affected by speech production since they were located on the nose (one point) and on eye glasses

TABLE 3. Interleaved Samples.

VCV TRIPLET	TRIPLET REPLICATIONS*	INTERLEAVED CV'S
Speaker 1		
ibi ipi imi	8	bipi,bipi,bipi,bipi,bipi,bipi,bipi,pimi
aba apa ama	2	bama,bapa,mama**
ubu upu umu	2	pumu,bupu,mumu
ibi aba ubu	3	bibi,bibi,biba
ipi apa upu	3	pipi,papa,papa,pupu,pipa,pipu,papu
Speaker 2		
ibi ipi imi	2	bipi,bimi,pimi,mibi,mimi
ipi imi ibi	2	mibi,pimi,pipi,bibi
apa ama aba	2	pama,paba,papa,mama
aba apa ama	2	bapa,bama,pama,baba,mama
ama aba apa	2	bapa,bapa,mapa,baba
umu,ubu,upu	2	mubu,bupu,bupu,bupu,mupu,mumu,bubu
ipi apa upu	2	pipi,papa,pipa,pipu,papu,papu

*This indicates how many replications there were for each type of triplet.

**Each /ma/ is from a different VCV replication; i.e. two different productions are interleaved. This is true for all non-contrasting samples. Contrasting samples are from a single triplet.

(two points). Any movement of these three points reflected global head movement rather than the effect of specific speech utterances.

Each measurement point was given a name for further reference and for use in the computer program used in experiment 2. Tables 4 and 5 identify the mnemonic name of each measurement point. Figures 8 and 9 show the locations of these points for speakers 1 and 2 respectively.Facial movements at the points of measurement were obtained as follows: the subjects' faces were first blackened using pancake make-up and the teeth were covered with a thin layer of black tooth wax. White, adhesive-backed paper dots, approximately 1.5 mm in diameter, were placed at the points determined by the interleaving method. These dots were called face makers.

The computer was programmed to identify white dots from a dark background thereby providing automatic tracking of facial movements as indicated by the dots. In order to facilitate the automatic identification of the dots, as large a contrast as possible between the white dots and black background was used.

Video Recordings

The speakers were seated facing the camera in the recording studio. Lighting was adjusted so that there was a maximum of contrast between the dark background and the face markers. It was necessary to avoid reflection from areas other than the dots so the walls and the subjects' glasses were covered with dark, non-reflective paper.

The speakers were able to maintain a steady head position by focusing on a distant point. A head rest was tried but not used since it caused more head movement than the above procedure.

Each recording was made in the same manner. The eight second video recording was activated by the computer. At the end of the eight seconds it was possible to see

TABLE 4: Face Markers for Speaker 1

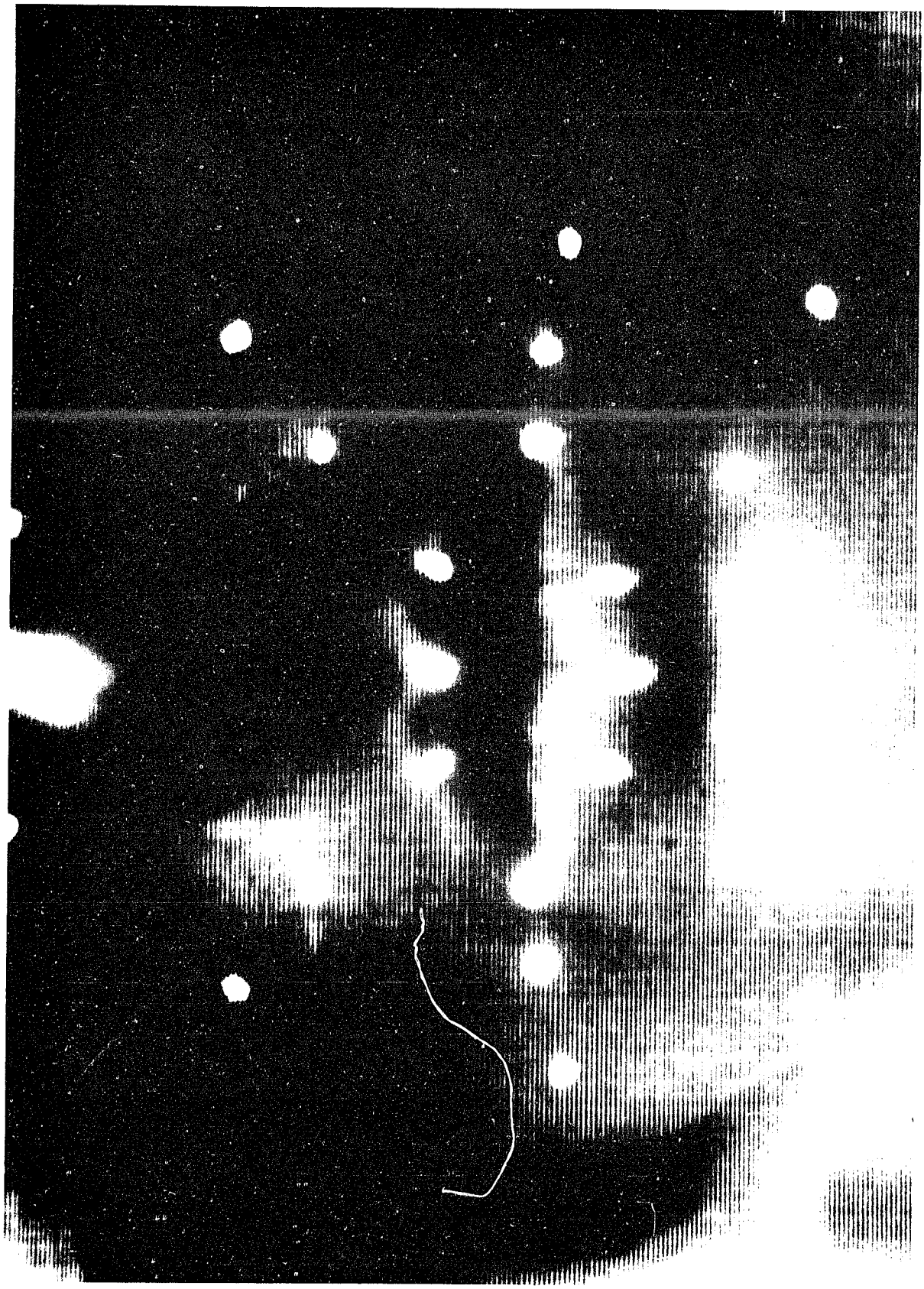
MULP	-	midpoint of upper lip
LULP	-	left upper lip
RULP	-	right upper lip
MLLP	-	midpoint of lower lip
LLLP	-	left lower lip
RLLP	-	right lower lip
CHIN	-	chin
LUCH	-	left upper cheek (superolateral to LULP)
LLCH	-	left lower cheek (inferior to LUCH)
RUCH	-	right upper cheek (superolateral to RULP)
RLCH	-	right lower cheek (inferior to RUCH)
LCR	-	left corner of mouth
RCR	-	right corner of mouth
LB	-	left bottom, between mouth corner and chin
RB	-	right bottom, between mouth corner and chin
L	-	left side of glasses
R	-	right side of glasses
N	-	nose, superior to points on glasses

TABLE 5: Face Markers for Speaker 2

MULP	-	midpoint of upper lip
LULP	-	left upper lip
RULP	-	right upper lip
MLLP	-	midpoint of lower lip
LLLP	-	left lower lip
RLLP	-	right lower lip
CHIN	-	chin
LLCH	-	left lower cheek (superolateral to LULP)
LUCH	-	left upper cheek (superolateral to LLCH)
RLCH	-	right lower cheek (superolateral to RULP)
RUCH	-	right upper cheek (superolateral to RLCH)
LCR	-	left corner of mouth
RCR	-	right corner of mouth
LB	-	left bottom, between mouth corner and chin
RB	-	right bottom, between mouth corner and chin
LJAW	-	left jaw
RJAW	-	right jaw
LNS	-	left near side (on cheek, lateral to LCR)
LFS	-	left far side (on cheek, lateral to LNS)
RNS	-	right near side (on cheek, lateral to RCR)
RFS	-	right far side (on cheek, lateral to RNS)
L	-	left side of glasses
R	-	right side of glasses
N	-	nose, inferior to points on glasses

Fig. 8. The location of the face markers are shown for Speaker 1.

Fig. 9. The location of the face markers are shown for Speaker 2.



the recording in a palindromic display. The utterances were re-recorded if there were any technical flaws due to computer or lighting problems. Once a recording was judged to be technically satisfactory it was stored on digital tape and the next group of utterances was then recorded on the drum.

An audio recording was made simultaneously in order to check the naturalness of the speech and the content of the recording.

Speech Materials

The utterances used here were also VCV's spoken as triplets (VCV, VCV, VCV). The vowel remained the same throughout a triplet and the three different consonants were homophenous. As before, each triplet was produced on a single breath group. The second syllable in each VCV was stressed. Pitch generally fell for the third VCV of each triplet. In order to balance for this effect each VCV appeared in each position within the triplet at least twice. This resulted in at least six replications of each VCV for each subject.

Table 6 lists the triplets recorded by both speakers. A complete set of three triplets, with each VCV in initial, medial and final position was recorded for each vowel. Since the available computer time was very limited, only a partial set of additional utterances could be recorded. For Speaker 1 there were five additional utterances and there were two for Speaker 2. Just as in the interleaving experiment, two replications of each triplet were recorded at a sitting.

Computer Processing

1. Identification of Face Markers

The face markers (i.e. the white dots) were located in terms of their x,y coordinates using a subroutine that read the intensity level of each picture element in

TABLE 6. Utterances Recorded for Face Marker Measurement

SPEAKER 1	SPEAKER 2
/aba//apa//ama/	/aba//apa//ama/
/ama//aba//apa/	/ama//aba//apa/
/apa//ama//aba/	/apa//ama//aba/
/ama//apa//aba/	/ibi//ipi//imi/
/aba//ama//apa/	/imi//ibi//ipi/
/ibi//ipi//imi/	/ipi//imi//ibi/
/imi//ibi//ipi/	/ipi//ibi//imi/
/ipi//imi//ibi/	/ibi//imi//ipi/
/imi//ipi//ibi/	/umu//ubu//upu/
/ibi//imi//ipi/	/upu//umu//ubu/
/ubu//upu//umu/	/ubu//upu//umu/
/umu//ubu//upu/	
/upu//umu//ubu/	
/ubu//umu//upu/	

each field. The face markers varied in size and consisted of five to fifteen contiguous pels. By choosing the correct threshold level for detection it was possible to program the computer to reject picture elements that were not face markers. Thus, the face markers were detected automatically from the stored picture elements.

Special care was given to rejecting areas of reflection that were not face markers but were equally bright. This was done interactively with the computer. A visual sequence was scanned and the dots were detected. If there were fewer dots detected than the number of face markers the threshold was lowered. If too many dots were detected this indicated that reflections were mistakenly being detected as face markers. In that case the threshold was raised. Threshold levels were varied according to the region of the face. In this way it was possible to discriminate between the dots and spurious reflections. The most troublesome area requiring the most interactive adjustment was in the region of the lips because of saliva. Figures 10, 11, and 12 (see below) demonstrate the lighting and threshold difficulties.

The x, y coordinates of the pels were listed in the order in which they were detected by the computer program. Pels which had adjacent coordinate values were grouped together by means of a sorting program. Each group of adjacent pels, so identified, constituted a face marker. This program also assisted in eliminating unwanted points of reflection; e.g. fields that did not conform to the standard pattern were identified and could be checked by eye for reflections or other erroneous estimates of face markers.

Since each face marker was composed of several pels, the average value of the pels' x and y coordinates, respectively, was used to identify the location of the face marker. A separate computer program took the list of average coordinates and identified them in terms of the articulatory location such as "left lower lip" (left side of the lower lip).

The location of a face marker was more difficult to measure in the case of streaks. Streaks resulted when a face marker moved rapidly within the time span of a field (1/60 of a second).

Streaking was primarily in the vertical direction and occurred for the vowels /i/ and /a/ but not for /u/ because there is a smaller range of movement for /u/. Most streaks consisted of eight to twelve pels in the vertical direction.

Due to the rapid movement, the brightness of the streaked face markers was reduced thereby lowering the contrast with the background. Since the velocity of movement during the field was not necessarily uniform, the brightness level was also not necessarily uniform throughout the streak. Thus, although the thresholds were set at a level to detect the streak, the less bright portions of the streak might not have been detected.

In cases where the precise length of the streak was difficult to measure, the error was no greater than half the size of the streak, since the averaged y coordinate value of an entire streak equals the midpoint of the streak. The worst error was 4-6 pels. (For Speaker 1, for example, 4-6 pels would be approximately 1cm.) The effect of error is that a face marker appeared to reach its next position sooner or later in time than it actually did.

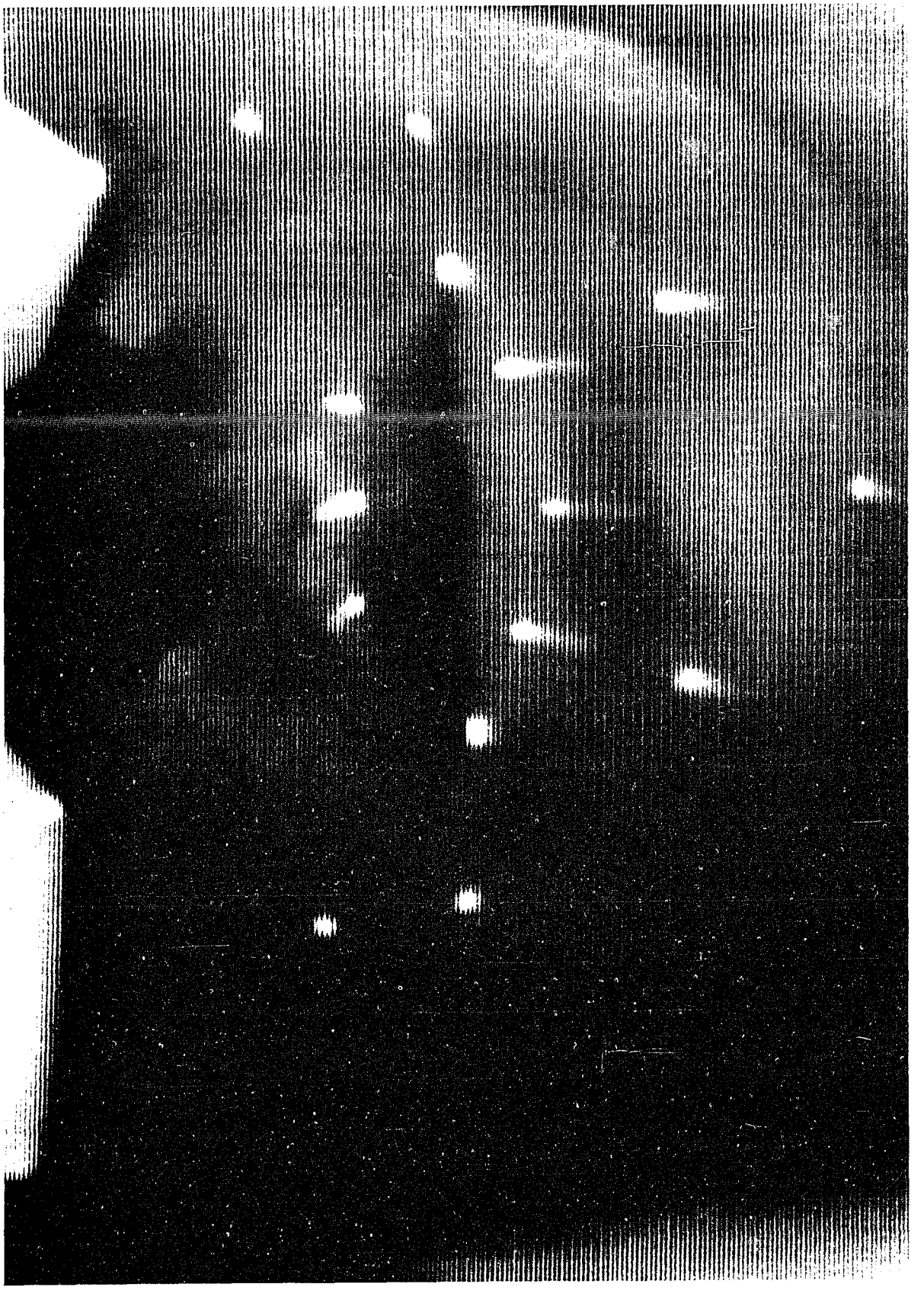
Note that the midpoint of the streak did not necessarily represent the same point in time for each frame since the velocity of the lips was not uniform.

Figures 10, 11, and 12 are three frames from the utterance /imi/. Figure 10 is two frames prior to closure, figure 11 is one frame prior to closure, and figure 12 shows closure. Note the streaking that occurred in regions of rapid movement. Streaks are most uniform in figure 10.

Fig. 10. Two frames prior to closure from the utterance /imi/. This frame illustrates lighting and threshold problems.

Fig. 11. One frame prior to closure from the utterance /imi/.

Fig. 12. The closure frame from the utterance /imi/.



2. Normalization

In order to compensate for differences in head position on different tapes and within a tape (due to head movement from frame to frame), a normalizing procedure was used. The computer program normalized the coordinates of each marker with respect to the calibration triangle (i.e. the two face markers on the glasses and the face marker on the nose). The dimensions of the triangle were unaffected by speech production and by head movement within the plane defined by the three points.

The calibration triangle was determined in the first field of each frame by calculating the distances between the nose and glasses face markers. Figure 13 shows how the origin of the new coordinate system is derived from the triangle and how the new (x',y') coordinates are obtained. This normalization compensated for any head movement within the plane defined by the three points of the triangle. Rotation of the head would involve movement in a third dimension (distance from camera) and could result in some error in the normalization procedure. Head movements were found to be small, however, with negligible rotation.

3. Tracking

Each eight second recording was reduced to 512 data files, one data file per field. Each data file consisted of a series of named arrays, each array corresponding to a face marker, with x and y coordinates of each marker in separate arrays. These arrays made it possible to track the values of a face marker from field to field.

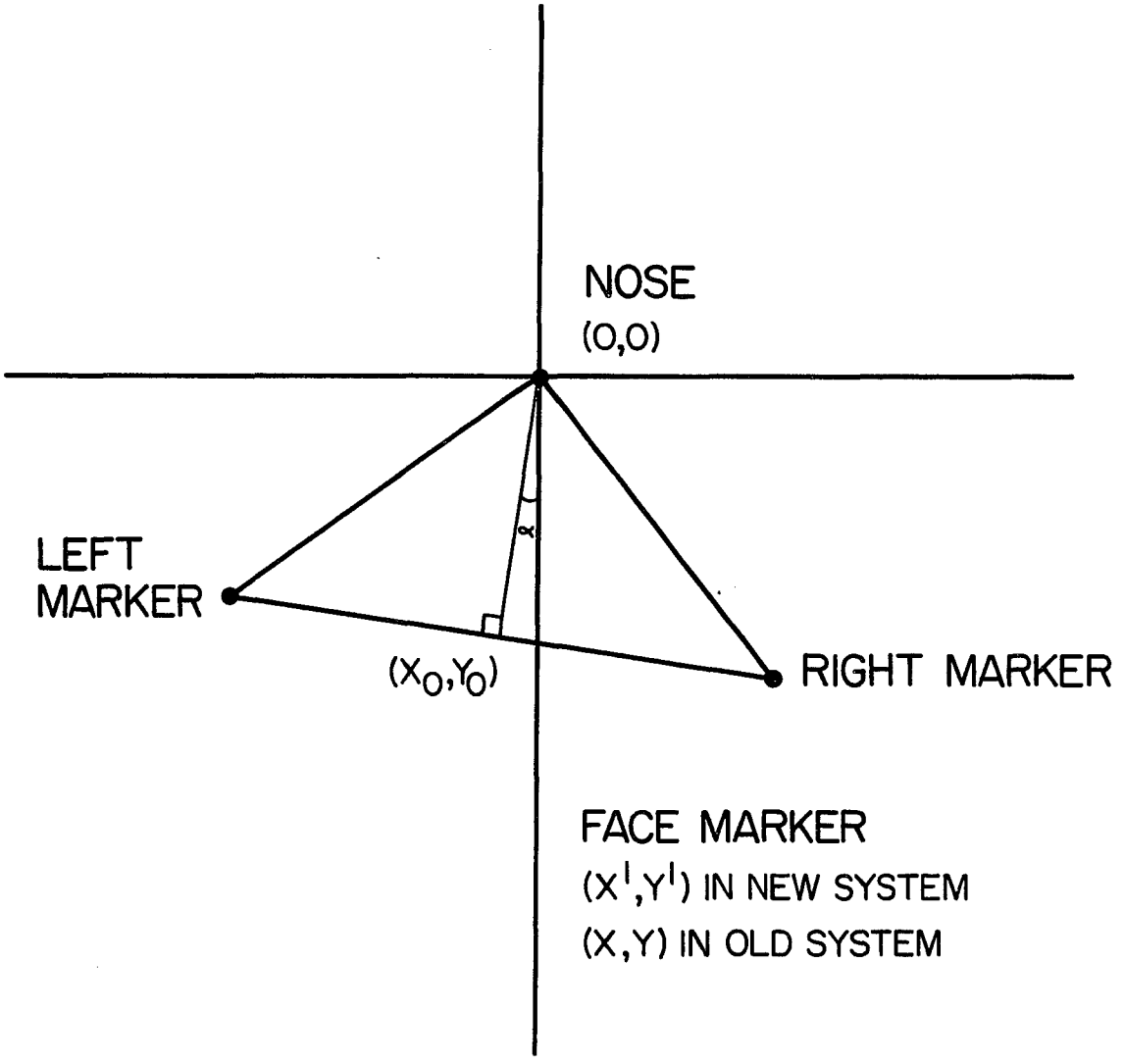
In order to track the facial movement during different utterances it was necessary to establish a time origin. This was done using the same procedure as in the interleaving study, i.e. that frame in which the first evidence of closure was apparent was defined as occurring at time=0. This point was used to establish a common time

Fig. 13. The calibration triangle that was used to derive the origin (x_o, y_o) of the new coordinate system.

$$x^1 = (x-x_o)\cos\propto + (y-y_o)\sin\propto$$

$$y^1 = (y-y_o)\cos\propto + (x-x_o)\sin\propto$$

$\propto$=angle between old and new y axis


NOSE
(0,0)
LEFT
MARKER
α
(X0,Y0)
RIGHT MARKER
FACE MARKER
(X',Y') IN NEW SYSTEM
(X,Y) IN OLD SYSTEM

referent for all utterances. The resolution of this measurement, as for any other in the time domain, was 1/60 of a second since the field rate was 60 fields per second.

Computer plots of the movement of markers were obtained for each VCV utterance, with field numbers along the abscissa and y coordinate values along the ordinate. The horizontal movement of the face markers was very small (typically less than 3 pels) so only the y coordinate values were plotted as a function of time. The y coordinates were plotted to the nearest pel.

