This is to certify that the
dissertation entitled

THE INFLUENCE OF COLOR AND VOWEL CONTEXT ON THE VISUAL PERCEPTION OF /p, b, m/

presented by

Willard Charles Hooks, Jr.

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Audiology

Major Professor's Signature

12-08-03

Date

THE INFLUENCE OF COLOR AND VOWEL CONTEXT ON THE VISUAL PERCEPTION OF /p, b, m/

VOLUME I

By

Willard Charles Hooks, Jr.

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Audiology and Speech Sciences

2003

ABSTRACT

THE INFLUENCE OF COLOR AND VOWEL CONTEXT

ON THE VISUAL PERCEPTION OF /p, b, m/

By

Willard C. Hooks, Jr.

Over the past 70 years, there have been few innovative strategies to enhance speechreading accuracy and efficiency. This study was designed to determine whether color and vowel context were significant influences on the speechreading ability of normal-hearing subjects. Traditional approaches to aural rehabilitation have emphasized strategies using visual aids, repetition, the manipulation of film and video production. However, there are no definitive approaches to speechreading training that optimize accuracy.

The /p, b, m/ viseme-cluster has been known to cause confusion in both normal-hearing and hearing-impaired groups. Several variables can influence the outcome of speechreading accuracy: (1) talker differences, (2) phonological factors, and (3) speechreading training.

In the case of speechreading, visual information from key regions of the face (lips, cheeks, eyes) must be directed to the color-sensitive foveal region of the eye for closer examination and for decision-making. Up until now, it has not been established whether viseme-cluster recognition could improve with the interaction of several independent factors—for example, color vision conditions,

vowel context, and the /p, b, m/ viseme-cluster. Erber (1974b) reported that confusion of /p/, /b/, and /m/ occurred for the vowels /i/, /ɑ/, and /u/. This investigator examined the hypothesis that the interaction of opponent colors and the vowels /i/ and /ɑ/ are the sources of variation for the accurate visual recognition of phonemes /p, b, m/.

Based on opponent color theory, it is physiologically impossible for humans (with normal color vision) to perceive simultaneously the opponent colors red and green, blue and yellow. To produce opponent lip-color conditions, the lips of an actor were cosmetically treated to create four color and one, natural experimental condition.

There were two experiments with the phonemes/p, b, m/: one for vowel-consonant-vowel (VCV) disyllables in the vowel context of /i/ and one for VCVs in the vowel context of /ɑ/. Six subjects, naïve to speechreading, viewed a video production of one talker speaking VCV in five color combinations. The subjects marked their perceived selection for each phoneme in a three-choice multiple-choice format for a total of 2,160 observations per experiment.

Based on logistic regression analysis, the results indicated that the null hypothesis was not rejected for either Experiment I or II. However, for some individuals, color was an important variable for visual recognition. The findings warrant more study of this different approach for aural rehabilitation. An established cross-modal perception between color vision and hearing could be key for some individuals to maximize visual speech perception.

Dedicated to my wife for her faith, insight, and encouragement

ACKNOWLEDGMENTS

I express my gratitude to Dr. Leo V. Deal, mentor, advisor, and chairperson of my dissertation committee for his guidance, encouragement, and belief in me. He provided me with an example of how to be a gentleman and a scholar during my study at Michigan State University (MSU).

My sincere thanks are extended to Dr. Richard Ball for his insightful guidance in the area of vision science and his moral support for the completion of this research. Thanks also are due to Dr. Mangala Sadasivan for her direction and expertise in the area of language processing. I thank Dr. Linda Lou Smith and Dr. Jill Elfenbein for their input on this research.

My sincere gratefulness is extended to Dr. Bradge Golding for his insight and advice in the area of spectral optics, and for technical assistance from the MSU Center for Sensor Materials. I am appreciative to his graduate student Ryan Kruz, for his proficiency and assistance in quantifiable color matching.

A special appreciation is due to Rich Tibbals for his recommendations and logistical support for the video production component of this research project. I am grateful to Gary Roettger, for his technical skill and knowledge during the video production. My gratitude is extended to Jonathan Cook and Eric Swanson for their work in the studio setup, lighting, cameras, and control-room switching.

I extend my thankfulness to Greta Giest for her assistance and logistical support from the makeup division of the MSU Theater Department. My thanks are also due to Emily Mickelson for her skillfulness, in the application of cosmetic

makeup during the video production phase of the project. I am grateful to Mark Olson for his assistance from the student acting division of the MSU Theater Department. I am appreciative to Amy Handleman for her participation and speaking skill during the video production phase of the project.

I express my gratitude to Dr. James Stapleton of the MSU Statistics and Probability Department for his insight and recommendations regarding the analysis of this research. A special appreciation is due to Fang Li for her statistical advice through the MSU Statistical Consulting Service.

My heartfelt thanks go to the Hooks and Fang families for their prayers and unvarying belief that I would complete this task. My deepest gratitude and thanks I extend to my wife Dr. Shi-Ruei S. Fang for her understanding, patience, and encouragement as we reached this goal.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF FORMS

KEY TO SYMBOLS OR ABBREVIATIONS

Chapter 1

SPEECHREADING: POINTS OF CONVERGENCE

The National Institute on Deafness and Other Communication Disorders (1995) reported that there are over 25 million hearing-impaired individuals in the United States of America. These individuals may have difficulty understanding speech especially in the presence of background noise. Many have probably become speechreaders without any training. However, there may be a substantial number of hearing-impaired individuals who require formal instruction. Since 1938, few new methods or new approaches to teaching speechreading have been introduced (Arnold, 1993). This chapter will serve to (1) provide an introduction and background to speechreading and (2) emphasize the need for a new approach by exploring the interaction between color perception and speechreading. Color-vision has been implicated as (1) a component for information processing (McGurk & MacDonald, 1976; Treisman, and Davies, 1973), (2) a contributor to depth perception (Triesman, 1986), and (3) a factor in motion detection (Livingstone & Hubel, 1987, 1988). As far as this investigator can determine, there has been no research to show a relationship between color vision and speechreading. Specifically, this investigation examined the visual consonant recognition of /p, b, m/ under the influence of vowel context and the opponent colors red-green and blue-yellow.

Auditory and Visual Input

An auditory system, compromised by hearing impairment reduces an individual's ability to interpret the message from audition alone. The visual channel, accordingly, assumes a greater role for speech perception (Erber, 1979; Massaro 1987; O'Neill, and Oyer, 1981). O'Neill (1954) suggested that the eyes alone might compensate for information lost from a hearing impairment.

The hearing-impaired person, by the very nature of the impairment, is forced to utilize speechreading as a major support for impaired auditory information (Hardy, 1970). Hardy related, in addition, that there was reason to suspect that there may be undetermined biologic factors that underlie the process of speechreading—for example, color vision, fast visual processing. Some of these issues are not well understood.

Aural Rehabilitation

Traditionally, aural rehabilitation (AR) involves the combined input of audition and vision to improve communication. Since sensory input is fundamental to communication, a clinician's initial approach to therapy goals should consider both visual acuity and residual hearing.

Auditory Training

Sanders (1982) defined auditory training as a systematic procedure designed to increase the amount of information that a person's hearing contributes to his total perception. The current practice of amplification—through hearing aids, cochlear implants, and certain assistive listening devices (ALD)—provides a variety of benefits to the person who is hearing-impaired. It also assumes that the hearing-impaired individual possesses residual hearing (the remaining usable hearing after a loss). The desired results of amplification include increased speech intelligibility and a heightened awareness of environmental sounds.

Amplification, however, is not always a remedy for those with a hearing loss. The interaction between the type and degree of loss can vary considerably—for example, individuals with a significant hearing loss may rely on visual information more than on auditory information. Additionally, the sound distortion of a hearing aid can produce limits on the improvement in the understanding of speech. Therefore, amplification is only one mechanism for resolving difficulties among the hearing-impaired; improved speechreading is another.

Generally, a hearing aid provides an increase in the intensity level of speech. The degree to which this increases intelligibility will be dependent on the ability of the auditory system to respond to the amplified signal can potentially increase the level of background noise as well as the loss of the signal. Recent

developments in programmable and digital hearing aids reduce the interference from extraneous sounds. However, even with an optimal hearing aid fitting, there will be some instances in which the hearing aid wearer does not benefit from amplification because of a low signal-to-noise ratio (S/N).

The S/N might correspond to a difference between the sound pressure level (SPL) of the desired signal and the SPL of the undesired noise. A converse relationship would be undesirable because the hearing aid would be perceived as too loud. The visual system offers a logical supplemental channel when hearing is compromised from the level of the signal, the background noise, or from either one separately.

Visual Training

Among normal-hearing people, speechreading develops as an unconscious skill. For people with normal hearing speechreading, then, functions to augment understanding of spoken language when auditory information is reduced. People who are hearing-impaired often experience reduced auditory input, speech becomes critical for the visual processing of becomes critical for accurate communication.

The term "speechreading" encompasses not only the lips but also extra-oral features such as the teeth, tongue, jaws, and cheeks (Erber, 1979; Rosenblum, Johansen, & Saldaña, 1996; Sanders, 1982). The scope of speechreading, then, must be regarded in a broad context. O'Neill and Oyer (1981, p. 2) defined speechreading as "the correct identification of thoughts

transmitted via the visual component of oral discourse." O'Neill and Oyer also suggested that "visual communication" yielded a more specific designation to distinguish the visual from the auditory channel as the main path for receiving information.

Commonly, the terms speechreading and visual communication are accepted in research literature and clinical practice. The observations of a talker's eyes (Vatikiotis-Bateson, Eigtsi, & Yano, 1994), lips (Rosenblum et al., 1996), or a combination of the lips, cheek, and mandibular facial regions (Marassa & Lansing, 1995; Preminger et al., 1998) have been customary considerations in speechreading training.

Speechreading

Scope of Traditional Approaches

Over the past 70 years, all the conventional approaches to teaching speechreading stressed the values of auditory and kinesthetic cues in varying degrees. Deland (1931/1968) and Hardy (1970) included the Mueller-Walle method (described by Bruhn, 1929), lipreading with auditory and kinesthetic cues (Kinzie and Kinzie, 1931), synthetic lipreading (Nitchie, 1905) and the Jena method explained by Bunger (1961).

Hardy (1970) noted the major differences in instructional objectives for the discrimination and recognition of speech-articulation lip positions. The Mueller-

Walle method stressed syllable drill as a framework on which to build sentences based on a specific speech-articulation movement or position, whereas the Nitchie approach emphasized practice materials to train a speechreading student to grasp the gestalt of speech rather than the individual components.

Sanders (1982) noted that a distinctive feature of the Jena method is that it employs rhythmic practice with speech-articulation movements. The desired outcome of the Jena method is a total perception of speech sounds by incorporating tactile (e.g., motor-kinesthetic movements with visual and auditory signals. Cora Kinzie (1920) combined the Mueller-Walle and the Nitchie methods to optimize visual communication with both synthetic and analytical approaches to speechreading. She also incorporated the analytical Jena method.

O'Neill and Oyer (1981) reported that the choice of "code or stimulus materials" is important in the development of a speechreading test. The rate of information transmission for stimuli (e.g., words, nonsense words, and sentences (Demorest & Bernstein, 1992 Johansson, Rönnberg, & Lyxell, 1991) may compromise the validity of the instrument. In recent years, investigations have sought to determine optimal methods for both testing and training (Erber, 1983; Alpiner, 1987; Arnold, 1993). Others explored the impact of lexical uniqueness of word stimuli (Auer & Bernstein, 1997) and of the presence of degraded conditions (e.g., lighting and background noise) (Erber, 1983; Rosenblum et al., 1996). No one has characterized definitively all of the factors that facilitate accurate speech perception through speechreading.

Visual Recognition

O'Neill (1954) reported that vision contributed 44.5 percent to the understanding of vowels and 72 percent to the understanding of consonants. Sanders (1982) divided consonant speech sound production into formative aspects and revealing aspects. The formative features are all of the position and movement patterns of articulators (e.g., tongue, teeth, lips, and jaws) involved in speech production. The revealing features correspond to visible aspects of the position movement of articulation. Extra-oral cues can be derived from the cheeks (Preminger, Lin, Payen, & Levitt, 1998).

The production of /k/ as in "kit" is formed by (1) closing the nasal passage with the soft palate, (2) blocking the oral passage with the back of the tongue and the soft palate, (3) building up pressure in the pharynx, and finally (4) exploding the air abruptly through the mouth. None of the formative steps in the production of /k/ are revealing. In comparison, the revealing) features of /p/ as in "pan" and /b/ as in "bad," differ in voicing (a formative characteristic) but have the same revealing features (lip movement patterns).

Homophenous Words

Deland (1931/1968) noted that Alexander Graham Bell coined the word "homophone" from "homonym" to characterize words that sounded alike—for example, /rain, rein, reign/. According to Deland, Bell used the word

"homophenous" to apply to words that "sounded alike to the eyes of the lip-reader." Davis and Hardick (1981) characterized phonemes that look alike (with a similar place of articulation) as homophemes. Similarly, they described words that are indistinguishable visually as homophenous words. Homophenous words are those associated with articulation movements that cannot be differentiated by means of visual cues. Later, Sanders (1982) noted that the adjective "homophenous" also refers to groups of phonemes (individual speech sounds), such as the initial phonemes in words—for example, "bad, mad, and pad."

Viseme

During the development of a test to evaluate homophenous words, Fisher (1968) created the term "viseme" to distinguish or indicate visible attributes of initial and final consonant sounds. Fisher characterized a viseme as "any individual and contrastive visually perceived unit" (p. 800). Among phonemes for example, the phonemes /p/, /b/, /m/ have the same place of articulation (front bilabial) and could be confused with each other but not with /f/ and /v/ which are labiodental. Thus /p, b, m/ is a different viseme than /f, v/. However, the manner of articulation for /p/ and /b/ is characterized as a stop. Similarly, the manner for /m/ is nasal. Fisher reported that the viseme pair of /f, v/ was distinguishable from each other by the place and manner of articulation.

Fisher (1968) investigated contrastive visemes using a group of 18 normal-hearing college students who watched a black and white film featuring six

adult male talkers. The results of the multiple choice test responses generated the following viseme-clusters, based on the consonant confusions within contrastive groups in the initial position (1) /p/, /b/, (/m/ /d/); (2) /f/, /v/;(3) (/k/, /g/); (4) /ʍ/, /w/, (/r/); and (5) /ʃ/, /t/, (/n/, /s/, /z/, /ʤ/, /j/, /h/). The parenthesized items correspond to directional confusion—for example, /m/ as a stimulus was confused with /b/, however, /b/ as a stimulus was not significantly confused with /m/. Fisher noted that particular confusion occurred for phoneme articulation movements not readily visible to the viewer.

Jeffers and Barley (1971) reported that under usual viewing conditions for a talker, vision would provide approximately one-half of the information available for the viewer. They found that the obscure movements of "teeth approximation" and "lips forward" were movements that contributed little to visual information for speechreading. Their conclusions were a similar viseme-clustering of Fisher (1968) with one exception: Jeffers and Barley considered /θ, ð/a separate visual entity. In a theoretical context of homophenous sounds and their corresponding visible appearance as a "viseme," several investigators agree that the /p, b, m/ cluster looks the same on the lips (Fisher, 1968; Jeffers & Barley, 1971; O'Neill, 1954; Owens & Blazek, 1985; Walden, Prosek, Montgomery, Scherr, & Jones, 1974).

Speech Movements

Jeffers and Barley (1971) They defined speechreading movement as "a recognizable visual motor pattern usually common to two or more speech sounds" (p. 42). The speechreader perceives the transitions from visible motor movements and not just the place of speech sounds.

Jeffers and Barley noted that most of the movements are labile. Only four out of fourteen speechreading movements can be considered stable: (1) lower lip to upper teeth, (2) lips puckered—narrow opening, (3) lips together, and (4) lips relaxed-moderate opening to lips puckered—narrow opening—for example, /j/ as in "yes." The rate of speech varied with the talker and transitional characteristics (e.g., transition from one phoneme to another phoneme).

In particular, Jeffers and Barley (1971) noted further that in isolation /p/, /b/, and /m/ are not seen as the same, a conclusion also reached by Owens and Blazek (1985). However, from the "lips together" position, there is a similarity in the appearance of /m/ to/p/ and /b/that is due essentially to transitional effects. Jeffers and Barley stated that for the formation of words sounds are not spoken or seen in isolation—for example, when spoken with a vowel, the release of movement for /p/ and /b/ is not seen because it is perceived as a part of the subsequent vowel sound. The slight pause between consonants "pan," "ban," "man" will appear all to look alike. This would render "Come to me" and "cup to me" equivalent visually. This suggests that vision cues alone may not provide sufficient information to distinguish among /p/, /b/, and /m/.

Jeffers and Barley (1971) categorized consonant speechreading movements for "ideal viewing conditions." They defined ideal conditions as those in which an expert observer can distinguish three-combination diphthong movements (e.g., /au/, /ɔɪ/, and /aɪ/). They reported / f, v/and /p, b, m/ as visible movements for almost all talkers, whereas the movements for /k, g/ are rarely visible for almost all talkers.

Speechreading Assumptions

Attention to the Lips

One premise of speechreading studies is that a speechreader's attention is devoted to the lips, tongue, and teeth for information (Binnie, Montgomery, and Jackson, 1977; Fisher 1968; Jorgensen, 1962; McGurk & MacDonald, 1976; Marassa & Lansing, 1995). Attention to the lips corresponds partially to the notion that certain viseme-clusters are always "homophenous"—for example, /p, b, m/ (Fisher, 1968; Owens & Blazek, 1985). However, the findings and conclusions from several studies indicate conditional confusion for viseme-clusters including /p/, /b/, /m/; /k/, /g/; /w/, /r/; /t/, /n/, /l/, /s/, /z/, /ʃ/, /ʤ/; and /hw/ (Demorest & Bernstein, 1992; Erber, 1974b; Jeffers & Barley, 1971; Kricos & Lesner, 1982; Pichora-Fuller, & Benguerel, 1991; Walden et al., 1974).

Using one talker producing vowel-consonant-vowel (VCV) utterances, Owens and Blazek compared the visual recognition of viseme-clusters by two groups of subjects (five normal-hearing and five hearing-impaired). The hearing-impaired subjects had post-lingual deafness. The comparative data analysis for /p/, /b/, and /m/ indicated performance varied from chance for both groups. A closer examination of the viseme-cluster of /p, b, m/ revealed the following accuracy levels for the normal-hearing subjects: /p/ 24 percent, /b/, 28 percent, and /m/ 76 percent. Recognition accuracy for the /m/ target appeared to be above chance. In the same context, the hearing impaired group's scores for /p/ (42 percent) and /b/ (58 percent) were at chance or better; however, accuracy for recognition of /m/ (6 percent) appeared to be below chance. Thus, although no statistical analysis was performed on these data, there is some evidence that for both subject groups recognition of the components of this viseme cluster was not a simple chance event.

A similar pattern was observed in the data for the /a/ context. The normal-hearing subjects' accuracy levels were 24 percent for /p/, 56 percent for /b/, and 76 percent for /m/. The hearing-impaired subject groups' accuracy levels were 58 percent for /p/, 64 percent for /b/ and 8 percent for /m/. It is important to note, however, that in the final analysis the investigators concluded that their data for a wide range of visemes provided strong indications that adults with normal hearing and adults with hearing impairment perform similarly in the recognition of visemes

Further, Owens and Blazek (1985) acknowledged that their analysis did not account for several variables (the talker, phonemic context, lighting conditions, viewing angle, and the statistical criterion to assess them). Therefore, it was not clear if these or other variables might have made a difference or provided for an influence in the perception of the /p, b, m/ viseme-cluster.

Lansing and McConkie (1999) investigated differential attention to the upper and lower regions of a talker's face in the context of word recognition and suprasegmental information processing. They tested the hypothesis that observers direct their gaze to those aspects of a stimulus pattern from which visual information was sought. Their findings contrasted with those of Vatikiotis-Bateson et al. (1994) who concluded that the foveal vision was directed to the eyes. There was evidence that suggested the lower region of the face was important for word recognition. However, they noted that the subjects raised their gaze to the eyes near the end of the talker discourse, as if to confirm the information. Lansing and McConkie concluded that more research for their "Gaze Direction Assumption" was needed.

Attention to Mouth Opening Size

The studies of Roback (1961) and Jorgenson (1962) investigated the ability of viewers to identify isolated homophenous words captured on motion picture film. First, Roback determined that there was statistical significance for

untrained viewers to select homophenous words accurately. Her findings suggested that, in spite of highly similar articulation appearance, homophenous words were not produced exactly alike on the lips.

In a subsequent study, Jorgenson used planimetry (the measurement of often irregular plane areas) to determine the horizontal planar dimensions of mouth opening (from corner-to-corner) during speech production. Jorgensen (1962) compiled 12 groups of four homophenous words each that had the greatest percentage of correct identification when looking at the independent variables of (1) mouth width, (2) speech production timing differences, and (3) visibility of the teeth.

In the final analysis, Jorgensen (1962) did not find statistical differences (with t-tests) for mouth width, timing differences, or visibility of the teeth. However, her subjective analysis of the data for mouth width indicated that the greatest amount of variability was for the production of four words (fade, feign, vain, and fete). Jorgensen concluded that, although not statistically greater than chance, the differences appeared to indicate that homophenous words were not made the same way. Her subjective analysis led to the conclusion that future research should address the quantifiable differences of homophenous sounds and words.

Attention to the Eyes

In the case of speechreading, visual information from key regions of the face must be directed to the foveal region of the eye for closer examination and for making a decision. The peripheral visual receptors are cone receptors and are located within the foveal region. It is within this region that the central vision is most acute. The cones also convey coding for color and both spatial and temporal resolution. Therefore, it is important physiologically that individuals rotate their eyes (horizontally and vertically) to align visual targets with the fovea (Palmer, 1999; Schwartz, 1999).

Vatikiotis-Bateson, Yano, Eigtsi, and Munhall (1994) provided evidence that as masking noise increased, it became more difficult to hear. As a consequence, the viewers of a talker's face shifted their acute foveal vision to the eyes from the lips. Their findings contradicted previous studies that assumed that the lips are most important for visual speech perception.

The specific aim of their research was to characterize the kinematics of eye motions and gaze for perception of the videotaped monologues of a talker. The data were mouth-to-eye-gaze transitions based on (1) angular tracking and attention and (2) fixation time. Vatikiotis-Bateson et al. (1994) reported that certain phonetic images—for example /p, b, m/—might be recognizable. However, their findings suggested that speechreading might be more dependent on the ability to "foveate" (focus on the foveal region) or to direct the gaze to strategic regions of the face, rather than just the lips.

Talker Differences

Empirical evidence has shown that talker differences can significantly influence accurate speechreading perception among normal-hearing subjects. Jeffers and Barley (1971) noted that individuals differ in the manner of forming sounds and in the precision of sound formation.

Kricos and Lesner (1982) collected data from the Utley Lipreading Test to determine whether selected talkers could influence the accuracy of visual recognition. The mean consonant recognition scores coupled with Utley Test of Lipreading scores varied significantly across six talkers of general American dialect for /θ, ð/, /t, d, n/ and, /p, b, m/.

The easiest talkers to speechread yielded a consonant recognition score of 75.2 percent and a mean Utley score equaling 43.3 percent. The most difficult talkers to speechread yielded a consonant recognition score of 63.8 percent and a mean Utley score equaling 27 percent. Therefore, phoneme production and visual intelligibility can vary as a function of the talker. The variance of subject performance suggested that a wide variety of talkers should be used for training students to speechread. Further, Kricos and Lesner concluded that not all talkers (those with normal articulation) would reveal the same viseme-clusters to those who attempt to speechread them.

Phonological Factors

Jeffers and Barley (1971) suggested six limiting phonologic factors of speechreading: (1) low visibility of speech sounds, (2) homophenous phonemes, (3) rapidity of normal speech, (4) co-articulation or transition effects, (5) individual differences in articulator movements, and (6) environmental limitations. Sanders (1982) included the following three limiting phonologic factors: (1) amount of context, (2) frequency bandwidth of the phoneme, and (3) the intensity of the speech signals.

Erber (1974b) reported on the perception of consonants within a VCV context by profoundly deaf children. Confusion of /p/, /b/, and /m/ occurred for the vowel contexts of/i/, /ɑ/, and /u/. Alveolar placement consonants such as /n/, /d/, /t/, /s/, and /z/ were confused more frequently within the /u/ C /u/ context than the intervocalic context of /i/, /ɑ/, and /ʌ/. The frequent occurrence of alveolar placed consonants in the English language creates a real problem for speech perception.

Erber (1974b) noted further that there was a bias toward the perception of /f/ rather than /v/ in the /f, v/ viseme-cluster within an /ɑ/C/ɑ/ context. Interestingly, this bias was not found in the context of other vowels. Erber could not resolve whether the reason for the differences in articulation perception was linked to the teacher of the students or to biases on the part of the subjects. In other words, the training method might have been a factor in the outcome.

Training Materials

Using the Central Institute for the Deaf (CID) sentences, Demorest and Bernstein (1992) examined the variability in speechreading performance among normal-hearing subjects. Videotape presentation of stimuli included both a male and a female talker. The researchers examined the significance of the talker and the test material with respect to the subjects' speechreading proficiency. Their findings suggested (in rank order) that the highest contribution to variance was from sentence materials, followed by the subjects themselves, and then the talkers. However, Demorest and Bernstein cautioned that the time of day could also have influenced their results.

Demorest and Bernstein (1992) attempted to account for the generalized sources of confusability for materials used in speechreading training. Using a multivariate analysis of variance (MANOVA), they determined that an interaction of sentence materials and the subject was the major source of variability for speechreading accuracy among normal-hearing adult subjects. They suggested that one should look at subject's response interacting with different talkers.
In a subsequent study, Demorest, Bernstein, and DeHaven (1996) addressed the generalizability of sentences, words, and nonsense syllables for normal-hearing subjects. Correlation data analysis suggested that learning effects occurred for subjects, in particular with specific talkers. The most significant improvement in performance was between words and sentences with a correlation coefficient of 0.8. Demorest et al. (1996) suggested that a high correlation for words and

sentences may mean that there are other viable approaches for training. They suggested that more research was needed to determine the sources of variability for both the word and the nonsense levels to account for residual error—for example, assumptions regarding the independence of variables.

Speechreading and Talker Training

Walden et al. (1974) suggested that improvement in visual consonant-recognition ability of hearing-impaired adults might be attributable to speechreading training. The findings included a significant difference between pretraining and posttraining performance on nine within-cluster visemes structured in consonant vowel (CV) syllables. Notably, a 40 percent change in performance was shown for /p, b, m/ (57.3% - 97%). Their findings show more evidence that in certain circumstances (variations in the talker, co-articulation effects) viseme-clusters are not always homophenous and that phoneme level training can improve accurate perception among normal-hearing subjects.

Pichora-Fuller and Benguerel (1991) investigated the effectiveness of computer-assisted speechreading training (CAST) for the correct identification of consonant viseme-clusters. Based on face-to-face discourse, CAST uses search rules to regulate feedback variables related to speech perception and linguistic redundancy. Pichora-Fuller and Benguerel noted that the amount of speech information that can be transmitted in the visual modality has not been established. More study of visual processing is needed to determine why some individuals can identify consonant viseme-clusters.