Movement between fields was defined as a change in y value of at least two units. A difference in y coordinates of only one unit in adjacent fields may be the result of interlacing and not a result of any movement. For example, if field 2 has lines 10 and 12 (mean=11), and field 3 has lines 11 and 13 (mean=12), then, although movement did not take place, there is a difference of 1 pel.

Experiment 3: Perceptual Study

Speakers

The speakers for the perceptual study were the same as in the previous sections. One speaker (Speaker 1) was very well known to the observers and the other was not. Both speakers were easy to speechread as reported by several deaf children and adults.

Observers

The observers for the perceptual study were one male and two females. They ranged in age from 24 to 29 years. Hearing was within normal limits for all subjects. All observers had performed well on a previous sentence speechreading test and all three observers spoke the same New York City dialect as the speakers in the study.

Speech Materials

Visual discrimination of the homophenous sounds /p/, /b/ and /m/ was tested in a CV (consonant-vowel) context with the vowels /i/, /a/ and /u/. The VCV disyllables used in Experiments 1 and 2 were not used here since preliminary trials showed that the initial movement into the first vowel distracted the speechreader from focusing on the consonant. It was necessary to minimize distraction since the intention of the study was to measure discrimination under the best circumstances.

The speechreading test was divided into three parts. Each part tested the two-way discrimination between /m/ and /b/, /m/ and /p/, and /p/ and /b/ respectively. The two-way discriminations are referred to as consonant contrasts and are designated M-B, M-P and P-B.

There were six subtests per contrast. Each subtest tested one vowel per contrast for one speaker for a total of eighteen subtests (three vowels, three contrasts, two speakers). The subtests were four alternative forced-choice tests. There were twenty-five of each of four stimulus alternatives for a total of 100 randomized items.

The stimuli were CV pairs (e.g. /pa/-/ba/). For one contrast and one vowel there were four different CV pairs. For example, for M-P and /a/ the four alternative stimuli (known as targets) were /pa/-/pa/, /ba/-/ba/, /pa/-/ba/ and /ba/-/pa/. The set of four targets for each subtest is called a stimulus set. For the three contrasts and three vowels there were nine stimulus sets. (See table 7.)

The observers were told the stimulus set for each subtest and had to choose which of the four targets was presented at each trial. There were a total of 1800 stimulus items seen by each observer (three contrasts x three vowels x four targets x 25 repetitions x two speakers).

TABLE 7. Stimulus Sets for the Perceptual Study

	CONTRAST M-B	CONTRAST M-P	CONTRAST P-B
	/ba/-/ba/	/pa/-/pa/	/pa/-/pa/
/a/	/ma/-/ma/	/ma/-/ma/	/ba/-/ba/
	/ba/-/ma/	/pa/-/ma/	/pa/-/ba/
	/ma/-/ba/	/ma/-/pa/	/ba/-/pa/
	/bi/-/bi/	/pi/-/pi/	/pi/-/pi/
/i/	/mi/-/mi/	/mi/-/mi/	/bi/-/bi/
	/bi/-/mi/	/pi/-/mi/	/pi/-/bi/
	/mi/-/bi/	/mi/-/pi/	/bi/-/pi/
	/bu/-/bu/	/pu/-/pu/	/pu/-/pu/
/u/	/mu/-/mu/	/mu/-/mu/	/bu/-/bu/
	/bu/-/mu/	/pu/-/mu/	/pu/-/bu/
	/mu/-/bu/	/mu/-/pu/	/bu/-/pu/

Video Recordings

The speakers were instructed to produce a series of 100 stimuli, saying each utterance twice (e.g. /m/-/pi/, /mi/-/pi/). There was a simultaneous audio recording in order to check the accuracy of the stimulus items. A subsequent analysis indicated that errors were made by both speakers, however, this was not a serious problem since there were a sufficient number of stimuli for every target category.

Six test tapes were prepared, one tape per vowel per speaker. Each tape contained three subtests, corresponding to the three contrasts (M-B, M-P, P-B) for a total of 18 recorded subtests. This is shown in Table 8.

During the recording of one subtest, twelve stimuli were lost due to an electrical malfunction. Unfortunately it was not possible to use the studio again to re-record the tape. Appendix A shows the actual number of stimuli recorded for each target in each subtest.

Apparatus

The video tapes for the perceptual study were prepared in a video recording studio under standard studio light. This studio was not the same as the one used for Experiments 1 and 2 since this experiment did not require computer processing.

A Sony video camera on external synchronization was used to record the utterances. The frame rate was 30 frames per second with interlace. The resolution was 525 lines.

The output of the camera went to a distribution amplifier. The output of the distribution amplifier and synchronization pulses went to a Dynair video and synchronization mixer. The output of the mixer (the video signal) was recorded on a Sony 1/2" reel to reel video recorder. The video signal was monitored on a Hewlett-Packard 6947A raster display.

TABLE 8. Test Tapes Used in the Perceptual Study

TAPE	VOWEL	ORDER OF CONTRASTS	SPEAKER
1	/a/	M-P, P-B, M-B	1
2	/i/	P-B, M-B, M-P	1
3	/u/	M-B, M-P, P-B	1
4	/a/	M-P, M-B, P-B	2
5	/i/	P-B, M-P, M-B	2
6	/u/	M-B, P-B, M-P	2

Since the observers were tested in a different laboratory, the video tapes were shown using a JVC video reproducer and video monitor (model TM 900).

Method

The test recordings were shown to the observers in random order. All recordings involving Speaker 1 were shown prior to those for Speaker 2.

The observers viewed the video recordings from a distance which they found comfortable. This was 26" for R.S., 36" for C.K. and 24" for L.S.

The subjects were required to write their response and could do so after the first presentation, however for most test items observers responded only after the repetition. The observers were reminded of the four targets at the start of the subtest.

Feedback on which was the correct response was provided after each test item. During the test, when observers were correct, they frequently reported the facial cue that they used to arrive at a choice.

Observers were briefly trained in the task by viewing the first 20 to 40 items of the first subtest seen by each observer. These items were then repeated as part of the regular test session, however this did not have an appreciable effect on the test results. The average score for the repeated items was .56 whereas the score for the same number of items that were not seen in practice (from the same subtest) was .55.

CHAPTER IV

RESULTS

Experiment 1: Interleaving

Each interleaved sample was viewed for each of the speakers and areas of jitter were noted. These observations are listed in tables 9 and 10. Note that in several samples showing jitter (e.g. /pa/-/ba/ for speaker 2), the jitter disappeared when the number of frames was reduced from 10 or 11 to 5. These first few frames consisted of the consonant closure and initial opening of the vowel.

For both speakers jitter was observed primarily in the cheeks, lower lip and chin.

Tables 11 and 12 show the occurrence of jitter as a function of the CV-CV contrast. The position of the CV's in the triplets is indicated. For speaker 2 all interleaved samples showed jitter for contrasting consonants. When two different productions of the same consonant were interleaved (table 12), there was no jitter for all samples for Speaker 1. For Speaker 2 there were 7 cases of no jitter, 1 case that was unclear due to head movement and four instances of some jitter. Tables 11 and 12 do not include the interleaved samples where the consonant was the same and the vowels were different (e.g. /pi/-/pu/). All seven interleaved samples of this type showed jitter.

Experiment 2: Measurement of Face Markers

A plot of the face marker movement was generated for each VCV in a triplet. In addition to these individual plots, the three plots per triplet were superimposed (by computer) using a common time origin as defined previously (i.e. t=0 corresponds to

TABLE 9. Jitter Observations for Speaker 1

TABLE 9. Jitter Observations for Speaker 1

INTERLEAVED SAMPLE	NUMBER OF FRAMES	LOCATION AND EXTENT OF JITTER
pi-bi	7	no jitter; pi more rounded than bi
pi-bi	11	lower lip; can remove by mismatching sequences in time
pi-bi	7	no jitter; pi more rounded than bi
pi-bi	3	no jitter
pi-bi	3	no jitter
pi-bi	3	no jitter
pi-bi	4	no jitter
pi-mi	21	jitter, not localized; pi more rounded than mi
pi-pi	9	no jitter; excessive head movement
bi-bi	7	no jitter; excessive head rotation
bi-bi	6	no jitter; difference in lip shapes between sequences
pa-ba	10	possibly lower lip; excessive head movement
ba-ma	8	possibly lower lip; excessive movement
ma-ma	6	no jitter; excessive sideways head movement
pa-pa	8	excessive head movement
pa-pa	7	no jitter
pu-mu	6	possibly cheek
pu-bu	8	possibly cheek
pu-bu	16	excessive head movement
mu-mu	5	no jitter
pa-pu	8	mouth area

Table 9 (cont'd)

INTERLEAVED SAMPLE	NUMBER OF FRAMES	LOCATION AND EXTENT OF JITTER
pi-pu	8	cheek, between mouth and chin, corner of upper lip
pi-pa	8	lower lip
bi-ba	8	cheek, slight jitter lower lip and jaw

TABLE 10. Jitter Observations for Speaker 2

INTERLEAVED SAMPLE	NUMBER OF FRAMES	LOCATION AND EXTENT OF JITTER
pi-bi	10	lower lip, chin
pi-mi	10	cheek (right of lip corner)
pi-mi	9	cheek, slight jitter
bi-mi	12	lower lip, chin
bi-mi	10	lower lip, chin; slight jitter
bi-mi	10	between mouth corner and chin, teeth; lip, possible jitter
pi-pi	11	no jitter
pi-pi	11	slight jitter, not localized; no jitter for 5 frames of same sample
bi-bi	10	no jitter
mi-mi	10	jitter, possibly due to slight head rotation
pa-ma	12	jitter, not localized; no jitter for 5 frames of same sample
pa-ma	8	cheek, slight jitter
pa-ma	9	cheek, slight jitter
pa-ba	10	possibly lower lip; excessive head movement
pa-ba	8	slight jitter, not localized
pa-ba	9	jaw, lower lip, chin
pa-ba	10	jitter, not localized; no jitter for 5 frames of same sample
ba-ma	9	jaw, lower lip, chin
pa-pa	11	slight jitter, not localized; no jitter for 5 frames of same sample
pa-pa	11	no jitter
ba-ba	10	no jitter

Table 10 (cont'd)

INTERLEAVED SAMPLE	NUMBER OF FRAMES	LOCATION AND EXTENT OF JITTER
ba-ba	9	excessive head movement, possibly no jitter
ma-ma	9	upper lip
ma-ma	8	no jitter
pu-mu	6	cheek
pu-bu	7	cheek, chin, between mouth corner and chin
pu-bu	6	cheek, to right of lip corner, slight jitter
pu-bu	6	cheek, to right of lip corner, slight jitter
bu-mu	6	cheek
mu-mu	8	no jitter
bu-bu	8	cheek
pa-pu	8	lower lip, chin, between mouth corner and chin, between nose and mouth corner,upper cheek
pa-pu	6	lower lip, chin, between mouth corner and chin, between nose and mouth corner, upper cheek
pi-pu	8	cheek primarily, some jitter in same pattern as pa-pu
pi-pa	8	lower lip, jaw under chin

TABLE 11. Number of Interleaved Samples with Contrasting Consonants Showing Jitter (J) or No Jitter (N). The number of samples that were Unclear (U) due to excessive head movement is indicated.

TRIPLET	INTERLEAVED SAMPLES								
	bi-pi			pi-mi			bi-mi		
	J	N	U	J	N	U	J	N	U
ibi ipi imi									
speaker 1	1	6		1					
speaker 2	1			1				2	
ipi imi ibi									
speaker 2					1		1		
proportion of samples showing jitter	2/8			2/3			3/3		
	ba-pa			pa-ma			ba-ma		
	J	N	U	J	N	U	J	N	U
aba apa ama									
speaker 1			1						1
speaker 2	1			1			1		
ama aba apa									
speaker 2	1		1	1					

Table 11 (cont'd)

TRIPLET	INTERLEAVED SAMPLES								
	ba-pa			pa-ma			ba-ma		
	J	N	U	J	N	U	J	N	U
apa ama aba									
speaker 2	1			1					
proportion of samples showing jitter	3/5			3/3			1/2		
	bu-pu			pu-mu			bu-mu		
ubu upu umu	J	N	U	J	N	U	J	N	U
speaker 1		1			1				
umu ubu upu									
speaker 2	1			1			1		
proportion of samples showing jitter	1/2			1/2			1/1		

TABLE 12. Number of Interleaved Samples with the Same Consonant Showing Jitter (J) or No Jitter (N). The number of samples that were Unclear (U) due to excessive head movement is indicated.

TRIPLET	INTERLEAVED SAMPLES								
	pi-pi			bi-bi			mi-mi		
	J	N	U	J	N	U	J	N	U
ibi ipi imi									
speaker 2									1
ipi imi ibi									
speaker 2		1			1				
ipi apa upu									
speaker 1		1							
speaker 2	1								
ibi aba ubu									
speaker 1					2				
proportion of samples showing no jitter	2/3			3/3			0/1		
	pa-pa			ba-ba			ma-ma		
	J	N	U	J	N	U	J	N	U
aba apa ama									
speaker 1								1	
speaker 2					1			1	
ama aba apa									
speaker 2					1				
apa ama aba									
speaker 2		1							1

Table 12 (cont'd)

TRIPLET	INTERLEAVED SAMPLES		
ipi apa upu			
speaker 1	1		
speaker 2	1		
proportion of samples showing no jitter	2/3	2/2	2/3
	pu-pu	bu-bu	mu-mu
ubu upu umu	J N U	J N U	J N U
speaker 1			1
umu ubu upu			
speaker 2		1	1
proportion of samples showing no jitter		0/1	2/2

that frame in which the first evidence of closure was apparent.) Three superimposed plots can be seen in figure 14.

Figure 14 shows the vertical movement of the lower lip. The abscissa represents time. The ordinate represents distance in pels in the digitized picture. The region of closure for the consonant is the upper part of each curve. Note that each division on the ordinate represents one pel or approximately 0.5 mm on the face. The vertical scale is approximately six times larger than life-size (i.e. 0.5 mm is represented by 3.0mm on the graph).

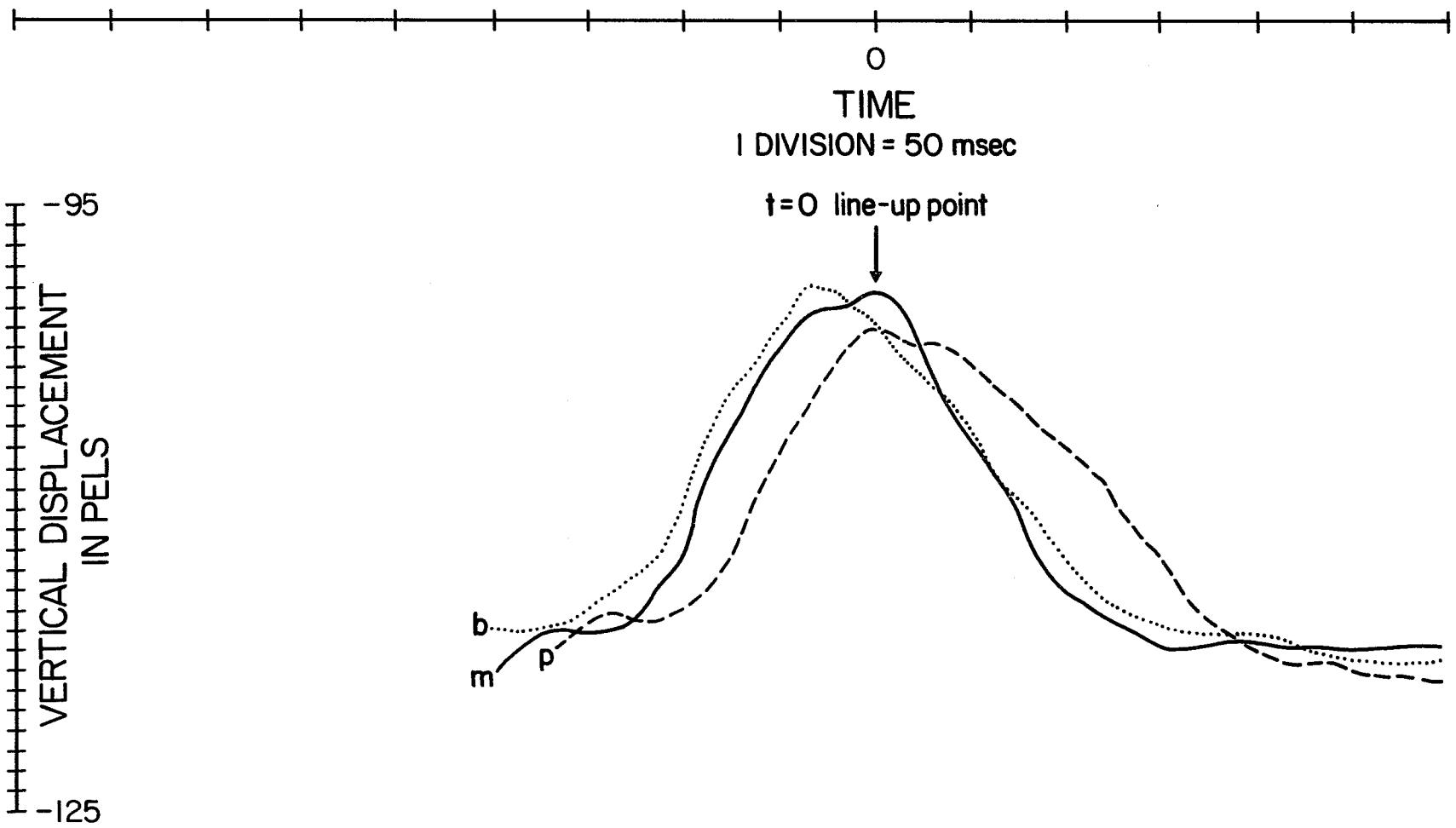
The face marker movement was analyzed by visually comparing pairs of individual plots. For each replication of a triplet (VC_1V VC_2V VC_3V) there were three two-way comparisons: VmV vs. VbV, VmV vs. VpV, VpV vs. VbV.

A total of 2,538 plots were obtained. In order to reduce this large body of data, those plots unlikely to show significant effects were eliminated. To begin with, plots were not considered for further analysis if the total movement was less than three pels. The precision of measurement was only one pel and it was not possible to make meaningful comparisons for movements of less than three pels. These eliminated plots were for the face markers LUCH, LLCH, RUCH and RLCH. All of these face markers were located on the cheeks, above the level of the lower lip.

A second group of face markers was eliminated because most of the two-way comparisons showed differences of less than 3 pels between corresponding points on the plots, even though the face markers did show movement. The face markers in this category are RNS, RFS, LNS, LFS, RCR, LCR, RJAW, LJAW, RB, LB.

Of the remaining face markers (LULP, RULP, MULP, LLLP, RLLP, MLLP, CHIN), visual inspection of the superimposed plots showed that the three face markers in midline locations showed the greatest differences. In addition, one off center face marker on the lower lip was chosen since the lower lip showed the most movement in

Fig. 14. An example of the vertical movement of the mid lower lip (MLLP) face marker for /p/, /b/ and /m/ utterances. The pel scale shows the displacement relative to the normalized origin (0,0). The normalized origin is described in the Normalization section of the Procedures chapter. Note that one pel is approximately equal to 0.5 mm.


0
TIME
1 DIVISION = 50 msec
t=0 line-up point
-95
VERTICAL DISPLACEMENT
IN PELS
-125
b
p
m

the jitter experiment. The data for this face marker were analyzed separately and are shown in the appendix. The four face markers that were analyzed in detail were: mid upper lip (MULP), mid lower lip (MLLP), left lower lip (LLLP) and chin (CHIN).

The basic curve shape observed for the lower lip and chin face markers (MLLP and CHIN) was a bell. Figure 15 shows typical /p/ and /b/ curves for MLLP. For the upper lip face marker (MULP) the basic curve shape was a trough (i.e. an inverted bell). Figure 16 shows typical curves for MULP. The general shape of the plots was similar for p, b and m (as in figure 14). There were slight vowel differences in that the plots were generally flatter for /u/ than for /i/ and /a/.

Three parameters of the curves were of interest. These were height, width and slope of the bell or trough curves. Differences between pairs of plots (i.e. /p/ vs. /b/, /p/ vs. /m/, /b/ vs. /m/) were measured by comparing these parameters. The bell height indicated upward movement of the lower lip and chin. Trough height indicated downward movement of the upper lip. Slope corresponded with the rate of movement of the lips and chin before and after closure. Left and right slope referred to movement towards and away from closure respectively. Width correlated with the duration of closure.

It was necessary to define the height, width and slope for these curves so that the pairs of plots could be compared in a consistent manner. In order to define the height it should be noted that the vertical face marker movement was plotted in terms of normalized coordinates. As described in the Normalization section of the Procedure Chapter, the coordinates of the face markers were normalized to a triangle formed between two face markers on the glasses and a face marker on the nose. The dimensions of this triangle were unaffected by both speech production and head movement within the plane defined by the three points. Any changes in the plane

Fig. 15. An example of /p/ and /b/ curves for the mid lower lip face marker (vowel /a/). Point H indicates the height of the curves. Displacement is measured relative to the normalized origin.