Lesner (1988) noted that variables such as facial cues, "extrafacial" cues, and cues related to rate and rhythm could influence the intelligibility among talkers. She concluded that talker training holds potential for reducing the variability among talkers.

Visual Speech Perception and Processing

Breeuwer and Plomp (1984) investigated speechreading and demonstrated differential contributions of vision relative to sound pressure and acoustic filtering. Vatikiotis-Bateson et al. (1994) reported that normal-hearing subjects spent more time fixating on the eyes rather than the lips during speechreading indicating divided attention. Based on an analysis of facial gaze patterns, the investigators suggested that subjects devoted most of their peripheral vision to the mouth and their foveal vision to the eyes. Thus, their attention to specific facial areas was allocated "asymmetrically" in a manner that was not expected. The investigators assumed that foveal vision would correspond to more attention to the mouth than to the eyes.

Vatikiotis-Bateson et al. (1994) studied spatial location and variability to determine the effects on a shift in gaze and attention to the face of a talker. The data analysis revealed that subjects directed most of their peripheral gaze to the perioral structures of the face, while the foveal vision was directed to the eyes. Vatikiotis-Bateson et al. hypothesized that the phonetically relevant information occurs over the whole face, not just the lips. Further, Vatikiotis-Bateson et al.

(1994) emphasized that an observer's attention to the perioral region, surrounding the lips, may be influenced by deformations or movements related to the vocal tract and respiration.

Later, Lansing and McConkie (1999) hypothesized that speechreading involves a differentiated distribution of gaze patterning across the face. This means that there were different facial regions associated with visual targeting. They tested this supposition on subjects without the benefit of audible cues but with the benefit of the facial expression r from a talker. They did not reject the null hypothesis; however, they concluded that there was sufficient evidence to support the visual gaze to the mouth region rather than the eyes or other regions. They surmised that the subjects directed their attention to the mouth to gain the majority of their linguistic cues.

Hopfield (1982), Kandel and Wurtz (2000), and Palmer (1999) noted that a "winner-take-all" visual perception strategy corresponds to attention to one target while rendering everything around it as background. Based on this theory visual speech perception may be mediated by a similar winner-take-all stratagem. Palmer (1999) noted that visual attention occurs (in part) because of covert selection at the neural level.

The fovea is the region of the eye that is most sensitive to color; it also contributes to motion detection at the retinal processing level. The greatest foveal neural excitation for color occurs for what we perceive as green-yellow, at the human photopic luminosity peak (Schwartz, 1999). Further, Thiele, Dobkins, and Albright (1999) reported that color contributes to motion detection in the

middle temporal (MT) cortical area. Croner and Albright, (1999) demonstrated that the MT is also active in the segmentation of audible noise and color. Several roadway signs and emergency vehicles (e.g., school crossings, fire trucks, and ambulances) have been changed to a green-yellow color (approximating 555 10^{-9} m).), more of our attention has been allocated to them. Thiele et al. (1999) suggested that differences in the allocation of attention might account for differential sensitivity in chromatic (color-derived) motion processing.

The juxtaposition of red with green has been shown to create a shift in gaze and focused attention based on oculomotor control and underlying neural centers—for example, posterior parietal lobe, superior colliculus, and pulvinar (Hurvich & Jameson, 1957; Livingstone & Hubel, 1987; 1984; Zeki, 1980). A similar shift in focused attention also occurs for a juxtaposition of yellow and blue, black and white. The "Opponent Process Theory," as described by Hurvich and Jameson (1957), states that three types of visual cone receptors respond in mutually opposite or exclusive directions (red-green and blue-yellow). Therefore, humans are forced into a pairwise polarized perception between red and green and between blue and yellow (Palmer, 1999).

Color-vision is important for perception of shape, form, texture, depth, and motion (Thiele et al., 1999). Brightness and reflectance are also factors for shape, form, texture, depth, and motion (Tessier-Lavigne, 2000c). However, juxtaposed or adjacent colors are critical to the overall visual perception of objects and surface details.

Statement of Problem

Owens and Blazek (1985) established that vowel variation and different talkers do not produce a significant difference in the correct recognition VCV nonsense syllables. Owens and Blazek investigated the visemes observed by those with a hearing-impairment and those with normal-hearing adult subjects for comparison. The stimuli were derived from 23 consonants and the vowels /i/, /ɑ/, /u/ and, /ʌ / to create intervocalic syllables.

Jeffers and Barley (1971) characterized "stability" as the production of consonant speechreading lip-to-teeth movements that are minimally affected by the speaker's habitual rate of speech, or by transitional movements from or to another sound. Owens and Blazek concluded that the viseme-clusters /p, b, m/ and /f, v/ were highly visible and stable, especially when coupled with the vowel /u/. Similarly, Binnie et al. (1974) found that the overall percentage correct intelligibility scores for visual recognition alone was poorer than chance when coupled with vowel variation. Owens and Blazek (1985) concluded that an absolute list of viseme-clusters was difficult to generate for several reasons: (1) talker, (2) viewing angle, (3) facial illumination, (4) vowel influence, and (5) phonemic context and length.

There was speculative empirical evidence that some conditions may exist whereby some "within group" visemes—for example, /p/, /b/, and /m/—are distinguishable from one another. Demorest and Bernstein (1992) noted that an important source of variability was possibly attributable to the test occasion (time

of day) whereas Kricos and Lesner (1982) found significant variance in consonant recognition performance based on differences among six talkers. Specific differences were noted for the similarity of phonemes between two talkers, one easy to speechread and the other more difficult. Kricos and Lesner found that /p/, /b/, and /m/ were the most recognizable, as a viseme-cluster, by normal-hearing subjects.

In a subsequent study, Lesner (1988) compared visible intelligibility with audible intelligibility and noted significant qualitative differences among talkers for selected features—for example, speaking rate, rhythm, and facial cues. Demorest & Bernstein noted that variability of speechreading accuracy was most significant for subject and stimulus material interaction.

The stimulus materials were sentence lists, a factor that was shown in relation to co-articulation effects and context (Binnie et al., 1977). They found that accuracy varied based on the vowel context. Erber (1974b) found that /p/, /b/, and /m/ were confused in all vowel contexts, primarily because of the talkers. Jeffers and Barley (1971) and Owens and Blazek (1985) agreed on limiting phonological factors—for example, vowel context might alter visual consonant recognition. Sanders (1982) reported that, in the context of stress, phonologic frequency, bandwidth, and intensity were factors in speechreading perception. Subsequently, Breeuwer and Plomp (1984) indicated variations in speechreading accuracy (among normal-hearing subjects) based on 1/3 octave bandwidth-derived sound pressure levels. They found that as sound pressure information was combined with frequency-selective information, accuracy

increased for phrases and sentences.

Jeffers and Barley (1971) and Owens and Blazek (1985) found that speechreading training yielded the greatest improvement for the viseme-cluster /p, b, m/, a cluster that is often considered homophenous (Binnie et al., 1974; Fisher, 1968; Owens and Blazek, 1985). Roback (1961) reported subjects who were capable of identifying correctly homophenous words with better than chance performance. Jorgenson (1962) reported measurable minute differences that implicate distinctions in the way homophenous words are made. Although there were differences in the mouth opening for words, Jorgenson did not find differences for individual phonemes. Taken together, it appears that the process of speechreading remains unclear by virtue of both phonologic and physiologic factors.

At the time of this writing, the literature on hearing loss and speechreading revealed no published studies that explored color vision's contributions to speechreading ability. However, a few studies—including Lansing and McConkie (1999); Marassa and Lansing (1995); and Vatikiotis-Bateson et al. (1994)—have shown that foveal vision is a factor in speechreading ability.

The dynamic processes of speech and visual attention to articulator movements (lips, teeth, and tongue) could be related to contributions from foveal (color) vision. Further, neither color vision nor even color blindness (termed medically as achromatopsia) has been associated with speechreading ability.

Following the findings of Hering (1878/1964), Hurvich and Jameson (1957) characterized opponent color theory with component neural mechanisms for red-

green, blue-yellow, and black/white. The theory explains how visual attention and color perception are constrained at the photoreceptor level (Tessier-Lavigne, 2000c). Hurvich and Jameson (1957) described and quantified putative color vision (cone) mechanisms as mutually opposite or exclusive. As a result, color stimulation can either excite or inhibit the neural response for pairwise opponency (pairing of opponent colors). However, in those instances, people do not experience colors being lost singly—for example, just red or green and just blue or yellow. Rather, a reduction in "redness," "blueness," and "greenness" or "yellowness" occurs as a function of allocated attention (Hering, 1878) and neural coding at the retinal and cortical levels (Livingstone & Hubel, 1987).

The conjunction of multiple sources of information is more informative and definitive than any single presented source alone (Massaro, 1987). Conditions could be created whereby color and a viseme-cluster could converge to produce multiple sources of information for overall speechreading perception.

Color-vision perception could relate to speechreading insofar that attention to form, texture, and motion appears to be essential to speechreading— possibly related to cone or color vision (Livingstone & Hubel, 1987, 1988; Nathans, 1989; Schwartz; 1999; Sharp & Philips, 1983; Tessier-Lavigne, 2000b, 2000c; Treisman, 1988,1986). It is also reasonable to speculate that if the lips were "colorized" in opponent colors (green-red and blue-yellow) that the visual recognition of /p, b, m/ might vary because of attention to one color over another. In a manner similar to the McGurk Effect, visual illusionary conflicts may occur for speech perception through color opponency.

It has not been established whether viseme-cluster recognition could improve with an interaction from several independent factors—for example color vision conditions, vowel context, and the /p, b, m/ viseme-cluster. Finally, in the visual perception of speech, a visual search of facial features may involve both "preattentive" and attentive processes. Tessier-Lavigne (2000a) referred to the preattentive search as one that scans for one unique attribute—for example color, texture, depth, and form. In comparison, the attentive process is "conjunctive" insofar that successive shifts in attention occur because of multiple and sometimes ambiguous attributes.

Logically, there are some questions that remain unanswered: (1) Why can some people distinguish between homophenous words/sounds pairs? (2) Could it be that during speechreading processing one attends to the articulator motion of the upper lip or lower lips for certain phonemes? (3) Can color optimize speechreading ability for training one to attend to either the upper or the lower lip? (4) Are there speech perception ambiguities for which we have not accounted in previous research? (5) In the presence of opponent colors, does a speechreader process information in an "all-or-none" model or an "attenuation" model? (5) Is there limited capacity to attend to strategic regions of the face during speechreading in a winner-take-all strategy? Therefore, this investigation was designed to answer the following questions regarding speechreading:

Questions

1. Is the interaction of opponent colors and the vowel /i/ the source of variation for the accurate visual recognition of phonemes /p, b, m/?

2. Is the interaction of opponent colors and the vowel /a/ the source of variation for the accurate visual recognition of phonemes /p, b, m/

3. Are the results of Experiments I and II comparable for the accurate visual recognition of phonemes /p, b, m/? The interaction of opponent colors and the vowel /i/ constitutes Experiment I while the interaction of opponent colors and the vowel /a/ constitutes Experiment II.

Chapter 2

REVIEW OF THE LITERATURE

This chapter surveys the historical foundation of aural rehabilitation, speechreading, and their relation to individuals with a hearing-impairment. Additionally, it reviews the theoretical framework for the assumptions underpinning information processing. An anatomic and physiologic overview of the human visual system illustrates the capacity and limitations of vision. This chapter includes a review of the classic research and recent interpretation of color vision theory and opponent process theory. The implications of color are then related to the visual perception of form, depth, and motion. This chapter establishes the theoretical framework and conceptual model for speechreading and the hypothetical influence from color perception.

Theoretical Framework

Information Processing Theory

Information processing theory provided the theoretical framework for this study. The term *information*, for communication purposes, was first proposed in the field of mathematics as a means of characterizing early telecommunication systems (Shannon & Weaver, 1962). An American Telephone and Telegraph Corporation (A T & T) mathematician, Claude Shannon, investigated the

limitations of a channel in transferring signals and the cost of information transfer via a telephone line. He developed a mathematical theory of communication, defining "information" as a purely quantitative measure of communicative exchanges. Although not intended for social science application, the information model provided a means of studying interfering factors in human communication and information processing (Losee, 1999). The essential components, characteristic of most information models, include the serial components of (1) message source, (2) transmitter, (3) channel with intervening noise, (4) receiver, and (5) destination.

Information Processing Models

Early in the development of information processing archetypes, Broadbent (1958) proposed a serial processing "all-or-none" model involving parallel contributions of speech: that is input from different types of sensory information as shown in Figure 1. Based on investigations of central effects during listening tasks with audition alone and during listening tasks with the benefit of viewing a talker—for example, listening to instructions from two voices at once and listening to instructions with visual cues from a talker—Broadbent concluded that (1) central effects outweigh sensory factors during the presentation of two concurrent messages; (2) the effects varied with the number of messages presented to a listener; (3) although not at random, some information-discarding resulted from a discriminating process. Broadbent suggested that some

information was considered and subsequently discarded because of limited capacity.

When based on the assumption of selective learning, Broadbent (1958) speculated that the "novel" nature of some similar information was filtered or selected more readily than dissimilar input—for example, an individual might be more attentive to an aperiodic acoustic signal rather than periodic acoustic signal. "Novel," in this context, refers to distinct characteristics of a message that divert or direct the attention of a listener. In other words, the unique portions of the message garner "attention resources" over ambiguous and concurrent input (vision and audition). Therefore, in Broadbent's serial processing model, salient features emerge as necessary details to gain and probably sustain attention resources for communication. Consequently, one selects one feature over another for an information source.

Schneider and Shiffrin (1977) mapped graphic alphanumeric two-dimensional representations "CM" and "JM" in different orientations (rotation of the characters). The investigators sought to determine the allocation of memory resources and accuracy of shape determination. In all cases, the CM pair required the least amount of memory resources based on processing time to make a decision. Schneider and Shiffrin concluded that one aspect of making a decision was linked to an all-or-none decision-making process.

The rejected information is related to an assumption of limited attention resources, processing speed, and salient information characteristics. Louder sounds garner more attention resources than softer ones, rendering them more

conspicuous to a listener. In some instances, vision may require more attention resources than does hearing for the examination of form, depth, color, and motion (Wickens, 1987).

It is reasonable to suggest that individuals with a hearing-impairment selectively discard information that interferes with pattern and message recognition. Two theoretical sensory processing systems—an "all-or-none model" (Broadbent, 1958) and an "attenuation model" (Triesman, 1960)—theorize that there are limited resources to allocate to memory and attention. These two models reinforce the supposition that parallel processing for audition and vision involves "tuning-in" and "tuning out" (Broadbent, 1958) or "turning down" (Triesman, 1960) as shown in Figure 1.

All-or-None Information Model

During speech input, other signals (vision, olfaction, audition, and taction) compete for our attention. Modeling information processing allows one to characterize the relationship of variables. Broadbent (1958) proposed a serial model with the major premise of a filter interposed in the communication chain. The minor premise was that the filter was selective for one information source. The theoretical selectivity occurs at a subconscious level based on sensory boundaries.

Schneider and Shiffrin (1977) explained that Broadbent's model incorporated channels. A channel was described as a source that conveyed

sensory input—for example, the two ears of a listener. Broadbent (1958) proposed a model that incorporated different information channels or conduits to direct the message signal through a selector, the precise characteristics of which are still unproved. This model illustrates the problem for capacity. The term capacity in communication theory corresponds to the limiting quantity of information that can be transmitted through a given channel at a given time. The capacity problem corresponds to serial processing (similar to a linear model).

Broadbent considered that speech entering the human communication system was analogous to an electromagnetic wave entering a telephone or radio receiver. In the design of more sophisticated electronic components, a computer serves to process information in both a serial and a parallel manner. The binary logic of a computer exploits the "all or nothing" logic for the execution of selected tasks. Therefore, the input from a nonattended channel will not be perceived in the presence of a competing channel (Schneider & Shiffrin, 1977).

In this vein, salient speech or linguistic units serve to reinforce Broadbent's notion that 'all-or-none" of the message continues in processing. In theory, salient aspects of communication with increased audibility and visibility are received with greater attention. It would appear that the auditory, visual, and tactile channels may interact for the various, categorical nonverbal and verbal cues conveyed in speech (O'Neill & Oyer, 1981). The distinctive or salient attributes gain our attention in the presence of competing stimuli.

Attenuation Information Model

Triesman (1960) proposed a parallel processing model involving the simultaneous input of sensory information (vision, hearing). Triesman's model incorporated an attenuation (reduction) function. She theorized that we "turn down" extraneous input to attend to intended messages. In other words, while attempting to attend to one sensory channel, other attention resources—for example, environmental noise and lighting conditions—may reduce our concentration.

Triesman (1969,1985) hypothesized a filter selector with a multilevel controlled processor functioned as an attenuator for the conscious management of multiple sensory input. She theorized that the concurrent encoding of different modalities at the input stage does not correspond to proportionate processing for the receiver. Instead, one modality or channel of information is attenuated in preference for another modality—for example vision over hearing. To the degree that all information is attenuated (vision, hearing), the succeeding responses are governed by other attention resources (see Figure 1).

Visual Information Processing Theory

Palmer and Kimchi (1986) reported that the typical "input-out" processor analogy for information processing (IP) was related theoretically to mental activity and a program running on a computer. Palmer (1999) commented that IP is the framework that supports most current theories of visual perception (Helmholtz,

1867; Hering, 1878; Hurvich & Jameson, 1957). The essential characteristics of most information models include visual information processing. Palmer and Kimchi (1986) listed five assumptions of information processing that relate to vision: (1) information description, (2) recursive decomposition, (3) flow continuity, (4) flow dynamics, and (5) physical embodiment.

Further, Palmer (1999) reported that (1) information description, (2) recursive decomposition, and (3) physical embodiment are the three most important assumptions for visual processing. The first assumption involves the formation of mental events—for example, visual perception derived from the retina of the eye. The second assumption concerns an iterative (repeating) process of deductive analysis to determine "what" the mental event was rather than "how" it occurred. The third assumption bridges the gap between abstract function and the information embodied in the states of a system (visual representations) and the system (processes)—that is representations exist in the abstract domain and the system that executes the processes is in the physical domain.

The Human Visual System

Retinal Visual Processing

Often likened to a camera, the human eye performs the function of focusing a visual image onto the photosensitive retinal lining after it passes serially through the cornea, aqueous humor, vitreous humor, and the lens. The retina is a structure found in vertebrates across the sub-phyla of mammalia,

amphibia, and reptilia. The retina is the sensory receptor neuroepithelium "lining" for the eye. The ciliary eye muscle (intraocular) must adjust the lens to project an image onto the retinal fovea (the most sensitive focal point) with minimal distortion or aberration (Tessier-Lavigne, 2000a).

Photopigments in the Rods and Cones

The human retina consists of two types of photoreceptors: rods and cones, named primarily because of their shape. The rod photoreceptors outnumber the cones about 20:1 (Schwartz, 1999; Tessier-Lavigne, 2000b). Sharp and Philips (1993) reported that there might be about 120 million rods situated around the fovea of the retina and 6 million cones in the foveal and parafoveal regions. More energy (in the form of photons) is required to excite the cones than the rods.

The rods and cones both contain molecular photosensitive pigments within the disks of their outer segment. Each disk contains approximately 10,000 molecules of pigment. The pigment in the rods is termed "rhodopsin"; it used to be called "visual purple." The molecules of rhodopsin absorb light quanta (photons in discrete packets of electromagnetic radiation) and convert the energy into electrochemical activity: one molecule of rhodopsin captures one photon of light.

From this point, visual processing begins the careful examination of objects, surfaces, shapes, brightness, distance, wavelength, and motion. The

sensory transduction of light energy (in the form of photons) into bioelectrical energy is modulated by specialized neurons (nerve cells) along visual pathways.

Photopic and Scotopic Vision

Photopic vision occurs under daytime or bright lighting circumstances. This vision is mediated by cone receptors. Scotopic vision occurs during dim or nighttime conditions and is mediated by rod receptors. The dichotomous function and structure of the retina (in vertebrates) is termed a "duplex retina" (Schwartz (1999). Photopic vision forms a basis for visual contrast, including brightness. Luminance (a constant of projected brightness from a surface) is also an important source of vision contrast.

Scotopic vision is more sensitive than photopic vision (in low light conditions) and yields greater spatial summation because of the retinal organization for many rods (greater than 15 rods per synapse to a single output ganglion cell). Spatial resolution (the process of multiple postsynaptic inputs from different sites) is better with the rods and can be attributed to (1) a smaller length constant and (2) a slower neural response for the cones. Rods can detect small amounts of light flickering less than 12 Hz. In contrast, the cones can resolve a flickering light down to approximately 55 Hz.

As a result of differentiated neural responses, the sensitivity for spatial resolution is not very good for cones; The temporal resolution is superior (Tessier-Lavigne, 2000b). Cones, although fewer in number, outperform the rods

on all visual tasks (especially for color). Hence, cones are characterized by higher visual acuity than rods (Schwartz, 1999). Typically, people perceive sharper detail under illuminated conditions than under darkened or dimly lighted situations. Cone or color vision contributes to the detailed analysis of focused attention along with orientation, size, and distance.

The human visual system is sensitive to electromagnetic radiation within a bandwidth of 400-700 nanometers (nm or 10^{-9} m). This bandwidth corresponds to a range of perceived colors from blue-to-red). The peak of the "human luminosity function" [V(λ)] corresponds to 555 nanometers. The human photopic luminosity function is shown in Equation 1.

$v(\lambda)$ = Human Photopic Luminosity Function
Where V, is a property of polychromatic radiation.
Polychromatic radiation is defined by: $V = \frac{\int \Phi_{e\lambda} V(\lambda)\, d\lambda}{\int \Phi_{e\lambda} d\lambda}$
Where $\Phi_{e\lambda}$ is the radiant flux of a specific wavelength (λ),
and v(λ) the spectral luminous efficiency of the wavelength. (1)
The $\int \Phi_{e\lambda}\, d\lambda$ term weighs the radiant flux of the radiation relative to the sensitivity of the observer.
$V(l)$ the spectral luminous efficiency of that wavelength (l)
Where wavelength is measured in nanometers.

Although related to brightness, v(λ) represents the results from a combination of experiments where sample wavelengths were matched to a color standard (Palmer, 1999; Schwartz, 1999). The peak of the function corresponds to the

wavelength that generates the greatest foveal neural excitation (Livingstone & Hubel, 1987; Kandel & Wurtz, 2000). Humans perceive the wavelength (λ) of 555 nm as a green-yellow color.

Color Vision Theory

Palmer (1999) traced the origins of the first color vision theories to Sir Thomas Young and Herman von Helmholtz. However, the pioneering research of both Helmholtz (1867) and Hering (1878) laid the foundation for color vision theory. The trichromatic theory was a combination of Young and Helmholtz. The trichromatic theory assumed three kinds of cone receptors that respond differentially to the wavelength of photons incident upon them (short, medium, and long wavelength receptors).

The trichromatic theory appeared to correspond to basic color matching and the mixture of hues to form secondary colors. However, the theory had a deficiency in its explanation of why certain color experiences are always lost in certain pairs: red-green, blue-yellow, and black-white. A physiologic and anatomic explanation has not been formulated to underpin the trichromatic theory (Palmer, 1999; Schwartz, 1999; Tessier-Lavigne, 2000c).

Hering (1878) observed human visual perception and noted that there was no series of colors (chromatic transitions) that appeared, simultaneously, both reddish and greenish or both yellowish and bluish. He concluded that the colors were mutually exclusive of each other. Therefore, he theorized that concomitant observation of opponent colors would be improbable.

Opponent Process Theory

Today, it is accepted that there are three classes of cones that mediate color vision (Palmer (1999). The cone types correspond to long wavelength cones (cones about 560 nm) "L," short wavelength cones (cones about 440 nm) "S," and medium wavelength cones (cones about 530 nm) "M". Palmer noted that these cones are designated misleadingly as blue, green, and red cones. De Valois and Jacobs (1968) measured cone responses in macaque and squirrel monkeys and found color-selective cells (L, S, M) that conformed to Hering's color processing theory.

Based on the work of De Valois (1968), De Valois, Abramov, and Jacobs (1966), and Jacobs (1996), Palmer (1999) described, algebraically, possible neural circuits for the dual process theory of color vision (see Figure 2). The R^+G^- opponent cone receptors are those derived algebraically by combining the excitatory input from L cones and the inhibitory input from M cones—for example, (L - M). It follows that the output of opposite green, red (G^+R^-) cells can be derived through the combined excitatory input from M cones with inhibitory input from L cones (M - L).

The excitatory (+) output of S cones and the summed inhibitory input (-) of M and L cone outputs generate the Y^+B^- components as follows: (S - (M+L) = S – M – L). The yellow, blue (Y^+B^-) opponent cone cells may be derived from the summed excitatory response of M and L cones and the inhibitory output of S

cones [(M + L) – S = M + L – S]. Finally, the white, black ($Wh^{-}Bl^{-}$) cells can be derived from the sum of the excitation of all three cone types (S + M + L) while the $Bl^{+}Wh^{-}$ cells can be determined from the summed inhibitory output from all three cone types (-S – M –L). De Valois et al. (1966) reported that opponent cells are not the same. There is significant variance between cells and their responses to different spectral regions. Consequently, excitation and inhibition occur for certain opponent cells.

Dual Process Theory

Hurvich and Jameson (1957) used colored lights to mix and measure the amount of color required to "lose" the opposite color. They proposed that the dual process theory incorporated two stages: (1) the trichromatic stage followed by (2) the opponent process stage. Both stages are known to occur in the retina (Palmer, 1999; Schwartz, 1999). The trichromatic stage involves the strength of blue-yellow and red-green relative to wavelength.

Hurvich and Jameson (1957) described the opponent process theory as postulated by Ewald Hering (1878/1964). Hering concurred with the Young-Helmholtz trichromatic theory in that three mechanisms were hypothesized for color perception. However, Hering assumed three *opponent* mechanisms (red-green, blue-yellow, and black-white). He characterized the receptors as three types that respond to light in opposite directions (red-green, blue-yellow, and black-white). The directionality was designated as either positive or negative. To

the extent that we see reds, greens, blues, and yellows, we do not see reddish-greens and bluish-yellows. Palmer (1999) noted that the two-stage model accounts for many facts, but residual problems suggest that more modification is needed. Furthermore, he indicated that one revision in the theory should account for why the shortest visible wavelengths (in the violet range) should appear reddish, especially if the red-green channel is defined by the difference between M and L cones.

Clearly, Hurvich and Jameson (1957) provided physiologic links between the human visual perceptual experience, the color-sensitive cone photoreceptors, and the coding for the opponent colors of red and green, blue and yellow. Hurvich and Jameson also indicated that the cones for black and white code in the domains of brightness and intensity. Their investigation indicated that it was physiologically impossible for individual and clusters of cones to generate parallel neural coding for the opponent colors of red and green, blue and yellow. Yet black and white coding is coterminous with opponent color stimulation in that one cannot perceive black and white at the same time (Hering,1878).