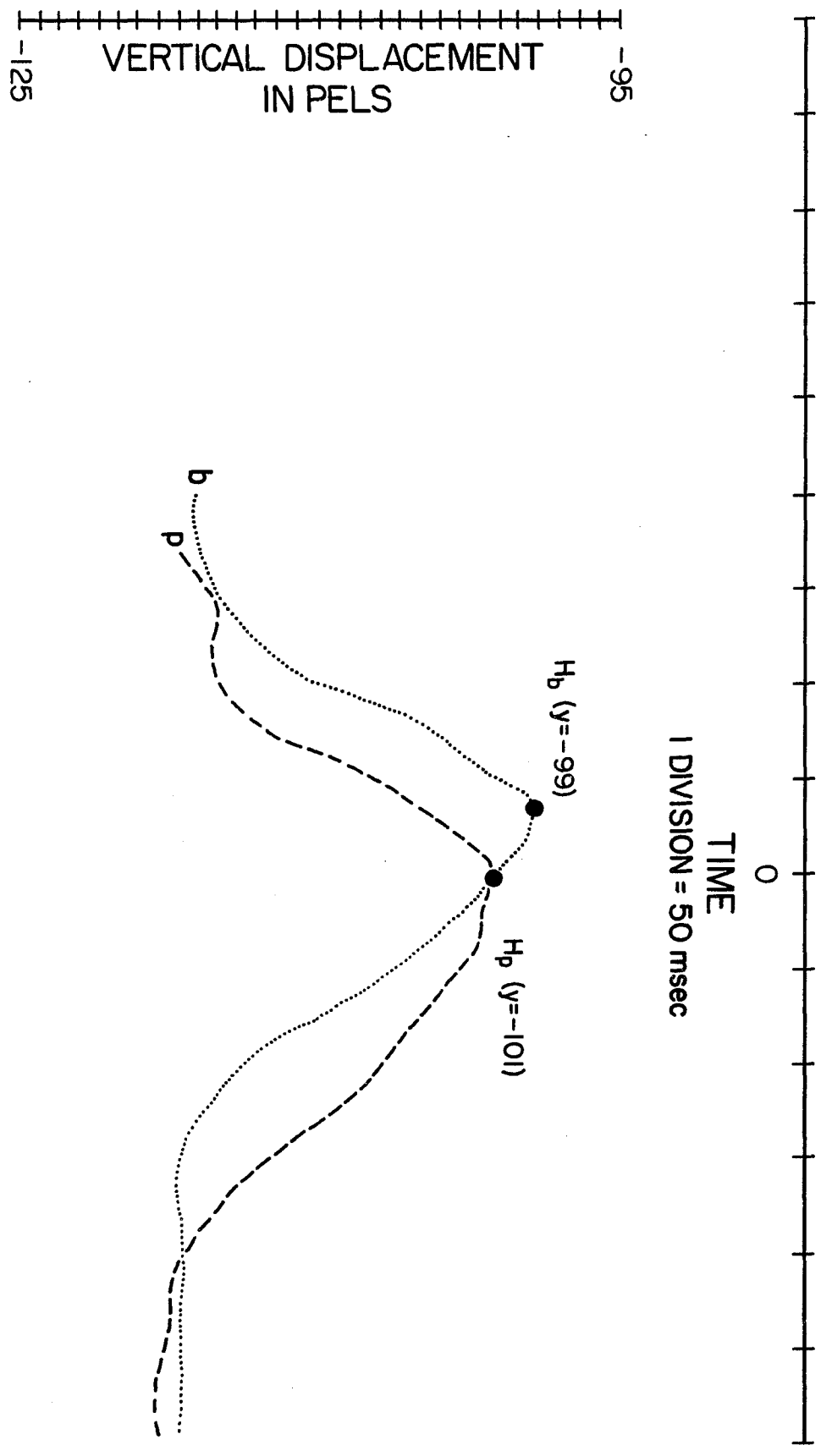
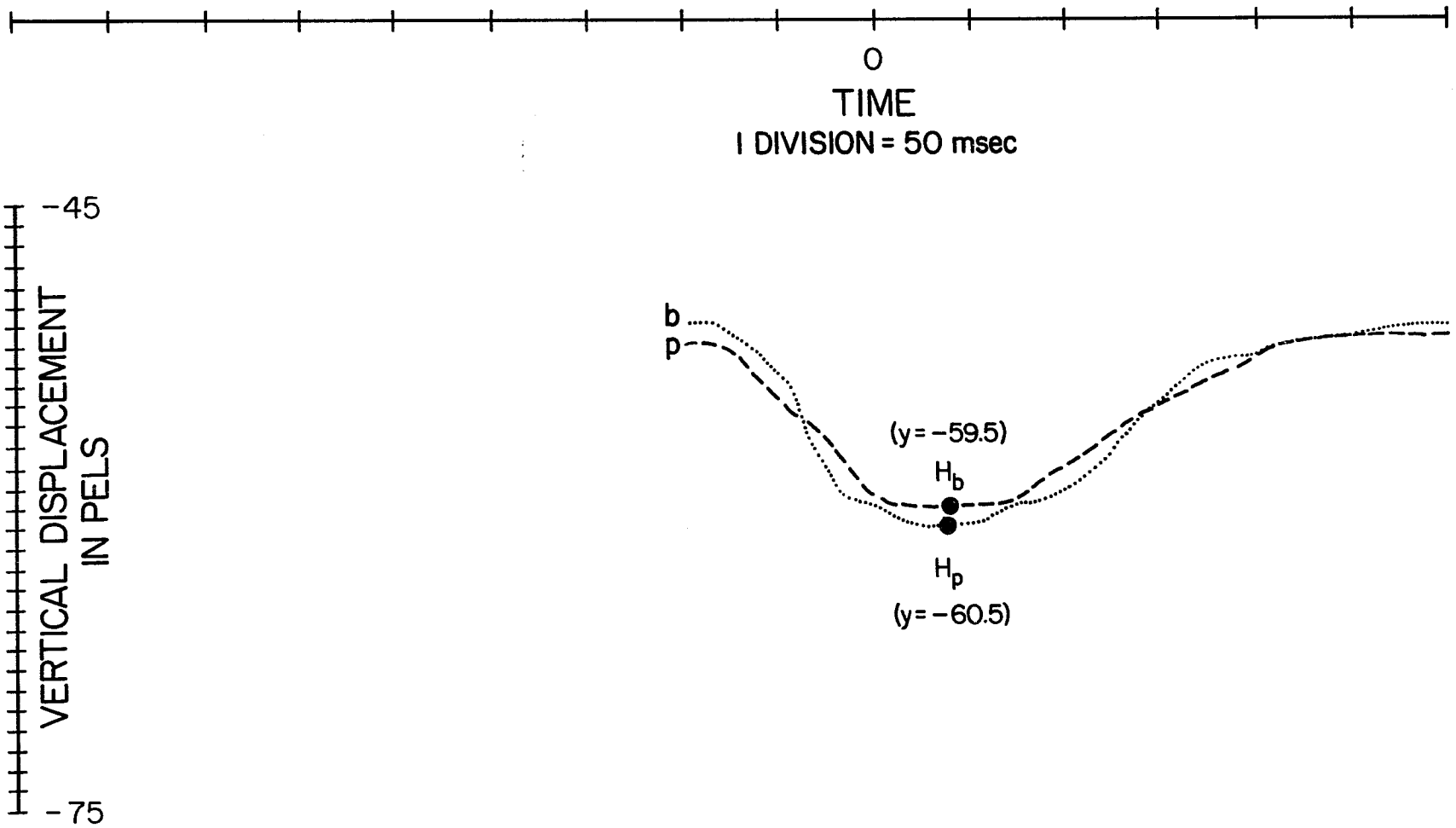

-125
VERTICAL DISPLACEMENT
IN PELS
-95
b
p
H_p (y=-99)
H_p (y=-101)
0
TIME
1 DIVISION = 50 msec

Fig. 16. An example of /p/ and /b/ curves for the mid upper lip face marker (vowel /i/). Point H indicates the height of the curves in pels as measured from the normalized origin.


0
TIME
1 DIVISION = 50 msec
-45
VERTICAL DISPLACEMENT
IN PELS
-75
b
p
(y=-59.5)
H_b
H_p
(y=-60.5)

defined by these three points were used to adjust the measured face markers. In doing so, a normalized coordinate system was set up.

Height was defined as the most extreme point on the curve (see figures 15 and 16) in the vicinity of closure. This corresponds to point H for each curve in figures 15 and 16. The heights of two curves were compared by comparing the values of the y coordinates at point H on each curve. Therefore, in figure 15, the b curve is higher than the p curve since the y value of point H_b(-59.5) is greater than the y value of point H_p (-60.5).

The slope was determined by fitting a straight line to each side of each curve. Each line was terminated at both ends where it deviated from the curve by more than three pels. Figure 17 shows the slope of the trough and figure 18 shows the slope of the bell.

The width was defined as the distance between the midpoints of the two lines used to determine the slopes for each curve. Figures 17 and 18 shows how width was determined.

The comparisons of interest were those considered in the perceptual study. In order to obtain data consistent with the perceptual study, only the direction of a difference was measured. This was done by tabulating which curve in each comparison showed the greater height, width or slope.

The data are shown in tables 13 through 18. Each table represents one curve comparison (slope, width, height) and shows the results for all three contrasts (M-B, M-P, P-B) and vowels (i, a u). The data are presented summed over the speakers since there were few differences between the speakers. Data are shown for the face markers MULP, MLLP and CHIN.

Note that only the left slope of the trough and the right slope of the bell curves were analyzed in detail. Both right and left slope comparisons were done for every pair

Fig. 17. This curve is an example of the mid upper lip face marker, /p/, vowel /i/. The slope was determined by fitting a straight line to each side of the curve. The line was terminated where it deviated from the curve by more than three pels. The dashed lines indicate three pel distances from the curve.

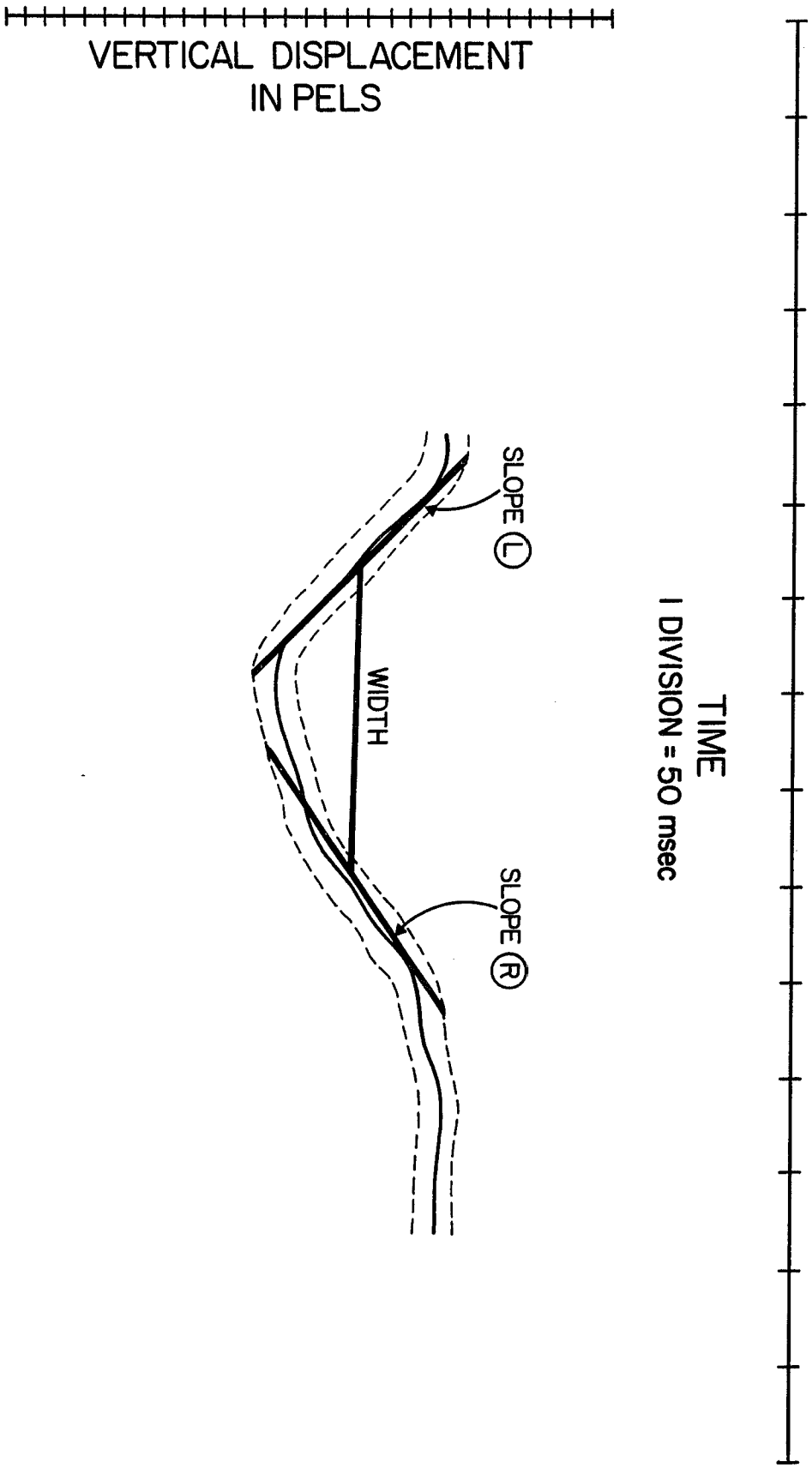
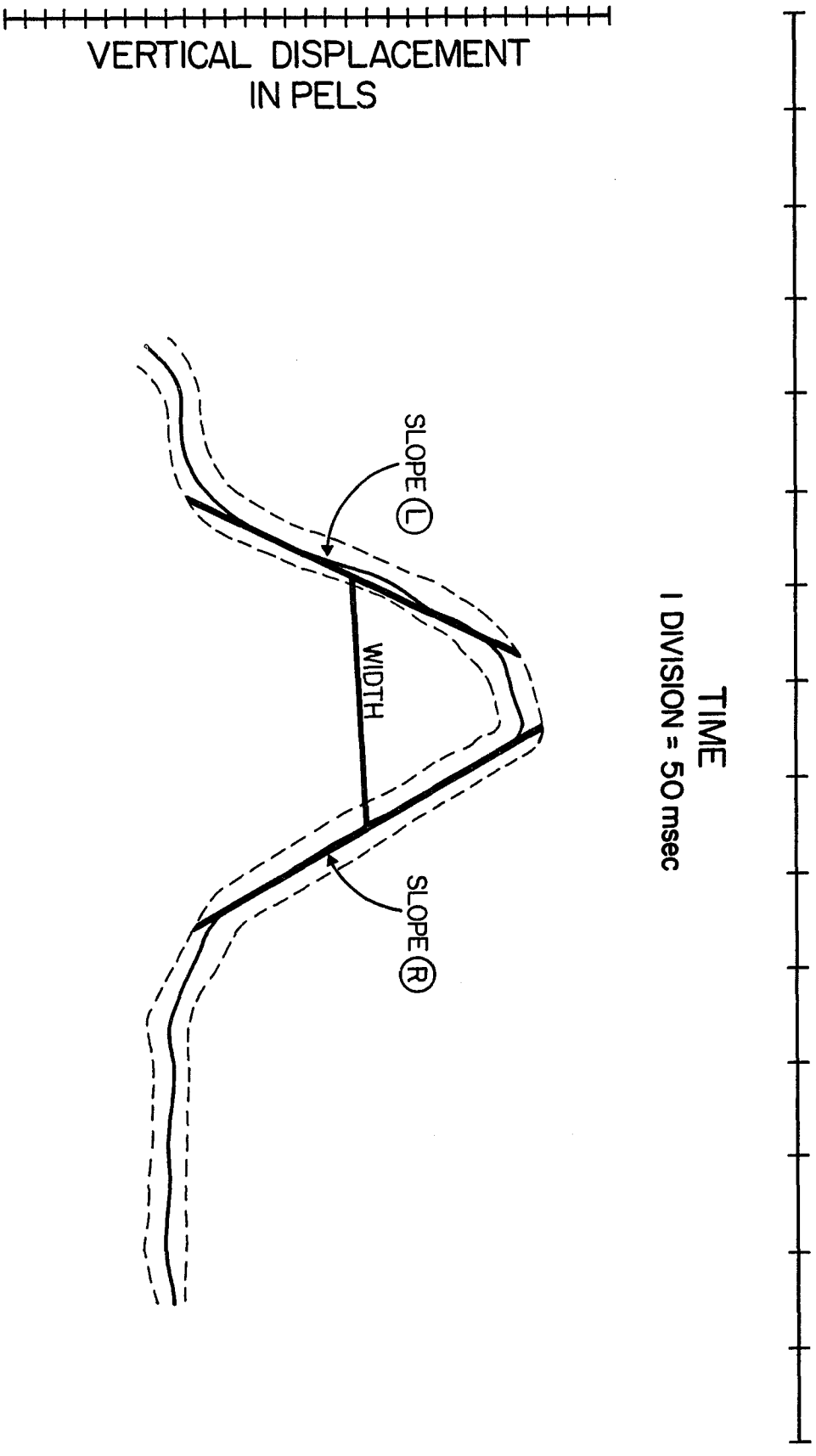

VERTICAL DISPLACEMENT
IN PELS
SLOPE (L)
WIDTH
SLOPE (R)
TIME
1 DIVISION = 50 msec

Fig. 18. This curve is an example of the mid lower lip face marker, /m/, vowel /a/. The slope was determined by fitting a straight line to each side of the curve. The line was terminated where it deviated from the curve by more than three pels. The dashed lines indicate three pel distances from the curve.


VERTICAL DISPLACEMENT
IN PELS
SLOPE (L)
WIDTH
SLOPE (R)
TIME
1 DIVISION = 50 msec

of curves however the comparisons indicated that the left slope measure of the trough showed greater /p/, /b/, /m/ differences than the right slope comparison. For the bell curves the right slope comparisons showed greater consonant differences than the left slope.

In tables 13 through 18 the notation /m/>/b/ means that the slope for an /m/ curve was greater (i.e. steeper) than the slope for a /b/ curve. The proportion of cases where the parameter of interest (e.g. slope) was greater is indicated. (For example, in table 13 the first entry in the first column indicates that in 0.11 of the 18 cases the m curve was steeper than the b curve.)

The tables indicate when the p,b,m differences were significant i.e. when the number of cases of /m/>/b/, for example, is significantly greater than the number of cases of /b/>/m/. A proportion was calculated by dividing the number of times that /m/ was greater than /b/ by the total number of comparisons. This proportion was then subtracted from the corresponding proportion for /b/>/m/. The difference was divided by the standard deviation of the difference. The quotient was compared to the normal distribution, for two tails, at both the .05 and .01 levels of significance.

A summary of the results of the paired comparisons is shown in table 19. The asterisks indicate which curve comparisons were significant for each face marker, consonant contrast and vowel.

The data will first be considered in terms of the observable differences that did occur. Then, the pattern for the vowels will be described.

The results for contrasts M-B and M-P were very similar. This finding is illustrated in table 20 which shows those cases in which a consistent difference occurred. The "m" entries in table 20 indicate that /m/>/b/ for the indicated slope, height or width comparison of /m/ and /b/ curves. Likewise, "b" entries indicate /b/>/m/. The same type of notation was used in subsequent tables. That is, the entry

TABLE 13. Incidence of Steeper Left Slope for /m/ vs. /b/, /m/ vs. /p/ and /p/ vs. /b/ Plot Comparisons of the Mid Upper Lip Face Marker

	MID UPPER LIP		
	/i/	/a/	/u/
Contrast M-B			
/m/>/b/	.11	.25	0
/m/=/b/	.44	.44	.64
/b/>/m/	.44*	.31	.36*
n	18	16	11
Contrast M-P			
/m/>/p/	.28	.06	.09
/m/=/p/	.44	.36	.45
/p/>/m/	.28	.63**	.45*
n	18	16	11
Contrast P-B			
/p/>/b/	.22	.56**	.18
/p/=/b/	.39	.50	.73
/b/>/p/	.39	.06	.09
n	18	16	11

*difference significant at .05 level

**difference significant at .01 level

TABLE 14. Incidence of Steeper Right Slope for /m/ vs. /b/, /m/ vs. /p/ and /p/ vs. /b/ Plot Comparisons of the Mid Lower Lip and Chin Face Markers

	Mid Lower Lip			Chin		
	/i/	/a/	/u/	/i/	/a/	/u/
Contrast M-B						
/m/>/b/	.63**	.79**	.45	.32	.50**	.44**
/m/=/b/	.37	.14	.18	.53	.43	.56
/b/>/m/	0	.07	.36	.16	.07	0
n	19	14	11	19	14	9
Contrast M-P						
/m/>/p/	.68**	.79**	.55	.42	.50**	.78**
/m/=/p/	.32	.14	.27	.37	.43	.11
/p/>/m/	0	.07	.18	.21	.07	.11
n	19	14	11	19	14	9
Contrast P-B						
/p/>/b/	.26	.29	.18	.21	.36	.22
/p/=/b/	.42	.43	.45	.58	.43	.33
/b/>/p/	.32	.29	.36	.21	.21	.44
n	19	14	11	19	14	9

*difference significant at .05 level

**difference significant at .01 level

TABLE 15. Incidence of Wider Trough for /m/ vs. /b/, /m/ vs. /p/ and /p/ vs. /b/ Plot Comparisons of the Mid Upper Lip Face Marker

	MID UPPER LIP		
	/i/	/a/	/u/
Contrast M-B			
/m/>/b/	.67**	.19	.27
/m/=/b/	.17	.50	.55
/b/>/m/	.17	.31	.18
n	18	16	11
Contrast M-P			
/m/>/p/	.61*	.56*	.64**
/m/=/p/	.17	.25	.27
/p/>/m/	.22	.19	.09
n	18	16	11
Contrast P-B			
/p/>/b/	.61**	.19	.09
/p/=/b/	.17	.25	.64
/b/>/p/	.22	.56*	.27
n	18	16	11

*difference significant at .05 level

**difference significant at .01 level

TABLE 16. Incidence of Wider Bell for /m/ vs. /b/, /m/ vs. /p/ and /p/ vs. /b/ Plot Comparisons of the Mid Lower Lip and Chin Face Markers

	Mid Lower Lip			Chin		
	/i/	/a/	/u/	/i/	/a/	/u/
Contrast M-B						
/m/>/b/	.42	.14	.45	.26	.14	.22
/m/=/b/	.21	.29	.45	.26	.21	.22
/b/>/m/	.37	.57**	.09	.47	.64**	.56
n	19	14	11	19	14	9
Contrast M-P						
/m/>/p/	.58	.36	.36	.32	.21	.11
/m/=/p/	.05	0	.36	.26	.29	.33
/p/>/m/	.37	.64	.27	.42	.50	.56*
n	19	14	11	19	14	9
Contrast P-B						
/p/>/b/	.32	.29	.36	.11	.14	.22
/p/=/b/	.05	.14	.36	.42	.14	.56
/b/>/p/	.47	.57	.27	.47**	.71**	.22
n	19	14	11	19	14	9

*difference significant at .05 level

**difference significant at .01 level

TABLE 17. Incidence of Higher Trough for /m/ vs. /b/, /m/ vs. /p/ and /p/ vs. /b/ Plot Comparisons of the Mid Upper Lip Face Marker

	MID UPPER LIP		
	/i/	/a/	/u/
Contrast M-B			
/m/>/b/	.50*	.69**	.55**
/m/=/b/	.39	.25	.45
/b/>/m/	.11	.06	0
n	18	16	11
Contrast M-P			
/m/>/p/	.67**	.94**	.91**
/m/=/p/	.28	.06	.09
/p/>/m/	.05	0	.0
n	18	16	11
Contrast P-B			
/p/>/b/	.33	0	0
/p/=/b/	.28	.31	.45
/b/>/p/	.39	.69**	.55**
n	18	16	11

*difference significant at .05 level

**difference significant at .01 level

TABLE 18. Incidence of Higher Bell for /m/ vs. /b/, /m/ vs. /p/ and /p/ vs. /b/ Plot Comparisons of the Mid Lower Lip and Chin Face Markers

	Mid Lower Lip			Chin		
	/i/	/a/	/u/	/i/	/a/	/u/
Contrast M-B						
/m/>/b/	.63**	.79**	.82**	.47	.79**	.33
/m/=/b/	.11	.14	0	.26	.14	.44
/b/>/m/	.26	.07	.18	.26	.07	.22
n	19	14	11	19	14	9
Contrast M-P						
/m/>/p/	.74**	.86**	1.00**	.53**	.71*	.78**
/m/=/p/	.21	0	0	.37	0	.22
/p/>/m/	.05	.14	0	.11	.29	0
n	19	14	11	19	14	9
Contrast P-B						
/p/>/b/	.16	.29	.09	.32	.36	.11
/p/=/b/	.37	.29	.27	.32	.21	.11
/b/>/p/	.47*	.43	.64**	.37	.43	.78**
n	19	14	11	19	14	9

*difference significant at .05 level

**difference significant at .01 level

TABLE 19. Summary of the Occurrence of Significant Differences

	left slope	right slope	bell/trough width	bell/trough height
CONTRAST M-B				
mid upper lip				
/i/	b>m*	n.e.	m>b**	m>b*
/a/		n.e.		m>b**
/u/	b>m*	n.e.		m>b**
mid lower lip				
/i/	n.e.	m>b**		m>b**
/a/	n.e.	m>b**	b>m**	m>b**
/u/	n.e.			m>b**
chin				
/i/	n.e.			
/a/	n.e.	m>b**	b>m**	m>b**
/u/	n.e.	m>b**		
CONTRAST M-P				
mid upper lip				
/i/		n.e.	m>p*	m>p**
/a/	p>m**	n.e.	m>p*	m>p**
/u/	p>m*	n.e.	m>p**	m>p**
mid lower lip				
/i/	n.e.	m>p**		m>p**
/a/	n.e.	m>p**		m>p**
/u/	n.e.			m>p**

TABLE 19. Continued

	left slope	right slope	bell/trough width	bell/trough height
chin				
/i/	n.e.			m>p**
/a/	n.e.	m>p**		m>p*
/u/	n.e.	m>p**	p>m*	m>p**
CONTRAST P-B				
mid upper lip				
/i/		n.e.	p>b**	
/a/	p>b**	n.e.	b>p*	b>p**
/u/		n.e.		b>p**
mid lower lip				
/i/	n.e			b>p*
/a/	n.e.			
/u/	n.e.			b>p**
chin				
/i/	n.e.		b>p**	
/a/	n.e.		b>p**	
/u/	n.e.			b>p**

* differences significant at .05 level

** differences significant at .01 level

n.e. not examined in detail (Preliminary analysis did not indicate a consistent pattern.)