Cortical Visual Processing

The visual system resolves component parts of an image—for example, form, contrast, and depth—to help generate a visual scheme. The experimental isolation of the visual pathways has been a challenge as demonstrated by experiments to (1) separate motion and color processing (Ramachandran & Gregory, 1978) and (2) define depth and form (Mishkin, Ungerleider, & Macko,

1983; Treisman, 1986). Tentative assumptions about color vision have been derived from primates—for example, man, apes, and monkeys (Jacobs, 1996), both monkeys and cats (Shapley & Perry, 1986), and from humans (Livingstone & Hubel, 1987). The latter study utilized psychophysical methodology to suggest a segregation or separation of cortical visual pathways for parallel processing. Livingstone and Hubel concluded that some visual attributes of form, shape, and depth are discarded at the neural level. In humans, these processes occur without the viewer making a conscious decision about the allocation of attention.

Kandel and Wurtz (2000) described the processing of visual information in multiple areas of the cerebral cortex. From the retinal photoreceptors, neural projections make synaptic connections onto retinal ganglion cells (the output cells of the retina). At this neural point, segregation occurs for large (M) cells and small (P) cells. Livingstone and Hubel (1987, 1988) reported that the M cells carry information about motion and stereopic depth, whereas the P cells convey information about color and form. The axonal projections of the ganglion cells constitute the optic nerve (cranial nerve II). The optic nerve then projects to the lateral geniculate nucleus (LGN) in the thalamus (the final "pre-striate cortex" relay for sensory information). The lateral geniculate nucleus projects to the primary visual cortex. This area corresponds to both the striate cortex (V1) and Broadman cytoarchitectonic area 17 (see Figure 3)

The work of Hubel and Wiesel (1968) highlighted the specificity of the visual receptive field for mapping from the eyes to the LGN. An anesthetized cat served as their animal model for mapping the cortical receptive field into three

cell classifications termed (1) simple cells, (2) complex cells, and (3) hypercomplex cells. The simple cells corresponded to what they called "edge detectors" because of the specific cellular response to spots of light. Primary central neural visual selectivity occurs in the LGN. There, the magnocellular ("magno") system of neurons serves to determine "where" objects are while the parvocellular ("parvo") neurons satisfy "what" of what is seen (Palmer, 1999; Shapley, Caelli, Grossberg, Morgan, & Rentschler, 1990).

Livingstone & Hubel (1988) characterized the primate "rentinogeniculocortical" pathway (from the retina to the cortex and beyond). Their experiments helped to determine the structure and functional segregation in the LGN. They found clear subdivisions of parvo and magno cells. Cells, in the magno layers, project to layer 4Cα, in area 17, or V1. The parvo cells project to layer 4Cβ of the LGN (Kandel & Wurtz, 2000).

Palmer (1999) reported that the layer 4Cβ blob subregion is rich in the enzyme cytochrome oxidase. The specificity of cytochrome oxidase (a stain for cell bodies) in the blob region allowed Livingstone and Hubel (1988) to distinguish the interblob from blob areas and to map four functionally specific pathways important to motion. Livingstone and Hubel (1987, 1988) found distinct functional pathways in rising hierarchy from the retinal output cells to the cerebral cortex. The hypothetical pathways are (1) color pathway, (2) form pathway, (3) binocular pathway, and (4) motion pathway. These parallel visual pathways course from the retina to the lateral geniculate nucleus (LGN) and to the layers of area 17— the striate cortex (V1-4Cβ, V4, 4Cα, 4B).

Livingstone and Hubel (1987) reported that over 80 percent of the parvo neurons show color opponency and are essentially color-sensitive. However, the magno neurons receive input from the red and green cones and possibly the blue cone photoreceptors. Here again the variance among opponent cells was demonstrated. The investigators found that the magnocellular response to color was suppressed by diffuse red light but not by white light. The latter finding suggested that the magnocellular neurons were more sensitive to reflective light contrast (luminance) than were the parvo neurons.

Overall, Livingstone and Hubel (1988) outlined physiological findings from primate and human perception studies relating to the magno and parvo systems. They presumed that these two systems were coupled with the blob and interblob pathways for color perception and shape discrimination. Livingstone and Hubel hypothesized that the magnocellular system is involved in "linking properties." The properties are related to (1) figure/ground discrimination, (2) collinearity (illusionary borders), and (3) linking by movement to relative and apparent motion effects. According to Gestalt psychologists, the linking properties help humans to make sense of images with orientation and motion features.

Palmer (1999) reported that visual images are constructed from the input of parallel pathways that process several features—movement, form, depth, and color. He suggested that for visual perception and attention, the neural associations are coordinated by an unspecified "binding mechanism" that may be related tangentially to the Livingstone and Hubel's linking properties.

Color Vision Limitations

Under normal circumstances, humans are not able to direct their attention to or simultaneously focus on opponent colors—for example, red-green. In juxtaposition, green-red and even blue-yellow combinations require that one attend to either one color over the other. Color opponent theory provides an explanation of the visual system and its limitations for concurrent perception of red and green without changes in hue and saturation. Likewise, typical humans also possess a blue-yellow opposition.

In an effort to demarcate separate neural pathways for color, Livingstone and Hubel (1987, 1988) found a limitation in "convergent sensitivity" (rotation of the eyes toward each other, to focus on an object). This occurred for the red-green and blue-yellow color contrast. Convergence in this context relates to changes in depth perception, or stereopsis. This finding was important because binocular vision (involving both eyes) is critical to depth perception. Humans do not view the world in images of discrete colors. Rather, visual perception encompasses reflected light across the visible spectrum (400-750 nm).

Color Deficiency

It is rare that a person is "color blind" and sees only variations of gray. Rather, the inability or difficulty in discriminating blue-yellow and red-green color combinations may be termed "color deficiency." Schwartz (1999) noted that a heterozygous female ($\overline{X}X$) is a carrier, whereas a male who has a defective

gene is color defective or deficient ($\overline{X}Y$). Therefore, a female must be homozygous ($\overline{X}\overline{X}$) to express the color defect because the color defect gene is recessive.

Molecular genetic evidence has indicated that red-green color anomalies are inherited and sex-linked (Nathans, 1989). Nathans reported that highly homologous genes for the L and M cones are arrayed from head-to-toe (in tandem) along the X chromosome. This arrangement of homologous genes led Neitz and Neitz (1994) to propose that an error might occur with the crossover of genetic information (during meiosis) when the pair of X chromosomes aligns for the exchange of genetic material. Neitz and Neitz reported that a specific error might occur in the misalignment of an L-cone gene coding with an M-cone gene coding. One of the chromosomes might lose its photopigment gene as in dichromacy (deficiency of two colors).

Schwartz (1999) reported a higher incidence for sex-linked, X-linked recessive genes for color deficiency for red-green in males and a lower incidence in females. The prevalence of red-green defects was 8 percent for males and only 0.40 percent for females (Nathans, 1989). Palmer (1999) reported that 8 percent of males and fewer than 1 percent of females have a form of color blindness or deficiency. He noted further that most color deficiency was differentiated into one of three types. The first two types, "protanopia," and "deuteranopia," characterize red-green color deficiency. A much rarer type of color deficiency (0.02% of males and 0.01% of females) distinguishes

"tritanopia," which is an inability to discriminate blues and yellows (Palmer, 1999; Nathans, 1989).

Winner-Take-All Network

Palmer (1999) elaborated on theoretical computational models that serve to explain correspondence between neural structures and perception. The reduction of specific aspects to the background is termed a "winner-take-all" (WTA) strategy for visual perception. A WTA network is defined as a set of synaptic connections that mutually inhibit each other with a common negative connection weight (synaptic strength). The network triggers the single unit with the highest activation (neural firing rate) to dominate all the other units. In the context of attention, the WTA network appears to correspond to figure-ground visual perception (Hopfield, 1982). Kandel and Wurtz (2000) noted that our eyes are accustomed to fixing on specific objects or targets. When this occurs everything else is reduced to background.

Visual Attention and Selectivity

Human Ocular Motor System

Palmer (1999) noted that eye movements are twofold: (1) fixation to position targeted objects of interest on the fovea where visual acuity is greatest, and (2) tracking to maintain the fixation of objects on the fovea despite the

motion of the object or the observer's head. Since the cone receptors are located in the highest density within the fovea region, the potential for neural excitation there is greater than in the retinal periphery. Consequently, eye movements are necessary to direct and redirect the eye to shift our gaze as we examine different aspects of a visual scene.

Palmer (1999) noted that a study of human visual attention would be incomplete without consideration of the ocular motor system. Human eye movement occurs through three pairs of muscles that function complementarily. Goldberg, Eggers, and Gouras (1993) described each eye as being coupled to four rectus muscles (superior, inferior, medial, and lateral) and two oblique muscles (superior and inferior). These muscles accomplish the rotation of the eye in three axes of space. Goldberg et al. (1993) reported that three degrees of freedom correspond to (1) the Y-axis for "forward-looking" line of sight vision, (2) the Z-axis for horizontal rotation of abduction and adduction movements, and (3) the X-axis for vertical plane rotation of elevation and depression.

However, not all possible torsional rotational movements are assumed within the Y, Z, and X planes. Goldberg et al. (1993) further classified eye movement according to specific functions. They subdivided the movements into the two classifications: (1) movements that stabilize the eye as the head moves (vestibulo-ocular and optikinetic movements), and (2) movements that keep the fovea on a visual target. The latter of the two classifications concern the eye movements of (1) saccades, (2) smooth pursuit, (3) vergence, and (4)

convergence. These four aspects of optikinetic movements are important functions in visual attention and selection.

Saccades

Goldberg et al. (1993) and Palmer (1999) defined saccades as very rapid, sudden, ballistic eye movements that function to bring new objects and surface features of interest to the fovea. Once a saccade has begun on a trajectory, it continues to its target without alteration in velocity or acceleration. Goldberg et al. reported further that a single saccade covers approximately 150-200 milliseconds (ms) during the planning and execution phases. Along the trajectory, the movement can attain a velocity of 900 degrees per second.

Smooth Pursuit

Smooth pursuit eye movements may be defined as those that track the position of a moving object to maintain its image in the foveal vision once it is established there. Goldberg et al. (1993) reported that the smooth pursuit system moves the eyes through space by calculating the velocity of a moving target while repositioning the eyes with the speed of the target. Smooth pursuit requires that you must first attend to the object or target to follow it. Accordingly, smooth pursuit is not an involuntary response.

Palmer (1999) distinguished saccades from smooth pursuit in several important respects: (1) "Smoothness pursuit" movements are not jerky and

abrupt like saccades. (2) "Feedback pursuit" movements require constant correction based on the visual feedback from the image—that is, the signals from brain-to-eye must be updated continuously to maintain target tracking of the image on the fovea. (3) "Speed pursuit" movements are slow in comparison to saccades. (4) "Acuity" is the image projected onto the fovea that must be held constant during pursuit. The targeted image remains clear while the images of untracked objects—including all stationary and moving objects—are perceived as smeared or unclear.

Vergence

Vergence refers to eye movements that control eye toward a target in determining depth. In vergence movement, the eyes are disconjugate (in different directions), whereas in pursuit the eyes are conjugate (in the same direction). If a moving object maintains pursuit and depth components, then the eye movements will include both conjugate and disconjugate components (Palmer, 1999).

Convergence

Convergence refers to the angle through which both eyes are rotated or turned in toward each other (Palmer, 1999). Schwartz (1999) noted that the degree of convergence varies with the distance of objects. For a fixed object or target, the angle of convergence is limited. Palmer noted that up to a distance of

a few meters, the convergence changes rapidly based on the following mathematical relationship: $d = \frac{c}{2 \tan (a/2)}$. The quotient d corresponds to the distance (in meters) between the eyes and a target. The numerator c refers to the convergence angle of both eyes. The denominator refers to the convergence asymptote (limiting value) of zero degrees (a/2), when the eyes are straight ahead (at a tangent for binocular vision: 2 tan).

Ocular Motor Neural Physiology

Goldberg et al. (1993) and Sharp and Phillips (1993) reported that saccades and smooth pursuit movements are organized in the pontine and mesencephalic reticular centers of the brain. The "center of control" lies within the paramedian pontine reticular formation (PPRF) of the pons. The specific control of saccades (in the brain stem) is in the superior colliculus (SC). Within the SC, the neurons are arranged in superficial intermediate and deep layers. These layers receive excitatory projections from the frontal eye field and posterior parietal cortex. Substania nigra inhibitory projections receive motor control from the frontal eye field to refine eye movement in saccades.

The medial temporal (MT) and medial superior temporal (MST) cortex both control smooth pursuit movements. It is important to note that the MT is associated with the "what" system (the ventral pathway) for object recognition. The input seems to project through the cerebellum and pons and then to the brain stem gaze centers. It is within the gaze center that the control to track a

moving target is generated. The second visual area adjacent to the striate cortex (area V2) appears to control vergence movements. The "where" system extends from the striate cortex extending to the parietal lobe (the dorsal pathway). It is here that spatial processing occurs.

Visual Attention and Perception

In the context of visual perception, Tessier-Lavigne (2000) reported that Gestalt psychologists generally ask selected questions like the following: (1) How is attention focused on one object in the visual field? and (2) What features of the object stand out from the background? As a result of psychophysical methodology, Triesman (1986) described two sequential visual perception processes: one preattentive and the other attentive. The preattentive process involves object detection and corresponds to (bottom-up) processing. The characteristics of color, orientation, size, and direction of movement are discerned in this process. The attentive (top-down) process is concerned with the dependent identification of a gestalt or whole image.

Triesman proposed a saliency map that considered color, orientation, size, and distance. She noted that such a map is essential during a search process. Tessier-Lavigne (2000a) reported that the search time is shorter when an individual is instructed to locate one different attribute from a scene or targeted surface. Conversely, the search time is longer when one must locate two or more attributes from a similar scenario. Triesman (1986) associated the faster

preattentive effect with parallel processing and the slower attentive process with serial processing.

Wickens (1992) reported that visual selective attention experiments on human performance exploit a small region of the visual field via the fovea. With a limit of only 2° of arc, pursuit and saccadic movements entail search through visual sampling until the target is acquired. This acquisition is termed spatial proximity. The physiologic limitations of the eyes necessitate sampling strategies whereby a subject can form a mental model or map of selected events. Such modeling optimizes performance from a predictable or expectant rate. A target-rich environment (indicator lights, metered values, colors) must be optimized by salient attributes that serve to increase the signal-to-noise ratio. Therefore, target-acquisition is achieved with a higher success rate when color, frequency, and intensity changes occur.

Audible Attention and Speech Perception

Massaro (1987) described top-down sources of information as those including semantic and syntactic constraints, whereas bottom-up sources included features and letters comprising words. Pickett (1999) reported that speech perception refers to essential aspects including the following: (1) first and second formants of vowels, F_1 and F_2; (2) acoustic cues or patterns related to consonant frication, place, and manner; and (3) interaction of formant transitions. Massaro characterized pattern recognition as an integration of information

(vision, audition) from multiple resources. He considered bimodal speech perception (integrated processing of audible and visible sources) to be essential to pattern recognition. Similarly, speechreading involves the ability to recognize visible patterns with corresponding audible markers, the combination of which creates a specific perception. Through masking, Preminger, Lin, Payen, and Levitt (1998) concluded that facial regions other than the mouth are critical to speech perception of consonant phonemes.

As a whole, it would be procedurally tedious and daunting to determine the exact neural correlates of speechreading relative to the visual system. However, it is logical to consider that color (corresponding to wavelength) could be a factor in accurate perception of certain definite articulatory movements, especially near the photopic luminosity peak.

Dyslexia

Models for dyslexia have been based either on a phonological model (Fitch, Miller, & Tallal, 1997; Shaywitz, 1996) or on a visual perceptual model (Livingstone, Rosen, Drislane, & Galaburda 1991; Skottun, Parke, & Lesley 1999; Stein, 1993). Livingstone and Hubel (1987) established that the magnocellular system carries information about motion and stereopsis (depth perception relative to lateral displacement). Their anatomical evidence, derived from the primate visual system, implicated fast temporal resolution in the magnocellular pathways and high spatial resolution in the parvocellular layers of

the lateral geniculate nucleus (LGN). The relationship between magno and parvo systems is not very clear. However, it is clear that the human visual system handles color and motion separately (Tessier-Lavigne, 2000a).

Through visual evoked potentials in subjects with dyslexia, Livingstone and Hubel (1987) noted specific physiological evidence for rapid visual processing in magnocellular system abnormalities that could be related to reading processing. The investigators postulated that a magnocellular system accounts partially for difficulties in reading based on suppression of the fixations performed by the parvocellular system. The findings suggested a relationship between variations in contrast and rapid processing difficulties associated with reading ability. The related anatomical indications suggested that the magnocellular system might be responsible for depth perception, hyperacuity, figural grouping, illusory border perception, and figure/ground segregation (Livingstone et al., 1991). Further, individuals with dyslexia have been shown to have problems of visual instability and visual localization, both of which are related to the magnocelluar pathway (Stein & Fowler, 1993).

The visual system has been considered foremost for functional magnetic resonance imaging (fMRI). Using this technique, one study has shown that there was no significant difference in the incidence of dyslexia in male and female subjects (Shaywitz, 1996). There continue to be disputes about differences among those with dyslexic tendencies. Livingstone et al. (1991) reported possible gender differences in the processing of information among individuals with dyslexia. Later, Lambe (1999) reported that through fMRI, categorical

gender differences have been seen. Females exhibit different patterns of functional activation during phonological processing from males. Lambe noted that animal brain studies corroborate rapid processing activation in males based on lesions from post-mortem studies. Female subjects have not demonstrated the same activation for rapid processing.

Shaywitz (1996) reported that phonologic outcomes of dyslexia include (1) feeling or perceiving illusory movement while reading and writing, (2) problems with vision not detected through eye exams, (3) being easily distracted visually, and (4) rapid naming difficulties. However, the rapidity of naming (labeling) corresponds to a limited association with dyslexia and remains disputed (Marshall, Snowling, & Bailey, 2001).

Finally, color vision has been associated with dyslexia, in that the rapid motion of saccades during reading is mediated by color-insensitive neurons in the magnocellular system (Schwartz, 1999). Livingstone and Hubel (1988) confirmed that people with dyslexia process reading with partial involvement of the color-sensitive parvocellular system. Skottun et al. (1999) noted that some categories of dyslexia might be the result of parvocellular suppression of the magnocellular system. Skottun et al. noted further that there was no definitive explanation to account for all the discrepancies that cause to reading difficulties for individuals with dyslexia.

Color Specification and Appearance

The CIE Color Specification System

A color standard for specifying color is maintained by the Commission Internationale de L'Eclairage (CIE). The highly theoretical CIE system specifies relative amounts of primaries (red, green, and blue) essential to match a color sample. The values (measured in wavelength) are mathematically imaginary: none of the primaries represents real color primaries. The calibration of television (TV) and video monitors is a primary purpose of the CIE system. Schwartz (1999) reported that the mixtures of three primary colors within the CIE system— red, green, blue (R, G, B)— are required to match a given sample. The collective objective description of "color quality" for R, G, and B is called chromaticity.

Schwartz (1999) described the CIE chromaticity diagram that is used to specify the relative amounts of R, G, and B in color matching functions. From matrix algebra, the color matching functions are then transformed to match three imaginary primaries (X, Y, and Z) and become the chromaticity coordinates within three-dimensional Euclidean space. Schwartz (1999) and Palmer (1999) agree that the CIE is highly theoretical. The mathematical treatment of color is a daunting task. This demonstrates why the CIE system and the corresponding diagrams are difficult to understand at best (Schwartz, 1999). However, in the television industry, there are devices (to be discussed in Chapter 3) that allow

one to approximate color matching for establishment and replication of production standards.

The 1931 chromaticity diagrams are based on a 2^0 CIE x, y, z color matching functions and are the benchmarks for colorimetry and photometry. The television monitors of most manufacturers act in accordance with this standard. Wyszecki and Stiles (1982) noted that television monitors are calibrated to 1931 2° CIE "Standard Colorimetric Observer Data" (hereafter referred to as "CIE 1931").

In image processing, video and the displays for computer graphics, "gamma" (symbolized by g) corresponds to a power function from luminance to voltage. The gamma function represents a numerical parameter that explains the nonlinearity of intensity reproduction of a cathode ray tube (Poynton, 2002). Poynton noted that gamma codes intensity into a "perceptually-uniform" domain to obtain the best perceptual performance from a limited number of bits in each of the red, green, and blue (or cyan, magenta, yellow, and black) components. Therefore, the perceived colors from a video camera can be reproduced on a cathode ray tube (CRT) or television monitor with minimal deviation.

The Munsell Color Appearance System

There is not a direct correspondence of wavelength to perceived color; rather, a system is required to describe the subjective experience of color. Munsell (1929) created a color system that allows one to express perceived color into three physical dimensions: (1) hue, related to the wavelength of a stimulus;

(2) value, a reference for reflectance and related to the lightness or brightness of a color sample; and (3) chroma, related to the saturation of the color sample. Hue, value, and chroma were symbolized as H, V, and C, respectively. A typical Munsell notation is written in the form of H V/C. Accordingly, the notation allows for a precise relation between adjacent colors on a continuum.

Munsell (1929) defined "color space" as a three-dimensional coordinate system wherein color phenomena are characterized as a single quantifiable point. The Munsell color space is arranged in a circle or a wheel with 100 arbitrary steps. Neutral colors were placed about a vertical plane and perpendicular to a "neutral axis."

Munsell (1929) designated a separate "value" scale from 0-10 for lightness from pure white to pure black. The "chroma" scale is the degree of departure of a color from the neutral color of the same value (color). The scale ranges from zero for neutral color to an arbitrary terminal number (positive integer). In practice, one can begin with a red that is characterized with a low chroma value and incrementally add more red until it becomes more vivid. As a whole, Munsell's color appearance system allows one to quantify the hue, value, and chroma for the desired color that is placed on series of color plates termed "chips." As such, a Munsell chip is a color plate with gradation from neutral to saturation.

Speech Perception

Massaro (1987) noted that visible speech influences the acquisition of phonology or speech sounds. However, with human communication, one cannot assume that one biologically distinct system exists for perception of phonetic information. There must be a corresponding neural link between visual and auditory processes that underpins the motor theory of speech perception (Liberman & Mattingly, 1985; Massaro, 1987). The categorical perception of speech relies on the articulation of the individual consonants. Vowels, on the contrary, are noncategorical but contribute to a continuum of acoustic cues leading to speech perception (Pickett, 1999).

From functional magnetic resonance imaging (fMRI) studies on children with language-learning impairment (LLI) Fitch et al. (1997) concluded that higher cortical order processes do not address how phoneme perception takes place. Fitch et al. (1997) reported that children with LLI show profound impairment on rapid auditory processing tasks, in a manner similar to the processing demonstrated by dyslexic children. The parvocellular system emerges as a neural linchpin for auditory-visual conveyance through auditory-visual association fibers.

Burnham (1998) reported that the development of auditory-visual-speech perception was based on three possible interactions: (1) information from one modality may enhance information in or direct attention to the other modality, (2)

information from the two modalities may be matched by the perceiver, and (3) information from the two modalities may be integrated by the perceiver. To be sure, facial kinematics interacts with the changing gaze and scan of an observer.

Although previous theoretical research suggests that the visual channel is in competition with other channels—for example, auditory (Triesman, 1969)—a few studies have sought to determine combinations that optimize speechreading performance. Testing to determine the dominant channel presents complications in planning a speechreading strategy because an examination with the specificity and sensitivity for speechreading ability for phonemes, words, and sentences has not been established.

Accordingly, Van Son, Bosman, Lamoré, and Smoorenberg (1993) investigated the role of temporal patterns of Dutch phonemes as a means to probe the contribution of auditory patterns to the perception of visible and invisible vowels. The outcome suggested that the perception of acoustically similar vowels and diphthongs received enhancement from the temporal pattern of F_1 and F_2 when coupled with visual speech (rather than audition alone).

Linguistic and lexical components are needed to decode words and sentences in a logical progression from parts to the whole. In contrast, it has not been clear whether to take a top-down perspective to speechreading (one where the individual performs from contextual and situation cues). Yoshinaga-Itano (1988) advocated a top-down methodology in contrast to the traditional bottom-up Mueller-Walle approach that serves to sequence linguistic and lexical rules with increasing complexity. With sentences related to stories, vocabulary size

did not appear to be a factor in speechreading performance (Lyxell & Rönnberg, 1991).

Massaro (1987) reported that communicating through watching the face and lips could occur with other multiple sources. Norton, Schultz, Reed, Braida, Durlach, Rabinowitz, and Chomsky (1977) noted that for individuals without sight or hearing, taction might become a possible modality that contributes to receptive communication. Norton et al. (1977) noted that the Tadoma method involved multiple sources of sensory input or information (auditory and tactile). He noted that individuals with a hearing-impairment received synchronous tactile and auditory stimulation from the placement of their hands on the neck and face of a talker.

Massaro suggested that the production of /b, p, m/ and /f, v/ were more precise for distinction than /d, t, g, k, n/ since the labial and labial-dental consonants are acquired more readily than the alveolar/velar consonants. The primary reason for the disparity is that the alveolar/velar consonants correspond to a larger viseme category. Another reason for the faster acquisition may be related to cues or patterns of consonant sounds—for example, Jeffers and Barley (1971) noted that /p, b, m/ serve as an "oral period" insofar that the phonemes mark the end of a thought unit. They reported a frequency of occurrence for /m/, /b/, and /p/ in the initial position and /m/, /p/, and /b/ in the final position respectively. They did not report on the rate or occurrence of /p/, /b/, and /m/ in the medial position of words.

Bimodal Speech Perception

Bimodal speech perception is governed primarily by an interaction between two modalities. Bimodal, in this instance, refers to sensory input—for example vision and audition. Massaro (1987) stated that bimodal speech perception conforms to the availability of information relating to audition and vision. When one adds environmental noise to a speech communication dyad (for example), the modalities of audition and vision do not correlate perfectly: that is, if hearing is compromised, vision cannot compensate absolutely for the loss.

For all the power or energy in a speech, one can only "feel" (Lyxell, Rönnberg, Andersson, & Linderoth, 1993) or see (Preminger et al., 1998) a limited amount related to all spoken phonemes. Empirical evidence supports the notion that the sensory input of taction contributes to speechreading ability but only with practice, as shown among normal subjects (Lyxell, Rönnberg, Andersson, and Linderoth, 1993). The investigators identified taction as one of the possible sensory inputs for efficient speechreading ability. They cautioned that the results were based on one method of vibrotactile transmission (MiniVib3®). Thus, with the exclusion of taction (touch), gustation (taste), and olfaction (smell), one could infer that the perception of human speech communication involves primarily the sensory inputs of audition and vision.

McGurk and MacDonald (1976) performed one significant exploration into the contribution of visual speech to listeners with a hearing-impairment. They created a set of short movies of visible speech coupled with dubbed audio tracks.

Through the manipulation of visible speech-articulation and contradictory dubbed speech, McGurk and MacDonald demonstrated that the face could shape or influence the perceived place of articulation. McGurk and MacDonald (1976) required the subjects to indicate their auditory perception of an audio-visual playback—for example, subjects were presented with the image of person speaking while they heard a contradictory dubbed speech. The visual speech perception of a person speaking /ga/ while hearing concurrently /ba/ corresponded to the visual illusion of /da/ being spoken. This misperception of place articulation was termed the McGurk effect. They noted further that in the absence of auditory input lip movements for /ga/ were misread as /da/. In the absence of one modality dominating the other, the viewer does not have a means of resolving the difference between the two sources of information.

Perspectives on Speechreading

In the main, speechreading studies have accented a gestalt viewpoint with an emphasis on form or the sum of the parts. However, the dynamic and continuous aspects identified with speechreading probably relate to brain organization and subsequent processes for form, depth, and perception. The inherent visual system strategies that help us to discern shapes, borders, and temporal and spatial differences contribute to the interpretation of visible speech or speech pattern recognition. Several studies have suggested certain definite neural correlates during visual speech perception, including the interaction of the

right and left cerebral hemispheres (Campbell, 1998) and hemispheric dominance related to visual fixation and speech (Smeele, Massaro, Cohen, & Sittig, 1998).

Experimental studies of normal human communication viewed through the lenses of psychology (Triesman, 1969, 1985), anatomy and physiology (Rönnberg, Andersson, Samuelsson, Söderfelddt, Lyxell, & Risberg, 1999), and education (Yoshinaga-Itano, 1988) highlight the complexities of speechreading processes. Speechreading occurs, with normal-hearing individuals, as a matter of course. One of the challenges in speechreading research has been to understand the underlying processes that correspond to speechreading efficacy. Efficacy may be connected to targeted attention, to specific facial regions, and to auditory signals that will benefit the hearing-impaired population.

Visual Contributions to Speechreading

Massaro (1987) and Sanders (1982) established that visual acuity is probably the primary consideration in speechreading testing and training procedures. Wickens (1992) noted that vision is inherently more efficient for attention than hearing. It follows that speechreading enables a hearing-impaired individual to access information quickly as a means of refining an audible message. When more information is available to the receivers, they can make decisions with greater precision and accuracy.

Vatikiotis-Bateson et al. (1994) studied the percentage of time that normal-

hearing subjects fixated visually on the eyes of a talker. The subjects viewed a videotape of a talker. As the S/N ratio of the tape decreased, the allocation of attention to the eyes and the eyebrows of talkers increased. Further, the subjects appeared to change their gaze pattern of targeting. Their findings showed that both English- and Japanese-speaking subjects consumed more time fixating on the eyes than on the lips. Even with the cultural differences (English-speaking and Japanese-speaking subjects) regarding eye contact with a talker, the two groups allocated equivalent attention to the oral and perioral regions.

Vatikiotis-Bateson et al. (1994) reported two possible explanations for the visual field differentiation: (1) If no major phonetic cues are dynamic, then the listener may be aided by viewing the oral region with the peripheral visual field. (2) The regions of the face not immediate to the oral aperture (mouth), are highly correlated to the oral aperture. They suggested that the coupling of facial muscles with the neck anterior region was a possible reason for secondary visual gaze.

In comparison, Lansing and McConkie (1999) drew a different conclusion from the investigation of Vatikiotis-Bateson et al. (1994). Lansing and McConkie suggested that the subjects were focusing their attention on the nose of the talker. They suggested that it was more natural to look at the gestalt rather than parts if the linguistic information has been compromised by a low S/N ratio. Lansing and McConkie investigated the effect of the linguistic markers of intonation and stress to probe the gaze pattern of normal-hearing subjects. They reported that in the absence of audible information, the eye-gaze of normal-

hearing subjects was directed to the mouth. The investigators noted further that the eye-gaze of the subjects rose consistently to the eyes of the talker near the end of sentence completion. In conclusion, Lansing and McConkie reported that more research for eye-gaze is needed to determine whether there are strategic targets on the face that will yield critical information for speechreading accuracy.

With a presumption that speechreaders focus on the lip region, Owens and Blazek (1985) investigated normal-hearing and hearing-impaired adults observing visemes. The researchers applied a 75 percent criterion of accuracy to designate a viseme-cluster in the context of /ɑ/C/ɑ/, /i/C /i/, /u/ C /u/, and /ʌ/C/ʌ/. They reported no significant difference between the performance of subjects with normal hearing and subjects with a hearing-impairment, in the correct identification of 23 consonant phonemes. The analysis of 23 consonants revealed an overall percentage correct score of less than 50 percent. Owens and Blazek (1985) selected VCV tokens in the context of /ɑ/, /u/, /i/, and /ʌ/ as an independent variable. Their findings indicated that, for both subjects with a hearing impairment and subjects with normal hearing, speechreading was easier when the phonemes /p/, /b/, /m/ were in the VCV context of /ɑ/, /i/, and /ʌ/. In comparison, the VCV /u/ context was the most difficult for both subject groups. For each subject group, the total number of correct responses was less than chance for all consonant phonemes combined.

Owens and Blazek (1985) suggested that there was a more marked difference in the accuracy of perception of consonants for the vowel /u/ than for the other vowels. The investigators suggested that a master list of viseme-

clusters for clinical training purposes might lead to a better understanding of increasing the accuracy of viseme perception.

Jeffers and Barley (1971) reported that the vowels /ɑ/ and /ʌ/ are made with the same distinguishable lip and jaw movement. However, with considerable jaw motion, /ɑ/ was rendered to look like a low front vowel and /ʌ/ to look like a high front vowel. Therefore, a decrease in visemes for consonants produced in the /u/ context might result because of the similar lip and jaw visibility.

The findings of Fisher (1968), Jeffers & Barley (1971), Binnie et al. (1974) and Erber (1974b) show that the place features of bilabials /p/, /b/, and /m/ and the labiodentals of /f/ and /v/ were consistently identified. The vowels /æ/ and /i/, in a VCV context, corresponded to fewer errors in consonant recognition.

Speechreading and Visual Communication Processes

For speechreading processing, most researchers include assumptions from a fundamental information-processing model (Palmer & Kimchi 1986; Shannon & Weaver, 1962). Correspondingly, O'Neill & Oyer (1981) stated that the variables involved in speechreading processes are (1) talker-sender effects, (2) code or stimulus, (3) speechreader-receiver, and (4) environment. They called for more controlled studies of environmental factors—for example, distractions, lighting conditions, viewing distance, and situational cues.

To provide the maximum benefit for those who rely on speechreading for the accurate perception of the spoken word, providers of assessment and

therapy must help patients in the decoding of this dynamic and fluent process. It has been established that the strategies essential to speech perception and speechreading involve the integration of information from multiple sources.

Speechreading Communication Model

To reiterate, O'Neill and Oyer (1981) reported that the "code or stimulus materials" are critical factors in the development of a speechreading test, for the speed of transmission may compromise the validity of the instrument. Traditional approaches of speechreading training methods—for example, the Jena method, the Mueller-Walle method, and the Kinzie method—all encompass the rudiments of information processing as reported by Broadbent (1958), Shannon and Weaver (1962), and Triesman (1969).

Selected aspects of speechreading include a spoken message with both audible and visual elements and environmental interference (O'Neill & Oyer, 1981). For those with hearing-impairment the visual channel then becomes virtually turned-up for both better speech reception and perception.

Across the spectrum of their divisions, hearing "disability" and "impairment" both influence the degree (severity) of loss and speech-identification proficiency. Investigators have described characteristic effects of hearing loss with varying dependency on vision for communication (Davis, 1970; Schow & Nerbonne, 1996). Berger (1972) noted that there is a negative correlation between speechreading ability and the age of onset for deafness. He reported further that, among children with prelingual deafness, speechreading is

more difficult without the benefit of linguistic and phonologic rules. It follows that speechreading is less difficult for those who have acquired language organization and phonologic patterns.

Physiological Limitations on Speechreading

Speechreading ability can be affected adversely by poor visual acuity (Sanders, 1982). Palmer (1999) noted several visual several conditions that affect visual acuity, including myopia (nearsightedness), hyperopia (farsightedness), stereopsis (depth perception), and presbyopia (loss of accommodation with age). Further along the visual pathways, deviations from the normal anatomy and physiology create adverse changes in perception in the visual field such as hemianopsia (a loss of half of the visual field).

Environmental Effects on Speechreading

O'Neill and Oyer (1981) considered the environment for speechreading to include (1) lighting conditions, (2) physical arrangements, (3) number of senders or talkers, and (4) physical distractions. A degradation in the message-signal may be attributable to a reduction in (1) speech or vocal intensity, (2) the viewing angle of the listener, (3) varying degrees of hearing loss, or (4) an addition of background noise.

Rosenblum et al. (1996) investigated the influence of point-light for speech perception and attention to extraneous or environmental factors that could

overload an individual's ability to follow a message by audition alone. In a study about the effect of lighting conditions, Erber (1974a) found that the classification of certain consonants required optimal optical circumstances, such as sufficient lighting. Later, Erber (1983) noted that good lighting conditions were sufficient for inexperienced "lipreaders" to identify the place of consonant articulation. It is clear that poor lighting leads to difficulties in accurate language and speech perception. The findings, related to lighting or illumination conditions, agree with the photopic advantage of the visual cone system.

Color Sensitivity and Speechreading

For the human visual system, the perception of some colors is easier than others—for example, red, green, yellow, and blue (Schwartz, 1999). In some circumstances, differentiated perception of color occurs in unfiltered light (Ramachandran & Gregory, 1978) and with illumination and brightness held at a constant level (Livingstone & Hubel, 1987). The fovea, located along the retinal neuroepithelium, is the region with the greatest density of cones. It is within this region that both spatial vision and color are most acute (Palmer, 1999).

A possible link between the color-sensitive foveal region and speechreading was indicated through the research of Vatikiotis-Bateson et al. (1994), who reported that viewers dedicated their foveal vision to a talker's eyes rather than their peripheral vision to the lips. Their findings were based on the amount of visual attention (fixation time) and visual gaze patterns.

The findings of Vatikiotis-Bateson and the studies of Marassa and Lansing (1995), Preminger et al. (1998), and Rosenblum (1996) led this investigator to a question: could opponent colors influence visual perception for speechreading? An exhaustive review of the research literature revealed that color had not been considered as an influence the visual perception of speech.

Through the digital masking of strategic areas, investigations attempted to isolate the location or facial region associated with essential cues for speechreading—for example, cheeks, lips, and chin (Preminger et al., 1998). Others attempted to pinpoint the precise location on one's face relating to motion (Marassa & Lansing, 1995) and to provide closer examination of a talker's face in the presence of noise (Rosenblum et al., 1996). More research is needed to determine the contribution of vision to speechreading and the inherent vision components of motion, depth, form, and color.

Chapter 3

INSTRUMENTATION AND PROCEDURE

Introduction

Images in This Dissertation are Presented in Color

To decode a spoken message, visual processing is an important component for those who must rely on speechreading. Vatikiotis-Bateson et al. (1994) indicated that there was uncertainty about the specificity of a speechreader's visual scan of, and attention to, the specific regions of a talker's face. They have found, although not conclusively, that the speechreader's foveal vision was directed to the talker's eyes [not the lips].

Since the retinal fovea is the most sensitive for focus and color (Palmer, 1999; Sharp & Philips, 1993) and motion detection (Livingstone & Hubel, 1987, 1988), the proposed experiment was designed to determine whether opponent color had any influence on the ability of subjects to distinguish among the phonemes /p, b, m/. It has been shown by Hurvich and Jameson (1957) and confirmed by De Valois and Jacobs (1968) that when viewing either the pairwise colors of red-green or blue-yellow, humans experience a change in color perception and shift in attention from one opponent color to the other: that is, one cannot direct a proportionate focus on two opponent colors.

The following research questions were posed:

1. Is the interaction of opponent colors and the vowel /ɑ/ the source of variation for the accurate visual recognition of the /p, b, m/ viseme-cluster?

2. Is the interaction of opponent colors and the vowel /i/ the source of variation for the accurate visual recognition of the /p, b, m/ viseme-cluster?

3. (3) Are the results of Experiments I and II comparable for the accurate visual recognition of phonemes /p, b, m/?

Subject Selection

Undergraduate female college students comprised the pool of subjects, six of whom were selected as speechreaders. The subjects ranged in age from 20 to 23 years. The mean age was 21.3 years with a standard deviation of 1.03. The subjects were financially compensated from the screening process through the data collection. Female subjects were selected based on empirical observations of a lower incidence for hearing loss. Lebo and Reddell (1972) reported a trend for men to develop hearing loss earlier than women do, based on certain noise-related occupational and recreational risks—for example

construction work, firearm hunting, and loud music. The subjects did not report any prior knowledge or experience with speechreading. Upon selection, the subjects verified their agreement to participate in this study by signing a subject consent form (see Form 1). The speechreaders were divided into two groups of three participants each similar to the design of Preminger et al. (1998). The subject pool preselection included five main aspects: (1) visual acuity screening, (2) color deficiency screening, (3) stereopsis, (4) hearing screening, and (5) dyslexia questionnaire.

Visual Acuity Screening

Vision screening for noncorrected and corrected vision to ≤ 20/20 was based on an eight-letter-per-line Snellen visual acuity chart at a distance of 20 ft (6.10 m). The light or luminous power that falls on a surface is termed "illuminance" (Schwartz, 1999). The typical Système Internationale d'Unités (SI unit) for illuminance is lumens per square meter or lux (symbolized lx). At the source, the Snellen chart luminance was maintained at ≥ 1.2×10^6 lx for the screening sessions. Used primarily as a quantitative measure of visual capability, the Snellen chart provided an index of visual health with respect to a single high level of black and white contrast (Ginsburg & Hendee, 1993). A Bernell Corporation Snellen E 20 "distance test chart" (number BC11931) was used for this experiment and based on a visual angle of one minute of arc. One minute (of arc) equals one-sixtieth of one degree.

Color Deficiency Screening

Protanopia and deuteranopia (red-green color deficiency) and tritanopia (blue-yellow color deficiency) occur at a lower incidence among females than males (Nathans, 1989; Schwartz, 1999). For that reason, the subjects selected for speechreading were female. A standard color vision test consisting of pseudoisochromatic plates was used because it was quick to administer and effectively identified individuals with color deficiency.

Typical pseudoisochromatic plates are useful for detecting red-green color deficiency. However, they are not capable of distinguishing between retinal photopigments and specific blue-yellow defects (Schwartz, 1999). Therefore, "pseudo-isochromatic" plates (Beck, 1965) were used to screen for blue-yellow color deficiency.

Specifically, the subjects were screened with 15 test plates under the illumination of a 40-Watt incandescent bulb. According to the recommended Beck protocol, the subject's line of sight was at right angles to the plates, with the eyes at a distance of approximately 30 inches (0.76 m) from the plates. The subjects were considered "normal" if they correctly identified 13 of 15 plates 87 or percent (see Form 2). The task required the subjects to report numbers that they perceived from the plates. An analysis of errors indicated color deficiency based on the items missed.

Stereopsis Screening

This screening test required the subject to view three black vertical bars from a Verhoeff Stereoptor® (American Optical Company, Southbridge, MA). This investigator provided directions prior to the screening task as follows: "You are going to be shown a stereoptor, a specialized viewer with three dark bars. For eight separate trials, I will prompt you by asking, "Which bar is the nearest?" or "Which bar is furthest?" You will indicate verbally either to the "left," "middle," or "right." This investigator recorded the responses of the subjects. A successful pass was noted for those who were able to identify depth correctly in six of the eight trials (see Form 3).

Hearing Screening

Subjects were screened with an audiometer to determine whether hearing was within the range of normal sensitivity. Although the proposed speechreading task did not require hearing, a subject with a significant hearing loss [41- ≤ 55 decibels (dB) hearing level (HL)] might have had an advantage in speechreading ability. Auditory screening was performed with a Grason-Stadler Incorporated (GSI) Model GSI-17 screening audiometer. The GSI-17 met American National Standards Institute (ANSI) specification S3.6-1989 and International Electrotechnical Commission (IEC) standardization. Pure tone air conduction signals were routed to each ear individually through Telephonics TDH 39P headphones seated in MX-41/AR cushions (ANSI 3.1-1987).

The subjects were seated in a minimally-treated acoustic environment (e.g., acoustic ceiling tile, low incidence of internal building traffic). The ambient room sound level (SL) was verified with a Radio Shack® digital sound level meter (Model 33-2055). Nábêlek (1985) noted that sound level meters measure sound based on frequency response curves (A, B, C, D, or E). An "A" frequency-weighting curve was selected based on its correspondence to normal human hearing thresholds (Dadson & King, 1952). The measured A-weighted sound level for the room was 57 decibels (L_A dB).

The subjects responded with a raised hand for threshold pure tone signals. Audiometric data were recorded onto a form for 0.5 kiloHertz (kHz), 1.0 kHz, 2.0 kHz, and 4.0 k kHz in each ear (see Form 4). The criterion for passing was based on ≤ 20 dB HL referenced to American National Standards Institute guidelines (re: ANSI-1989). This investigator was screened in the subject hearing screening environment (after the subject screening) to establish a baseline for comparison to threshold responses in an acoustically treated room.

To establish a normalized audiogram for comparison to the subject hearing screening environment, the investigator was tested at a different site. The investigator's hearing thresholds were obtained within an Acoustic Systems® sound enclosure, Model RE-142. The sound enclosure was in accordance with American Society for Testing Materials (ASTM) standard E-596. A Fonix® FA-10 diagnostic audiometer routed pure tone signals to the investigator's ears through EA-3, 50 Ohm audio inserts.

The investigator's threshold responses were compared to his thresholds from the subjects' screening site. The following threshold shifts were noted -5 dB at 0.5 kHz, 1.0 kHz, and 2.0 kHz. Therefore, a correction factor of –5 dB HL was subtracted from all subject screening thresholds at 0.5 kHz, 1.0 kHz, and 2.0 kHz (see Table 1). The subjects met or exceeded the screening criterion of 20 dB HL or less at the octave frequencies noted above.

Dyslexia Screening

The investigator determined that it was necessary to screen for dyslexia based on empirical findings from a phonological perspective. Some individuals with dyslexia report several problems that might include the following: (1) difficulty with reading when letters and background are different colors, (Livingstone et al., 1991), (2) problems with binocular vision while reading (Stein& Fowler, 1993), and (3) auditory processing deficits (Livingstone et al., 1991). Shaywitz (1996) reported phonologic and visual correlates of dyslexia for (1) perceiving illusory movement while reading and writing, (2) experiencing problems with vision not detected through eye exams, (3) being distracted visually, and (4) having rapid naming difficulties. However, rapidity of naming corresponds to a limited association with dyslexia and remains disputed (Marshall et al., 2001).

This investigator selected informal adult dyslexia screening questions from sample questions from the Initial Dyslexia Screening Test (University of Bradford,

2002). There were 20 items that composed a questionnaire (see Form 5), of which attempted to highlight problem areas associated with dyslexia (as noted by Livingstone et al., 1991; & Shaywitz, 1996). This investigator included subjects with no more than three confirming responses to questions 1, 3, 6, 8, 13, 17, and 20.

Only candidates with (1) English as the first language, (2) normal visual acuity, (3) fundamental red-green and blue-yellow color discrimination, (4) normal-hearing, and (5) no overt signs or behaviors of dyslexia were considered as subjects for the study. Each of the six subjects engaged in four 40-minute sessions over a three-week period.

Talker

After orientation and training, one actor was selected from a pool of candidates who was a native United States of America English speaker, female, Caucasian, and a student. She was compensated financially for her participation in this research project. The rationale for using a female talker was based on two main considerations: (1) Cosmetic application and desired appearance was optimized on relatively light skin complexion. (2) A female talker wearing lip-color was considered less distracting than a male to the subjects viewing the talker. One talker was used as a means of controlling for one source of variability. Kricos and Lesner (1982) collected data from the Utley Lipreading Test and found that different talkers could influence the accuracy of visual recognition.

Their findings agreed with Jeffers and Barley (1971), who noted that individuals differ significantly in the manner in which they formed sounds and in the precision of sound formation.

The talker for this investigation was selected from a pool of four Michigan State University Theater Department actors. The professor who managed the student actors considered the pool proficient in the (1) enunciation, (2) phrasing, and (3) rate of speaking of Midwestern United States of America (USA) English dialect. Therefore, possible confounding variables—for example, pronunciation differences from other USA dialects (e.g., southern, eastern) and variations from a foreign accent—were considered as controlled variables.

The actor (hereafter referred to as "the talker") was videotaped in a room with a darkened background. The video image encompassed a face-front, "head and shoulder" view of the talker. Her face was illuminated with Strand Century® lighting (standard television studio lights). The temperature rating for the studio lights was 3200° Kelvin (5300° Fahrenheit). The studio lighting orientation was "down" to eliminate any shadows from the talker's face. A light meter was used to quantify the luminance of the studio and the light incident on the face of the subject. The level of studio illumination during the taping, was 2.37 x 103 lx (220 ft candles) while the illumination at the face of the actor was 1.72 x 103 lx (160 ft candles).

The talker was videotaped with a digital video camera (DVCAM). The viewing distance between the DVCAM and the talker was one meter. Backlight was incorporated into the studio setup to render an image that would be lifelike to

the viewer. The talker's script was composed of VCV disyllable tokens derived from the /p, b, m/ consonant visemes cluster. The talker spoke six disyllables. She was prompted to speak with 21.59 x 27.94 cm (8.5 x 11 inch) handheld cue cards (held in front of her). All phoneme disyllables were recorded in a serial manner. In other words, for each color condition, the talker spoke all the VCVs in three videotaped takes.

Experiment Stimuli

The /p, b, m/ viseme-cluster was selected based on the following reasons: (1) the cluster was considered stable (Jeffers & Barley, 1971); (2) the cluster was readable for visibility as shown in previous research (Binnie et al., 1974; Erber 1974b; Jeffers & Barley, 1971); and (3) the lip-colors could be applied and remain without a loss of cosmetic integrity from lip movements.

There were five blocks of VCV tokens derived from the /p, b, m/ consonant viseme-cluster. The VCVs were presented for to the five lip-color conditions described in the next section. Olsen and Matkin (1979) reported that, at least for hearing-impaired listeners, more vowel errors occur for open-set materials in contrast to closed-set characteristics. Further, a presentation of intervocalic (VCV) construction served to minimize consonant confusion errors. Accordingly, the presentation for the experiments was in a closed-set format.

The phoneme /i/ is characterized as a high-front vowel, tense and unround in speech production. In comparison, /ɑ/ is characterized as a low-back vowel, lax and unround in production. Overall, these phonemes conformed

to speech-articulation movements with high visibility (revealing aspects) and differentiation. The vowel /u/ was not selected because of its potential adverse influence on /p/, /b/, /m/ (Jeffers & Barley, 1971; Binnie et al., 1974; & Erber 1974b). As such, VCV disyllables were derived from the phonemes of /p/, /b/, /m/. There were two separate vowel contexts of /i/ and /a/ for each disyllable.

The videotape of the VCV disyllables was edited to create different orders of consonant phonemes and lip-color conditions. Simple combinatorial analysis was used to generate 10 distinct combinations of the three consonants grouped in five color conditions. The different combinations served to eliminate the likelihood of learning effects from repeated or replicated trials for each of the six subjects. The defining equation for combination was ${}_nC_r$, where C corresponded to the number (*n*) of three phonemes grouped (*r*) at a time.

The stimulus presentations were counterbalanced for Experiments I and II (see Table 2 and Table 3). The test sessions (for data collection) were structured to accommodate two subject groups of three subjects each. The subjects were required to view a silent video of the talker speaking VCV syllables derived from the /p, b, m/ viseme-cluster. The subjects were permitted to choose only from the selected disyllable in multiple-choice format. The subjects marked their perceived selection for each phoneme in a five-choice, multiple-choice format. A separate response form was provided for Experiments I and II (see Form-6).

The Experiment Room Environment

The testing occurred in a quiet room 30.35 m^2 (326.7 sq. ft). The ambient illumination from fluorescent lighting was 10 candela per square meter (cd/m^2) at the point of the video monitor (see Figure 4). Three subjects were seated in individual table armchairs while facing a video monitor, a Sony® "microblack" Trinitron® tube with a diagonal measurement of 20-inch (50.8 cm). The Sony® model KV20TS29 (serial number 8168115) met the American Television and CIE 1931 standards. A video home system (VHS) analog tape of the digital master tape was used for playback on a video cassette recorder (VCR) Sharp® model VC H952U (serial number 608836354). To obtain the highest resolution, the tape was run in the "standard play" normal speed mode at 2.01 meters per minute (6.6 ft/min).

The linear subject-to-monitor distance was 1.52 m (5 ft) to minimize the effect of viewing angle on convergence (Palmer, 1999; Schwartz, 1999). The video monitor positioned at an azimuth of 0° for the subject in the middle position (see Figure 4). Within an arc of ± 15° relative to monitor, three subjects were seated for the video presentation. The separation corresponded to an intersubject distance of 0.73 m (2.6 ft). A 1.39 x 1.21 m (55 x 48 in.) opaque partition was placed between the subjects. In this manner, the partitions were to the left and right of the subject seated in the middle position. Therefore, there were 15° of separation between the subjects seated on either side of the middle position.

The monitor height was 0.76 m (2.5 ft). The seating for the three subjects was rotated (once every two trials to counterbalance the viewing angles (±15° relative to the video monitor). As a result, the subjects rotated counterclockwise three times during the six trials for viewing within an arc of ± 15° relative to monitor (see Figure 4).

There were six 20.32 x 27.94 cm (8.5 x 11 inch) cue cards attached to the forward wall of the room as a reminder of the task per experiment. The following corresponded to the text of the six cue cards: /ipi/, /ibi/, /imi/, /apa/, /aba/, and /ama/.

Visual Stimuli

For the purposes of this investigation, color precision and accuracy was accomplished in four ways: (1) measurement of Munsell color chips for reflective light, (2) color sample matching for cosmetic makeup, (3) calibration of DVCAM with a vectorscope, and (4) CIE chromaticity diagram analysis.

Cosmetic Color

The talker's lips were the targets of focus under five conditions: (1) natural—no lip-color, (2) upper-lip red/lower-lip green-yellow, (3) upper-lip green-yellow/lower-lip red, (4) upper-lip blue/lower-lip yellow, and (5) upper-lip yellow /lower-lip blue. The color combinations in conditions 2-5 were based on the characteristics of normal human visual perception and were selected because

they were likely to cause a change in color perceptual experience—that is, one of the opponent colors would appear more saturated than the other color. Consequently, the juxtaposition of the lip-colors may have created a shift in attention and focus from one of the opponent colors.

The lip-colors were obtained from cosmetic products of the Ben Nye Company. These colors met the United States Food and Drug Administration guidelines. The Food, Drug and Cosmetic (F D & C) Act (2001) defined cosmetics as "articles intended to be applied to the human body for cleansing, beautifying, promoting attractiveness, or altering the appearance without affecting the body's structure or functions." Included in this definition are skin creams, lotions, perfumes, lipsticks, fingernail polishes, eye and facial makeup preparations...."

Removal of the cosmetics was achieved through the application of cold cream, acceptable by the F D & C (Office of Cosmetics Fact Sheet, 1995). The cosmetic remover was applied to the lips and facial tissue was used to remove it. Mild hypoallergenic soap and water safely removed any residual cosmetic agents.

The specific Ben Nye Company cosmetic colors included CL-13 Fire Red, CL-5 Yellow CL-32 Lime Green (green-yellow), CL-19 Blue, and CL-29 Black. In the context of the Munsell color system, black was used to decrease the "value" or to darken (decrease "lightness") the color. Lip-colors were placed on the talker's lips by a senior theater student who was skilled in color matching and the brush application of lip makeup.