TABLE 20. Direction of Observable Differences for Contrasts M-B, M-P and P-B. The notation m indicates m>b or m>p, b indicates b>m or b>p, and p indicates p>m or p>b.

	Contrast M-B			Contrast M-P			Contrast P-B		
	/i/	/a/	/u/	/i/	/a/	/u/	/i/	/a/	/u/
<u>mid upper lip</u>									
higher trough	m	m	m	m	m	m		b	b
wider trough	m			m	m		p	b	
steeper left slope	b		b		p	p		p	
<u>mid lower lip</u>									
higher bell	m	m	m	m	m	m	b		b
wider bell		b							
steeper right slope	m	m		m	m				
<u>chin</u>									
higher bell		m		m	m	m			b
wider bell		b				p	b	b	
steeper right slope		m	m		m	m			

indicates which consonant in the two-way comparison shows the particular curve characteristic to a greater extent (e.g. steeper slope).

For the upper lip, /m/ was greater than /p/ or /b/ for height and width and /b/>/m/ and /p/>/m/ for the left slope. For the lower lip and chin, for both right slope and height, /m/ was greater than /p/ or /b/, but for width /b/>/m/ and /p/>/m/.

Table 20 also shows the results for the P-B contrast. For the upper lip /b/ was greater than /p/ for trough height. For the lower lip /b/ was greater than /p/ for bell height. For the chin /b/ was greater than /p/ for bell height and width. Note that for many of the P-B comparisons there were no observable differences (shown by no entry in the table).

In all three contrasts, the greatest number of observable differences occurred for the vowel /a/. Observable differences occurred with similar frequency for /i/ and /u/.

Table 21 presents the pattern of observable differences with respect to height, slope and width. In contrast P-B there were no observable differences for /u/ for right/left slope or bell/trough width.

For bell height, M-B and M-P showed observable differences for /a/ for all face markers, but P-B did not show any differences.

For bell/trough height and /u/ there were observable differences for all face markers and all three contrasts, with the exception of CHIN in contrast M-B.

There were no observable differences for /u/ for right slope for all three contrasts, with the exception of observable differences for CHIN for M-B and M-P.

For bell width there were no observable differences for /i/ for M-B and M-P for both face markers.

TABLE 21. Direction of Observable Differences for Height, Slope and Width Measures of the Mid Upper Lip, Mid Lower Lip and Chin Face Marker Plots. The notation m indicates m>b or m>p, b indicates b>m or b>p, and p indicates p>m or p>b.

	Contrast M-B			Contrast M-P			Contrast P-B		
	/i/	/a/	/u/	/i/	/a/	/u/	/i/	/a/	/u/
TROUGH HEIGHT									
mid upper lip	m	m	m	m	m	m		b	b
BELL HEIGHT									
mid lower lip	m	m	m	m	m	m	b		b
chin		m		m	m	m			b
LEFT SLOPE									
mid upper lip	b		b		p	p		p	
RIGHT SLOPE									
mid lower lip	m	m		m	m				
chin		m	m		m	m			
TROUGH WIDTH									
mid upper lip	m			m	m	m	p	b	
BELL WIDTH									
mid lower lip		b							
chin		b					p	b	b

Table 22 shows the observable differences when the data are summed over vowels and speakers. For MLLP right slope and bell height. For P-B the results were b>p for right slope and bell height. There were no observable differences for bell width for MLLP.

Figures 19-22 show examples of the m, p and b differences in slope, width and height. Figure 19 shows m>p for trough width for MULP for /i/ (i.e. the trough for m is wider than the trough for p). Figure 20 is an example of p>m for left slope. Figure 21 shows m>b for trough height. In figure 22 b>p for right slope for CHIN. An example of m>p for bell height for MLLP is shown in figure 23.

Tables 23, 24 and 25 show the hierarchy of observable differences for contrasts M-B, M-P and P-B respectively. The entries in the table are listed in descending incidence of observable differences. For example, in table 23, the top row shows that for MULP the trough height measure resulted in observable differences for all three vowels.

In all three contrasts, the bell/trough height measures for MLLP and MULP resulted in a high number of observable differences.

For CHIN and MLLP bell width measures resulted in few differences for contrasts M-B and M-P. In contrast P-B, MLLP bell width showed few differences in contrast P-B, but CHIN bell width resulted in differences for /i/ and /a/. Unlike M-B and M-P, the P-B contrast showed few differences for the right slope measure for MLLP and CHIN.

The triplets were recorded so that VmV, VpV, VbV each appeared in the initial, medial and final position in the triplet. An analysis of the effect of the position of VmV in the triplet was done for the MLLP face marker. This analysis focussed on VmV since /p/-/b/ differences were minor compared to /m/-plosive differences.

The results of this analysis are shown in table 26. For Speaker 1 the medial position shows the highest proportion of m>plosive. For Speaker 2 the highest proportion of m>plosive is for the initial position. Speaker 2 shows a trend towards the lowest proportion for the final position, however for Speaker 1 the lowest proportion tends to be for initial position.

The expectation was that the final position would show the smallest /m/-plosive differences. The data showed a small effect for only one speaker.

TABLE 22. Observable Differences for Mid Upper Lip (MULP), Mid Lower Lip (MLLP) and Chin, Summed over Vowels

	LEFT SLOPE	TROUGH HEIGHT	TROUGH WIDTH	BELL HEIGHT	BELL WIDTH	RIGHT SLOPE
MULP	b>m	m>b				
	p>m	m>p	m>p			
		b>p				
MLLP				m>b		m>b
				m>p		m>p
				b>p		
CHIN				m>b	b>m	m>b
				m>p	p>m	m>p
					b>p	

Fig. 19. The /m/ trough is wider than the /p/ trough. The vertical scale is shown in pels. An arbitrary reference is used since relative and not absolute displacements are of interest here.

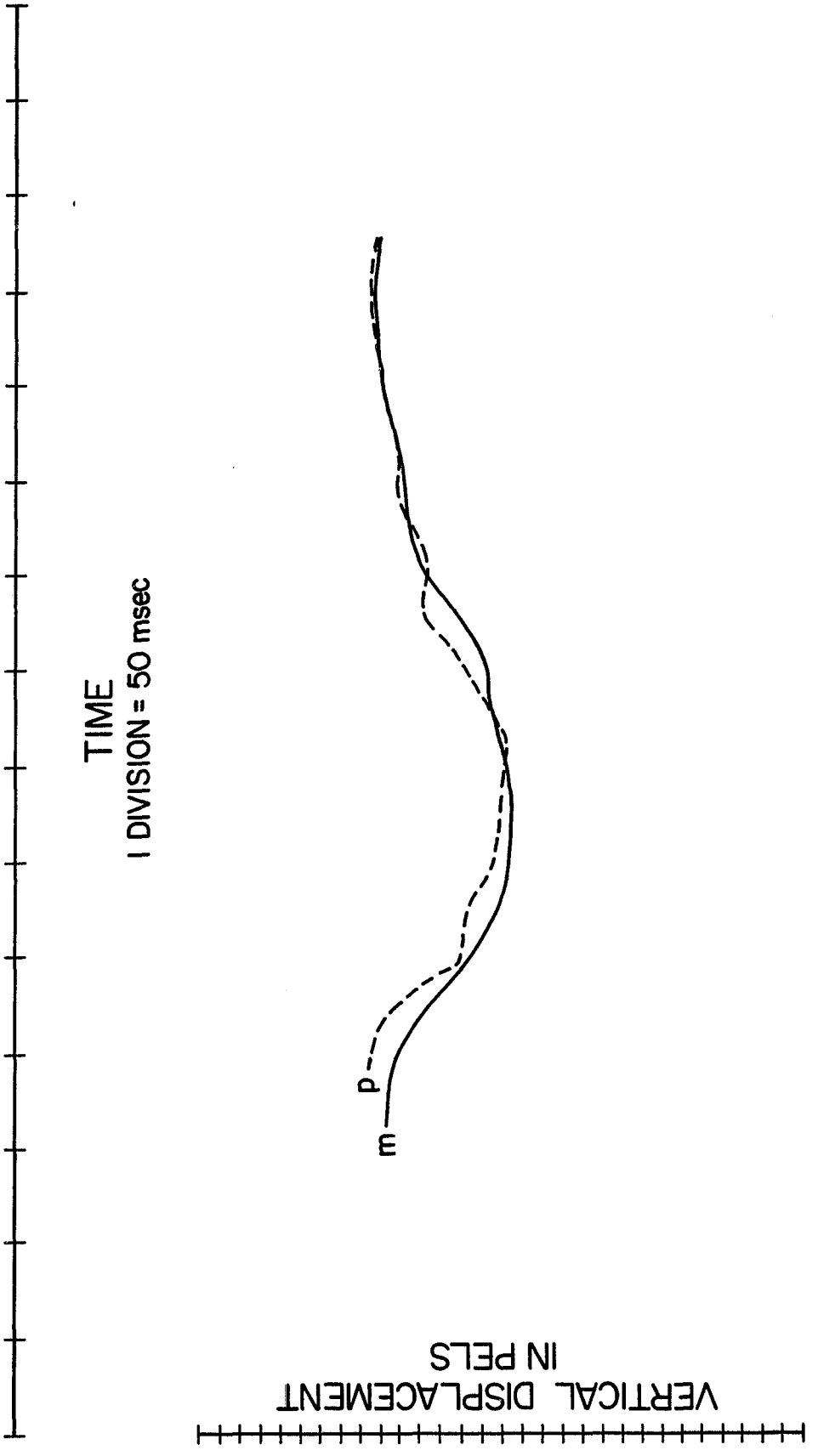

TIME
1 DIVISION = 50 msec
VERTICAL DISPLACEMENT
IN PELS
p
m

Fig. 20. The left slope of the /p/ curve is greater than the left slope of the /m/ curve for the MULP face marker.

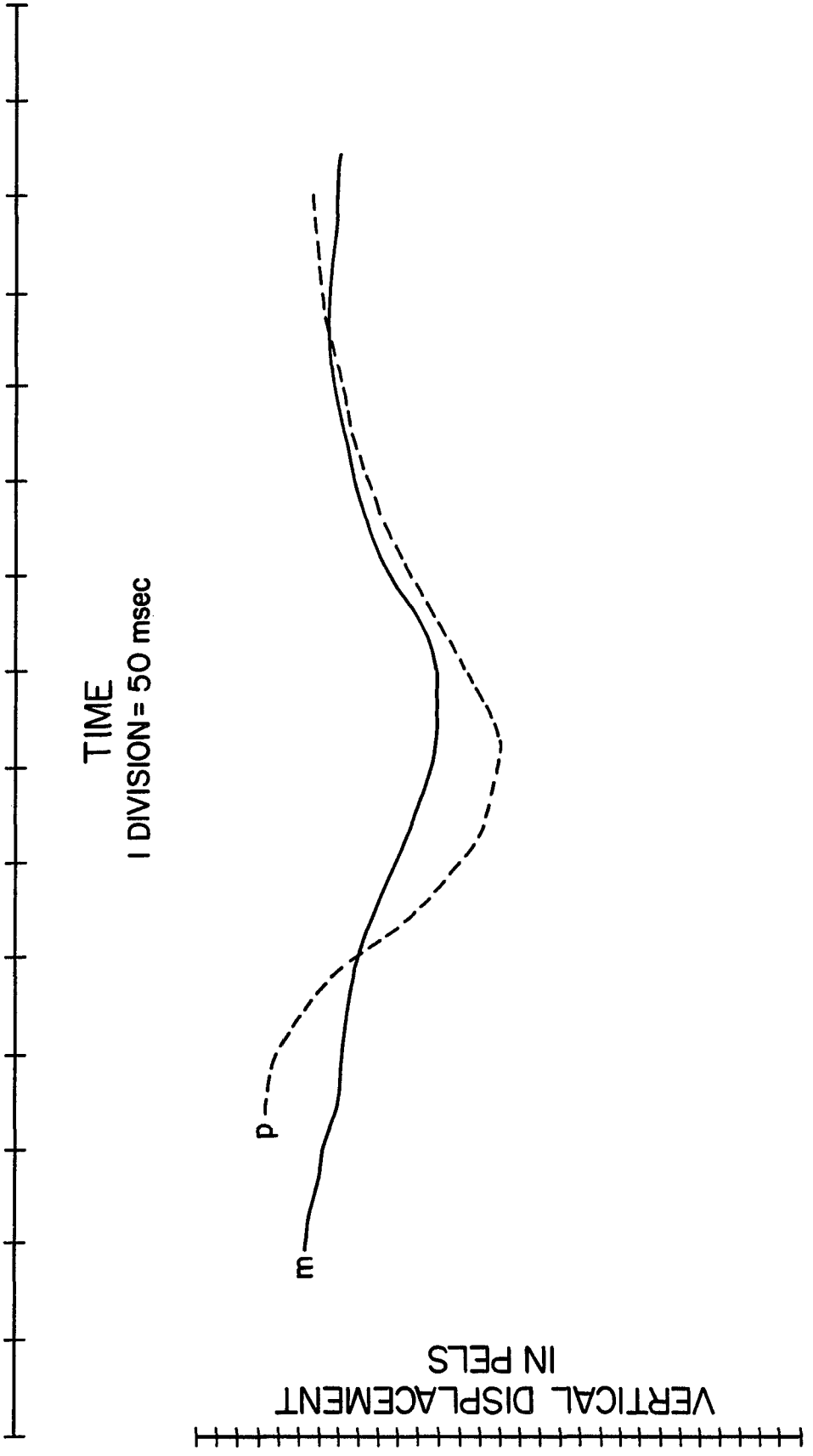

TIME
I DIVISION = 50 msec
VERTICAL DISPLACEMENT
IN PELS
p
m

Fig. 21. The height of the /m/ trough is greater than the height of the /b/ trough for the MULP face marker.

VERTICAL DISPLACEMENT
IN PELS
b
m
TIME
1 DIVISION = 50 msec

Fig. 22. The right slope of the /b/ curve is steeper than the right slope of the /p/ curve for the CHIN face marker.

VERTICAL DISPLACEMENT
IN PELS
p
b
TIME
1 DIVISION = 50 msec

Fig. 23. The height of the bell curve is greater for /m/ than for /p/ for the MLLP face marker.

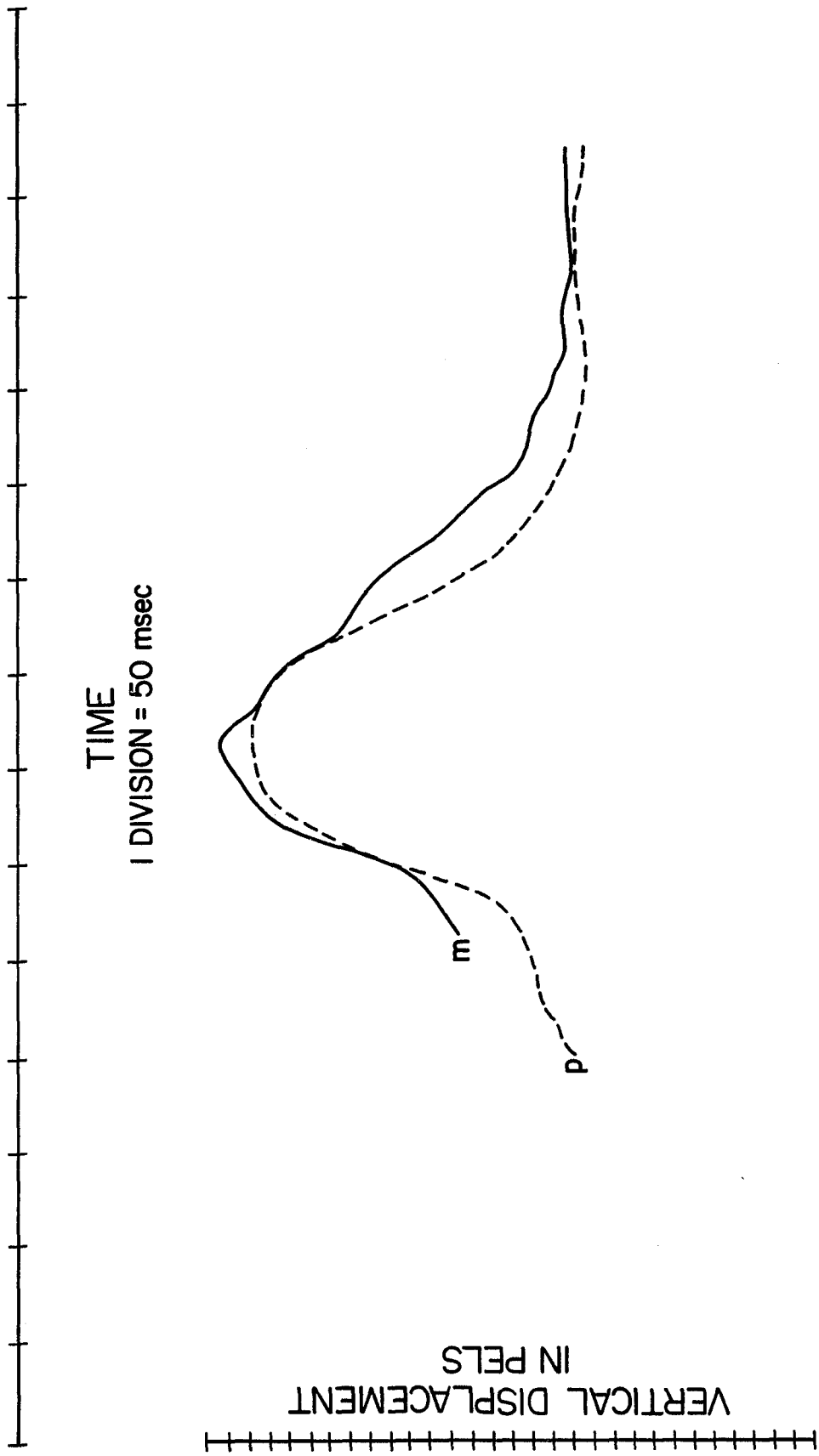

TIME
1 DIVISION = 50 msec
VERTICAL DISPLACEMENT
IN PELS
m
p

TABLE 23. Hierarchy of Observable Differences for Contrast M-B. The notation m indicates m>b and b indicates b>m. MULP and MLLP indicate mid upper lip and mid lower lip respectively.

	CONTRAST M-B		
	/i/	/a/	/u/
MULP-TROUGH HEIGHT	m	m	m
MLLP-BELL HEIGHT	m	m	m
MULP-LEFT SLOPE	b		b
MLLP-RIGHT SLOPE	m	m	
CHIN-RIGHT SLOPE		m	m
MULP-TROUGH WIDTH	m		
MLLP-BELL WIDTH		b	
CHIN-BELL HEIGHT		m	
CHIN-BELL WIDTH	b		

TABLE 24. Hierarchy of Observable Differences for Contrast M-P. The notation m indicates m>p and p indicates p>m. MULP and MLLP indicate mid upper lip and mid lower lip respectively.

	CONTRAST M-P		
	/i/	/a/	/u/
MULP-TROUGH HEIGHT	m	m	m
MLLP-BELL HEIGHT	m	m	m
CHIN-BELL HEIGHT	m	m	m
MULP-TROUGH WIDTH	m	m	
MULP-LEFT SLOPE		p	p
MLLP-RIGHT SLOPE	m	m	
CHIN-RIGHT SLOPE		m	m
CHIN-BELL WIDTH			p
MLLP-BELL WIDTH	no observable differences		

TABLE 25. Hierarchy of Observable Differences for Contrast P-B. The notation b indicates b>p and p indicates p>b. MULP and MLLP indicate mid upper lip and mid lower lip respectively.

	CONTRAST P-B		
	/i/	/a/	/u/
MULP-TROUGH HEIGHT		b	b
MLLP-BELL HEIGHT	b		b
CHIN-BELL WIDTH	b	b	
MULP-LEFT SLOPE		p	
CHIN-BELL HEIGHT			b
MULP-TROUGH WIDTH	p	b	
MLLP-BELL WIDTH	no observable differences		
MLLP-RIGHT SLOPE	no observable differences		
CHIN-RIGHT SLOPE	no observable differences		

TABLE 26. Proportion of m>b, m>p for MLLP, Summed over Vowels. Data are shown for the initial, medial and final positions of VmV in a triplet.

		SPEAKER 1			SPEAKER 2		
		initial	medial	final	initial	medial	final
steeper right slope	m>b	.71	1.0	.75	.80	.75	.36
	m>p	.75	1.0	.80	.75	.60	.67
higher bell	m>b	.73	.73	.67	1.0	1.0	.64
	m>p	.90	.91	1.0	1.0	.89	.60
	mean	.77	.91	.81	.89	.79	.57

Experiment 3: Perceptual Study

The data of the perceptual study were analyzed in two parts: an analysis of correct responses and an analysis of error responses. A separate ANOVA (analysis of variance) was carried out for each contrast. Since three contrasts were considered (M-B, M-P, P-B), three ANOVA'S were performed for each set of data (correct and incorrect responses) for a total of six ANOVA'S.

It was necessary to do separate ANOVA'S for each contrast since each contrast has its own response set. Each response set consists of four targets. Table 27 shows the targets for each contrast. Note that target BB occurred at level one in the M-B contrast but BB occurred at level two in the P-B contrast.

Since the data were in the form of proportions, an arcsine transformation was used. This transformation of the data was necessary since the variance of an estimated proportion varies as a function of the value of the proportion. A basic assumption underlying the F-tests used in the analysis of variance is that the error variance be the same for each observation. The effect of the arcsine transformation on proportional data is to make the error variance roughly constant for the transformed data, i.e., the error variance in arc sine units for an observation based on n trials is approximately 1/n. The transformation used was $y=2\arcsine\sqrt{p}$ where p is the proportion correct and y is the transformed score (Brownlee, 1965).

Analysis of Correct Responses

A four factor ANOVA was performed on the percent correct responses for each contrast. The results for each contrast are shown in Table 28. Statistically significant effects (at the .01 level of significance) are underlined.

In the M-B contrast the only statistically significant effect was Vowel. None of the interactions was statistically significant.

TABLE 27. Stimulus Targets for each Consonant Contrast in the Perceptual Study

CONTRAST	TARGETS			
M-B	BB	MM	BM	MB
M-P	PP	MM	PM	MP
P-B	PP	BB	PB	BP

TABLE 28. Significance Levels Obtained in Analysis of Variance of Correct Responses (Arcsine Transformed Data) for M-B, M-P and P-B Contrasts. The asterisk (*) indicates significance at the .01 level.

Factor	M-B	M-P	P-B
TARGET (T)	.064	.003*	.012
VOWEL (V)	.001*	.001*	.026
SPEAKER (S)	.929	.122	.752
OBSERVER (O)	.293	.023	.223
T X V	.127	.002*	.025
T X S	.240	.050	.003*
V X S	.107	.004*	.857
T X O	.392	.138	.089
V X O	.327	.155	.844
S X O	.122	.271	.563

In the M-P contrast, Target and Vowel were statistically significant main effects. The Target-Vowel interaction and the Vowel-Speaker interaction were also statistically significant.

In the P-B contrast none of the main effects was statistically significant. The interaction of Target and Speaker was significant.

Third order interactions were not statistically significant in any of the contrasts.

The effects of each factor are shown in the following tables and graphs. The data shown are transformed back from arcsine units to proportions.

Figure 24 shows the effects of Vowel for each contrast. Scores for /u/ were higher than scores for /a/ and /i/ in both the M-B and M-P contrasts. The /a/ scores were better than the /i/ scores. In the P-B contrast /u/ was only slightly better than /i/ and /a/. The effect of the vowel was statistically significant for the M-B and M-P contrasts, but not for the P-B contrast (in which overall performance was poor).

Figure 25 shows the effect of target for each contrast. Target was a statistically significant effect for the M-P contrast only and as can be seen from the figure, the target PM score was much lower than the scores for the other three targets.

Figure 26 shows the vowel-target interaction. This interaction was significant for the M-P contrast which showed higher target scores for the vowel /u/ than for the other vowels. A similar pattern was observed for the M-B contrast, but the interaction was not statistically significant.

Figure 26 also indicates which scores were significantly better than chance. In contrast M-B the only score not better than chance was for target MM and /i/ (i.e. /mi-mi/ stimuli). The score for the PM target and /i/ was the only score not better than chance in the M-P contrast (/pi-mi/ stimuli). As seen in figure 26, several scores in the P-B contrast were not significantly better than chance. Appendix C lists the scores for the individual subtests and indicates which were better than chance.

Fig. 24. The proportion of correct responses as a function of vowel and consonant contrast.

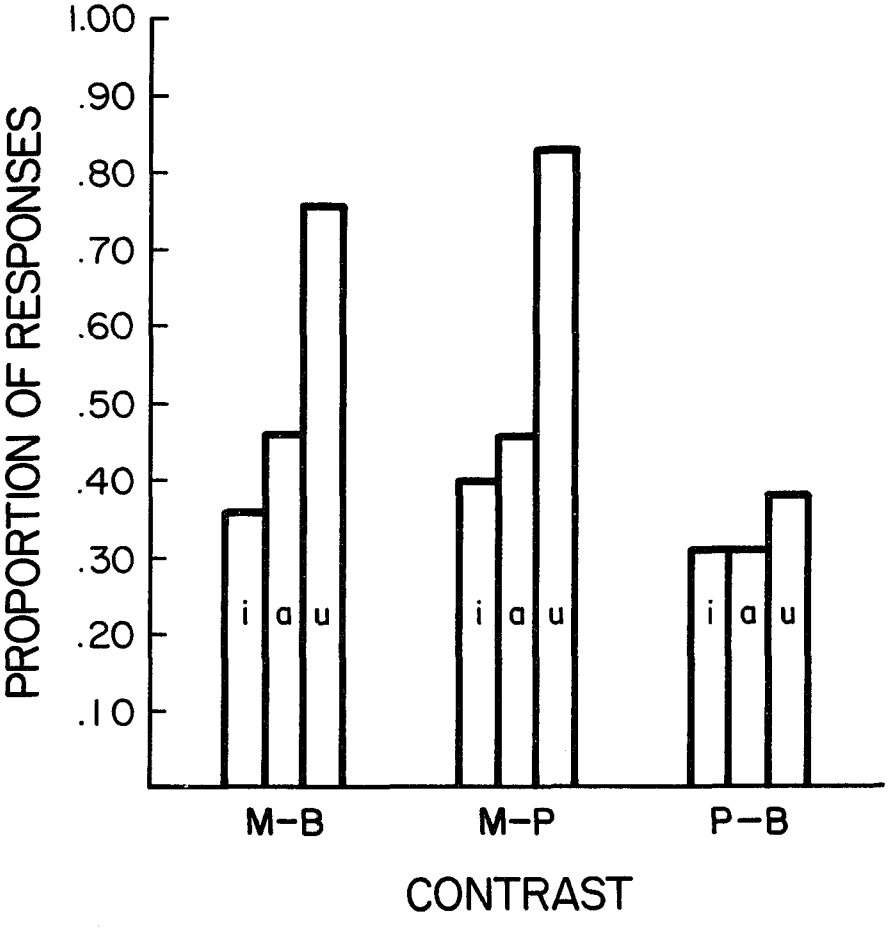

1.00
.90
.80
.70
.60
.50
.40
.30
.20
.10
PROPORTION OF RESPONSES
i
a
u
i
a
u
i
a
u
M–B
M–P
P–B
CONTRAST

Fig. 25. The effect of Target for each contrast.

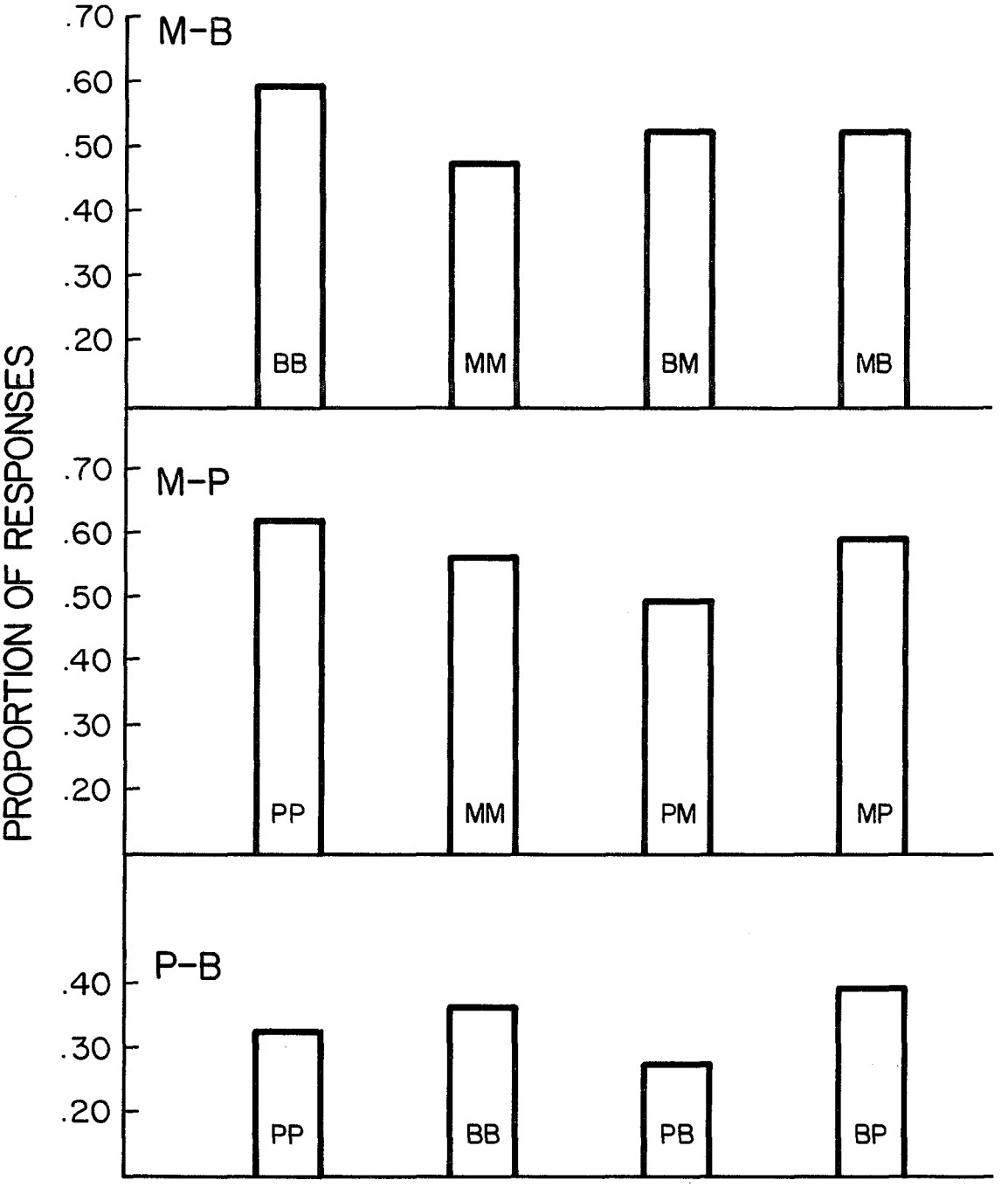

PROPORTION OF RESPONSES
M-B
.70
.60
.50
.40
.30
.20
BB
MM
BM
MB
M-P
.70
.60
.50
.40
.30
.20
PP
MM
PM
MP
P-B
.40
.30
.20
PP
BB
PB
BP

Fig. 26. The Vowel-Target interaction. The horizontal line indicates the proportion that is significantly different from chance.

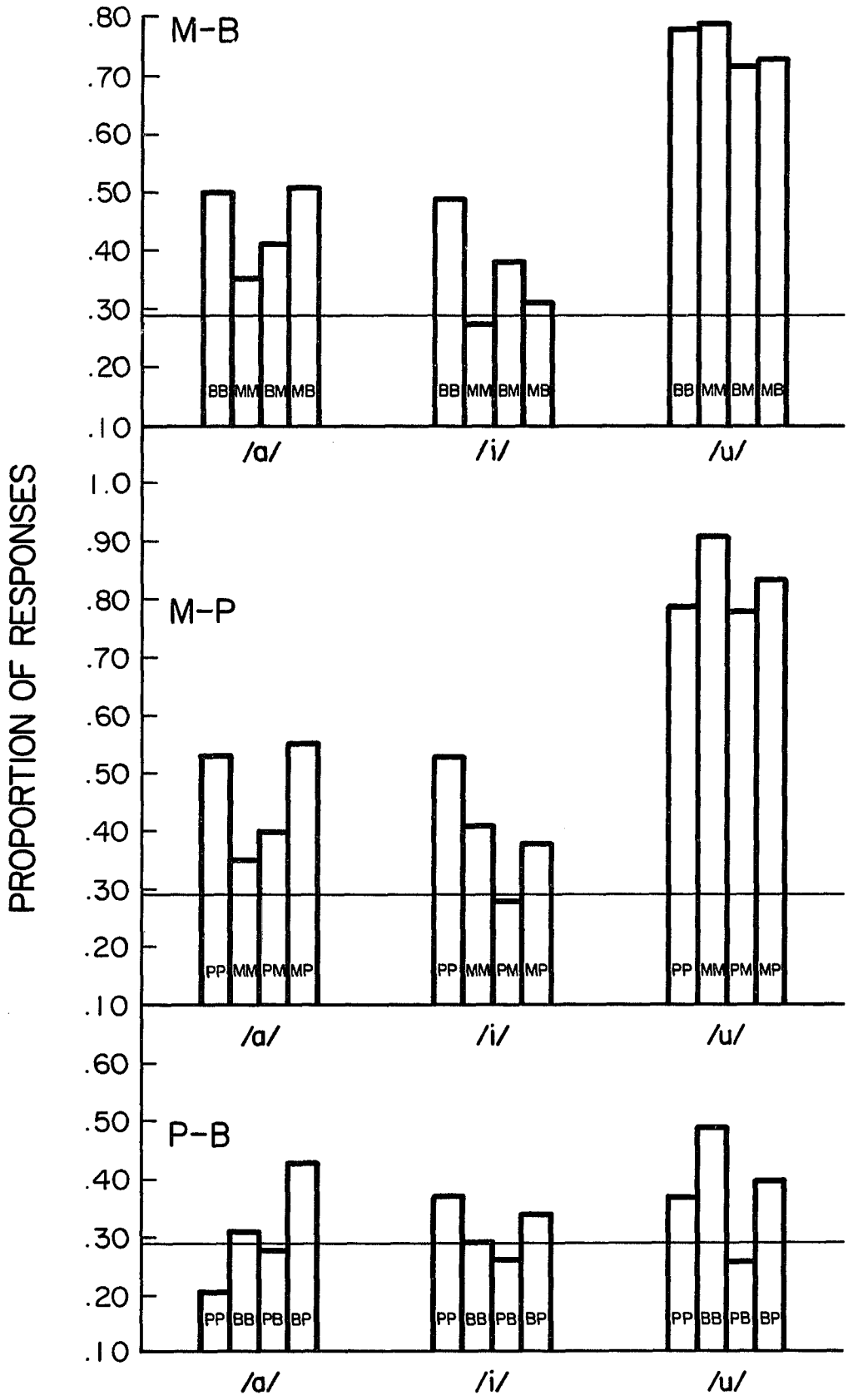

PROPORTION OF RESPONSES
M-B
.80
.70
.60
.50
.40
.30
.20
.10
BB MM BM MB
/a/ /i/ /u/
M-P
1.0
.90
.80
.70
.60
.50
.40
.30
.20
.10
PP MM PM MP
/a/ /i/ /u/
P-B
.60
.50
.40
.30
.20
.10
PP BB PB BP
/a/ /i/ /u/

Table 29 shows the interaction between Target and Speaker for each contrast. Differences between speakers are within 10% for all targets in M-B and M-P. In P-B there is a statistically significant interaction in which Speaker 2 scores were better than Speaker 1 scores for BB and PP targets but Speaker 1 scores were better for PB and BP targets.

Table 30 shows the interaction between Speaker and Vowel. The interaction was significant for the M-P contrast. There was a relatively large difference between speakers for /u/ but not for the other vowels. The data suggest a similar interaction for the /a/ vowel for the M-B contrast. However, a 9% difference in the vicinity of a 50 percent score (as in the case of the /a/ vowel) is not statistically significant whereas a 10% difference in the vicinity of an 80 percent score is significant. The smallest differences between speakers as a function of vowel were observed for contrast P-B.

Analysis of Error Responses

A five factor ANOVA was performed for each contrast (M-B, M-P, P-B). The factors were Target, Vowel, Speaker, Observer and Error Type. The three error types were as follows: incorrect first consonant, incorrect second consonant and both consonants incorrect. Table 31 shows the possible incorrect responses for each stimulus.

The results of the ANOVA for each contrast are shown in Table 32. In the M-B contrast both Vowel and Error Type were statistically significant factors. There were no significant interactions. Vowel and Error Type were also statistically significant in the M-P contrast. In addition, the Error Type-Vowel interaction was statistically significant. In the P-B contrast there were no statistically significant main factors or interactions.

TABLE 29. Proportion of Correct Responses for each Target for Speaker 1 and Speaker 2, for Contrasts M-B, M-P and P-B.

		CONTRAST M-B			
	Target	BB	MM	BM	MB
SPEAKER 1		.55	.50	.54	.52
SPEAKER 2		.63	.45	.50	.52

		CONTRAST M-P			
	Target	PP	MM	PM	MP
SPEAKER 1		.64	.53	.47	.64
SPEAKER 2		.59	.58	.50	.53

		CONTRAST P-B			
	Target	PP	BB	PB	BP
SPEAKER 1		.27	.29	.31	.44
SPEAKER 2		.36	.43	.22	.33

TABLE 30. Proportion of Correct Responses for each Vowel for Speaker 1 and Speaker 2, for Contrasts M-B, M-P and P-B.

	Contrast M-B		
	/a/	/i/	/u/
SPEAKER 1	.41	.37	.79
SPEAKER 2	.50	.35	.72

	Contrast M-P		
	/a/	/i/	/u/
SPEAKER 1	.47	.37	.88
SPEAKER 2	.45	.43	.78

	Contrast P-B		
	/a/	/i/	/u/
SPEAKER 1	.31	.31	.37
SPEAKER 2	.30	.32	.39

TABLE 31. Alternative Incorrect Responses for each Target in Contrasts M-B, M-P and P-B. The incorrect response is either incorrect in the first, second or both consonants of the two consonant target.

	Target	First Consonant Incorrect Response	Second Consonant Incorrect Response	Both Consonants Incorrect Response
	BB	MB	BM	MM
CONTRAST	MM	BM	MB	BB
M-P	BM	MM	BB	MB
	MB	BB	MM	BM
	Target			
	PP	MP	PM	MM
CONTRAST	MM	PM	MP	PP
M-P	PM	MM	PP	MP
	MP	PP	MM	PM
	Target			
	PP	BP	PB	BB
CONTRAST	BB	PB	BP	PP
P-B	PB	BB	PP	BP
	BP	PP	BB	PB

TABLE 32. Significance Levels Obtained in Analysis of Variance of Error Responses (Arcsine Transformed Data) for M-B, M-P and P-B Contrasts. The asterisk (*) indicates significance at the .01 level. Speaker and Observer interactions were not significant.

Factor	M-B	M-P	P-B
ERROR TYPE (E)	.007*	.001*	.642
TARGET (T)	.394	.113	.051
VOWEL (V)	.001*	.001*	.117
SPEAKER (S)	.876	.291	.789
OBSERVER (O)	.686	.255	.617
E X T	.016	.001*	.267
E X V	.988	.006*	.098
T X V	.639	.082	.172

Figure 27 shows the proportion of responses in each error category. (The data in the figures and table have been transformed back from arcsine units into proportions.) Error Type was a significant factor for contrasts M-B and M-P. The pattern of the incorrect responses was the same for the two contrasts: The number of incorrect first and second consonant errors was nearly equal and there were 4-5% fewer both consonant incorrect responses. In contrast P-B the proportion of responses in each of the error categories was approximately the same.

Figure 28 shows the relationship between Vowel and Error Type. (Vowel was a statistically significant main factor for M-P and M-B). In contrast M-P the Vowel-Error Type interaction was significant. For /a/ only 11% of the errors were in the both consonant incorrect category; there were twice as many responses in each of the other two categories. For /i/ and /u/ there was a more equal distribution of error types. There was no significant interaction in contrasts M-B and P-B.

The interaction of Error Type and Target is seen in figure 29. In Contrast M-P the interaction was statistically significant. For targets MM and PM the second consonant error was the most frequent error. For targets PP and MP the first consonant error was the most frequent.

In contrasts M-B and M-P there were more second consonant errors for PM (/pV-mV/) and BM (/bV-mV/) and more first consonant errors for MP (/mV-pV/) and MB (/mV-bV/). That is, there were more errors for /m/. Note that the total number of incorrect responses was greater for MM than for either BB or PP.

Figure 29 also indicates that /mV-bV/ was the least likely response to the stimulus /bV-mV) and vice versa. Similarly, /mV-pV/ and /pV-mV/ were less frequently confused with each other than with the other stimuli.

Observer biases in favor of the stop response are evident in figure 30, which shows the distribution of incorrect responses (without regard to error type). BB is the

Fig. 27. The proportion of responses in each of the error categories: first consonant incorrect (1), second consonant incorrect (2), both consonants incorrect (B).

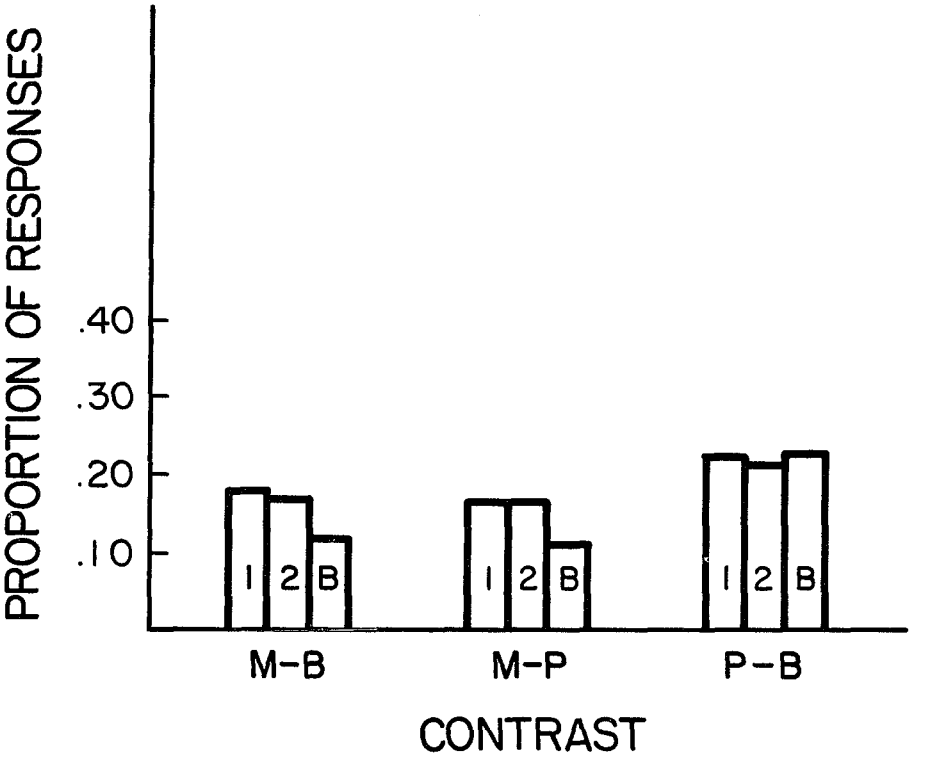
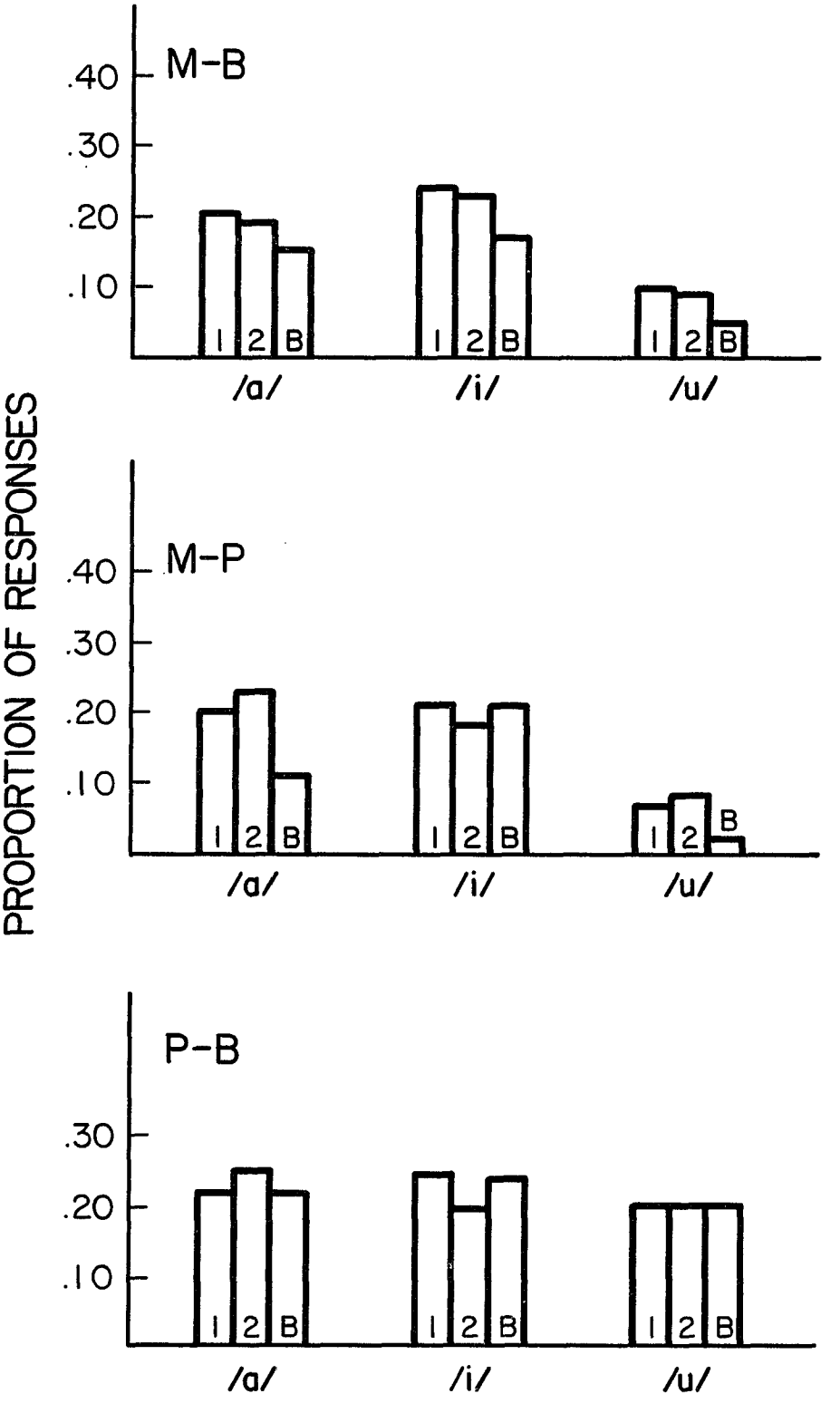

PROPORTION OF RESPONSES
.40
.30
.20
.10
1
2
B
M-B
M-P
P-B
CONTRAST

Fig. 28. Proportion of responses for each vowel and contrast for each of the three Error-Type categories: first phoneme incorrect (1), second phoneme incorrect (2), both phonemes incorrect (B).