All makeup was applied under the television studio lighting prior to the video recording. The "natural" color condition was recorded first to prevent any color deviation from the color application. The red, yellow, blue, and green-yellow colors were spread onto the lips in one-to-two thin coats. The green-yellow color was closest to "chartreuse." However, to be consistent with Munsell notation, "green-yellow" was used for technical description purposes. A cream lip "concealer" was applied to the lips to neutralize the red color of the lips for green, yellow, and blue. The concealer was especially important for the application of green-yellow. The color application preserved the natural crevices and creases of the lip surface. The thin coat application was important in all lip-color conditions that comprised more than one color. The application of the colors to the lips followed these sequences: (1) red upper/green-yellow lower, (2) green-yellow upper/red lower, (3) blue upper/yellow lower, and (4) yellow upper/blue lower as shown in Figure 5 through Figure 8.

Cosmetic Color Sample Matching

Color chip samples were obtained from The New Munsell Student Color Set (Long & Luke, 2001). The Munsell color samples were compared to a color scale from the humanly visible electromagnetic spectrum between 400 and 700 nm (Palmer, 1999).

The color matching was performed with the assistance of a senior MSU Theater Department makeup artist and a physics graduate student from the MSU

Center for Sensor Materials. The mixture of cosmetic color that was eventually applied to the actor's lips corresponded to the matching above.

Munsell Color Chips

There are different editions of the Munsell Color System. Each edition is available in either glossy or matte finishes from Gretag Macbeth of New Windsor, NY. The New Munsell Student Color Set was the reference for this investigation. The color chips (each approximately 20x13 mm) were placed on a constant hue chart in an array of rows and columns. Colors were coded for assignment to red, yellow, blue, and green.

Munsell notation incorporates a standard procedure for color scaling according to Long and Luke (2001). Munsell notation is a mathematical expression by which hue, value, and chroma are distinguished by an alphanumeric numeral assignment.

The investigator performed monochromatic color scaling to designate uniform intervals according to the three dimensions of "hue, " "value," and "chroma." These dimensions corresponded to "lightness," "darkness, " and "color strength" respectively. All of the cosmetic lip-colors were verified and scaled according to Munsell notation. In the notation "5R 5/14," for example, the first number (5R) represents the hue (R) indicates red, the second number (5/) represents the value, and the third number (/14) represents the chroma as shown in Table 4.

Munsell-to-CIE 1931 Conversion

It was necessary to quantify the colors since the visual perceptual experience of color corresponds to wavelength. A financially compensated physics graduate assisted in the calculation of wavelength. The analysis for wavelength was performed at the MSU Center for Sensor Materials. This site was an optically controlled environment to match the studio lighting conditions of the video production. Newhall, Nickerson, and Judd (1943) reported that the color chips from the Munsell Book of Color correspond to the Commission Internationale de L'Eclairage (CIE) tristimulus values or coordinates (y, x, z).

There is no algorithm to convert unequivocally from Munsell notation to CIE 1931. However, to derive an approximation of the wavelength corresponding to the color samples (within the precision of nanometers), the following steps were performed: (1) Munsell notations were assigned for the color chip (samples) used for cosmetic color matching (red, yellow, green-yellow, and blue) as shown in Table 4. (2) The color samples were analyzed for lighting with the assistance of the MSU Center for Sensor Materials. (3) Tabled values from Wyszecki and Stiles (1982) provided the x and y values from the CIE 1931 chromaticity diagram. (4) The CIE 1931 luminance factor (Y) was calculated from the Munsell value (V) factor that corresponds to darkness as shown in the Equation 2.

$$Y = 1.2219\,V - 0.23111\,V^2 + 0.23951\,V^3 - 0.021009\,V^4 + 0.0008404\,V^5$$
$$\left(V^2 = V\ squared, V^3 = V\ cubed, etc.\right) \quad (2)$$

In this equation Y corresponds to photopic luminance factor for a specific primary—for example red, green, and blue and where V corresponds to the Munsell notation value. The Munsell-to-CIE 1931 values were indexed to a chromaticity diagram (Wyszecki & Stiles, 1982) to obtain the approximate wavelength (in nanometers).

DVCAM Color Calibration

Specific color estimation and experimental replication were determined with a vectorscope. A vectorscope is an instrument that provides a primary video measurement to calibrate the dispersion of chromatic levels (displayed as a histogram of colors). Prentiss (1971) described the vectorscope display as one composed of a distribution of 10 color bars gated at 30^0 intervals (30, 60, 90, 120, 150, 180, 210, 240, 270, and 300 degrees). A standard rainbow color oscillator is crystal-controlled for the frequency of 3.58 megaHertz (Mhz), a frequency that contributes ultimately to the raster of 10 color bars. The phase or hue ranges from yellow-orange to bright green and includes red, blue, magenta, and cyan. After the talker's lip-color application, the composite video image was calibrated

with the vectorscope. The specific vectorscope for this experiment, was a (Magni® WVM-178) video monitoring system. The coordinates for the color distribution were recorded for future reference and replication.

After the cosmetic application to the talker's lips, a Sony® digital video camera (DVCAM®) focused on the talker's face from two angles (0° and 30°). An analog third camera (Sony® BETACAM®) focused on the talker at a 45° angle. The recorded images from the 0° angle were the main focus of this investigation. The imaging, from 30° and 45°, was preserved for possible future research.

Procedure

A Sony® imager/studio (DVCAM®) format was used to videotape the subject on a 0.5-inch master cassette tape. The digital video camera was the selected model for video production (see Table 5). A desired feature of the DVCAM format is 4:1:1 compression ratios, corresponding to a superior broadcast quality playback image. A second benefit is the precision of editing. The DVCAM format provided optimal resolution for the illumination of the talker's face.

The video images were recorded directly onto a Sony® DVCAM™ digital medium (PDV-34ME). The videotape was edited for a counterbalancing of the phoneme disyllable and color condition sequences. An eight-second interstimulus interval was inserted between the disyllables. An 11 percent gray

reflectance monitor image coupled with a 1000 Hz signal level served as a warning that another stimulus phoneme was to follow subsequently.

The frequency of 1000 Hz has been considered the easiest for normal-hearing individuals to perceive based on the human hearing sensitivity data of Dadson and King (1952). The color gray served as neutral point for the intervals between the disyllable presentations. The sequence followed each of five intervocalic disyllables per experimental condition.

The raw video footage was recorded onto 0.5 inch format Sony® DVCAM® digital media. The Sony® model PDV-34ME master tapes were selected because of their broadcast quality reproduction. This medium also offered a major advantage of preservation and integrity of playback compatibility with digital video (DVD) format.

A studio recorder (Sony® DSR 1,800) was coupled to an Avid® system (computer-based editing instrumentation for selecting frames, indexing, and timing). All video production and editing was performed by the MSU College of Communication Arts and Sciences Information Technology (CASIT) laboratory. using the Avid® system. The camera technicians, video technicians, and video editor, who worked at the CASIT laboratory, were financially compensated. The digital media were edited and dubbed onto a TDK® Digital Source Performance® (DSP) VHS tape for playback for the subjects. A digital master was preserved as well as the raw footage from the 30° and 45° angles.

Data Collection

Three subjects were tested simultaneously per experimental session. Over a two-week period, all six subjects were tested. Each subject was assigned an identification (ID) number (S1, S2, S3, S4, S5, and S6). Seating was assigned and rotated counterclockwise for "A," "B," and "C." Position "A" corresponded to -15°, position "B" (in the middle) was at 0°, and position C corresponded to -15° relative to the video monitor. The subjects changed their seating position after every two trials to provide a different viewing angle from each seating position (± 15° relative to 0°). Therefore, for each experiment, the subjects viewed the video monitor within 30° of arc (see Figure 4).

The subjects were given a response form (see Form 1) that verified their ID and relative seating position. Each trial for Experiment I and Experiment II consisted of six presentations of intervocalic disyllables (including a repeat of trial one) separated by an interval of eight seconds (s) (5 seconds for a decision + 3 seconds for an audible warning). The warning tone was the only sound on the videotape. The subjects were given an opportunity for a break only between VCV—color combination trials.

The investigator read a brief introduction to the subjects: "You will be watching a person speaking without sound. The talker will be speaking in disyllables (two syllables) in which either the vowel /i/ or /ɑ/ was in the initial and final position. The consonant will be either a /p/, /b/, or /m/. Each disyllable (token) will be presented in one of five lip-color conditions: (1) natural-

no lip-color (2) upper-lip red/lower-lip green-yellow (3) upper-lip green-yellow/lower-lip red (4) upper-lip blue/lower-lip yellow (5) upper-lip yellow /lower-lip blue). Your task will be to check one box that corresponds to your perception of the spoken syllable. You will be given five seconds to check your answer, followed by a three-second warning signal to prepare you for the next presentation. The tone will be the only sound that you will hear during the video presentation. Your assigned seating position will change for every two trials as indicated on your response form. This investigator will assist you between trials. For each session, you will view the experiments twice. In the end, you will have viewed each experiment four times."

Each trial was approximately 2 minutes and 11 seconds long (see Table 5). The sequence for all six trials covered approximately 16 minutes per experiment. The total duration for the video presentation (Experiments I and II) was 33 minutes and 27 seconds. With a repeat of the trials four times, each subject generated 720 observations for Experiments I and II. To forestall fatigue, the subjects were given two sessions to complete viewing the experiments four times. The test presentation corresponded to a closed-set; that is, the subjects were permitted to choose only from the selected phoneme in multiple-choice format as shown on Form 6. The subjects marked their perceived selection for each phoneme in a three-choice multiple-choice format. Each experimental condition test session generated 90 observations in six trials. There was a separate response form for the four viewings of each experimental condition as shown in Form 6.

At the end of each experiment, the subjects were given an opportunity to take a break and to rotate their seating for the next experiment. This investigator retrieved the response forms after each experiment to note the sequence of the viewing—for example, Time 1, Time 2, Time 3, and Time 4. In addition, the investigator verified whether there were any missing data.

Chapter 4

RESULTS

Prior research in speechreading has indicated that the phonemes /p/, /b/, and /m/ are not recognized as discreet patterns (Jeffers & Barley, 1971). However, Owens and Blazek (1985) noted that transitional effects—for example, vowel contexts from a "lips closed" position—rendered /p/, /b/, and /m/ very similar in appearance. Erber (1974a) reported that consonant confusability occurred with profoundly deaf children speechreading /p/, /b/, and /m/ for the VCV context of /i/, /a/, and /u/.

Owens and Blazek (1985) reported further that the overall speechreading recognition of visemes, including the /p, b, m/ cluster, was no better than chance for hearing-impaired and normal-hearing subjects. These findings appeared to corroborate the research of Binnie et al. (1974); however, the investigations of Kricos and Lesner (1982) and Lesner (1988) suggested that the talker could make /p/, /b/, and /m/ easier to discern when consideration was given to speaking rate, rhythm, and facial cues.

In the context of opponent color theory (Helmholtz, 1867; Hering, 1878; Palmer, 1999; Schwartz, 1999), several assumptions could be made: First, the juxtaposition of the upper and lower lip-colors of (1) red with green-yellow and (2) blue with yellow would influence the accuracy of speech perception of /p/, /b/, and /m/. The peak of the human photopic luminosity function ($\lambda\nu$) corresponds to the greatest neural excitation in the human foveal region. For this reason, a

second assumption was made that green-yellow (in opposition to red) might add to or detract from the attention of the viewer during a speechreading task. Third, it was assumed that yellow (in opposition to blue) would also add to or detract from the attention during the visual recognition task of speechreading but to a lesser degree than a green-yellow and red opposition.

This study sought to (1) provide an introduction and background to speechreading and (2) emphasize the need for a new therapeutic approach by exploring the interaction between color perception and speechreading. Color-vision has been associated with (1) a component for information processing (McGurk & MacDonald, 1976; Treisman & Davies, 1973), (2) a contributor to depth perception (Triesman, 1986), and (3) a factor in motion detection (Livingstone & Hubel, 1987, 1988). The data were analyzed using software from the SAS Institute including Statistical Package for Social Science (SPSS) PROC® logistic (SAS Institute, 2002b) and JMP® 5.0 (JMP, 2002a).

Design of Experiment and Analysis

There were six counterbalanced trials for each experimental condition. The first trial for each experimental vowel context (/i/ and /ɑ/) was repeated in trial six for an index of reliability. Trial six was not factored into the combinatorial equation since it was a replica of trial one as shown in Table 2 and Table 3.

The subjects are termed statistically as "cases." The population sample of this investigation corresponded to six subjects. An "observation" corresponded

to one choice or judgment of a phoneme from the /p, b, m/ viseme-cluster during the six trials of each experimental condition. Each trial consisted of 15 VCV tokens. Accordingly, Experiments I and II consisted of 2,160 observations each [(6 trials x (15 VCVs) (4 viewings of experiments x 6 subjects)]. Therefore, 2,160 observations for Experiment I (/i/vowel context) and 2,160 observations for Experiment II (/a/vowel context) produced 4320 observations from the sample of six subjects.

For Experiments I and II, lip-color and vowel context were independent variables. The lip-colors were at five levels: (1) natural-no—lip-color, (2) upper-lip red/lower-lip green-yellow, (3) upper-lip green-yellow/lower-lip red (4) upper-lip blue/lower-lip yellow (5) upper-lip yellow/lower-lip blue. The data were derived from the dependent variable, visual recognition. Data were coded as a 1 "correct" response or a 0 "incorrect" response. The data were analyzed for 12 blocks of 3 phoneme disyllables across the five lip-color conditions. The main parameters of vowel, color ,and phoneme are shown in Figure 9.

The lip-color conditions were symbolized as (1) NAT for natural upper and natural lower, (2) R-G for red upper with green-yellow lower, (3) G-R for green-yellow upper with red lower, (4) B-Y for blue upper with yellow lower, and (5) Y-B for yellow upper with blue lower.

Logistic Regression

Motulsky noted that continuous data are random values about a mean and follow a normal or "parametric" distribution. Binary data are two possible values

that follow a non-normal or "nonparametric distribution". Comparable to the logic of linear regression, in logistic regression the data (dependent variable) are binary "1" or "0," and correspond respectively to "success or failure," "present or absent," or even "survival or death."

Linear regression finds an equation that can predict an outcome variable Y of continuous data from at least one of the independent variables. On the other hand, logistic regression finds an equation that provides the best prediction of an outcome variable (one that is binary)—for example, the presence of or the absence of disease (Motulsky, 1995).

Within the theoretical framework of this study, logistic regression was used as a statistical tool to determine mainly the influence of the five independent variables (lip-color) and vowel context on the outcome variable visual recognition of /p/, /b/, and /m/.

Assumptions of Logistic Regression

Similar to linear regression, several conditions must be satisfied for a logistic regression model to be legitimate (Agresti, 1996; Motulsky, 1995; Wright, 1995). Collectively, the conditions include the following: (1) Statistical independence is presupposed for all cases or values of the dependent and independent variables. (2) The dependent variable (the data) is qualitative (nominal or ordinal). (3) The data do not follow a normal distribution; that is the distribution is binomial or bimodal. In some cases, the logistic model is applied to

a multinomial distribution. 4) The categories must be mutually exclusive and collectively exhaustive—for example, a subject cannot be in the categories of a speaker of French and a speaker of Chinese. A case—for example, a subject, vowel, or color—cannot be represented in more than one outcome category at a time. (5) There must be an assumption of specificity. This requires that the model is composed of only relevant predictors for visual recognition.

Motulsky (1995) noted that one basic assumption of the logistic regression model is that each independent variable contributes independently to the odds ratio (also termed "logit"). The logistic regression equation involves a logit transformation of the probability that a success occurs. For each possible set of color condition values there was a probability *(p)* that a successful identification of the phoneme would occur as shown in Equation 3.

$$Y = \log\left(\frac{p}{1-p}\right).$$

Where the logistic regression predicts a nonlinear transformation of p (3)

$$Y = a_0 + \underset{R}{\beta_G}\, x_1 + \underset{G}{\beta_R}\, x_2 + \underset{Y}{\beta_B}\, x_3 + \underset{B}{\beta_Y}\, x_4 + \underset{NAT}{\beta_{NAT}}\, x_5 + e$$

Types of Variables

Agresti (1996) noted that categorical data consist of frequency counts of observations that correspond to response categories. He distinguished the two

primary types of categorical variables according to the class of measurement scale (nominal and ordinal). Agresti explained that for a categorical variable assigned to an ordinal scale, the dependent variables follow a progression—for example, normal, mild, moderate, and moderate-severe. The ordering of the data is appropriate for indicating a prognosis, trend, or learning effect. Alternatively, for data that are assigned to a nominal or "unordered" scale, the listing of categories is irrelevant to order—for example, religious affiliation (Christian, Islamic, Jewish, other).

For this study, the dependent variable (also referred to as the response variable) was coded for a nominal scale. The "visual recognition" category was assigned two levels of response with "correct" or "incorrect." The explanatory or independent variables were (1) vowel, at two levels ascribed to /ɑ/ and /i/; (2) color, at five levels of presentation; (3) angle, at three levels associated with the subject viewing angle (15°, 0°, and –15°); (4) subject, at six levels; (5) order, at six levels, corresponded to the sequence of trials per experiment; (6) phoneme, at three levels, corresponded to one of the three perceived phonemes of /p/, /b/, and /m/; (7) the /p, b m/ viseme-cluster response, at the seven levels of possible observations; and (8) time, at four levels was associated with the number of times each subject viewed Experiment I and Experiment II.

There were six subjects for this investigation. Wright (1995) suggested a minimum of 50 cases per predictor. Motulsky (1995) suggested that the number of events for each independent variable should be 5 to 10. He noted further that,

with logistic regression, one should not count the number of subjects. Rather, an investigator should count the number of least likely outcomes.

To answer the research questions, this investigator analyzed the data based on PROC® logistic procedures. According to the SAS Institute (2002b), the PROC® logistic procedure provides (1) model-selection methods—for example, forward, backward, and stepwise selection of explanatory variables; (2) regression diagnostics—for example, measures of leverage, influence, and residuals for each observation; (3) a choice of link function—for example, logit and probit analyses; and (4) parallel-lines tests for ordinal response variables.

Model Statistics

A statistical model is inherently incomplete, since the inclusion of all possible variables may be impractical or mathematically illogical. The "whole model" refers to the mathematical and conceptual representation of the experimental variables considered for this investigation. Logistic regression analysis was applied to the whole model in several steps. The specific design of the experiment (DOE) was written using Proc® logistic (SAS, 2002b) to include the following for analysis: (1) color, (2) order, (3) time, (4) phoneme, (5) vowel, (6) /p, b, m/ viseme-cluster response (7) angle, and (8) subject (see Figure 9). The research questions for this investigation generated categorical data based on eight variables.

Model Chi-square

Agresti (1996) noted that the function $-2\log(\ell_0/\ell_1)$ is the deviation Chi-square for the "likelihood-ratio" test statistic for independence among design variables. The likelihood ratio is the most common criterion for model fit and is used to calculate the "alpha, beta" (symbolized as α, b) combination that the data would most likely be when $b = 0$. In the end, an investigator using logistic regression only uses the significance level criterion ($p = 0.05$) to determine if the model fit. The variables that are not significant at the 0.05 probability criterion are excluded.

It might be tempting to force an analysis of nonsignificant design variables. However, Hosmer and Lemeshow (1989) advise against the inclusion of all irrelevant variables, because it leads to an increase in the standard errors caused by the common variance shared with other design variables. In other words, when independence (among the variables) is not satisfied, then effects may be assigned to irrelevant variables.

Chi-square Test Statistic

The test statistic for logistic regression independent variable coefficients is Chi-square (symbolized χ^2). The Chi-square distribution is a theoretical distribution. Similar to the *t* and *F* distribution, the shape of the distributions changes with the degrees of freedom (symbolized *df*). As the *df* increase for χ^2, it becomes easier to reject the null hypothesis. The opposite function is true for a

t-test: as the *df* decrease, it becomes easier to reject the null hypothesis (Agresti, 1996; Spatz & Johnston, 1976). Therefore, examination of logistic regression analysis involves an understanding of the *df* in respect to the null hypothesis.

Interpreting the Logistic Regression Tables

Analysis of effects.

The analysis of effects is a table item that corresponds to the statistical significance of calculated estimates. The table values correspond to the effect of two or more variables in interaction with each other or with a third variable. As with multiple regression, significant interactions could occur when the impact of one independent variable depends on the value of another independent variable. This does not mean that the effects are additive as in other forms of applied regression (Rogers, 1995; Wright, 1995).

Association of predicted probabilities.

As a whole, this table summarizes how well the model predicts the following values: (1) The "percent concordant and the percent discordant" are percentages of accuracy in classification (PAC). For binary data, the possible pairs of observations correspond to two response levels (1, 0). However for calculation purposes, 1 represents correct and 2 represents incorrect. If the

observations corresponding to the higher order value have a lower predicted value, then the pair is labeled concordant. Otherwise, the pairs of observations are labeled discordant (Wright, 1995). (2) Kendall's Tau-a is a nonparametric measure and equivalent to Spearman's rho. Kendall's Tau-a ranges from –1 through 0 to + 1 (JMP, 2002b).

Chi-square.

The Chi-square tabled items are measures of variation. Almost invariably, the Chi-square likelihood ratio is considered the more robust and sensitive for testing significance of individual effects (Hosmer & Lemeshow,1989). The method of "maximum likelihood" is the most common method for determining parameter estimates. It is twice the $-2\log(\ell_0/\ell_1)$, the –Log likelihood for difference model. One favorable aspect of the method of maximum likelihood is that it generates estimators with small variances. Consequently, an investigator can account for more error and focus on the variables that are the most significant for explanation (Agresti, 1990; Rodgers, 1995; and SAS Institute, 2002b).

Chi-square Wald statistic.

According to Agresti (1990, 1996), the Wald Chi-square statistic is a method that is used to test the hypothesis H_0 : b = zero about individual logistic

parameters for each independent or predictor variable (e.g., vowel, subject, and color). Its usefulness increases with large sample sizes. The Wald statistic tests for a specific effect (logit) relative to a constant of zero. It is a common practice to drop the independent variables where the outcomes do not meet the criterion for significance.

Coefficient for the predictor variable.

Generally, the whole model table contains several specialized terms that help characterize the data analysis for comparison, prediction, and deduction. However, to interpret the logistic regression coefficients (predictor variables), one needs to understand the concept of "odds." In this study the data were categorical with a label of "visual recognition." The data were coded with a "1" for correct and a "0" for incorrect. Therefore, the data are dichotomous.

For a dichotomous variable, as noted above, the odds of membership in the "correct" group are equal to the probability of those in the "incorrect" group (Wright, 1995). He noted further that when the odds are equal to 1 (.50/.50), the proportion for each group is .50. However, when the odds are equal to 4 (.80/.20), the proportion or probability shifts to .80, or if the odds equal .33 (.25/.75), the probability is .33.

Unlike a probability (with a range of zero to one), odds can range from zero to infinity. One can see that when the odds equal 1, both correct and incorrect are likely. As the odds increase, the more likely an observation is a

member of the target or correct group. Alternatively, as the odds decrease, the less likely an observation is a member of the target or correct group.

Logistic regression outcomes, when applied to a log-linear model, are expressed usually as an odds ratio. According to Agresti (1996) and Hosmer and Lemeshow (1989), logistic regression is used to fit a model to a binary or twofold response—for example, correct (event), incorrect (nonevent). The response variable (data) corresponds to "Y." The predictor variable corresponds to "X." For each set of possible values for the predictor variable, there is a probability (p) that a success occurs. The "Y " term is the logistic (also called logit) transformation of p so that $Y = \log(p/(1-p))$. In other words, rather than considering probabilities that cover a range from zero to one, logistic regression requires the logarithm of the odds (which can take any value, positive or negative).

Degrees of freedom (*df*).

The degrees of freedom correspond to the number of predictor variables or levels minus one. For a model with one predictor, the probability for the likelihood ratio statistic is derived from a Chi-square distribution with one degree of freedom (Agresti, 1996; Wright, 1995).

As a rule, the number of degrees of freedom is invariably equal to the number of observations minus the necessary relations from the observations (Spatz & Johnston, 1976). They noted that the number of degrees of freedom is

equal to the number of original observations minus the number of parameters that are estimated from the observations.

In the context of this investigation, the degrees of freedom correspond to the levels of the parameter estimators (color, vowel, subject, etc.) minus one—for example, for six subjects *df* = five. In logistic regression, the *df* will vary, based on the "relevant" parameters (those included in the model) and the number of levels assigned to them. For the standard error of the mean (symbolized $s_{\bar{X}}$), one degree freedom is subtracted because the mean, $\bar{X}$ is used as an estimate of the population mean, symbolized by μ.

Interaction.

Interaction, in a log-linear model, requires the concurrent specification of at least three variables—for example, vowel, phoneme, and color. In the tables of this study an interaction is noted with a "*" as in "density and temperature * pressure." The partial association of any two interactions is different for cases in the context of different categories (Rodgers, 1995).

Intercept (α).

This is an element of the raw score within the regression equation that is an indicator of the criterion score when all the predictors equal zero. The intercept always equals zero in the standard form.

Model Chi-square test.

The model Chi-square is the nonparametric analog to Duncan's multiple-range test as it functions as a significance test. The Duncan is a procedure by which an *F* ratio tests the significance in analysis of variance (ANOVA) for continuous data (Motulsky, 1995). The model Chi-square functions to test the null hypothesis that, with the exception of the constant, all population logistic regressions coefficients are zero. With a model Chi-square probability ≤ 0.05, the null hypothesis is rejected. That means that the presence of the independent or predictor variables—for example, angle, color, phoneme, and vowel— makes no difference in the prediction of the dependent variable (in logistic regression). Rejection of the null hypothesis (H_0 : $b = 0$) does not mean that all of the independents were found significant. The model Chi-square measures the improvement in fit that was made from the "initial log likelihood."

Parameter.

A parameter is an unidentified or unknown constant that quantifies an aspect of a population and is reported in the SAS output. This tabled category lists each term or parameter in the logistic model. There is an intercept and slope (*b*) for each term, at each level of the response (Agresti, 1996; Rodgers, 1995).

Probability of greater Chi-square (Prob>ChiSq).

The designation refers to the observed significance probability, or *p* value for the Chi-square test. It is the probability of getting by chance a calculated Chi-square value larger than the one computed. Models are considered significant if the probability is equal to or below the criterion of 0.05 (SAS Institute, 2002b; Wright, 1995). Statistical significance is also designated with an asterisk (*) to symbolize the level of significance—for example, *** corresponds to p, < 0.0001.

Stepwise selection.

The PROC® logistic application is capable of stepwise selection of relevant parameters (JMP, 2002b). Stepwise selection for the design of the experiment begins with no independent variables in the model. Individual independent variables are added to the model one-by-one until no other variable meets statistical significance. Therefore, the variables with the strongest contribution are included.