M-B
.40
.30
.20
.10
1 2 B
/a/
/i/
/u/
M-P
.40
.30
.20
.10
1 2 B
/a/
/i/
/u/
P-B
.30
.20
.10
1 2 B
/a/
/i/
/u/
PROPORTION OF RESPONSES

most frequent error response in contrast M-B. Figure 29 indicates that this was true for all targets. In contrast M-P PP was the most frequent response for targets PM and MP.

In contrast P-B there was a nearly equal distribution of incorrect responses across the response categories (figure 30).

Subjective Impressions

All three observers reported that for the vowel /u/, they were able to differentiate the /p/ and /b/ stimuli from /m/ stimuli on the basis of the rapid cheek movement that was present for /p/ and /b/.

One observer associated a lower cheek indentation with /ba/ in a /ma/-/ba/ comparison.

Fig. 29. The interaction of Error-Type and Target. The three error types are indicated by 1, 2, B and the actual observer response is shown in each bar.

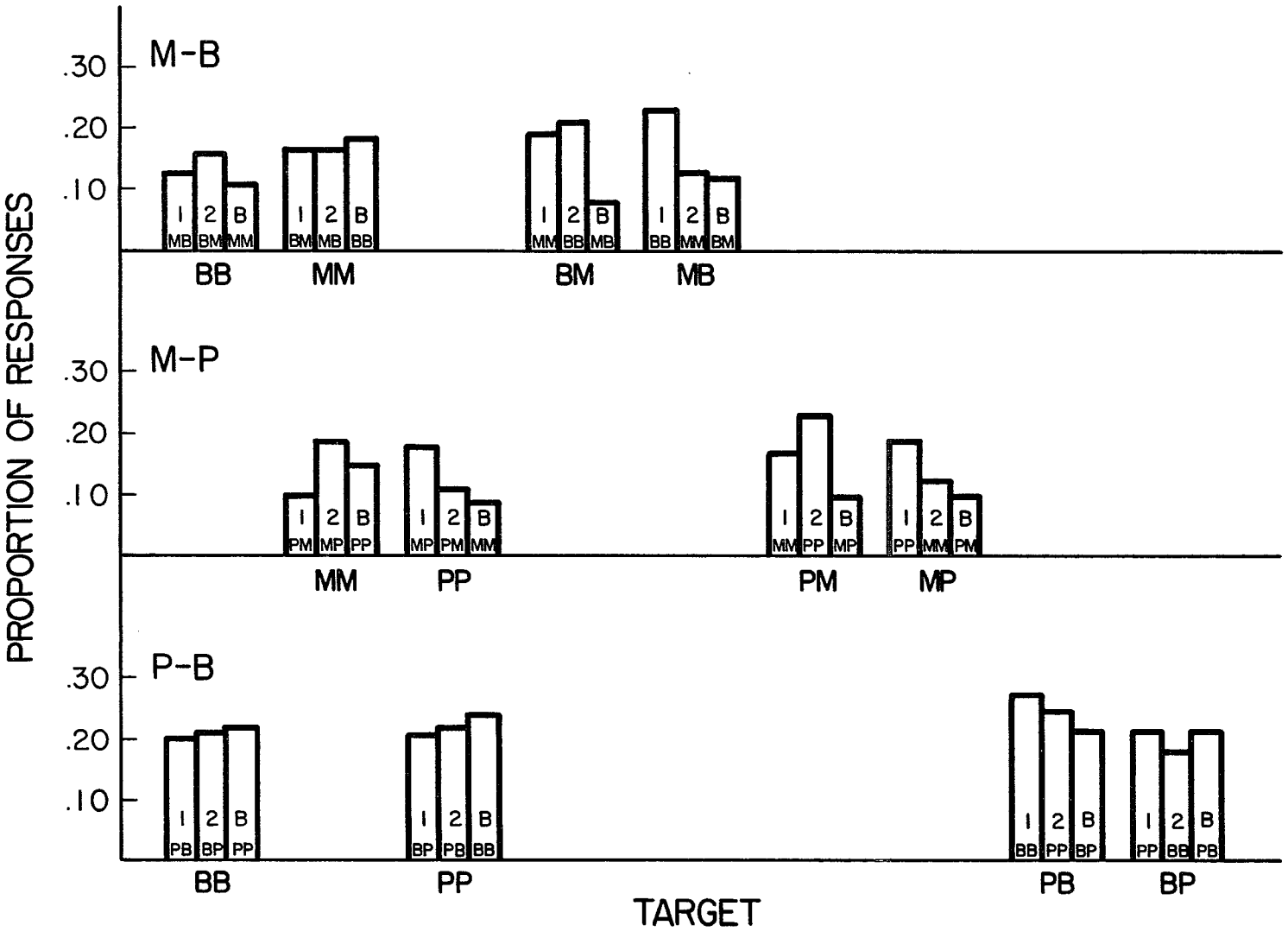
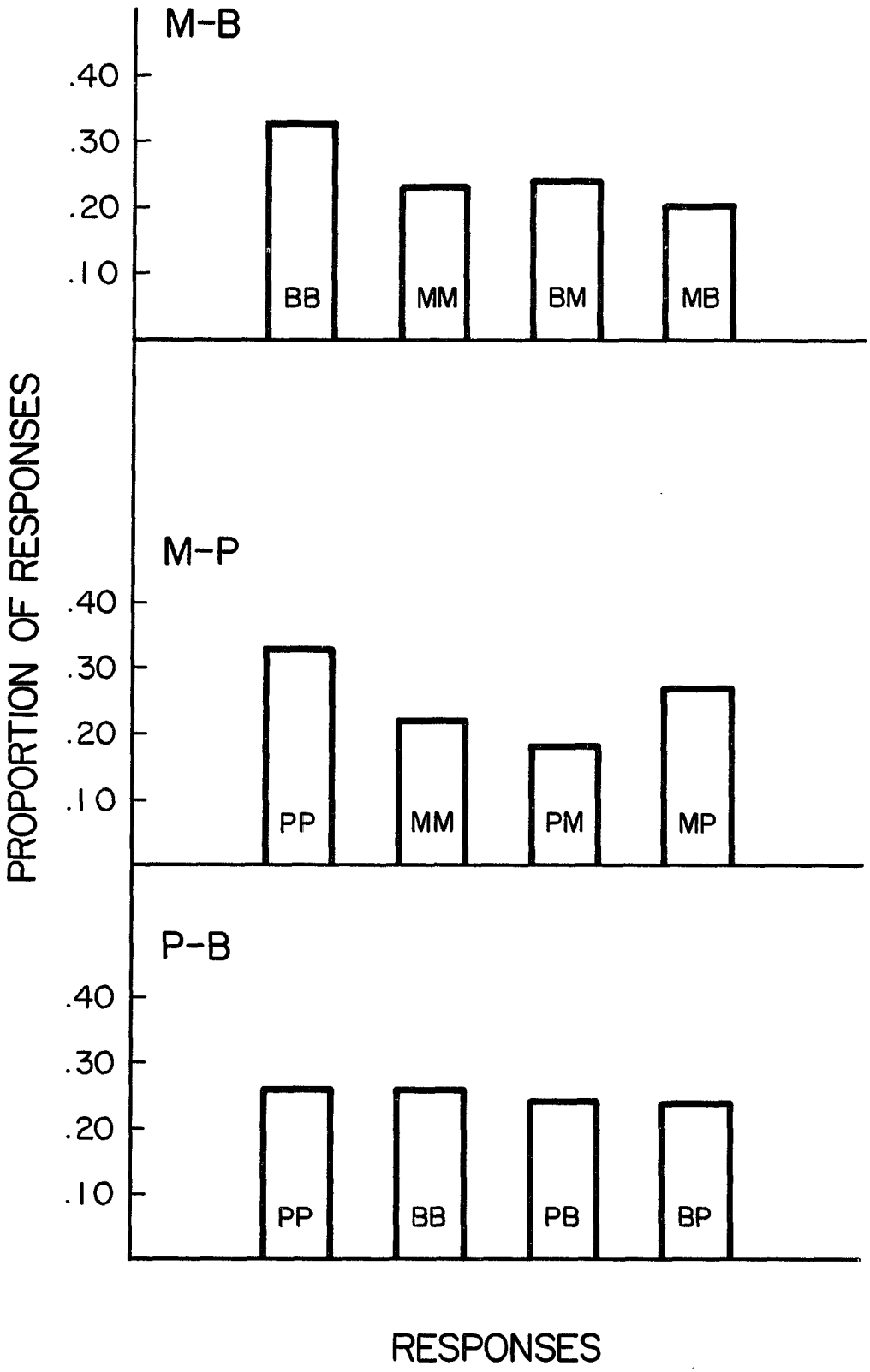

M-B
.30
.20
.10
1 MB
2 BM
B MM
1 BM
2 MB
B BB
1 MM
2 BB
B MB
1 BB
2 MM
B BM
BB
MM
BM
MB
M-P
.30
.20
.10
1 PM
2 MP
B PP
1 MP
2 PM
B MM
1 MM
2 PP
B MP
1 PP
2 MM
B PM
MM
PP
PM
MP
P-B
.30
.20
.10
1 PB
2 BP
B PP
1 BP
2 PB
B BB
1 BB
2 PP
B BP
1 PP
2 BB
B PB
BB
PP
PB
BP
PROPORTION OF RESPONSES
TARGET

Fig. 30. The distribution of incorrect responses for each contrast.


M-B
.40
.30
.20
.10
BB
MM
BM
MB
M-P
.40
.30
.20
.10
PP
MM
PM
MP
P-B
.40
.30
.20
.10
PP
BB
PB
BP
PROPORTION OF RESPONSES
RESPONSES

CHAPTER V
DISCUSSION

This investigation examined novel approaches to the study of the speechreading signal. The three approaches were: interleaving, point measurement and perceptual measurement. Each of these approaches provided unique information that could not be obtained by other means. These techniques were evaluated in the context of examining the differences between the so-called homophenous sounds /p/, /b/ and /m/.

Of particular interest was the precision of the point measurement in time and space. The video recordings were computerized which allowed for a powerful analysis of the speechreading signal. The spatial resolution of the video images was 256 pels by 256 lines per frame, with one pel or video line approximately equal to 0.5 mm. The face markers were approximately 1.5 mm in diameter however variations in the reflected light and streaking due to rapid movement increased the apparent size of the markers during speech production. The number of pels per marker, therefore, varied from five to fifteen in most cases and the marker location was defined as the mean of the y coordinate values. This was rounded to the nearest whole number for the plotting of the curves. Changes in light intensity that might have indicated anterior/posterior movement would not have affected the value of the marker coordinates if the mean value remained the same.

Using an average value for the x and y coordinates resulted in some place distortion in that the extreme positions of the articulators were replaced by an average position. This was preferable to the loss of accuracy in the time domain that would have resulted if the streak endpoint had been used as the coordinates for the streaked markers. If the endpoints were used, an extreme position would appear to have

occurred earlier than it actually did. Use of the mean would be inaccurate in the case of direction reversals within a field, however a frame by frame visual analysis indicated that such rapid changes in direction were unlikely.

The 30 frame (60 field) per second rate limited the timing information available in this study. Consonant timing differences less than 17 msec could not be measured. For example, preliminary comparisons of closure duration between /p/, /b/, and /m/ did not yield differences, however the differences may have been less than 17 msec. When timing differences did exist the resulting pattern was inconsistent. For example, there was no pattern to the differences between the time line-up points of two curves with the bell/troughs was aligned. (The line-up points of the VCV's had to be ignored in order to compare the bells and troughs of the curves.)

The horizontal movement data was not analyzed in this investigation. Since /i/ shows a great deal of horizontal movement there may have been differences between consonants for this movement for /i/. The cheek movement associated with the vowel /u/ may have contained a horizontal component as well.

Horizontal movement of the corner of the upper lip was measured by Brooke and Summerfield (1983). They reported that a comparison of /aba/, /apa/ and /ama/ yielded differences no greater than the intra-token differences. (These data were based on three tokens of each utterance.)

A key assumption in the use of video recordings is that the essential information for speechreading is contained in the video signal. A major limitation of video recordings is that, using present technology, the picture is two dimensional. The lack of three dimensional movement may have been a problem for the vowel /u/ for both lip and cheek movement, however this could be avoided in future work by incorporating profile views of the face. The advantage of video recordings is that they can be obtained under better conditions than would occur in ordinary viewing in that the lighting can be arranged to fall directly on the face of the speaker and the camera is

focussed on the face, limiting the number of distracting stimuli. Erber (1974) found that illumination was a significant factor in speechreading. In this study it was particularly important since the jitter depended on light reflection from the face. Fujimura (1961) pointed out that optical methods are beneficial in speech research since the method of measurement does not interfere with the articulation of speech.

In this study computerized video recordings were used. An important advantage of computerized video recordings is that they permit the use of the interleaving technique and the precise measurement of the movement of markers on the face. The interleaving technique is valuable for demonstrating differences between video sequences. It is extremely sensitive to any differences between two utterances. This includes spatial and temporal differences which may not be perceptible in normal lipreading. The jitter isolates the potential source of difference between two visual signals. If there is a difference in the utterance it would show up as jitter, however not all jitter would indicate phonemically relevant differences. In this study the interleaving technique showed that jitter occurred in the region of the lower lip, chin and cheeks. There was very little jitter, if any, on the upper lip. Of interest in this study were those differences between facial movements that were phonemically important (e.g. differences between pa and ba).

The interleaving technique may also be used to derive other information about the speech signal. For example, in this study, there was no jitter for sequences of a few frames, however when additional frames were added to the same sequence the jitter was increased. This was not studied in detail within this experiment, however it would provide information concerning the importance of the following vowel on consonant discrimination. Interleaving may also be used with varying frame rates, in order to determine the effect of the time window on visible differences. In addition, the interleaving technique could be used for the precise alignment of sounds if better equipment were available.

One purpose of this study was to determine whether the methodology used in these experiments was effective for studying the speechreading signal in great detail. Although there were some problems and limitations, all three experiments resulted in significant data.

Several other video systems for tracking articulatory data have been developed (e.g. McCutcheon, Fletcher and Hasegawa, 1977; Sonoda and Wanishi, 1982; Brooke and Summerfield, 1983). These systems appear to be more efficient for tracking face marker movement than the system used in the present study. McCutcheon, Fletcher and Hasegawa used a 100 Hz frame rate and were able to record and store data for up to 20 minutes (compared with 8 seconds in the present study). The resolution of the system was 0.3 mm. An "adaptive gate network" eliminated spurious points and produced a coordinate value for each face marker.

The system described by Sonoda and Wanishi used light-emitting diodes (LEDs) and a position-sensitive device which generates coordinate values in real time from the light of the LED. That is, the LEDs are the marker points on the lips, jaw, etc. The great advantage of this system is that the coordinate values are recorded continuously. The resolution was 0.1 mm.

Brooke and Summerfield used video recordings, however, their instrumentation eliminated some of the problems encountered in this study. The use of a shutter camera prevented streaking since the exposure time was very short for each frame. Their method of digitizing coordinate values for the points on the face (i.e. pressing a key when a cursor on a screen intersected with a marker) eliminated the problem of spurious points of reflection and the need for averaging coordinate values. The limit of resolution was 1.0 mm.

Point Measurement

A primary focus of this investigation was the measurement of differences between /p/, /b/ and /m/. The data in the point measurement study show the differences in lip and chin position for /p/, /b/ and /m/ during the production of VCV utterances.

Three aspects of lip and chin movement were found to be important. These were: vertical displacement (height), rate of displacement (slope) and duration (width). As noted in the Results chapter, the bell height comparison indicated upward movement of the lower lip and chin. Trough height indicated downward movement of the upper lip. Slope corresponded with the rate of movement of the lips and chin before and after closure. Left and right slope referred to movement towards and away from closure respectively. Width correlated with the duration of closure.

Of the above measurements, the vertical displacement (trough/bell height) was the most revealing in that it indicated differences between /m/, /p/ and /b/ for upper and lower lip and chin across vowels and subjects.

The following pattern was seen when the curves were compared with respect to height: The lower lip and chin move further upward for /m/ than for /b/ or /p/. The lower lip moves further upward for /b/ than for /p/. The upper lip moves less downward for /m/ than for /p/ and /b/. That is, the chin and both lips are displaced further upward for the production of /m/ than for /p/ and /b/.

Another revealing finding was that the rate of movement away from closure for the lower lip and chin is greater for /m/ than for /p/ or /b/. For speaker 1 the rate of movement was greater for /b/ than for /p/ for the left lower lip face marker. There was no difference for the chin.

The velocity of movement of the upper lip towards closure is faster for /p/ and /b/ than for /m/.

A third observation was that there were only subtle durational differences measured between /p/, /b/ and /m/. The results show that trough/bell width is the least

salient difference between the consonants. Since the eye has poor sensitivity to temporal differences, subtle differences in duration would not be expected to have great significance in the speechreading process, even though temporal relationships are critical to auditory speech perception (e.g. voicing). However, it may be possible to measure greater temporal differences between consonants in the speechreading signal when there is better temporal resolution and in other than VCV contexts.

The results of the width measure did show differences between the upper and lower lip for the M-B and M-P contrasts. For the lower lip the duration between the points of maximum opening for the vowels was longer for /b/ than for /m/ (i.e. there was a wider bell for /b/). For the upper lip the reverse was true: /m/ showed the longer duration. The /m/ trough was also wider than the /p/ trough for the upper lip. For the upper lip the difference in width did not occur as frequently for /m/ and /p/ comparisons. These results were consistent with the finding (for Speaker 2) that the differences in width between /p/ and /b/ were in the direction b>p for bell width for the lower lip and chin. There were no differences in width for Speaker 1 for contrast P-B.

There was a consistent relationship between the slope, width and height measures. (Note that the velocity of movement into closure was compared for the upper lip, but that the velocity away from closure was compared for the lower lip.) The height and slope results show that when the lip moves a greater distance (/m/ lower lip, /p/ and /b/ upper lip) the velocity of movement is faster, and the duration between the points of maximum vowel opening is shorter. That is, since for all three consonants the lower lip must reach the same open position for the following vowel, the velocity is greater for /m/ since the lower lip starts the movement towards the vowel from a higher position. Similarly, the distance into the closure position is shorter for /m/ for the upper lip since the upper lip is in a higher position for /m/ during closure, and therefore the velocity into closure is slower for /m/. Sussman, MacNeilage and Hanson (1973) also found that the lower lip and jaw lowered fastest for /m/ and slowest for /p/, with /b/ in between, however their results for the lower lip were vowel dependent and did

not hold for /æ/. As in the present study, their results showed that /m/ was produced with less downward movement of the upper lip and slower velocity than /p/ and /b/ (i.e. p>b>m for upper lip velocity and m>b>p for lower lip velocity). They attributed the upper lip results to the fact that /m/ requires less articulatory preparation prior to closure than the plosives since there is less pressure build-up for /m/.

Sussman, MacNeilage and Hanson also commented on the relationship between increasing velocity and increased distance to be covered. They pointed out that in their study, jaw movement was faster for vowels requiring a lower jaw position.

These results for /m/, /b/, and /p/ may be related to the aerodynamic properties of the consonants. The velocity away from closure and the height of the lips and chin are in the direction m>b>p, and the relationship of the amount of pressure build-up during production is m<b<p. It may be that the build up of pressure pushes downward on the lips in anticipation of the opening for the vowel. Thus, the /m/ closure can be in a higher position and would then require a higher velocity to reach the vowel opening.

Another explanation for the faster lower lip movement for /m/ may be that the lip is freer to move for the nasal since the muscles are not involved in the pressure build up required for the plosive. Sussman, MacNeilage and Hanson (1973) also reported greater coarticulation effects for /m/.

Although nasal-plosive differences occurred for all three vowels, the most consistent differences were seen for the vowel /a/. This may be due to the greater distance of movement with /a/, or it may be that the differences are just easier to measure due to the greater vertical displacement.

The vowel /u/ showed fewer differences than the other vowels. As previously stated, the lack of three dimensional measurement may account for the lower incidence of differences in lip and cheek movement for /u/. There were significant differences for the measure of vertical displacement, even for the /p/-/b/ comparison, however the measure of "right slope" (i.e. rate of displacement after closure) for the lower lip was

never significant for /u/. Since the /u/ curves were very shallow, differences in the rate of movement may have been very difficult to detect even though displacement differences were observable.

The vertical displacement of the midpoint of the lower lip was the most consistent measure across contrasts, subjects and vowels. The next most consistent measures were the vertical displacement of the midpoint of the upper lip and the rate of displacement of the midpoint of the lower lip. Thus, when the data are summed over all three contrasts, the midpoint of the lips was most important in distinguishing between consonants.

Perceptual Study

The perceptual study showed that observers were able to discriminate between consonants with the same place of production. The proportion of correct responses was significantly better than chance for twenty-two out of twenty-four targets for contrasts M-B and M-P. In contrast P-B the data for approximately half the targets resulted in proportions significantly better than chance.

It is important to consider the vowel environment when describing consonant discrimination performance. Performance for the vowel /u/ was better than 75% for M-B and M-P contrasts however scores for /i/ and /a/ were in the 35-50% range. While /i/ and /a/ scores were statistically significant, it is clear that discrimination with the vowel /u/ was much easier than with /i/ and /a/. Performance for /u/ was better than for /i/ and /a/ in the P-B contrast as well, but scores for P-B were in the 30-40% range.

The 50% score is defined here as the just noticeable difference point between two phonemes. (In the four alternative forced-choice task chance is equal to 25%.) Using this criterion, the results of the perceptual study suggest that the traditional grouping of /p/, /b/ and /m/ into one viseme may not be correct. In this study /m/ was contrastive with both /p/ and /b/. The overall scores for contrasts M-B and M-P were 52.6% and 56.1% respectively. On the other hand, /p/ and /b/ were sufficiently

confused with each other to be considered a single viseme. The percent correct for P-B was 33.5%. Thus the grouping of /p,b,m/ might better be described as two visemes: /p,b/ and /m/.

In the analysis of error responses, the error type "both phonemes incorrect" (e.g. a response of /pa/-/ma/ for target /ma/-/pa/) was the least likely error for most vowel and target conditions for contrasts M-P and M-B. These results indicate that observers did not tend to use a same-different criterion in which the two consonants within a stimulus were compared without regard to order.

The lower incidence of the error type "both phonemes incorrect" for contrasts M-B and M-P compared with contrast P-B further indicates that the /p/-/b/ distinction is much more difficult than the plosive-nasal distinction. For "both phonemes incorrect" two CV's were perceived incorrectly whereas for "first phoneme incorrect" and "second phoneme incorrect" only one CV was incorrectly perceived.

Comparison of Interleaving, Measurement and Perceptual Experiments

This study was notable in that three different types of data were linked together to describe the speechreading signal. The interleaving data specified the important locations for measurement. Following that, the measurements obtained at those locations showed differences in articulatory movement for /p/, /b/ and /m/. The perceptual data indicated that subjects do perceive differences between /p/, /b/ and /m/.

The interleaving technique was very sensitive to differences however it can only indicate the location of differences. The cause of the jitter (i.e. differences in displacement or rate of displacement, for example) could not be identified in this interleaving study. In future studies it would be possible to synthesize differences in displacement or rate and then observe whether jitter occurred in each case. In the

present study the point-measurement analysis was designed to identify the source of the differences between utterances.

The interleaving data showed differences in the lower lip primarily for /i/ and /a/ and differences in the cheeks primarily for /u/. Cheek movement was also reported by observers in the perceptual study. They noticed that the presence (for plosives) or absence (for /m/) of cheek movement enabled them to discriminate between the nasal and plosives for the vowel /u/.

The perceptual data showed primarily nasal versus stop differences and relatively small voiced-voiceless differences. The nasal-stop differences were largest for the vowel /u/.

The point-measurement data showed that the nasal-stop differences occurred in the upper lip, lower lip and chin. The point-measurement differences occurred most frequently for the vowel /a/ and least frequently for the vowel /u/. Voiced-voiceless differences were less observable than plosive-nasal differences. The most consistent measure across all contrasts (M-B, M-P, P-B) was the vertical displacement of the midpoint of the lower lip. The vertical displacement of the midpoint of the upper lip and the rate of displacement of the midpoint of the lower lip were slightly less consistent.

The point-measurement and interleaving studies differed in that there was no upper lip jitter in the interleaving study although there were significant point-measurement differences in upper lip movement between consonants. The absence of upper lip jitter may be explained by the differences between the jitter and point measurement experiments in utterance type and timing. The utterances in the jitter experiment consisted of the closure and following vowel portion of the VCV. (The portion of the consonant prior to closure was eliminated since it was difficult to view the jitter in the palindromic display if the first vowel of the VCV was included.) The

point measurement utterances did include the vowel portion prior to closure and consonant differences in rate of movement occurred during that portion of the VCV.

The timing difference between the two studies may explain why jitter did not occur although the trough height measure showed consonant differences for the upper lip. The curves in the point measurement study were compared without regard to the original line-up point (t=0), whereas the interleaved sequences were interleaved precisely according to the visually pre-determined t=0. In fact, until the time origin was ignored, the differences in trough height were obscured.

Another explanation for the lack of upper lip jitter may be that the absolute differences between consonants was less than or very close to the just noticeable difference for visual perception of lip movement. However, since the curves were compared on a relative basis (i.e. which slope is steeper, which trough is higher), the data do not indicate whether the absolute differences (which ranged from approximately 0.5 mm to 4.0 mm) were greater for lower lip measures than for upper lip measures. If this difference between the upper and lower lip did exist, then it would explain why jitter existed for the lower lip and not the upper lip.

Based on the point-measurement data, the movement of the midpoint of the lips was the most important cue for distinguishing between consonants for all three vowels. However, when all three sets of data are considered it is clear that the cheek movement was most important for /u/ although it could not be measured in the point-measurement study. Cheek jitter was seen in the interleaving study and cheek movement for plosives was reported by observers in the perceptual study. Differences in lower lip movement were more important for the vowels /a/ and /i/. This was evident in both the point-measurement and interleaving studies.

The absence of observable cheek movement in the point-measurement study was an important difference between point measurement and both the interleaving and

perceptual studies. The perceptual study and the interleaving study both involved visual perception rather than measurement of movement. Cheek movement was reported by observers in the perceptual study and jitter was seen in the cheeks in the interleaving experiment; i.e. the differences in movement were visible to observers watching the interleaving and perceptual tapes. The point-measurement study measured only one dimension of the movement and differences were not apparent.

In both point measurement and perceptual studies scores for /i/ were poorer than scores for /a/. The major difference between the point measurement and perceptual studies occurred for the vowel /u/. The perceptual study indicated that for /u/ subjects were able to discriminate the bilabial nasal from the stops, however in the point-measurement data the differences between consonants were less frequent for /u/ than for the other vowels, with most consistent differences for /a/. Two factors may account for this: Differences in cheek movement which may have occurred for /u/ were not demonstrable in the point-measurement data. The second factor is that the cheek movement may have been a subtle change in reflected light that could not be measured using dots on the face as in the point measurement study.

Although point measurement (and interleaving) identified consonant differences in lip movement for /a/, observers did not report using this as a cue whereas they did report using cheek movement for the vowel /u/. The cheek movement was described as either present or absent whereas differences in lip movement were probably more subtle and could not be described by observers. Since scores for /a/ and /i/ were better than chance, observers were no doubt using some cue, but could not identify it as they could for /u/.

Therefore, although the observers in the perceptual study were able to discriminate between /p/, /b/ and /m/ utterances, it was not possible to identify the physical basis for their discrimination. The point-measurement study demonstrated physical differences between utterances, however it was not possible to identify

specific cues that were directly related to the perceived differences. On the other hand, it was clear that /p/ and /b/ were difficult to differentiate both in measurements and in perception, whereas the nasal-plosive discrimination occurred consistently in both the point-measurement and perceptual studies. In addition, in both studies, consonants were more easily differentiated in the /a/ context than in the /i/ context. Thus, when there were more measurable differences between consonants, there was also better perceptual performance.

Coarticulation is a particularly important aspect of speechreading. The cues available to the speechreader are very much dependent on how the consonants and vowels interact. Coarticulatory effects were manifested in slightly different ways in the three different procedures. In the point-measurement study, the greatest number of differences between the consonants occurred for /a/. For example, in contrast M-B, nearly every difference in bell width between /m/ and /b/ was with the vowel /a/.

If the point-measurement curves are compared in the region immediately around closure (i.e. for a duration of approximately 150-200 msec), the effect of the vowel is seen even though the lip positions are far from the extreme vowel opening. The curves for /u/ were much flatter than the curves for /i/ and /a/.

The location of the jitter when two consonants were interleaved varied as a function of the vowel. For /i/ and /a/ there was jitter in the lower lip and some cases of jitter in the cheek. For /u/ the jitter was seen in the cheek only.

The most important manifestation of coarticulation is that shown in the perceptual study since it is the perceptual data that is most relevant to actual speechreading and speechreading training. Differences between /p/, /b/ and /m/ were quite evident in the /u/ context, but were difficult to distinguish in /i/ and /a/ contexts.

Few studies have examined coarticulatory effects in speechreading, however Erber (1971,1974) found better speechreading performance when consonants were

paired with /a/ rather than /i/ or /u/. Benguerel and Pichora-Fuller (1982) found that some phonemes were less affected by phonetic context than other phonemes. Montgomery, Walden and Prosek (1987) showed differences in vowel lipreading performance as a function of consonant visibility, with poorer performance occurring in the higher visibility context.

Speaker/Observer Differences

In all aspects of this investigation there were few differences between speakers and between observers. Moreover, the results of one speaker or observer never contradicted the results of the others. This consistency in the data indicates that it is reasonable to study the speechreading signal and speechreading perception using a small number of subjects.

Note that both speakers were chosen because they were known to be easy to speechread. Both speakers also were from Queens, New York. It would be of interest to do an interleaving study with a large group of speakers, to see whether there are differences in the location of jitter for different faces and for different accents or dialects.

Comparison with other Research

The differences between /i/, /a/ and /u/ observed in the point measurement data are similar to those reported by Fromkin (1964) and Brooke and Summerfield (1983). That is, the curves for /i/ and /a/ were alike, while the curves for /u/ showed much less movement.

One major difference between this perceptual study and previous speechreading research is that in this study there were 100 productions per speaker (for two speakers) to test a single phoneme contrast (e.g. pi-bi). In previous research each stimulus was

recorded no more than ten times (e.g. Walden et al., 1981) or only once with the same production presented five times (e.g. Binnie, Jackson and Montgomery 1976). Other studies have used large groups of subjects (e.g. 35 subjects in Walden et al., 1981), whereas there were only three subjects in this experiment. An advantage of high numbers of tokens per stimulus type is that it is possible to measure small differences on a statistical basis.

Although it was not done in this study, another advantage of multiple tokens is that it would be possible to look at the correct and incorrect responses and determine whether subjects erred on particular tokens. That is, if certain stimuli were "easier", it would be interesting to compare the physical characteristics of "easy" and "difficult" stimuli.

The observers in the perceptual study showed a response bias in favor of /b/. This bias toward the /b/ response is also seen in the data of other studies (e.g. Erber (1974), Walden, et al. (1977), and Owens and Blazek (1985)).

The observers in this study tended not to confuse the nasals and non-nasals which was expected in light of the stop-nasal differences in production. This tendency was also seen in Erber's (1972, 1974) studies. However, this was not seen in the data of some other studies (e.g., Walden et al. (1977) and Binnie, Jackson and Montgomery (1976)). The difference in the results may be due to speaker characteristics. The speakers in the present study were both particularly easy to speechread. In addition, the speechreading task in this study was a very sensitive test and therefore maximized the chances for good performance.

The data obtained by Fujimura (1961) for intervocalic consonants for approximately 20 msec after the onset of lip opening support the finding in this study that the rate of lip movement following closure is greater for /m/ when compared with /p/ and /b/. Differences between intervocalic /p/, /b/ and /m/ were also shown in the

measure of upper and lower lip separation. The separation was greater for /m/ up to approximately 100 ms. (At 100 ms /p/, /b/ and /m/ showed the same value.)

As mentioned previously, Brooke and Summerfield (1983) did not find significant differences between /ama/, /aba/ and /apa/ for either horizontal or vertical measurements. Small differences may have been significant had there been a larger sample size. They used only three productions of each VCV in their analysis. Brooke and Summerfield also tested subjects in an identification task of VCV bisyllables. The subjects were unable to correctly identify /p/, /b/ and /m/ for both natural and synthetic stimuli. Subjects performed well in the present investigation since the task was discrimination rather than identification. The results of speechreading tests are very dependent on the particular task employed.

Implications for Speechreading Research

The perceptual limits for speechreading consonants, vowels and larger speech segments are unknown. Psychophysical data for the visible speech signal is needed in order to identify cues that are perceptually relevant.

One drawback to this study is the inability to directly compare perceptual and physical data since it was not possible to use identical stimuli for all parts of the study. It would have been preferable to compare interleaved and point-measurement CV results with the CV perceptual data since CV and VCV stimuli differ in supraglottal pressure (Arkebauer, Hixon and Hardy, 1967; Flege, 1983; Lisker, 1970; Malecot, 1955). This difference may result in different jitter and/or measurement patterns for CV's and VCV's.

In order to synthesize the visible speech signal it is necessary to have quantitative data. The studies which have quantified articulatory movement have thus far concentrated on vowel differences and on characteristic movements for all bilabials.

These studies have emphasized the differences in movement between different articulators (e.g. lips vs. jaw) for a class of phonemes, rather than the differences between phonemes. As the synthesis technique becomes more refined the relative differences between /p/, /b/ and /m/ can be quantified and incorporated both for research and training purposes.

Implications for Speechreading Training

The perceptual data showed the effect of the vowel on consonant discrimination. The best scores were obtained for /u/. That is, the perception of the differences between /m/ and /p/ or /m/ and /b/, depended on the vowel in the CV. This is an important consideration for for speechreading training. Based upon the data obtained in the perceptual study it is reasonable to train students to differentiate between /mu/ and /pu/ and between /mu/ and /bu/.

The perceptual study also indicates that it may be beneficial for teachers of the deaf to produce plosives with cheek movement. Erber and DeFilippo (1978) suggested that temporary exaggerated articulation that emphasizes a phonemic difference may be useful in speechreading training for the deaf.

The point measurement study indicated several cues that may be important for speechreading training. These were: vertical displacement of the lips and chin, rate of displacement and duration of the displacement. The vertical displacement differences were most consistent of the three measures, however the actual differences were very small and are unlikely to be a useful cue for speechreading. (The vertical displacements ranged from approximately 0.5 to 2.5 mm.) In contrast, the rate of displacement is likely to be perceptually salient although the measured differences in rate of movement between /p/, /b/ and /m/ were not as consistent as the vertical displacement differences.

Speechreading training methods have not emphasized analytical training. Analytic training may be effective, however, if sufficient cues are identified. It is interesting to note, however, that analytical information is valuable for all kinds of speechreading training. Jeffers and Barley (1971), who advocate a synthetic approach to speechreading, stress the importance of analytic knowledge for the speechreading teacher. The teacher can devise lessons in stages of difficulty and student errors can be interpreted analytically.

Intensive speechreading training using cues discussed here may improve performance in a controlled therapy situation but may not effect overall speechreading performance; the efficacy of any training procedures would have to be determined by further study. In addition, the cues that are salient in a CV or VCV context may not be evident in running speech. On the other hand, the effectiveness of some cues may not be apparent until the speechreading signal is combined with auditory or tactile information. That is, the increase in speechreading performance seen with the addition of auditory or tactile cues may be even greater when visual speechreading cues are emphasized as well.

The differences between /p/, /b/ and /m/ produced by the speakers in this study made it possible to discriminate between those consonants. It is arguable that the speech of the two speakers was in some way idiosyncratic and that the differences observed here may not occur for most speakers. On the other hand, the implication of the results in this study is that even if these differences are not naturally occurring in the speech of some talkers, it is possible to produce these phonemes so that differences do exist. That is, parents and teachers of the deaf may be trained to produce speech that contains more salient cues for speechreading discrimination.

This study showed the feasibility of computerized video recordings. This would be beneficial as well for speechreading training for clinic, school and home use. The

exact stimuli can be duplicated and recordings can be computer modified to enhance the signal or vary time and space characteristics of the signal. This study used a forced choice method which resulted in a high performance level. Video recorded tapes of forced choice tasks with feedback would provide systematic speechreading training. This can be done on the phonemic level as in this study, or on the word and sentence level. In this study the differences between the two speakers was insignificant. This also suggests that good performance with one speaker on a training tape may generalize to various speakers.

Conclusions

1. Each of the three methods showed consistent differences between /p/, /b/ and /m/. The three methods provided three different ways for examining contrasts between homophenous sounds. Each method was more sensitive than the others in some way. For example, point measurement in this experiment was not very sensitive to cheek movement, however it was very sensitive to upper lip movement. Interleaving showed the existence of lower lip and cheek cues but was not sensitive to upper lip movement. Of the three methods the most important from a practical standpoint was the perceptual method because it demonstrated optimal discrimination performance.

2. The perceptual study indicated that the nasal/plosive distinction was more perceptible than the /p/-/b/ distinction. It is necessary, therefore, to reconsider whether /p/, /b/ and /m/ form a single viseme.

3. Vowel context was clearly an important factor. All three methods showed coarticulatory effects.

4. While these studies did not reveal a single physical correlate for the good performance on the perceptual task, they showed that when there were more physical differences there was better perceptual performance.

5. This investigation showed that the common notion of homophenous sounds should not be used to set expectations for speechreading performance. The point measurement study revealed differences between so-called homophenous sounds. The perceptual study showed that untrained speechreaders were able to perceive differences in a controlled test situation.

6. Video recordings proved to be effective for revealing subtle differences. It is not necessary to use a live test situation for subjects to perceive speechreading differences.

7. It is possible to study speechreading with few subjects since there were few speaker/observer differences.

CHAPTER VI

SUMMARY

There were two purposes to this investigation. The first was to compare the visible articulatory movements associated with three English consonants that were considered homophenous: /p/, /b/ and /m/. The movement patterns were examined in three vowel contexts (/i/, /a/, and /u/) and for two different speakers.

The second purpose of the investigation was to measure the ability of subjects to visually discriminate these so-called homophenous sounds using a forced choice discrimination task.

The study consisted of three experiments: In the first experiment the points of measurement were determined using a technique called interleaving. In the second experiment these points were used to measure facial movements. The third experiment of the study consisted of a speechreading task. A set of homophenous utterances was used as test material in all three parts of the study.

In the interleaving and point measurement experiments the bilabial consonants /p/, /b/ and /m/ were combined with the vowels /i/, /a/ and /u/ to form VCV disyllables in which the same vowel was in the initial and final position. The VCV's were spoken in groups of three, called triplets. The utterances were spoken by two speakers.

The utterances were video recorded and digitized using a computer system designed for video processing. The video signals could then be processed or could be viewed using a palindromic display format. The use of a palindromic display eliminated discontinuities in the video display.

In the interleaving study, two video recorded sequences were combined into one sequence. Odd frames of one sequence were alternated with even frames of the other sequence to make a new video recording. The interleaved recording showed articulatory movements that appeared quite natural except for areas of jitter in those regions of the face that were not in identical positions at the same time during the production of the two utterances. The areas of jitter indicated the locations of interest for detailed measurement.

Consonants in the initial and final position within the triplets were interleaved with each other and with the consonants in the medial position. Some of the interleaved samples consisted of the two different productions of the same CV. This was done to test whether there was any observable jitter for identical phonemes.

For both speakers jitter was observed primarily in the cheeks, lower lip and chin. When two different productions of the same consonant were interleaved there was no jitter for all samples for Speaker 1 and no jitter in seven out of eleven samples for Speaker 2.

In the second experiment, the points of measurement were chosen so as to sample those regions of the face that showed jitter in the interleaving experiment. There were fifteen points of measurement for Speaker 1 and twenty-one points of measurement for Speaker 2.

The speakers' faces were blackened and white dots (called face markers) were placed at the points determined by the interleaving method. The computer was programmed to identify white dots from a dark background, thereby providing automatic tracking of facial movements as indicated by the dots.

A plot of the face marker movement was generated for each VCV utterance. The movement of each face marker was analyzed by visually comparing pairs of curves. Differences between pairs of plots were measured by comparing the height (vertical

displacement), width (duration) and slope (rate of displacement) of the curves. The data obtained from three face markers were analyzed in great detail. These points were located in the mid upper lip, mid lower lip and the chin.

The major findings of the point measurement study were as follows: The lower lip and chin move further upward for /m/ than for /b/ or /p/. The lower lip moves further upward for /b/ than for /p/. The upper lip moves less downward for /m/ than for /p/ and /b/. That is, the chin and both lips are displaced further upward for the production of /m/ than for /p/ and /b/. A second finding was that the rate of movement away from closure for the lower lip and chin is greater for /m/ than for /p/ or /b/. The rate of movement of the upper lip towards closure is faster for /p/ and /b/ than for /m/. A third observation was that there were only subtle durational differences measured between /p/, /b/ and /m/.

The vertical displacement of the midpoint of the lower lip was the most consistent measure, followed by the vertical displacement of the midpoint of the upper lip and the rate of displacement of the midpoint of the lower lip. Thus, the midpoint of the lips was most important in distinguishing between /p/, /b/ and /m/.

In the third experiment three observers participated in a speechreading study. The two-way discrimination between /m/ and /b/, /m/ and /p/, and /p/ and /b/ was tested in a CV context with the vowels /i/, /a/ and /u/. The utterances were spoken by two speakers. The stimuli were video recorded and consisted of CV pairs (e.g. /pa/-/ba/), called targets, presented in a four alternative forced-choice test.

The data were analyzed in two parts: an analysis of correct responses and an analysis of error responses.

The results of the speechreading study showed that observers were able to discriminate between consonants with the same place of production. The proportion of correct responses was significantly better than chance for twenty-two out of twenty-

four targets for /m/ vs. /b/ and /m/ vs. /p/ discrimination tests. For /p/ vs. /b/ the data for approximately half the targets resulted in proportions significantly better than chance.

Performance was affected by the vowel environment. The scores for the vowel /u/ were better than 75% for /m/ vs. /p/ or /b/. For /i/ and /a/ the scores were in the 35-50% range. Subjects reported that for /u/, cheek movement during /p/ and /b/ utterances enabled them to discriminate between /m/ and /b/ and between /m/ and /p/.

The results suggest that the traditional grouping of /p/, /b/ and /m/ into one viseme may not be correct since /m/ was contrastive with both /p/ and /b/. However, /p/ and /b/ were sufficiently confused with each other to be considered a single viseme. Thus the grouping of /p,b,m/ might better be described as two visemes: /p,b/ and /m/.

In each of the three experiments there were consistent differences between /p/, /b/ and /m/. Of the three methods the most important from a practical standpoint was the speechreading test because it demonstrated optimal discrimination performance.

Advances in computerized video technology will result in more analytical studies of the speechreading process. In future research of visible articulatory movement the face markers should be tracked using a system better designed for dot detection during movement. A mathematical comparison of the plots of face marker movement would be more precise than the visual comparison of curves used in the present investigation.