Cross-Classification Analysis

The data generated from this investigation were qualitative rather than quantitative. Rodgers (1995) described qualitative data as those that do not correspond to quantifiable aspects for variables of counts and amounts. The JMP "design of the experiment" (DOE) required logistic regression analysis as a

means of determining the relevant predictors to the whole model Chi-square test (JMP, 2002b). Further analysis was performed to derive the significant effects from the relevant predictors. There is a way to ascertain whether a relationship might exist between two nominal variables through contingency table analysis. A Relationship, in this context, refers to an index of association rather than interaction (JMP, 2002b).

Contingency Table Analysis

Rodgers (1995) described association as any lack of independence in the distribution of two variables, seen with an odds ratio different from 1.0 or the log of the odds ratio different from 0.0. Generally, a contingency table, also called a cross-classification table, is used in the analysis of "association" between two categorical variables (Wright, 1995). Agresti (1996) noted that a contingency table analysis is a means of examining qualitative data with respect to two variables.

Friendly (1994) stated that a contingency table for hair color and eye color might include combinations for several levels of hair color (brown, blond, red, and black) and several levels for eye color (brown, blue, hazel, green). Thus, a contingency table is composed of frequency counts of observations that occur in each combination of levels for the variable response categories.

A two-way contingency table consists of cells containing the frequency counts of the outcome or data. As with logistic regression, certain assumptions

must be satisfied for contingency table analysis. Rodgers (1995) and Motulsky (1995) explained that the cross-classification of categories must be (1) mutually exclusive and exhaustive and (2) each observation can only be a member of one category.

The contingency table is a part of Chi-square analysis when there are two cross-classified variables. However, when a third variable is added for analysis, the Chi-square statistic is no longer an appropriate test for association or independence. Accordingly, with the addition of three or more variables, logistic regression is appropriate for the analysis in multivariate qualitative data.

In this investigation, the test statistic for a contingency table was the Pearson Chi-square test. The null hypothesis for independence states that the likelihood of the response in one group for one variable is the same for other groups into which the response falls (SAS Institute, 2002b). Two variables—for example, hair color and eye color— are considered "statistically independent" when the distributions of Y are identical at each level of X.

Mosaic Plots

A contingency table analysis can be used to derive meaningful relationships across categories. Statistical models and designs for categorical data do not allow an investigator to display data in a manner similar to quantitative data—for example, sigmoidal plots, log relationships, and scatter plots. Hartigan and

Kleiner (1984) created a "mosaic plot" to depict cross-classified data. With refinement from Friendly (1994), the mosaic plot (or mosaic display) has emerged as a method of displaying meaningful categorical data for proportions and counts that are derived from contingency tables.

Generally, a mosaic plot is any plot that represents data, in which each count is represented by a small rectangle of areas whose size proportional to a frequency count. In computing each (X, Y) proportion, the Y counts are divided within each X level by the total of the level. This proportion estimates the response probability for each (X, Y) level (JMP, 2002b). Mosaic plots and contingency tables analyses comprise several graphic and cell qualities that serve to summarize qualitative data.

Interpreting Contingency Tables and Mosaic Plots

Contingency Tables

The JMP® software (JMP, 2002b) generates an output of contingency tables that feature the following analytical indexes and tests: (1) "Prob>ChiSq" indicates the probability of obtaining by chance a Chi-square value greater than the one computed, if no relationship exists between the response and factor. (2) "Source" lists the names of the three sources of uncertainty, called "Model," "Error," and "C Total." (3) "R square (U)" is the portion of the total uncertainty attributed to the model fit. An R^2 of zero means that there is no gain from fixed

background response rates. A high R^2 value is atypical of categorical data analysis. (4) "Error" does not hold much meaning except that the value becomes negative if the contingency table holds many categories of sparse data. (5) "Likelihood ratio" Chi-square test is twice the negative log-likelihood for the model test for the hypothesis. The criterion for statistical significance is *p*, 0.05. (6) "Pearson Chi-square" is another test of the hypothesis that there is no difference between the observed and expected frequency counts. (7) "Count" refers to the cell frequency, margin total frequencies, and grand total (total sample size). (8) "Total percentage" corresponds to the percentage of cell counts and margin totals to the grand total. (9) "Col percentage" relates to the percentage cell count in its column total.

Mosaic Plots

Friendly (1994) described a mosaic plot as a depiction of the contingency table for the response probability of each (X, Y) proportion. At a glance, a mosaic plot is comparable to adjacent divided vertical bar graphs (see Appendix B for examples of mosaic plots).

The "x-axis " represents the proportions of the relative sizes of the variable totals. The x axis might represent one variable with multiple levels The right y-axis represents proportions along the relative sizes of each group for the combined number of levels. The left y-axis corresponds to a scale along the

response probability. The whole axis ranges from 0.0 to 1.0 (the sample total). Typically, color is used to distinguish variables with five or more levels.

Friendly (1994) explained that the width of each column is proportional to frequency for each variable, whereas the height is proportional to the conditional frequency of each row. Friendly described complete independence for a mosaic plot when the areas of the rectangles are proportional to the frequencies. When this occurs, all the "tiles" in each row have the same height. Upon an examination of the pattern, one can determine the relative proportion for each variable and the combinations of levels.

Correspondence Analysis

Greenacre (1988) described correspondence analysis as an exploratory multivariate technique that converts a matrix of nonnegative data into a graphical display. The mathematical treatment for this technique has been used in France and in Germany over the past thirty years. In recent years, the analysis has increased in use in the United States of America. The display depicts the rows and columns of the matrix as individual points. A correspondence analysis generates a plot derived from a matrix of counts divided by the total frequency. A Euclidean algorithm converts a matrix of nonnegative data into points that represent rows and columns. The analysis shows which rows of the frequency table have similar patterns (JMP, 2002b). Thus, correspondence analysis

becomes useful for tables with large numbers of levels in situations where the numbers might be confusing.

The coordinates for the correspondence plot axes are labeled "c1," "c2," and "c3" for three-dimensional space. According to JMP (2002b), a two-dimensional plot is interpreted as follows: (1) higher response levels tend to be both negative in c1 and positive in c2, (2) moderate levels of response are negative in c2 and neutral in c1, and (3) the lowest response values are positive dimensions of c2. When there are sufficient observations to generate three levels of correspondence variables, a three-dimensional correspondence plot can be produced with "c3." The three-dimensional space corresponds to Euclidean space with the vectors of Y, Z, and X. The array of labeled vectors indicates the direction and magnitude of the responses relative to the centroid (mean center).

Descriptive Statistics

The descriptive statistical analysis calculation was based on five trials for a total of 1,800 observations; a sixth trial was a repeat of trial one. Intervocalic (VCV) tokens were derived from the /p, b, m/ viseme-cluster—for example /ipi/, /ibi/, or /imi/. There were seven possible responses per multiple choice selection; one of which was "correct" for either /p/, /b/, or /m/. The six possible phoneme substitutions ("incorrect" responses) were as follows: (1) p/b, where /b/ was the target and /p/ was the perceived phoneme; (2) p/m, where /m/ was the target and /p/ was the perceived phoneme; (3) m/p, where /p/

was the target and /m/ was the perceived phoneme; (4) m/b, where /b/ was the target and /m/ was the perceived phoneme; (5) b/p, where /p/ was the target and /b/ was the perceived phoneme; and (6) b/m/, where /m/ was the target and /b/ was the perceived phoneme.

Distribution of /p, b, m/ Viseme-cluster Responses

For Experiment I (in the /i/ context), the frequency distribution of the responses from the /p, b, m/ viseme-cluster indicated a count of 329 for /p/, 235 for /b/ and 284 for /m/ (see Table 6). The distribution for both the p/b and b/p substitutions was 188. The distribution for b/m was comparable to m/b with counts of 175 and 174 respectively. The m/p substitution represented the lowest count at 91 for Experiment I.

In Experiment II (in the /a/ context), the /p, b, m/ viseme-cluster frequency distribution indicated a count of 276 for /p/, 210 for /b/, and 257 for /m/ (see Table 6). The /b/ and b/m distributions were comparable with counts of 210 and 212 respectively. For both experiments, the highest counts were indicated for /p/ with 605 and the lowest counts were indicated with 214 for the m/p substitution (see Figure 10).

Visual Recognition by Experiment

The visual recognition of the phonemes from the /p, b, m/ viseme-cluster varied between the two experiments, as shown in Table 7. For Experiment I, the total for correct visual recognition indicated 849 observations and the total for incorrect visual recognition indicated 951 observations. In Experiment II, the total for correct visual recognition indicated 739 observations while the total for incorrect visual recognition indicated 1,061 observations. The overall results of visual recognition by experiment were derived from multiple factors including vowel, color condition, and the phonemes from the /p, b, m/ viseme-cluster.

Distribution of Color Condition by Phoneme

Tables corresponding to the distribution of visual recognition by color condition are found in Tables 8 through Table 11. Figures corresponding to the distribution of visual recognition by color condition are found in Figure 11 through Figure 20. The visual recognition of the phonemes from the/p, b, m/ viseme-cluster varied across the five color conditions of the investigation. The observations from six subjects indicated differentiation of visual recognition.

In Experiment I the distribution of color condition was differentiated by phonemes and visual recognition. Among the five color conditions, the NAT condition had the highest overall count for correct visual recognition (216). This count included 76 for /p/, 67 for /b/, and 73 for /m/ as shown in Table 8 and

Figure 15. The count was also 76 for /p/ in the G-R condition. The G-R and B-Y conditions had the same counts for correct (161). The Y-B condition had the highest individual correct count with 84 for /p/. The Y-B counts of 44 for /b/ and 37 for /m/ were among the lower counts in comparison to the other color conditions.

Among the five color conditions, the NAT had the lowest count of incorrect visual recognition (144). This count included 25 for /p/, 56 for /b/, and 63 for /m/. The R-G condition had the highest count of incorrect visual recognition (214) with 61 /p/, 88 for /b/, and 65 for /m/. The G-R and B-Y conditions had the same counts for correct (199) and comparable to the Y-B condition (195).

Among the five color conditions of Experiment II, the R-G condition had the highest count of phonemes for correct visual recognition (161) as shown in Table 9. The R-G phonemes were comparable with 52 /p/, 53 for /b/, and 56 for /m/. The NAT condition had a similar count for phonemes in the correct category (159). However, the phonemes were not similar in count with 56 /p/, 38 for /b/, and 65 for /m/. The G-R and Y-B conditions had the same counts of phonemes for correct visual recognition (142). The B-Y condition had the lowest number of phonemes in the correct category (135). The distribution of phonemes, across the five color conditions, was comparable for incorrect visual recognition—ranging from 199-to-225 (see Table 9).

Distribution of Visual Recognition by Color Condition

For Experiment I, the distribution of visual recognition by color condition indicated the following rank order from highest-to-lowest for correct visual recognition: 216 for NAT, 165 for Y-B, 161 for both B-Y and G-R, and 146 for R-G (see Table 10). Only the NAT condition indicated a higher count for correct than incorrect.

For Experiment II, the distribution of visual recognition by color condition indicated the following rank order from highest-to-lowest for correct visual recognition: (1) 161 for R-G, 159 for NAT, (2) 142 for both G-R and Y-B, and (3) 135 for B-Y (see Table 11). In every color condition, there was a higher count for incorrect than correct. The NAT condition was high in the ranking of both experiments for correct. The descriptive data indicated that the color conditions did not appear to promote correct visual recognition for the /p, b, m/ viseme-cluster.

Review of the Research Questions

The three research questions of this study addressed a possible interaction of vowel, color, and the phonemes /p, b, m/. The main data analysis targeted the presumed interaction. However, the data were also analyzed to account for unpredicted interactions with other variables, sources of error, or both. Therefore, the whole model fit included color, order, time, phoneme, angle,

and subject— plus error (as shown in Equation 3). The purpose of this test was to determine which of the seven model components were statistically significant. The PROC® logistic calculation factored only the significant variables within the model. In this manner, more of the error could be explained.

Question 1

Is the interaction of opponent colors and the vowel /i/ the source of variation for the accurate visual recognition of phonemes /p, b, m/?

Experiment I Group Data Analysis

The model Chi-square test was performed to analyze the maximum likelihood estimates. Based on the PROC® logistic procedure backward stepwise selection was conducted for the relevant parameters (SAS Institute, 2002b). The model was analyzed for the following effects: (1) color, (2) vowel, (3) time, (4) phoneme, and (5) subject. The parameter estimates for "order" and "angle" were not significant at the 0.05 statistical criterion level. The analysis of the vowel also did not assume independence with the other parameters at the criterion level.

Perhaps the simplest way to view the vowel parameter (for both experiments) is to consider that each vowel corresponded to an independent variable and to a separate slope. The intercept of each independent variable was zero. Geometrically speaking, the slopes for /i/ and /ɑ/ had the potential

to intercept with the slope of the dependent variable (visual recognition) and the slopes of the other relevant parameters (color, time, subject, etc.). The Chi-square test statistic was a calculation of the likelihood for independence among the parameters. In some cases, a significant interception did not occur for the vowel parameter.

Experiment I Model Fit Test

The initial Chi-square tests the global null hypothesis that the beta coefficients are zero. For Experiment I, the probability was greater than chance of obtaining a Chi-square value greater than the one computed by the model (p = < 0.0001). The Wald Chi-square statistic was 143.44, df = 23 (color, vowel, subject, phoneme, and time). A statistical significance (p = < 0.0001) was shown for interaction of "color by phoneme" (color *phoneme). The likelihood ratio was 161.48. This did not mean that there was statistical significance among the individual independent variables. As a result, further analysis was performed to determine the influence and statistical significance of individual parameters and their effects.

Experiment I Maximum Likelihood Estimates

An analysis of the maximum likelihood estimates indicated statistical significance for selected parameters relative to color and phoneme including R-G

(see Table 12). The PROC® logistic analysis used the Y-B color condition as α while *b* corresponded to individual parameters that met the model fit analysis. The degrees of freedom characterized the family of Chi-square distributions that are seen in logistic regression. For tests of independence, contingency tables were generated.

The results of the whole model Chi-square test indicated that the phoneme individual parameters for/p/ and /m/ were not influential predictors for phoneme recognition. It followed that these predictors were also not statistically significant for /p/ (p = 0.9739) and /m/ (p = 0.9673) for the model. Moreover, individual parameter estimates for the individual color conditions of Y-B, G-R, and NAT were also not statistically significant.

The R-G condition was statistically significant (p = 0.0379) along with an interaction of color by phoneme (color *phoneme) in the B-Y color condition p = 0.0149), as shown in Table 12. One salient aspect of the logistic data analysis was the strength of the R-G color influence (0.0379). The statistically significant influence did not appear to correspond to accuracy for R-G (see Figure 11). The data analysis suggested that there was a global influence of color for the subjects, at least for the R-G color condition as shown in the Experiment I analysis. There were no significant interactions for the other predictors with the vowel /i/ and visual recognition.

Experiment I Analysis of Effects

The analysis of effects indicated statistical significance among specific global parameters for all six subjects. The parameter /i/ was not a relevant predictor for visual recognition. Significant effects were indicated for Experiment I (see Table 13). The analysis for effects indicated statistical significance for phoneme ($p < 0.0001$) and subject ($p < 0.0001$). A significant interaction was indicated for color*phoneme ($p = 0.0203$). As an independent parameter, the global effect of color was not significant ($p = 0.1390$).

There was an assumption that the R-G and the G-R conditions would be influential because of red (616.7 +/- 0.3 nm) opposed to green (565.8 +/- 0.2 nm). The calculated color samples correspond to measurements within an optically controlled environment. To reiterate, the greatest color sensitivity, for humans, corresponds to 555 nm. The significance of the R-G (red upper lip/green lower lip) condition might have been an indicator of shifts in gaze in the speech perception task.

It would appear that the interaction of color and perceived phoneme was a source of variation for the perception of the /p, b, m, / viseme-cluster. However, the research question was based on the hypothesis of an interaction of "color by vowel" (color *vowel). In this context, the null hypothesis was not rejected for the interaction of vowel, phoneme, and lip-color since vowel was not a relevant parameter. This analysis suggested that, in certain definite color conditions, the

influence was significant for the visual perceptual task of speechreading /p, b, m/.

Question 2

Is the interaction of opponent colors and the vowel /ɑ/ the source of variation for the accurate visual recognition of phonemes /p, b, m/?

Experiment II Group Data Analysis

Experiment II Model Fit Test

For Experiment II, the global null hypothesis (H_0 : $b = 0$) was tested. The analysis indicated that the probability was greater than chance of obtaining a Chi-square value greater than the one computed by the model ($p = < 0.0001$). The Wald Chi-square statistic was 70.37, $df = 10$ (phoneme, subject, time). The likelihood ratio was 73.22. The statistical significance was analyzed for independence among the individual beta coefficients.

Through stepwise selection, the variables of color, vowel, order, and angle were excluded from the model because they did not meet the probability criterion at the 0.05 level for statistical significance. The statistically significant maximum likelihood estimates included the following: (1) phoneme, (2) subject, and (3) time

(corresponding to the four occasions to view Experiment II). The three estimates were the relevant parameters for a prediction of visual recognition.

Experiment II Maximum Likelihood Estimates

With /p/ selected as the intercept, phoneme was significant for /p/ ($p < 0.0001$) and /b/ ($p < 0.0001$), as shown in Table 14. The Chi-square test statistic indicated significance for the parameters of Time 1 ($p = 0.0065$) and Time 2 ($p = 0.0257$). For all six subjects, Time 1 and Time 2 were separated by an interval of approximately 30 minutes at the first of two test sessions. It is tempting to conclude that there might have been a learning effect; however, the estimators for Time 3 and Time 4 did not appear to support that notion.

The analysis of maximum likelihood also indicated that the subjects were a statistically significant influence on the perception of the phonemes from the /p, b, m/ viseme-cluster, when coupled with color conditions. Statistically significant values were noted for S1 ($p = 0.0078$), S2 ($p = 0.0015$), S4 ($p = 0.0058$), and S5 ($p = 0.0190$). In comparison, the analysis revealed that parameter estimates for S3 were not statistically significant. For S6 the predictor coefficient (beta) was interpreted as negligible or near zero rather than "not significant." An analysis of individual subject data will be discussed in a subsequent section.

Overall, the analysis suggested that the parameters of subject and time exerted significant influence on the visual perceptual task of speechreading /p,

b, m, /. In particular, /p/ and /b/ emerged as independent predictors for visual recognition.

Experiment II Analysis of Effects

An examination of the effects for Experiment II indicated statistically significant effects for phoneme (p = < 0.0001) and subject (p = < 0.0001). To a lesser extent, statistical significance was noted for the parameter of time (p = 0.0182) as shown in Table 15. The analysis indicated that the influences on visual perception did not include an interaction (color *vowel). Further, the color and vowel conditions of Experiment II did not produce interactions.

The parameter of time was a significant effect. The influence of viewing the experiments four times emerged with statistical significance for Time, as shown in Table 14 and Table 15. It was not clear why this predictor was significant for a design parameter. Subsequent discussion in Chapter 5 explores a few possibilities. The null hypothesis was not rejected for the specific assumed interaction of vowel, phoneme, and lip color.

Question 3

Are the results of Experiments I and II comparable for the accurate visual recognition of phonemes /p, b, m/? The interaction of the /p, b, m/ viseme-cluster, opponent colors, and the vowel /i/ constituted Experiment I, whereas

the interaction of the /p, b, m/ viseme-cluster, opponent colors, and the vowel /ɑ/ constituted Experiment II.

Experiments I and II Group Analysis

To answer this question (1) vowels, /i/ and /ɑ/ were combined in one model in addition to the other relevant design variables: (2) order, (3) time, (4) phoneme, (5) color by vowel interaction, and (6) color by phoneme interaction. This analysis encompassed both experiments for 3,600 observations (2 x 1,800 observations). The whole model fit global null hypothesis was rejected. The probability was greater than chance of obtaining a Chi-square value greater than the one computed by the model (p = < 0.0001) with df = 32.

Significant likelihood estimates were indicated for the analysis of maximum likelihood estimates. This analysis was the strongest evidence for color influence with the opponent colors of Y-B as the intercept (p = < 0 .0001),, as shown in Table 16. Color parameter estimates were significant for NAT (p = < 0.0001) and B-Y (p = 0.0237). A significant interaction "color by phoneme" (color *phoneme) was indicated for NAT /m/ (p = < 0.0001) and to a lesser degree B-Y /m/ (p = 0.0212). A significant influence on visual recognition was indicated for /b/ (p = < 0.0001).

Only the observed probability for the vowel /ɑ/ indicated a greater Chi-square value than the one estimated at the < 0.05 level. This finding did not occur for the vowel /i/ in the model fit test. In other words, there was a greater

than chance probability that the vowel /ɑ/ was a statistically significant and independent of any other member group. Specifically, there was a significant interaction indicated for vowel and color—for example, natural by /ɑ/ (NAT * /ɑ/) with (p = 0.0030) and "red-green by vowel" (R-G * vowel) /ɑ/) with (p = 0.0019). The "subject" predictor was also indicated as a significant parameter estimator for the combined analysis of the vowels /ɑ/ and /i/ (see Table 16). As statistically significant parameters, the following subjects influenced the overall analysis for effects: S1, S2, S3, S4, and S5. Subject 6 (S6) was not included in the whole model fit, because it did not meet the criterion for statistical significance (p < 0.05).

Analysis of Effects for Experiments I and II

The Wald Chi-square test analysis of model effects indicated significance for color (p = < 0.0001) and visual recognition (p = 0.0001) as shown Table 17. The interaction of color*phoneme was significant (p = 0.0063), as was the interaction of color*vowel (p = 0.0046). Overall, the effect for subject was significant (p =< 0.0001). Statistical significance was indicated for the parameter of time (p = 0.0003). Taken together, there appeared to be significant interaction for the vowel /ɑ/ and color to influence the perception of /p, b, m/.

In comparison, the vowel /i/ did not meet the statistical criterion for whole model fit with the combined analysis. It is not clear why the vowel was not significant in both the Experiment I and the combined analysis of the vowels /ɑ/

and /i/. In a contrasting finding, the vowel /ɑ/ was not included in the model fit for Experiment II but became significant in the combined analysis. Geometrically, a significant intercept was made with color and with phoneme.

Interestingly, the accuracy of the responses appeared to emerge only in the NAT condition for the vowel context of /i/, as shown in Figure 15. A similar finding was not shown in the NAT condition for the vowel /ɑ/, as shown in Figure 20. An examination of the individual subjects offered additional insight for answering the research questions.

The results indicated the following: (1) Degrees of freedom were larger for Experiment I (*df* = 23) than Experiment II (*df* = 8). (2) Both model fit tests indicated a rejection of the null hypothesis. (3) The vowel /ɑ/ was a statistically significant predictor for visual recognition in the context of an interaction of vowel, lip-color, and the phoneme responses from the /p, b, m/ viseme-cluster. However, the vowel /i/ was not shown as a significant influence on the outcome. It appeared that the vowel /ɑ/ exerted more influence on the outcome.

Individual Subject Analysis

The individual data corresponded to 360 observations for Experiment I and 360 observations for Experiment II for a total of 720 observations for further examination. Comparisons and distinctions were drawn from this analysis relative to the separate results of both experiments in the following section.

The logistic regression analysis of the individual subjects was combined

for both Experiment I and Experiment II to determine the significant influences for accurate visual perception. These data were reported according to subject identification number (S1, S2, S3, S4, S5, and S6). The specific parameters included the following: (1) color, (2) vowel, (3) time, (4) phoneme, and (5) order. The analysis of S5 revealed a "questionable" model fit. As noted in the PROC® logistic output "the maximum likelihood estimate may not exist."

For a few subjects the analysis revealed minimal effects. For S5, the only parameter effect indicated from the maximum likelihood estimate was the perception of "phoneme" (/p, b, m/). In comparison, the analysis of S6 data indicated "order" as the only significant influence for visual recognition (only in the vowel /a/ context).

Individual Subject Maximum Likelihood Estimates

The analysis of specific parameter coefficients indicated differentiated influences among the six subjects. Subject 2 (S2) emerged prominently in the color conditions of Y-B ($p = 0.0162$), B-Y ($p = 0.0005$), G-R ($p = 0.0231$), and NAT ($p = < 0.0001$) as shown in Table 18. To a lesser degree, S3 was significant for Y-B ($p = < 0.0001$), G-R ($p = 0.0461$), and NAT ($p = 0.0105$). For S4, only the NAT condition was statistically significant ($p = < 0.0001$). Statistically significant influences were not indicated for S5 for any of the relevant parameter estimates.

Statistically significant estimators for /p/ and /b/ were revealed for S1 at (p = < 0.0001) and (p = 0.0344) respectively (see Table 18). For /b/ statistical significance was indicated for S2 (p = 0.0004) and S4 (p = < 0.0001). The estimator for /m/ (p = 0.0395) was seen for S2. The vowel /ɑ/ was a statistically significant influence for S2 (p = 0.0315) and S6 (p = 0.0036).

The Effect of Color

The individual subject analysis of effects indicated statistical significance for the influence of color on S2 (p = < 0.0001), S3 (p = 0.0423), and S4 (p = 0.0003), as shown in Table 19. The overall analysis of effects indicated that S2 experienced a significant interaction between vowel and color (p = 0.0370). For S4, there was a significant interaction indicated (color*phoneme) (p = 0.0486). As it turned out, the effect of the R-G color condition was not significant in any of the subjects.

The Effect of Phoneme: /p, b, m/

The effect of phoneme was a significant influence for S1 (p = 0.0150), S2 (p = 0.0017), S4 (p = < 0.0001), S5 (p = 0.0398), and S6 (p = < 0.0001), as shown in Table 19. In the analysis of S3, the phoneme effect was negligible.

The Effect of Vowel

For S2, significant interaction was indicated for vowel ($p = 0.0315$) and an interaction of color*vowel ($p = 0.0370$) as shown in Table 19. Vowel was a significant effect for S6 ($p = 0.0036$). Vowel did not generate significant effects for S1, S3, S4, and S5. The results of the individual subjects indicated that there was a differentiated influence on the recorded observations that might have been dependent on extraneous variables.

Association of Predicted Probabilities

The analysis of PAC is a means of summarizing how well the logistic regression model calculated or identified correctly the parameters. For Experiment I, 64.5 percent of the parameters were concordant, whereas 34.8 percent were discordant. The tie for concordant and discordant was 0.7 percent. For Experiment II, 59 percent of the parameters were concordant and 38.3 percent were discordant. The tie for concordant and discordant was 2.7 percent (see Table 20). In both Experiments I and II the model predicted more values correctly than not.

Kendall's Tau-a is one of several "Kendall Tau" nonparametric measures of association (correlation) for the predictive ability of the model. It is a nonparametric alternative to Spearman's rho. Kendall's Tau-a ranges from –1 through 0 to + 1. Kendall Tau-a is based on the number of concordant and

discordant pairs of observations and incorporates a correction for tied pairs (JMP, 2002b).

For Experiment I, Kendall's Tau-a was at 0.148 and for Experiment II, the value was 0.104. Both values were statistically significant. The association of predicted probabilities adds to the validity of the model and the findings of this study because the model correctly identified the parameter estimators with better than chance performance at the criterion (p= <0.05) level of significance.

Cross-Classification Analysis of Data

Clearly, the logistic regression model correctly identified two-thirds of the parameters. However, a closer examination of the Chi-square distributions provided more information regarding the association between the /p, b m/ viseme-cluster responses and the distribution of color. Contingency tables and mosaic plots were generated to compare the distribution of the /p, b m/ viseme-cluster responses across color for Experiments I and II and to summarize the response with counts and proportions.

The null hypothesis for independence states that the likelihood of an observation falling into one group, for one variable, is independent of the other group into which the observation falls (JMP, 2002b) Therefore, this investigator probed independence between the distribution of two variables: (1) color condition, and (2) /p, b m/ viseme-cluster response. The color conditions were

represented at five levels while the visemes cluster response was represented at nine levels.

Experiments I and II Contingency Table

A contingency table analysis was performed to examine the data for /p, b m/ viseme-cluster responses by color condition (see Table 21). The test statistic indicated a Chi-square likelihood ratio of 188.860, a value that corresponded to statistical significance (p = < 0 .0001), as shown in Table 22. The Pearson Chi-square test (hereafter referred to as the "Pearson") capitalizes on the property that frequency counts tend to follow a normal distribution with large samples. In this analysis, the Pearson was 179.923, a value that was statistically significant (p = < 0 .0001). Therefore, the null hypothesis was rejected, indicating that the /p, b m/ viseme-cluster response rates were significantly different from those for color.

Experiments I and II Mosaic Plot by Color

A mosaic plot was generated for the /p, b m/ viseme-cluster response by color contingency table described above (see Table 21 and Figure 21). The plot indicated the relative sizes of the proportionate values (shown in rectangular tiles) for the nine levels of the /p, b m/ viseme-cluster response (right y-axis) and depicted the relative independence. The width of each column is proportional

to frequency for each variable (color). The height is proportional to the conditional frequency of each row (phoneme for the /p, b, m/ viseme-cluster). Complete independence corresponds to the same height for variable tiles representing the same level of response.

The mosaic plot indicated an over-representation of /p/ in the G-R, NAT, and Y-B conditions along the right y-axis for row proportion. Both over-and under-representation was seen for /m/ across the color conditions. However, independence was seen for /m/ in the NAT and G-R color conditions. A similar pattern of independence was depicted for /m / across the five color conditions. In comparison, the response rates were similar for p/m, /b/, and b/m across the color conditions.

Experiments I and II Correspondence Analysis by Color

A two-dimensional correspondence analysis was performed to indicate similarity among different response patterns. To restate, a correspondence analysis generates a plot derived from a matrix of counts divided by the total frequency. A Euclidean algorithm converts a matrix of nonnegative data into points that represent rows and columns.

For the correspondence analysis of the /p, b m/ viseme-cluster response by color, the /m/ and m/p were the highest values and in the same direction as NAT. This finding agreed with the depiction shown in Figure 22. This analysis indicated that the lowest representation was for p/b, at the extreme. The NAT

condition was indicated as a moderate score with m/b. The neutral values were located along the vertical zero coordinate.

Experiment I Contingency Table

Contingency table analysis for viseme-cluster response by color was performed for the vowel /i/ context (see Table 23). The test statistics indicated a Chi-square likelihood ratio of 164.597, a value that corresponded to statistical significance (p = < 0 .0001) as shown in Table 24. The Pearson value was 156.679 and met the statistical criterion for significance of 0.05 (p = < 0 .0001). The null hypothesis was rejected for the same response rates between the /p, b, m/ viseme-cluster response and color.

Experiment I Mosaic Plot

A mosaic plot was generated for a phoneme by color contingency table data (see Figure 23). The plot indicated similar response rates for /m/ in the NAT and R-G conditions. The response rates were essentially the same for /b/ across every color condition except NAT, where an over-representation is shown.

Experiment I Correspondence Analysis

The correspondence plot coincided with the contingency table data (see Figure 24). The moderate and neutral values were attributable to G-R and p/b

while the pattern of the lowest scores was seen for m/p, b/p, B-Y and R-G. The p/b and p/m responses were prominent with a correspondence to G-R in Experiment I.

Experiment II Contingency Table

A contingency table analysis for phoneme responses by color was conducted for the vowel /ɑ/ context (see Table 25). The test statistics indicated a Chi-square likelihood ratio of 72.075, a value that corresponded to statistical significance (p = < 0 .0001), as shown in Table 26. The Pearson value of 71.066 was statistically significant at the 0.05 level (p = < 0 .0001). The null hypothesis was rejected for the same response rates between the phoneme and color. This finding suggested that there were differentiated responses for the phonemes by color condition.

Experiment II Mosaic Plot

A mosaic plot for phoneme by color condition for the vowel /ɑ/ indicated the proportionate response rates across color conditions (see Figure 25). The response rates for /b/ in the R-G condition were over-represented relative to overall proportionate response for /b/. For /b/, the response rates were very similar. The response rates for the NAT condition indicated prominent representation for /p/ and /m/. This observation might be related to the visual

recognition seen in Figure 24. The mosaic plot was consistent with rejection of the null hypothesis for independence of response rate.

Experiment II Correspondence Analysis

The two-dimensional correspondence plot coincided with the contingency table data (see Table 25). The highest scores were assigned to m/b with a directional correspondence to NAT, as shown in Figure 26. The points along the vertical neutral line indicated a clustering that corresponded to under-representation among the several phoneme substitutions. In comparison, /b/ and /p/ emerged, but in an opposite direction from the emergence of /m/ .

Taken together, the cross-classification analysis appeared to strengthen the supposition that, in certain definite circumstances, there is more than a chance likelihood that color influences the perception of the /p, b, m/ viseme-cluster. It would have been unwieldy and inappropriate to perform contingency tables on all the variables modeled in the logistic design of experiment. However, the cross-classification analysis provided an analysis of the counts and frequency distribution of the three main variables stated in the research questions (vowel, phoneme, and color).

Other Predictor Variables

Angle

The angle for viewing the monitor was irrelevant to the whole model fit based on the Chi-square goodness of fit test statistic. The viewing angle could have been a relevant variable had the subject been seated at a distance of 0.5 meters (2.5 ft) or less. Palmer (1999) noted that beyond a distance of one meter, the effect of angle on visual convergence becomes negligible. For this reason, it was not too surprising that angle was not relevant to the statistical model. It is not likely that the effect of this predictor would be different had the analysis been performed with continuous data.

Order

The were five distinct trials for the test presentations in Experiments I and II. The subjects viewed trial 1 twice for a total of six trials. Trials 1 and 6 were identical in presentation. The variable order corresponded to the six trials viewed by the subjects. There is no direct test for reliability for nonparametric data, such as Cronbach α. This test has been used in the analysis of continuous data in parametric studies. However, to test the model fit for the effect of order, logistic regression was used to test whether or not "order 1" was closer to "order 6." The odds ratio point estimate for Trial 1 versus Trial 6 was 1.013. The null hypothesis ($H_0 = 0$) "no difference" between the trials was not rejected. The presentation of

Trail 1 and Trial 6 was considered very close or "reliable." Therefore, the effect of all orders (trials) was insignificant (p => 0.05) and. was not a relevant predictor for visual recognition.

Time

The parameter of "time" had the potential to show effects related to the occasion when the subjects were tested. Each subject was tested in two sessions (two viewings of the Experiments per session) covering a span of two weeks. In other words, Time 1 and Time 2 were in the first week, while Time 3 and Time 4 were in the second week.

The data for this study were collected in the late afternoon/early evening between 17:00-18:15 central daylight savings time (CDT). The data analysis indicated that Time 1 and Time 2 (both in week one) were significant influences, as shown in Table 14 and Table 15. Time 3 and Time 4, (both in week two) were not significant. The Time 4 term was irrelevant and was not factored in the Chi-square whole model test. Collectively, these results suggested that the subjects might have performed with possible differences of mood.

The individual subject data pointed to S2 and S4, who were influenced significantly by the Time 1 and Time 2 terms (see Table 17). It turned out that only S2 and S4 were also influenced significantly by the color and phoneme terms of the logistic calculation. The data do not appear to show any association for fatigue or learning effects. However, less motivation to complete the second

session might be inferred. It is logical to suggest that the novelty of the task might have been less for Time 3 and Time 4 and therefore, the impetus for the task would have been less.

Evaluation of the Conceptual Model

The investigator of this study probed the interaction of color and vowel context on the visual perception of /p, b, m/. The three research questions codified three independent (predictor) variables (phoneme, color, and vowel context) that could change the dependent (outcome) variable of visual recognition. The outcome was dichotomous and generated binary data (correct and incorrect).

As such, the distribution was not normal. Rather, the recorded observations produced a binomial or bimodal distribution of 1 and 0. The data were categorical, a restriction that violated the assumptions for parametric analyses—for example, analysis of variance (ANOVA), linear regression, or a t-test.

The accuracy of recognition data was collected. However, these data were not analyzed beyond descriptive terms because of the following: (1) The data were not continuous. (2) DOE could not infer accuracy from the research questions. (3) Specific accuracy required a different DOE for continuous, normally distributed data.

Because of this, logistic regression analysis was selected based on the type of data that were generated and certain assumptions about the population distribution including the following: (1) The dependent variable (outcome) did not follow a normal distribution. (2) A normal distribution of errors was not assumed. (3) A linear relationship between the predictors and outcome (visual recognition) was not essential. (4) The dependent and independent variables can be nominal.

To account for the probability or likelihood that other variables could change the outcome, a statistical model was formulated to include other design variables apart from phoneme, color, and vowel context. Those other variables were viewing angle, time, order of trials, and subject.

Certain assumptions were made about the outcome: (1) Color and vowel context could influence the perception of /p, b, m/. (2) Opponent colors could be a significant influence on the visual perception of /p, b, m/. However, one could not predict, with certainty, whether the main design variables (phoneme, color, and vowel context) would prevail as relevant or become irrelevant. It is important to note that the initial Chi-square test generates the likelihood ratio based on $[-2\log(\ell_0/\ell_1)]$ to determine the relevancy of variables. Further, inclusion of irrelevant variables leads to increases of the standard error and an attribution of the variance to the irrelevant or extraneous variables (Wright, 1995).

Within the framework of Experiment I, color and phoneme were relevant to the whole model fit, but the vowel /i/ was not. In contrast for Experiment II, subject, time, and phoneme were relevant predictors for visual recognition, yet

color, phoneme, and vowel /ɑ/ were not relevant to the whole model fit test (see Table 12 and Table 14). Consequently, the null hypothesis ($H_0 : b = 0$) was not rejected for the interaction of color and vowel on the perception of /p, b, m/ in the context of the vowel /i/. Similarly, the null hypothesis was not rejected for the interaction of color and vowel on the perception of /p, b, m/, in the context of the vowel /ɑ/.

These data analyses suggest that an interaction might have been realized for different vowel contexts for both experimental conditions— however, not in the manner assumed by this investigator. Additionally, the phoneme perceived by the subjects was a significant influence on visual recognition in certain color conditions, a finding that was a basic underpinning of this study.

Chapter 5

DISCUSSION

This chapter was written in six parts: summary of research, discussion of major findings, conclusions of the study, implications for audiology, implications for future research, and limitations of the study.

Summary of the Research

The specific aim of this study was to test the visual perception by color of the sounds/p, b, m/ in the context of /i/ and /ɑ/. Prior experiments have probed the influence of visemes and consonant recognition among hearing-impaired speechreaders (Rönnberg et al., 1999) and between normal-hearing hearing-impaired populations (Owens & Blazek, 1985).

The study done by Binnie et al. (1974) noted that, individually, /p/, /b/, /m/ were not formed with the same place and manner. Jeffers and Barley noted that /p/, /b/, /m/ are sometimes difficult to recognize because of "transitional" or co-articulation effects, the results of which create the perception of similarity for /m/ and /p/ and /p/ and /b/. The VCV contexts of /i/ and /ɑ/ were selected based on empirical findings (Jeffers & Barley, 1971; Owens & Blazek, 1985; Preminger et al., 1998). These investigations suggested distractibility and ambiguity for accurate speech perception. However, they did not report why these assumed distractions occurred.

One aspect of the conceptual model for this study was based on two information processing theories: (1) "all-or-none" (Broadbent, 1958) and (2) "attenuation" (Treisman, 1969). Both theoretical constructs deduce that environmental factors might hinder perception or decisions (see Figure 1). O'Neill and Oyer (1981) noted that environmental factors (attention to peripheral vision, poor room lighting, distractions, situational cues) could either degrade the message or distract the speechreader from an accurate perception of speech.

In an alternate view, specific distractibility has been shown in speechreading (Marassa & Lansing, 1995; Preminger et al., 1998; Vatikiotis-Bateson et al., 1994). These studies explored speechreader judgments that might not have been made at the conscious level. Rather, the cases might have been influenced by physiologic or neurologic variables. To the extent that foveal vision was implicated, higher order neural visual centers—for example, the superior colliculus, pulvinar, and the middle temporal (MT) cortical area—may have also been involved in the decisions. These decisions could also be influenced by a bias toward the familiarity of the /p, b, m/ viseme-cluster. Thus, one could speculate that aspects of pattern recognition could occur during speechreading (Massaro, 1987)

With controls for environmental factors such as room lighting, distance, and viewing angle, this investigation probed phonologic (vowel context) influences on the visual perception of /p, b, m/ viseme-cluster, described in previous research (Fisher, 1968; Jeffers & Barley, 1971; Owens & Blazek, 1985). In addition, the introduction of color vision was assumed to exert an influence on

speech perception, based on opponent color theory (Helmholtz, 1867; Hering, 1878; Hurvich & Jameson, 1957) and "all-or-none" information processing theory (Broadbent, 1958).

Discussion of Major Findings

The Influence of Vowel Context

The data analysis indicated that the vowel /i/ was not an influence for the visual perception of /p, b, m/, according to the Chi-square model fit analysis at the < 0.05 level. Further, an interaction of vowel and color was not indicated. In the context of overall results, these findings compare well with the previous findings of Owens and Blazek (1985). Owens and Blazek did not indicate statistically significant group differences between normal-hearing and hearing-impaired subjects for the /i/ and /a/ VCV context. For the normal-hearing subjects the /p, b, m/viseme-cluster indicated maximum accuracy (76%) for the /i/ VCV context for /m/and minimum accuracy (24%) for /p/ in the /a/ VCV context (Owens & Blazek, 1985). The findings from this investigation indicated a maximum accuracy of 52 percent for the /i/ vowel context and the /m/ phoneme. Based on the vowel influence alone, the overall percentage accuracy for this investigation were less than chance. The maximum percentage for /a/ was 49 percent for /p/.

It is important to note that the comparison of this investigation with that of Owens and Blazek (1985) is a partial descriptive statistical account of the visual recognition of consonant phonemes. They acknowledged study limitations including an (1) analysis for the effect or influence of lighting conditions, (2) viewing angle, and the (3) statistical criterion to assess them. In comparison, this investigation sought to determine influences on phonemes with logistic regression, an analysis of that probes the likelihood that estimates of influence are better than chance. Further, since a model represents a partial representation of reality, the logistic model analysis incorporated relevant predictors for visual recognition, to the exclusion of others.

The findings of this study indicated that /i/ was not a significant influence in the context of the other whole model design variables (color, phoneme, and time). This finding may be similar to Binnie et al. (1974). They found that the overall correct intelligibility scores for visual recognition alone were less than chance when coupled with vowel variation. Vatikiotis-Bateson et al. (1994) noted, in descriptive terms, that /i/ was a probable influence corresponding to easier speechreading for some observers. They based their assumption on the inherent "lip spreading" visual effect during the production of /i/. However, they could not account for the allocation of both attention and targeting of foveal vision to the perioral and eye region in the vowel context of /i/.

The data analysis from this study also did not indicate a statistically significant influence for vowel context of /ɑ/. The Chi-square whole model fit

test at the criterion level α, > 0.05 indicated a low likelihood that /ɑ/ was a real influence in this study.

The data analysis of S2 and S6 indicated a significant influence for the vowel /ɑ/ from the Chi-square model fit, as shown in Table 18. In particular, there was an effect for an interaction of color and vowel for S2, a finding that was assumed, overall, for DOE. Logistic regression analysis, for S2 and S6 indicated that the null hypothesis was rejected for the vowel /ɑ/. That is, the population coefficient differed significantly from zero.

For some subjects (S1, S3, S4, and S5), vowel context was not an influence on visual recognition: (1) S1 was influenced by the phonemes /p/ and /b/. (2) S3 was influenced by color (NAT, Y-B, G-R). (3) S4 was influenced by (Time 2), the NAT color condition, an interaction of NAT*/m/, and the phoneme /b/. (4) S5 was not influenced by any of the relevant parameters (vowel, phoneme, color, time, and order)

When /i/ and /ɑ/ were compared together in one analysis, /ɑ/ emerged as a statistically significant estimates and influence for effects (see Table 16 and Table 17. The individual data analysis also indicated that /i/ was not a significant influence on the perception of /p/, /b/, or /m/. The parting of the lips, as in the production of /ɑ/, may have been influential for the /p, b, m/.

The "spreading of the lips," noted by Vatikiotis-Bateson et al. (1994), might have been a distraction. In the /i/ lip posture, the opponent color pairs were more approximated than in the /ɑ/ lip posture. If the subjects shifted their eyes toward the talker's eyes, then color would have been "more likely" a positive

influence on the perception of the targeted phonemes (p = 0.1390), as shown in the positive likelihood estimates in Table 12 and the analysis of effects in Table 13. Therefore, a distraction might have occurred independent of the vowel /i/.

The Influence of Color

The analysis of data from Experiment I seemed to indicate influences on the visual perception of /p/, /b/, /m/ from the presentation of opponent colors as shown in Table 12. The most likely candidates for color influence for the phoneme /m/ were the B-Y condition and the NAT condition. The analysis of maximum likelihood estimates for /i/ indicated a significant probability that the B-Y color condition was a major influence. There was also a significant interaction indicated for the perception of phoneme (/p/, /b/) in the B-Y condition as shown in Table 12 and Table 13. However, this investigator noted that the individual subject analysis indicated that subjects S1, S5, and S6 were not influenced significantly by color (see Table 18).

When the vowel contexts of /i/ and /a/ were analyzed together for all subjects, color emerged as a statistically significant effect, with an interaction as shown in Table 16. There was a high likelihood that the vowel was /a/ for the NAT and R-G conditions (see Table 17). This finding was interesting because (1) the mouth opening is larger for /a/ than for /i/ and (2) the opponent colors were not as approximated. It is logical to consider that the NAT and R-G

conditions would be incongruous with the assumption that color might be an influence on the NAT condition.

The interaction of /ɑ/ with the two color conditions might have related to the findings of Binnie et al. (1974) and Jeffers and Barley (1971), who noted that the production of /ɑ/ creates transitional effects. However, these transitional effects were also noted with the vowel /i/.

One cannot ignore the influence of the vowel in interaction with color. It turns out that the color could have been influential, but not necessarily for accuracy of the perception as shown in Figures 11 through 20. There, the descriptive statistics show that there are numerically more errors for visual recognition in nearly every color condition except one; NAT as shown in Figure 15

The Influence of Phoneme

In Experiment I (/i/ vowel context), the /p, b, m/viseme cluster was not an influence for visual recognition under the five color conditions. Owens & Blazek (1985) reported that /m/ was robust for visual perception in the /i/ vowel context. The current study was not a direct comparison to the results of Owens & Blazek The findings from this investigation suggest that the interaction of color and the vowel /i/ may have not been an influence on the visual perception of /p, b, m/. In Experiment II (/ɑ/ vowel context), the interaction of color and the vowel /ɑ/ was a significant influence only for /p/ and /b/.

Conclusions of the Study

In Experiment I, the logistic regression model fit for six cases indicated significant effects of phoneme, time, and subject. There also was an interaction for color and the perceived phoneme from the /p, b, m/ viseme-cluster. For both Experiment I and Experiment II, "subject" was an exceedingly statistically significant term (p = < .0001). This suggested that the subjects were influenced by the vowel /ɑ/ and interacted with the /p, b, m/ cluster. The analysis suggested that (1) S1, S5, and S6 experienced difficulty speechreading the talker, (2) the /i/ vowel context was difficult, or (3) there was a combination of the two.

The first possibility was consistent with Kricos and Lesner (1982) and Lesner (1988), who indicated that the talker could make a significant difference on speechreading performance. This observation about the talker was also noted by Demorest and Bernstein (1992). They noted further that other variables might influence variability for speechreading—for example, day-to-day variations and the mood of the subject at the time of testing.

One subjective observation was made for the talker for this study. The bottom lip may have been more salient than the top lip based on the asymmetry between the two. The anatomic shape of a talker's lips may influence the overall visual perception for speechreading. There may be a subtle but essential common component that links the visual and auditory neural centers to fill in the gaps or to lay a template for pattern recognition.

A second possibility might have been related to the approximation of the lips bringing opponent colors together. Further, the transitional effects for /i/ are inherently not as dynamic as /ɑ/ for the /p, b, m/ viseme-cluster. This observation has been corroborated by investigators who noted mouth opening for /i/ (Jeffers & Barley, 1971; Jorgensen, 1962; Vatikiotis-Bateson et al., 1994, 1998).

Implications for Audiology

- Over the past 70 years, there have been few innovative solutions for the enhancement of speechreading accuracy and efficiency. Empirical research has produced strategies using photo flashcards, repetition and drill, manipulation of signal-to-noise ratio, and slow-motion films and video.

- The contribution of color vision to speechreading is not clear, but this research has suggested that it is an influence.

- Generally, for individuals with a significant hearing loss, visual attention is heightened. The findings from this investigation represent a first approximation toward more research in audiology and other disciplines to determine whether color vision is related to speechreading with those with a hearing-impairment.

- For some individuals in this study, color was an important variable for visual recognition. For S2, color was influential with corresponding accuracy. This subject was also influenced by the presentation within the first two trials. However, color might not influence all subjects in the same way, as shown by their differentiated responses. In particular, S5 was not influenced significantly by color, phoneme, or the vowel.

- In the future, as audiologists assess patients for rehabilitative options, the application of color in speechreading training videos might be coupled, in some way, with filtered lenses (eye glasses) to enhance accurate speechreading.

- Clinical trials could be conducted to determine the variables that are practical and meaningful for the patient. As a component of an assessment for aural rehabilitation, a clinician might have the potential to probe the patient's response to selected variables—for example, color, phoneme, and vowel—to determine the feasibility and efficacy of speechreading training.

- For aural rehabilitation, an established cross-modal perception between color vision and hearing could be key for selected individuals to maximize the most from visual speech perception.

Implications for Future Research

Hering (1878) observed and noted that there was no series of colors (chromatic transitions) that appeared both reddish and greenish or both yellow and bluish. To restate, this investigator assumed that the subject observations would be influenced by opponent color pairs that attract or detract from attention to the lips.

Hubel and Livingstone (1987, 1988) reported that retinal coding was just part of the picture for understanding how the human visual system processes color. There are both retinal and cortical influences on the perception of color (Croner & Albright, 1999). In particular, we are just beginning to understand how color vision might be related to the perception of form (Mishkin, et al., 1983) and motion (Ramachandran & Gregory, 1978). Empirical findings have even implicated the contributions of color vision to speechreading (Vatikiotis-Bateson et al., 1994) and the reading of text (Fitch et al., 1997; Livingstone et al., 1991).

The perception of opponent colors was assumed to be a distraction under neural control rather than that of conscious or environmental influences. In this context, this investigator concluded that the selector or model referred to as the "all-or-none" model for information processing might be a closer fit to the outcome of Experiments I and II. He based this on the theoretical underpinning of the opponent color theory, as described by Hering (1878). Hering noted that it was not possible to maintain coterminous attention to opponent color pairs.

Therefore, foveal vision would be directed to one color or the other, or even elsewhere.

Some possible future research might include the following: (1) an investigation using the experimental design of this investigation with subjects who are hearing-impaired and who are also experienced speechreaders for the visual perception of /p, b, m/; (2) an investigation to explore the influence of opponent colors with different talkers; (3) an eye-gaze experiment to probe the influence of opponent colors on the perioral regions (eyes, cheeks); (4) a study to determine whether the placement of the "green-yellow" color (approximating 555 nm) influences the perception of the /p, b, m/ viseme-cluster; (5) an investigation to quantify changes in eye-gaze using other identifiable viseme-clusters (Fisher, 1968; Jeffers & Barley, 1971; Owens & Blazek, 1985); (6) an experiment to determine whether a 45° angle of the talker's lips would interact with speechreading performance; and (7) a study involving subjects with a diagnosis of protanopia or deuteranopia (red-green color deficiency). Further, it might be noteworthy to determine whether their speechreading performance differs from those with normal color vision. If lip-color is to be studied again, a talker with symmetrical lip structure might yield different results.

Empirical research led to speculation (Helmholtz, 1867; Hering, 1878) and subsequent verification (De Valois & Jacobs, 1968) that there are collections of visual sensory receptors (retinal cones) that appear to convey color-coding for red, blue, and yellow. More recent research findings support these assumptions (Livingstone & Hubel, 1987, 1988). These studies have explored the possibility

for the contribution of color to motion sensitivity (Croner & Albright 1999; Thiele et al., 1999). The contribution of color to motion sensitivity for speechreading is yet another area that may shed some light on "where" we look for information and "when" to look.

Limitations of the Study

- The investigator recognized that the research questions could not answer questions of statistical differences between color and vowel context conditions.
- The research design was modeled to determine influences and variation that may arise from one talker.
- The conclusions can only be applied to a normal-hearing population of younger women.
- The inclusion of irrelevant variables would be a violation of logistic regression analysis, leading to incorrect attribution of variance. As such, certain variables were factored out of the model.
- One subject (S5) might have been different enough to influence the model fit for vowel and color for Experiment II.
- A failure to meet the model fit for vowel and color introduced restrictions on overall interpretation and conclusion for Experiment I and Experiment II.
- To make certain global statements regarding statistically significant logistic coefficients, corresponding correlations must also be statistically

significant. The data from this study were coded categorical or nominal and were not appropriate for parametric correlation data analysis.

- The anatomical features of the talker's lips were more prominent for the bottom lip. This may have unknowingly created a skewed subject observation.

THE INFLUENCE OF COLOR AND VOWEL CONTEXT ON THE VISUAL PERCEPTION OF /p, b, m/

VOLUME II

By

Willard Charles Hooks, Jr.

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Audiology and Speech Sciences

2003

APPENDICES

APPENDIX A

TABLES

Table 1

Subject Audiometric Thresholds

Note. The subjects' auditory thresholds were obtained within an ambient acoustic environment of approximately 57 dBA SPL. A correction factor of –5 dB HL subtracted from 500 Hz, at 1000 Hz, and at 2000 Hz, based on a comparison of archetypical thresholds, within an acoustically treated environment. The pure tone average (PTA) corresponds to the threshold values at 0.5 kHz, 1.0 kHz, and 2.0 kHz.