This study looked at a frontal view of the face. Future work in this area could include profile views which would show lip protrusion and anterior-posterior chin and cheek movements.

In addition to the description of consonant and vowel differences in the speechreading signal, the point measurement technique may be used to study the effects of suprasegmental changes or distortion (e.g. smiling) on the speechreading signal.

The point measurement approach may also be used to study speech production. For example, the speech of deaf speakers may be analyzed using this method. Deviant lip, cheek and jaw movements may then be modified in order to increase intelligibility.

Future research in speechreading perception should be done with various groups of speechreaders. It is important to quantify differences in performance between good and poor speechreaders. The results with trained expert speechreaders may be used to set expectations for speechreading training programs.

Speechreading teachers and speakers in speechreading studies may be trained to exaggerate differences identified in point measurement studies. Assessments of speaker intelligibility for speechreading would indicate which differences are salient. Further investigation would reveal whether teachers trained to produce such differences are more effective in a speechreading training program.

APPENDICES

APPENDIX A
NUMBER OF STIMULI PER TARGET IN EACH RECORDED SUBTEST

CONTRAST M-B

	Targets			
	BB	MM	BM	MB
Speaker 1				
/a/	25	24	25	26
/i/	25	25	25	25
/u/	25	25	25	24
Speaker 2				
/a/	25	25	25	25
/i/	25	25	25	25
/u/	25	25	25	24

CONTRAST M-P

	Targets			
	PP	MM	PM	MP
Speaker 1				
/a/	25	25	25	25
/i/	24	25	26	25
/u/	25	25	25	25
Speaker 2				
/a/	25	25	25	25
/i/	24	26	26	24
/u/	25	25	25	25

CONTRAST P-B

	Targets				
	PP	BB	PB	BP	
Speaker 1					
/a/	24	25	27*	24	*indicates a
/i/	25	25	24	26	deviation greater
/u/	25	25	25	25	than + or - 1 from
Speaker 2					25 stimuli per
/a/	26	25	25	24	target
/i/	25	24	25	25	
/u/	24	24	18*	22*	

APPENDIX B

CURVE COMPARISON DATA

The data in the following nine tables indicate the number of curve comparisons showing greater slope, width or height for /m/ vs. /b/, /m/ vs. /p/ and /p/ vs. /b/.

Table 1. Contrast M-B; Slope

	/i/		/a/		/u/							
	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	i	a	u	Σ
RIGHT SLOPE												
MLLP												
m>b	7	5	7	4	5	0	19	9	12	11	5	28
b>m	0	0	0	1	1	3	1	4	0	1	4	5
n	10	9	8	6	6	5	24	20	19	14	11	44
	**	**	**	*	**	**	**	**	**	**		
LLLP												
m>b	4	5	4	4	1	1	9	10	9	8	2	19
b>m	1	0	0	1	2	2	3	3	1	1	4	6
n	10	9	10	6	6	5	26	20	19	16	11	46
		**	**	*			*	*	**		**	**
CHIN												
m>b	4	2	3	4	3	1	10	7	6	7	4	17
b>m	1	2	0	1	0	0	1	3	3	1	0	4
n	10	9	8	6	4	5	22	20	19	14	9	42
			*	*	**		**		**	**		**
LEFT SLOPE												
MULP												
m>b	1	1	2	2	0	0	3	3	2	4	0	6
b>m	5	3	4	1	2	2	11	6	8	5	4	17
n	9	9	10	6	6	5	25	20	18	16	11	45
	*						**		*		*	**

*Significant at .05 level

**Significant at .01 level

Table 2. Contrast M-B; Width

BELL WIDTH

	/i/		/a/		/u/							
	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	i	a	u	Σ
MLLP												
m>b	3	5	2	0	2	3	7	8	8	2	5	15
b>m	6	1	4	4	3	2	13	7	7	8	5	20
n	10	9	8	6	6	5	24	20	19	14	11	44
		*			**					**		
LLLP												
m>b	2	4	2	0	5	2	9	6	6	2	7	15
b>m	6	4	4	6	1	0	11	10	10	10	1	21
n	10	9	10	6	6	5	26	20	19	16	11	46
	*				**	**				**	**	
CHIN												
m>b	1	4	2	0	1	1	4	5	5	2	2	9
b>m	6	3	4	5	3	2	13	10	9	9	5	23
n	10	9	8	6	4	5	22	20	19	14	9	42
	**				**		**			**		**

TROUGH WIDTH

MULP												
m>b	5	7	1	2	1	2	7	11	12	3	3	18
b>m	1	2	4	1	2	0	7	3	3	5	2	10
n	9	9	10	6	6	5	25	20	18	16	11	45
	*	**						**	**			

*Significant at .05 level
**Significant at .01 level

Table 3. Contrast M-B; Height

BELL HEIGHT

	/i/		/a/		/u/							
	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	i	a	u	Σ
MLLP												
m>b	7	5	6	5	3	3	16	13	12	11	9	29
b>m	2	3	1	0	2	0	5	3	5	1	2	8
n	10	9	8	6	6	5	24	20	19	14	11	44
	**		**	**		**	**	**	**	**	**	**
LLLP												
m>b	3	3	4	4	3	2	10	9	6	8	5	19
b>m	2	1	3	0	1	0	6	1	3	3	1	7
n	10	9	10	6	6	5	26	20	19	16	11	46
				**				**		*	*	**
CHIN												
m>b	6	3	6	5	1	2	13	10	9	11	3	23
b>m	2	3	1	0	1	1	4	4	5	1	2	8
n	10	9	8	6	4	5	22	20	19	14	9	42
	*		**	**			**	*		**		**

TROUGH HEIGHT

MULP												
m>b	5	4	7	4	3	3	15	11	9	11	6	26
b>m	0	2	1	0	0	0	1	2	2	1	0	3
n	9	9	10	6	6	5	25	20	18	16	11	45
	**		**	**	*	**	**	**	*	**	**	**

*Significant at .05 level

**Significant at .01 level

Table 4. Contrast M-P; Slope

	/i/		/a/		/u/							
	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	i	a	u	Σ
RIGHT SLOPE												
MLLP												
m>p	8	5	7	4	5	1	20	10	13	11	6	30
p>m	0	0	1	0	0	2	1	2	0	1	2	3
n	10	9	8	6	6	5	24	20	19	14	11	44
	**	**	**	**	**		**	**	**	**		**
LLLP												
m>p	6	5	8	4	2	1	16	10	11	12	3	26
p>m	0	0	1	1	1	1	2	2	0	2	4	4
n	10	9	10	6	6	5	26	20	19	16	11	46
	**	**	**	*			**	**	**	**		**
CHIN												
m>p	5	3	3	4	4	3	12	10	8	7	7	22
p>m	2	2	0	1	0	1	2	4	4	1	1	6
n	10	9	8	6	4	5	22	20	19	14	9	42
			*	*	**		**	*		**	**	**
LEFT SLOPE												
MULP												
m>p	2	3	1	0	0	1	3	4	5	1	1	7
p>m	4	1	7	3	3	2	14	6	5	10	5	20
n	9	9	10	6	6	5	25	20	18	16	11	45
			**	*	*		**			**	*	**

*Significant at .05 level

**Significant at .01 level

Table 5. Contrast M-P; Width

	/i/		/a/		/u/							
	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	i	a	u	Σ
BELL WIDTH												
MLLP												
m>p	4	7	2	3	1	3	7	13	11	5	4	20
p>m	6	1	6	3	2	1	14	5	7	9	3	19
n	10	9	8	6	6	5	24	20	19	14	11	44
		**	*				**	**				
LLLP												
m>p	2	4	0	1	3	1	5	6	6	1	4	11
p>m	5	4	7	4	1	1	13	9	9	11	2	22
n	10	9	10	6	6	5	26	20	19	16	11	46
			**	*			*			**		*
CHIN												
m>p	4	2	2	1	0	1	6	4	6	3	1	10
p>m	6	2	4	3	2	3	12	8	8	7	5	20
n	10	9	8	6	4	5	22	20	19	14	9	42
					*	**			**		*	*
TROUGH WIDTH												
MULP												
m>p	5	6	5	4	4	3	14	13	11	9	7	27
p>m	2	1	2	1	1	0	5	3	4	3	1	8
n	9	9	10	6	6	5	25	20	18	16	11	45
		*		*	*	**	**	**	*	*	**	**

*Significant at .05 level
**Significant at .01 level

Table 6. Contrast M-P; Height

	/i/		/a/		/u/							
	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	i	a	u	Σ
BELL HEIGHT												
MLLP												
m>p	9	5	7	5	6	5	22	15	14	12	11	37
p>m	0	1	1	1	0	0	1	2	1	2	0	3
n	10	9	8	6	6	5	24	20	19	14	11	44
	**	*	**	**	**	**	**	**	**	**	**	**
LLLP												
m>p	5	5	7	5	4	4	16	14	10	12	8	30
p>m	0	1	3	1	0	0	3	2	1	4	0	5
n	10	9	10	6	6	5	26	20	19	16	11	46
	**	*		**	**	**	**	**	**	**	**	**
CHIN												
m>p	7	3	6	4	4	3	17	10	10	10	7	27
p>m	0	2	2	2	0	0	2	4	2	4	0	6
n	10	9	8	6	4	5	22	20	19	14	9	42
	**		*		**	**	**	*	**	*	**	**
TROUGH HEIGHT												
MULP												
m>p	8	4	10	5	5	5	23	14	12	15	10	37
p>m	0	1	0	0	0	0	0	1	1	0	0	1
n	9	9	10	6	6	5	25	20	18	16	11	45
	**		**	**	**	**	**	**	**	**	**	**

*Significant at .05 level

**Significant at .01 level

Table 7. Contrast P-B; Slope

RIGHT SLOPE

	/i/		/a/		/u/							
	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	i	a	u	Σ
MLLP												
p>b	2	3	2	2	1	1	5	6	5	4	2	11
b>p	3	3	4	0	3	1	10	4	6	4	4	14
n	10	9	8	6	6	5	24	20	19	14	11	42
LLLP												
p>b	0	2	1	2	0	1	1	5	2	3	1	6
b>p	3	1	7	1	1	2	11	4	4	8	3	15
n	10	9	10	6	6	5	26	20	19	16	11	46
	*		**				**			*		*
CHIN												
p>b	1	3	1	4	0	2	2	9	4	5	2	11
b>p	3	1	2	1	2	2	7	4	4	3	4	11
n	10	9	8	6	4	5	22	20	19	14	9	42
				*	*							

LEFT SLOPE

MULP												
p>b	2	2	4	5	0	2	6	9	4	9	2	15
b>p	2	5	1	0	0	1	3	6	7	1	1	9
n	9	9	10	6	6	5	25	20	18	16	11	45
				**						**		

*Significant at .05 level

**Significant at .01 level

Table 8. Contrast P-B; Width

BELL WIDTH

	/i/		/a/		/u/							
	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	i	a	u	Σ
MLLP												
p>b	4	2	4	0	2	2	10	4	6	4	4	14
b>p	3	6	3	5	1	2	7	13	9	8	3	20
n	10	9	8	6	6	5	24	20	19	14	11	44
		*		**				**				
LLLP												
p>b	4	2	3	1	2	1	9	4	6	4	3	13
b>p	1	4	2	5	1	2	4	11	5	7	3	15
n	10	9	10	6	6	5	26	20	19	16	11	46
				**				*				
CHIN												
p>b	2	0	2	0	1	1	5	1	2	2	2	6
b>p	3	6	5	5	1	1	9	12	9	10	2	21
n	10	9	8	6	4	5	22	20	19	14	9	42
		**		**				**	**	**		**

TROUGH WIDTH

MULP												
p>b	5	6	2	1	1	0	8	7	11	3	1	15
b>p	3	1	5	4	2	1	10	6	4	9	3	16
n	9	9	10	6	6	5	25	20	18	16	11	45
		**		*					**	*		

*Significant at .05 level
**Significant at .01 level

Table 9. Contrast P-B; Height

BELL HEIGHT

	/i/		/a/		/u/							
	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	Spkr1	Spkr2	i	a	u	Σ
MLLP												
p>b	2	1	0	4	0	1	2	6	3	4	1	8
b>p	6	3	5	1	3	4	14	8	9	6	7	22
n	10	9	8	6	6	5	24	20	19	14	11	44
	*		**	*	*	*	**		*		**	**
LLLP												
p>b	0	2	2	3	0	1	2	6	2	5	1	8
b>p	3	4	6	2	2	4	11	10	7	8	6	21
n	10	9	10	6	6	5	26	20	19	16	11	46
	*		*			*	**		*		**	**
CHIN												
p>b	3	3	1	4	1	0	5	7	6	5	1	12
b>p	4	3	4	2	3	4	11	9	7	6	7	20
n	10	9	8	6	4	5	22	20	19	14	9	42
						**	*				**	

TROUGH HEIGHT

MULP												
p>b	2	4	0	0	0	0	2	4	6	0	0	6
p>m	5	2	9	2	4	2	18	6	7	11	6	24
n	9	9	10	6	6	5	25	20	18	16	11	45
			**		**		**			**	**	**

*Significant at .05 level

**Significant at .01 level

APPENDIX C

PROPORTION CORRECT FOR EACH TARGET WITHIN EACH SUBTEST

CONTRAST M-B

	Target			
	BB	MM	BM	MB
Subtest				
Observer B				
Speaker 1				
/a/	.48	.42	.48	.58
/i/	.52	.28	.36	.44
/u/	.72	.92	.76	.88
Speaker 2				
/a/	.72	.24	.44	.52
/I/	.44	.20	.20	.24
/u/	.84	.72	.92	.75
Observer C				
Speaker 1				
/a/	.44	.25	.56	.42
/i/	.60	.16	.60	.20
/u/	.56	.88	.68	.83
Speaker 2				
/a/	.48	.32	.36	.48
/i/	.48	.24	.52	.28
/u/	.84	.60	.60	.54
Observer L				
Speaker 1				
/a/	.32	.21	.40	.42
/i/	.48	.44	.20	.20
/u/	.80	.92	.80	.71
Speaker 2				
/a/	.56	.68	.52	.64
/i/	.40	.32	.40	.52
/u/	.92	.72	.56	.67

__underlined scores are significantly better than chance

CONTRAST M-P

	Target			
	PP	MM	PM	MP
Subtest				
Observer B				
Speaker 1				
/a/	.56	.28	.44	.76
/i/	.67	.40	.35	.48
/u/	.84	.96	.88	1.00
Speaker 2				
/a/	.56	.36	.20	.56
/I/	.50	.39	.42	.50
/u/	.76	.84	.92	.84
Observer C				
Speaker 1				
/a/	.64	.32	.40	.48
/i/	.54	.20	.15	.32
/u/	.80	1.00	.76	.96
Speaker 2				
/a/	.48	.36	.48	.40
/i/	.46	.58	.50	.50
/u/	.76	.84	.60	.52
Observer L				
Speaker 1				
/a/	.40	.32	.40	.60
/i/	.58	.32	.15	.24
/u/	.72	1.00	.72	.92
Speaker 2				
/a/	.56	.44	.48	.48
/i/	.42	.58	.12	.25
/u/	.84	.84	.80	.76

__underlined scores are significantly better than chance

CONTRAST P-B

	Target			
	PP	BB	PB	BP
Subtest				
Observer B				
Speaker 1				
/a/	.08	.36	.37	.50
/i/	.24	.28	.25	.42
/u/	.44	.40	.32	.56
Speaker 2				
/a/	.15	.40	.20	.63
/i/	.36	.33	.20	.44
/u/	.29	.50	.17	.32
Observer C				
Speaker 1				
/a/	.21	.08	.41	.58
/i/	.36	.12	.25	.31
/u/	.12	.32	.24	.44
Speaker 2				
/a/	.35	.32	.12	.21
/i/	.40	.29	.28	.28
/u/	.33	.75	.33	.27
Observer L				
Speaker 1				
/a/	.21	.32	.26	.38
/i/	.32	.36	.38	.38
/u/	.48	.40	.36	.40
Speaker 2				
/a/	.23	.40	.32	.25
/i/	.52	.33	.20	.20
/u/	.58	.58	.17	.41

__underlined scores are significantly better than chance

REFERENCES

Arkebauer, H., Hixon, T., and Hardy, J. 1967. Peak intraoral air pressure during speech. Journal of Speech and Hearing Research, 10, 196-208.

Bell-Berti, F. and Harris, K.S. 1981. A temporal model of speech production. Phonetica, 38, 9-20.

Benguerel, A-P., and Pichora-Fuller, M.K. 1982. Coarticulation effects in lipreading. Journal of Speech and Hearing Research, 25, 600-607.

Binnie, C.A., Jackson, P.L., and Montgomery, A.A. 1976. Visual intelligibility of consonants: A lipreading screening test with implications for aural rehabilitation. Journal of Speech and Hearing Disorders, 41, 530-539.

Binnie, C.A., Montgomery, A.A., and Jackson, P.L. 1974. Auditory and visual contributions to the perception of consonants. Journal of Speech and Hearing Research, 17, 619-630.

Borden, G.J., and Harris, K.S. 1984. Speech Science Primer Second Edition, Baltimore: Williams and Wilkins.

Brooke, N.M., McGrath, M., and Summerfield, Q. 1983. Analysis, synthesis, and perception of visible articulatory movements. Journal of Phonetics, 11, 63-76.

Brooke, N.M., McGrath, M., and Summerfield, Q. 1984. Visual speech perception experiments using a video speech synthesizer. Paper presented at the 108th meeting of the Acoustical Society of America, Minneapolis.

Brownlee, K.A. 1965. Statistical Theory and Methodology in Science and Engineering Second Edition, New York: John Wiley and Sons.

Erber, N.P. 1971. Effects of distance on the visual reception of speech. Journal of Speech and Hearing Research, 14, 848-857.

Erber, N.P. 1972. Auditory, visual, and auditory-visual recognition of consonants by children with normal and impaired hearing. Journal of Speech and Hearing Research, 15, 413-422.

Erber, N.P. 1974a. Discussion: Lipreading Skills (Chapter 3) In R.E. Stark (Ed.) Sensory Capabilities of Hearing-Impaired Children. Baltimore: University Park Press.

Erber, N.P. 1974b Effect of angle, distance and illumination on visual reception of speech by profoundly deaf children. Journal of Speech and Hearing Research, 17, 99-112.

Erber, N.P. 1979 Real-time synthesis of optical lip shapes from vowel sounds. Journal of the Acoustical Society of America, 66, 1542-1544.

Erber, N.P., and De Filippo, C.L. 1978. Voice/mouth synthesis and tactual/visual perception of /pa,ba,ma/. Journal of the Acoustical Society of America, 64, 1015-1019.

Erber, N.P., Sachs, R.M., and De Filippo, C.L. 1979. Optical synthesis of articulatory images for lipreading evaluation and instruction. In D.L. McPherson (Ed.), Advances in Prosthetic Devices for the Deaf. Rochester, N.Y.: Nat. Tech. Inst. Deaf.

Fisher, C.G. 1968. Confusions among visually perceived consonants. Journal of Speech and Hearing Research, 11, 796-802.

Flege, J.E. 1983. The influence of stress, position, and utterance length on the pressure characteristics of English /p/ and /b/. Journal of Speech and Hearing Research, 26, 111-118.

Franks, J.R., and Kimble, J. 1972. The confusion of English consonant clusters in lipreading. Journal of Speech and Hearing Research, 15, 474-482.

Fromkin, V. 1964. Lip positions in American English vowels. Language and Speech, 7, 215-225.

Fujimura, O. 1961. Bilabial stop and nasal consonants: A motion picture study and its acoustical implications. Journal of Speech and Hearing Research, 4, 233-247.

Harris, K.S., Lysaught, G..F., and Schvey, M.M. 1965. Some aspects of oral and nasal labial stops. Language and Speech, 8, 135-147.

Heider, F., and Heider, G. 1940. An experimental investigation of lipreading. Psychological Monographs, 52, 124-153.

Jackson, P.L., Montgomery, A.A., and Binnie, C.A. 1976. Perceptual dimensions underlying vowel lipreading performance. Journal of Speech and Hearing Research, 19, 796-812.

Jeffers, J., and Barley, M. 1971. Speechreading. Springfield, Il: Charles C. Thomas.

Joergenson, A. 1962. The measurement of homophenous words. Masters Thesis, Michigan State University.

Lisker, L. 1970. Supraglottal air pressure in the production of English stops. Language and Speech, 13, 215-230.

Lubker, J.F., and Parris, P.J. 1970. Simultaneous measurements of intraoral pressure, force of labial contact, and labial electromyographic activity during production of the stop consonant cognates /p/ and /b/. Journal of the Acoustical Society of America, 47, 625-633.

Malecot, A. 1955. An experimental study of force of articulation. Studia Linguistica, 9, 35-44.

McCutcheon, M.J., Fletcher, S.G., and Hasegawa, A. 1977. Video-scanning system for measurement of lip and jaw motion. Journal of the Acoustical Society of America, 61, 1051-1055.

Montgomery, A.A., Walden, B.E., and Prosek, R. 1987. Effects of consonantal context on vowel lipreading. Journal of the Acoustical Society of America, 30, 50-59.

Owens, E., and Blazek, B. 1985. Visemes observed by hearing-impaired and normal-hearing adult viewers. Journal of Speech and Hearing Research, 28, 381-393.

Sonoda, Y., and Wanishi, S. 1982. New optical method for recording lip and jaw movements. Journal of the Acoustical Society of America, 72, 700-704.

Sussman, H.M., MacNeilage, P.F., and Hanson, R.J. 1973. Labial and mandibular dynamics during the production of bilabial consonants: Preliminary observations. Journal of Speech and Hearing Research, 16, 397-420.

Tatham, M.A.A., Daniloff, R.G., and Hoffman, P.R. 1985. Electromyographic invariance of lip closure for /p/-/b/. In R.G. Daniloff (Ed.), Speech Science, (pp 279-311). San Diego: College-Hill Press.

Tatham, M.A.A. and Morton, K. 1973. Electromyographic and intraoral air pressure studies of bi-labial stops. Language and Speech, 16, 336-350.

Walden, B.E., Erdman, S.E., Montgomery, A.A., Schwartz, D.M., and Prosek, R.A. 1981. Some effects of training on speech recognition by hearing impaired adults. Journal of Speech and Hearing Research, 24, 207-216.

Walden, B.E., Montgomery, A.A., and Prosek, R.A. 1987. Perception of synthetic visual consonant-vowel articulations. Journal of the Acoustical Society of America, 30, 418-424.

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. Journal of Speech and Hearing Research, 20, 130-145.

Walden, B.E., Prosek, R.A., and Worthington, D.W. 1974. Predicting audio-visual consonant recognition performance of hearing-impaired adults. Journal of Speech and Hearing Research, 17, 270-278

Walden, B.E., Prosek, R.A., and Worthington, D.W. 1975. Auditory and audiovisual feaature transmission in hearing-impaired adults. Journal of Speech and Hearing Research, 18, 272-280.

Woodward, M.F. 1957 Linguistic methodology in lip reading. In John Tracy Research Papers IV. Los Angeles: John Tracy Clinic.

Woodward, M.F., and Barber, C.G. 1960. Phoneme perception in lipreading. Journal of Speech and Hearing Research, 3, 212-222.