Table 1

Subject Audiometric Thresholds

RIGHT EAR

	0.5 kHz	1.0 kHz	2.0 kHz	4.0 kHz	PTA
S1	20	10	0	0	15
S2	5	5	5	5	5
S3	15	0	5	10	6.6
S4	10	5	0	-10	5
S5	15	10	5	-5	10
S6	15	10	5	0	10

A correction factor of –5 subtracted from 0.5 kHz, -5 at 1.0 kHz, and -5 at 2.0 kHz

LEFT EAR

	0.5 kHz	1.0 kHz	2.0 kHz	4.0 kHz	PTA
S1	10	10	-10	0	6.6
S2	5	5	5	5	10
S3	15	10	0	5	13.3
S4	10	5	0	-10	10
S5	10	5	0	5	10
S6	20	10	5	-5	16.7

A correction factor of –5 subtracted from 0.5 kHz, -5 at 1.0 kHz, and -5 at 2.0 kHz

Table 2

Experimental Condition I Stimuli

Note. Table of six trials for color conditions in the /i/ VCV context for the phonemes /p, b, m/, for five lip-color conditions, where natural =1, red-green = 2, green-red = 3, blue-yellow = 4, and yellow-blue = 5. Subject positions (1, 2, 3) are separated by 15° of arc. The asterisk marks a repeated trial number 1 for reliability.

Table 2

Experimental Condition I Stimuli

Trial 1	Natural (1)	Red (2) Green	Green (3) Red	Blue (4) Yellow	Yellow (5) Blue
	/ipi/	/imi/	/ibi/	/ibi/	/imi/
	/ibi/	/ibi/	/imi/	/ipi/	/ipi/
	/imi/	/ipi/	/ipi/	/imi/	/ibi/

Trial 2	Red (2) Green	Green (3) Red	Blue (4) Yellow	Yellow (5) Blue	Natural (1)
	/imi/	/ibi/	/ibi/	/imi/	/ipi/
	/ibi/	/imi/	/ipi/	/ipi/	/ibi/
	/ipi/	/ipi/	/imi/	/ibi/	/imi/

Trial 3	Green (3) Red	Blue (4) Yellow	Yellow (5) Blue	Natural (1)	Red (2) Green
	/ibi/	/ibi/	/imi/	/ipi/	/imi/
	/imi/	/ipi/	/ipi/	/ibi/	/ibi/
	/ipi/	/imi/	/ibi/	/imi/	/ipi/

Trial 4	Blue (4) Yellow	Yellow (5) Blue	Natural (1)	Red (2) Green	Green (3) Red
	/ibi/	/imi/	/ipi/	/imi/	/ibi/
	/ipi/	/ipi/	/ibi/	/ibi/	/imi/
	/imi/	/ibi/	/imi/	/ipi/	/ipi/

Trial 5	Yellow (5) Blue	Natural (1)	Red (2) Green	Green (3) Red	Blue (4) Yellow
	/imi/	/ipi/	/imi/	/ibi/	/ibi/
	/ipi/	/ibi/	/ibi/	/imi/	/ipi/
	/ibi/	/imi/	/ipi/	/ipi/	/imi/

*Trial 6	Natural (1)	Red (2) Green	Green (3) Red	Blue (4) Yellow	Yellow (5) Blue
	/ipi/	/imi/	/ibi/	/ibi/	/imi/
	/ibi/	/ibi/	/imi/	/ipi/	/ipi/
	/imi/	/ipi/	/ipi/	/imi/	/ibi/

Table 3

Experimental Condition II Stimuli

Note. Table of six trials for color conditions in the /ɑ/ VCV context for the phonemes /p, b, m/, for five lip-color conditions, where natural =1, red-green = 2, green-red = 3, blue-yellow = 4, and yellow-blue = 5. Subject positions (1, 2, 3) are separated by 15° of arc. The asterisk marks a repeated trial number 1 for reliability.

Table 3

Experimental Condition II Stimuli

*Trial 1	Natural (1)	Red (2) Green	Green (3) Red	Blue (4) Yellow	Yellow (5) Blue
	/apa/	/ama/	/aba/	/aba/	/ama/
	/aba/	/aba/	/ama/	/apa/	/apa/
	/ama/	/apa/	/apa/	/ama/	/aba/

Trial 2	Red (2) Green	Green (3) Red	Blue (4) Yellow	Yellow (5) Blue	Natural (1)
	/ama/	/aba/	/aba/	/ama/	/apa/
	/aba/	/ama/	/apa/	/apa/	/aba/
	/apa/	/apa/	/ama/	/aba/	/ama/

Trial 3	Green (3) Red	Blue (4) Yellow	Yellow (5) Blue	Natural (1)	Red (2) Green
	/aba/	/aba/	/ama/	/apa/	/ama/
	/ama/	/apa/	/apa/	/aba/	/aba/
	/apa/	/ama/	/aba/	/ama/	/apa/

Trial 4	Blue (4) Yellow	Yellow (5) Blue	Natural (1)	Red (2) Green	Green (3) Red
	/aba/	/ama/	/apa/	/ama/	/aba/
	/apa/	/apa/	/aba/	/aba/	/ama/
	/ama/	/aba/	/ama/	/apa/	/apa/

Trial 5	Yellow (5) Blue	Natural (1)	Red (2) Green	Green (3) Red	Blue (4) Yellow
	/ama/	/apa/	/ama/	/aba/	/aba/
	/apa/	/aba/	/aba/	/ama/	/apa/
	/aba/	/ama/	/apa/	/apa/	/ama/

Trial 6	Natural (1)	Red (2) Green	Green (3) Red	Blue (4) Yellow	Yellow (5) Blue
	/apa/	/ama/	/aba/	/aba/	/ama/
	/aba/	/aba/	/ama/	/apa/	/apa/
	/ama/	/apa/	/apa/	/ama/	/aba/

Table 4

Munsell Notation -to-CIE 1931 Conversion

Color sample	Hue	Value	Chroma	Wavelength
Red	5R	5	14	616.7 +/- 0.3 nm
Yellow	5Y	8.5	12	576.0 +/- 0.3 nm
Green/Yellow	5GY	6	8	565.8 +/- 0.2 nm
Blue	5PB	4	8	475.6 +/- 0.2 nm

Note. The table includes the Munsell notation for red, yellow, green-yellow, and blue with the corresponding calculated wavelength approximation.

Table 5

Video Production Specifications

Note. The table corresponds to video equipment used in during video productions and related measurements and conditions for recording editing, and playback. The video production and recording was performed in MSU Telecommunication Department Studio D.

Table 5

Video Production Specifications

Camera Sony® HFD	Medium Close Up Head On Angle = 0° Distance = 1.3 meter
Ambient Illumination	2.37 x 10^3 lx (220 ft candles)
Studio Lighting	9 standard ceiling-mounted studio lights (3200° Kelvin)
Studio Illumination (on actor's face)	2.37 x 10^3 lx (220 ft candles)
Video Recording Equipment	Sony® DSR 1,800
Makeup Color Matching	Munsell color chip samples and Ben Nye creme color liner (CL-13 Fire Red, CL-32 Lime Green CL-19 Blue, and CL-Yellow, and CL-29 Black)
Recording Media	Sony® DVCAM® digital medium (PDV-34ME Sony® BETACAM SP® metal tape (BCT-30Ma)
Color Camera Calibration	Magni® WVM-718 vectorscope
Raw video footage duration	12:27:51 (minutes, seconds, tenths of sec)
Final edited video duration	33:27:42 (minutes, seconds, tenths of sec)
Duration per trial	02:11:75 (minutes, seconds, tenths of sec)

Table 6

Distribution of /p, b, m/ Viseme-cluster Responses

Phonemes	Combined Count	Exp. I /i/	Exp. II /ɑ/
b	445	235	210
b/m	387	175	212
b/p	386	188	198
m	541	284	257
m/b	412	174	238
m/p	214	91	123
p	605	329	276
p/b	338	188	150
p/m	272	136	136
Total	*3,600	1,800	1,800

Note. The table corresponds to the frequency distribution of all 9 possible phoneme responses for the /p, b, m/ viseme-cluster for all six subjects in Experiment I (/i/ vowel context) and in Experiment II (/ɑ/ vowel context). The six possible phoneme substitutions ("incorrect" responses) included the following: (1) p/b, (2) b/m, (3) m/p, (4) p/m, (5) b/p, and (6) b/m.

Table 7

Distribution of Visual Recognition Split by Experiment

Visual Recognition	Experiment I	Experiment II
Correct	849	739
Incorrect	951	1,061
Total	1,800	1,800

Note. The table corresponds to the descriptive statistics for all six cases for Experiment I (/i/ vowel context) and Experiment II (/a/ vowel context). The dependent variable was visual recognition within the /p, b, m/ phoneme cluster.

Table 8

Distribution of Color Condition by Phoneme for Exp. I

Correct	B-Y	G-R	NAT	R-G	Y-B	Total
/p/	48	76	76	45	84	329
/b/	41	41	67	43	44	236
/m/	72	44	73	58	37	284
Total	161	161	216	146	165	
Incorrect						
/p/	68	87	25	61	82	323
/b/	72	74	56	88	73	363
/m/	59	38	63	65	40	265
Total	199	199	144	214	195	1,800

Note. The table corresponds to the descriptive statistics for all six cases for Experiment I (/i/ vowel context). The dependent variable was phoneme recognition.

Table 9

Distribution of Color Condition by Phoneme for Exp. II

Correct	B-Y	G-R	NAT	R-G	Y-B	Total
/p/	45	56	56	52	64	273
/b/	32	47	38	53	38	208
/m/	58	39	65	56	40	258
Total	135	142	159	161	142	
Incorrect						
/p/	63	69	38	50	69	289
/b/	84	87	67	81	92	411
/m/	78	62	96	68	57	361
Total	225	218	201	199	218	1,800

Note. The table corresponds to the descriptive statistics for all six cases for Experiment II (/ɑ/ vowel context). The dependent variable was phoneme recognition.

Table 10

Visual Recognition Split by Color for Exp. I

COLOR	Correct	Incorrect	TOTAL
B-Y	161	199	360
G-R	161	199	360
NAT	216	144	360
R-G	146	214	360
Y-B	165	195	360

Note. The table corresponds to the frequency distribution of five color conditions in Experiment I (/i/ vowel context) and for all six cases.

Table 11

Visual Recognition Split by Color for Exp. II

COLOR	Correct	Incorrect	TOTAL
B-Y	135	225	360
G-R	142	218	360
NAT	159	201	360
R-G	161	199	360
Y-B	142	218	360

Note. The table corresponds to the frequency distribution for five color conditions in Experiment II (/ɑ/ vowel context).

Table 12

Experiment I Maximum Likelihood Estimates

Parameter	Designator	*df*	Pr > Chi-Sq
Intercept	Y-B	1	0.9681
Color	B-Y	1	0.2105
Color	G-R	1	0.7320
Color	NAT	1	0.183
Color	R-G	1	0.0379*
Phoneme	p	1	0.9739
Phoneme	m	1	0.9673
Color*Phoneme	B-Y	1	0.0149*

Note: An analysis of the maximum likelihood estimates indicated statistical significance for selected parameters relative to color and phoneme including (1) Y-B, (2) G-R, (3) /p/, (4) /b/, and (5)/m/. Interaction (*) was indicated for color and phoneme.

*$p < 0.05$

Table 13

Experiment I Analysis of Effects

Effect	*df*	Wald Chi-square	Pr > Chi-Sq
Color	4	42.1400	0.1390
Phoneme	2	32.6171	< 0 .0001 ****
Color*Phoneme	8	18.1197	0.0203 *
Time	3	10.6879	0.0135 *
Subject	5	48.2105	< 0 .0001 ****

Note. This table corresponds to the analysis of significant parameter effects for color, phoneme, time, and subject. These parameters indicate influences on the perception of phonemes. Interaction (*) was indicated for color and phoneme.

*$p < 0.05$ ***$p < 0.001$ **** $p < 0.0001$

Table 14

Experiment II Maximum Likelihood Estimates

Parameter	Designator	*df*	Pr > Chi-Sq
Intercept	p	1	< 0.0001 ****
Phoneme	b	1	< 0.0001 ****
Time	1	1	0.0065 **
Time	2	1	0.0257 *
Time	3	1	0.3763
Subject	1	1	0.0078 ***
Subject	2	1	0.0015 ***
Subject	4	1	0.0058 ***
Subject	5	1	0.0190 **

Note. This table indicates the parameter estimates that were selected from the whole model fit test for an analysis of effects.

*$p < 0.05$ ** $p < 0.01$ ***$p < 0.001$ **** $p < 0.0001$

Table 15

Experiment II Analysis of Effects

Effect	*df*	Wald Chi-square	Pr > Chi-Sq
Phoneme	2	32.2406	< 0.0001 ****
Time	3	10.0431	0.0182 *
Subject	5	25.7482	< 0.0001 ****

Note. This table corresponds to the analysis of significant parameter effects for phoneme, time, and subject. These parameters indicated influences on the perception of phonemes.

*$p < 0.05$ **** $p < 0.0001$

Table 16

Experiments I and II Maximum Likelihood Estimates

Note. This table indicates the parameters that selected from the model fit test for analysis of effects. Interaction (*) was indicated for color by vowel and color by phoneme.

Table 16

Experiments I and II Maximum Likelihood Estimates

Parameter	Variable	*df*	Wald Chi-square	Pr > Chi-Sq
Intercept	Y-B	1	44.84	< 0.0001 ****
Color	B-Y	1	05.1183	0.0237 *
Color	NAT	1	29.2821	< 0.0001 ****
Phoneme	/b/	1	40.1025	< 0.0001 ****
Color *Phoneme	B-Y /m/	1	05.3107	0.0212 *
Color *Phoneme	NAT /m/	1	14.1699	< 0.0001 ****
Subject	1	1	08.7068	0.0032 *
Subject	2	1	25.7033	< 0.0001 ****
Subject	3	1	03.6815	0.0550 *
Subject	4	1	08.5004	0.0036 **
Subject	5	1	14.2940	0.0002 ***
Vowel	/ɑ/	1	12.9171	0.0003 ***
Color*Vowel	NAT /ɑ/	1	08.8134	0.0030 ***
Color*Vowel	R-G /ɑ/	1	09.6735	0.0019 ***

*$p < 0.05$ ** $p < 0.01$ ***$p < 0.001$ **** $p < 0.0001$

Table 17

Experiments I and II Analysis of Effects

Effect	*df*	Wald Chi-square	Pr > Chi-Sq
Color	4	30.1836	< 0.0001 ****
Phoneme	2	42.6246	< 0.0001 ****
Color*Phoneme	8	21.3403	0.0063 ***
Time	3	18.4837	0.0003 ***
Subject	5	48.2105	< 0.0001 ****
Color*Vowel	4	15.0295	0.0046 ***

Note. This table corresponds to the analysis of significant parameter effects for color, phoneme, time, and subject. These parameters indicate influences on the perception of phonemes. Interaction (*) was indicated for color and phoneme and color and vowel.

***$p < 0.001$ **** $p < 0.0001$

Table 18

Individual Subject Maximum Likelihood Estimates

Note: This table indicates the influence of selected parameters for six subjects. These data show that the observations of S2 were influenced by vowel and vowel context. The model fit of S5 was questionable. Only S6 was influenced significantly by order.

Table 18

Individual Subject Maximum Likelihood Estimates

Subject	Parameter	*df*	Wald Chi-square	Pr > Chi-Sq	
S1	/p/	1	41.1213	< 0.0001	****
	/b/		4.4733	0.0344	*
	/m/		0.4818	0.4876	
S2	/b/	1	12.7751	0.0004	****
	/m/		4.2389	0.0395	*
	Y-B		5.779	0.0162	*
	B-Y		12.2665	0.0005	***
	G-R		5.1602	0.0231	*
	R-G		3.0521	0.0806	
	NAT		53.5083	< 0.0001	****
	Time 1		7.9403	0.0048	***
	Time 2		18.1390	< 0.0001	*****
	Vowel /ɑ/		4.6256	0.0315	*
S3	Y-B	1	32.7747	< 0.0001	***
	G-R		3.9774	0.0461	*
	NAT		6.5545	0.010	*
S4	/b/	1	18.3294	< 0.0001	****
	NAT		20.2202	< 0.0001	****
	NAT*/m/		11.1771	0.0008	***
	Time 2		8.0827	0.0045	**
S5	Phoneme /p/	1	0.0021	0.9639	
	Phoneme /b/		0.0011	0.9733	
	Phoneme /m/		0.0016	0.9677	
S6	Order 2	1	5.0472	0.0247	*
	Order 3		4.4863	0.0342	*
	/b/		21.6926	< 0.0001	****
	Vowel /ɑ/		8.4758	0.0036	***

*$p < 0.05$ ** $p < 0.01$ ***$p < 0.001$ **** $p < 0.0001$

Table 19

Individual Subjects Analysis of Effects

Subject	Effect	Wald Chi-square	*df*	Pr > Chi-Sq
S1	Phoneme	8.3940	2	0.0150 *
S2	Color	55.6915	4	< 0.0001 ****
	Time	24.6819	3	< 0.0001 ****
	Phoneme	12.7836	2	0.0017 ***
	Vowel	4.6256	1	0.0315 *
	Color*Vowel	10.2102	4	0.0370 *
S3	Color	9.8901	4	0.0423 *
S4	Color	21.0123	4	0.0003 ***
	Time	8.1093	3	< 0.0001 ****
	Phoneme	20.6697	2	< 0.0001 ****
	Color*Phoneme	15.5930	8	0.0486 *
S5	Phoneme	8.3212	2	0.0398 *
S6	Order	13.8300	5	0.0167 *
	Phoneme	25.6675	2	< 0.0001 ****
	Vowel	8.4756	1	0.0036 **

Note. The table provides an overview of the parameter effects that correspond to the most likely influences on the observations for each subject. Interaction (*) was indicated for color and phoneme and color and vowel.

*$p < 0.05$ ** $p < 0.01$ ***$p < 0.001$ **** $p < 0.0001$

Table 20

Association of Predicted Probabilities

	Experiment I	Experiment II
Percent Concordant	64.5	59.8
Percent Discordant	34.8	38.3
Ties	0.7	2.7
Concordant—Discordant		
Kendall'sTau-a	0.148*	0.104*

Note. This table summarizes how well the model predicts the with the following values: (1) The percent concordant and discordant are percentages of accuracy in classification (PAC). Kendall's Tau-a is a nonparametric equivalent to Spearman's rho.

* Significant Kendall Tau-a is based on a range from –1 through 0 to +1.

Table 21

Contingency Table: Viseme-cluster Response by Color

Note. This table indicates the count, total percentage, column percentage, and row percentage. The right column shows the total count and total percentage per color condition. The bottom right indicates the total for the sample.

Table 21

Contingency: Viseme Cluster Response by Color

	Color	b	b/m	b/p	m	m/b	m/p	p	p/b	p/m	
Count	B-Y	73	60	96	130	84	53	93	87	44	720
Total %		2.03	1.67	2.67	3.61	2.33	1.47	2.58	2.42	1.22	20.00
Col %		16.40	15.50	24.87	24.03	20.39	24.77	15.37	25.74	16.18	
Row %		10.14	8.33	13.33	18.06	11.67	7.36	12.92	12.08	6.11	
Count	G-R	88	93	67	83	62	38	132	95	62	720
Total %		2.44	2.58	1.86	2.31	1.72	1.06	3.67	2.64	1.72	20.00
Col %		19.78	24.03	17.36	15.34	15.05	17.76	21.82	28.11	22.79	
Row %		12.22	12.92	9.31	11.53	8.61	5.28	18.33	13.19	8.61	
Count	NAT	106	67	55	138	107	52	133	25	37	720
Total %		2.94	1.86	1.53	3.83	2.97	1.44	3.69	0.69	1.03	20.00
Col %		23.82	17.31	14.25	25.51	25.97	24.30	21.98	7.40	13.60	
Row %		14.72	9.31	7.64	19.17	14.86	7.22	18.47	3.47	5.14	
Count	R-G	95	72	99	114	85	48	97	59	51	720
Total %		2.64	2.00	2.75	3.17	2.36	1.33	2.69	1.64	1.42	20.00
Col %		21.35	18.60	25.65	21.07	20.63	22.43	16.03	17.46	18.75	
Row %		13.19	10.00	13.75	15.83	11.81	6.67	13.47	8.19	7.08	
Count	Y-B	83	95	69	76	74	23	150	72	78	720
Total %		2.31	2.64	1.92	2.11	2.06	0.64	4.17	2.00	2.17	20.00
Col %		18.65	24.55	17.88	14.05	17.96	10.75	24.79	21.30	28.68	
Row %		11.53	13.19	9.58	10.56	10.28	3.19	20.83	10.00	10.83	

Table 22

Tests for Viseme-cluster Response by Color

Source	*df*	-LogLike	R Square (U)
Model	32	94.4300	0.0122
Error	3560	7664.4930	
C Total	3592	7758.9230	
N	3600	1800	

Tests	Chi-Square	Prob >ChiSq
Likelihood Ratio	188.860	< 0.0001 ****
Pearson	179.923	< 0.0001 ****

Note. This table indicates the significance of the Chi-square test statistic. The table indicates that there is significant independence between color and phoneme responses for Experiments I and II.

**** $p < 0.0001$

Table 23

Viseme-cluster Response by Color for Experiment I

Note. This table indicates the count, total percentage, column percentage, and row percentage. The right column shows the total count and total percentage per color condition. The bottom right indicates the total for the sample. The table summarizes the proportion of the nine levels of phoneme responses for the /p, b, m/ viseme-cluster.

Table 23

Viseme Cluster Response by Color for Experiment I

	Color	b	b/m	b/p	m	m/b	m/p	p	p/b	p/m	
Count	B-Y	41	28	44	72	29	30	48	52	16	360
Total %		2.28	1.56	2.44	4.00	1.61	1.67	2.67	2.89	0.89	20.00
Col %		17.45	16.00	23.40	25.35	16.67	32.97	14.59	27.66	11.76	
Row %		11.39	7.78	12.22	20.00	8.06	8.33	13.33	14.44	4.44	
Count	G-R	41	43	30	44	25	13	76	54	34	360
Total %		2.28	2.39	1.67	2.44	1.39	0.72	4.22	3.00	1.89	20.00
Col %		17.45	24.57	15.96	15.49	14.37	14.29	23.10	28.72	25.00	
Row %		11.39	11.94	8.33	12.22	6.94	3.61	21.11	15.00	9.44	
Count	NAT	67	33	23	73	41	22	76	10	15	360
Total %		3.72	1.83	1.28	4.06	2.28	1.22	4.22	0.56	0.83	20.00
Col %		28.51	18.86	12.23	25.70	23.56	24.18	23.10	5.32	11.03	
Row %		18.61	9.17	6.39	20.28	11.39	6.11	21.11	2.78	4.17	
Count	R-G	42	29	60	58	46	19	45	32	29	360
Total %		2.33	1.61	3.33	3.22	2.56	1.06	2.50	1.78	1.61	20.00
Col %		17.87	16.57	31.91	20.42	26.44	20.88	13.68	17.02	21.32	
Row %		11.67	8.06	16.67	16.11	12.78	5.28	12.50	8.89	8.06	
Count	Y-B	44	42	31	37	33	7	84	40	42	360
Total %		2.44	2.33	1.72	2.06	1.83	0.39	4.67	2.22	2.33	20.00
Col %		18.72	24.00	16.49	13.03	18.97	7.69	25.53	21.28	30.88	
Row %		12.22	11.67	8.61	10.28	9.17	1.94	23.33	11.11	11.67	
Total x	5	235	175	188	284	174	91	329	188	136	1800
Color		13.06	9.72	10.44	15.78	9.67	5.06	18.28	10.44	7.56	

Table 24

Tests for Exp. I Viseme-cluster Response by Color

Source	*df*	-LogLike	R Square (U)
Model	32	82.2985	0.0214
Error	1760	3766.4381	
C Total	1792	3848.7366	
N	1800	1800	

Tests	Chi-Square	Prob >ChiSq
Likelihood Ratio	164.597	< 0.0001 ****
Pearson	156.679	< 0.0001 ****

Note. This table indicates the significance of the Chi-square test statistic. The table indicates that there is significant independence between color and phoneme responses in the context of /i/.

**** $p < 0.0001$

Table 25

Viseme-cluster Response by Color for Experiment II

Note. This table indicates the count, total percentage, column percentage, and row percentage. The right column shows the total count and total percentage per color condition. The bottom right indicates the total for the sample.

Table 25

Viseme Cluster Response by Color for Experiment II

	Color	b	b/m	b/p	m	m/b	m/p	p	p/b	p/m	
Count	B-Y	32	32	52	58	55	23	45	35	28	360
Total %		1.78	1.78	2.89	3.22	3.06	1.28	2.50	1.94	1.56	20.00
Col %		15.24	15.09	26.26	22.57	23.11	18.70	16.30	23.33	20.59	
Row %		8.89	8.89	14.44	16.11	15.28	6.39	12.50	9.72	7.78	
Count	G-R	47	50	37	39	37	25	56	41	28	360
Total %		2.61	2.78	2.06	2.17	2.06	1.39	3.11	2.28	1.56	20.00
Col %		22.38	23.58	18.69	15.18	15.55	20.33	20.29	27.33	20.59	
Row %		13.06	13.89	10.28	10.83	10.28	6.94	15.56	11.39	7.78	
Count	NAT	39	34	32	65	66	30	57	15	22	360
Total %		2.17	1.89	1.78	3.61	3.67	1.67	3.17	0.83	1.22	20.00
Col %		18.57	16.04	16.16	25.29	27.73	24.39	20.65	10.00	16.18	
Row %		10.83	9.44	8.89	18.06	18.33	8.33	15.83	4.17	6.11	
Count	R-G	53	43	39	56	39	29	52	27	22	360
Total %		2.94	2.39	2.17	3.11	2.17	1.61	2.89	1.50	1.22	20.00
Col %		25.24	20.28	19.70	21.79	16.39	23.58	18.84	18.00	16.18	
Row %		14.72	11.94	10.83	15.56	10.83	8.06	14.44	7.50	6.11	
Count	Y-B	39	53	38	39	41	16	66	32	36	360
Total %		2.17	2.94	2.11	2.17	2.28	0.89	3.67	1.78	2.00	20.00
Col %		18.57	25.00	19.19	15.18	17.23	13.01	23.91	21.33	26.47	
Row %		10.83	14.72	10.56	10.83	11.39	4.44	18.33	8.89	10.00	
Total x	5	210	212	198	257	238	123	276	150	136	1800
Color		11.67	11.78	11.00	14.28	13.22	6.83	15.33	8.33	7.56	

Table 26

Tests for Exp. II Phoneme Response by Color

Source	*df*	-LogLike	R Square (U)
Model	32	36.0373	0.0093
Error	1760	3859.0138	
C Total	1792	3895.0511	
N	1800	1800	

Tests	Chi-Square	Prob >ChiSq
Likelihood Ratio	72.075	< 0.0001 ****
Pearson	71.066	< 0.0001 ****

Note. This table indicates the significance of the Chi-square test statistic. The table indicates that there is significant independence between color and phoneme responses in the context of /ɑ/.

**** $p < 0.0001$

APPENDIX B

FIGURES

Figure 1. Information processing models. This figure shows two theoretical information processing models: an "all-or-none" model (Broadbent, 1958) and an "attenuation model" (Triesman, 1960). These two models reinforce the supposition that parallel processing for audition and vision involves a "tuning-out" or "turning down" selected information. From "Controlled and Automatic Human Information Processing: I. Detection, Search, and Attention," by W. Schneider and R. Shiffrin, *Psychological Review 84* p. 5. Copyright 1977 by the American Psychological Association. Used with permission by the American Psychological Association.

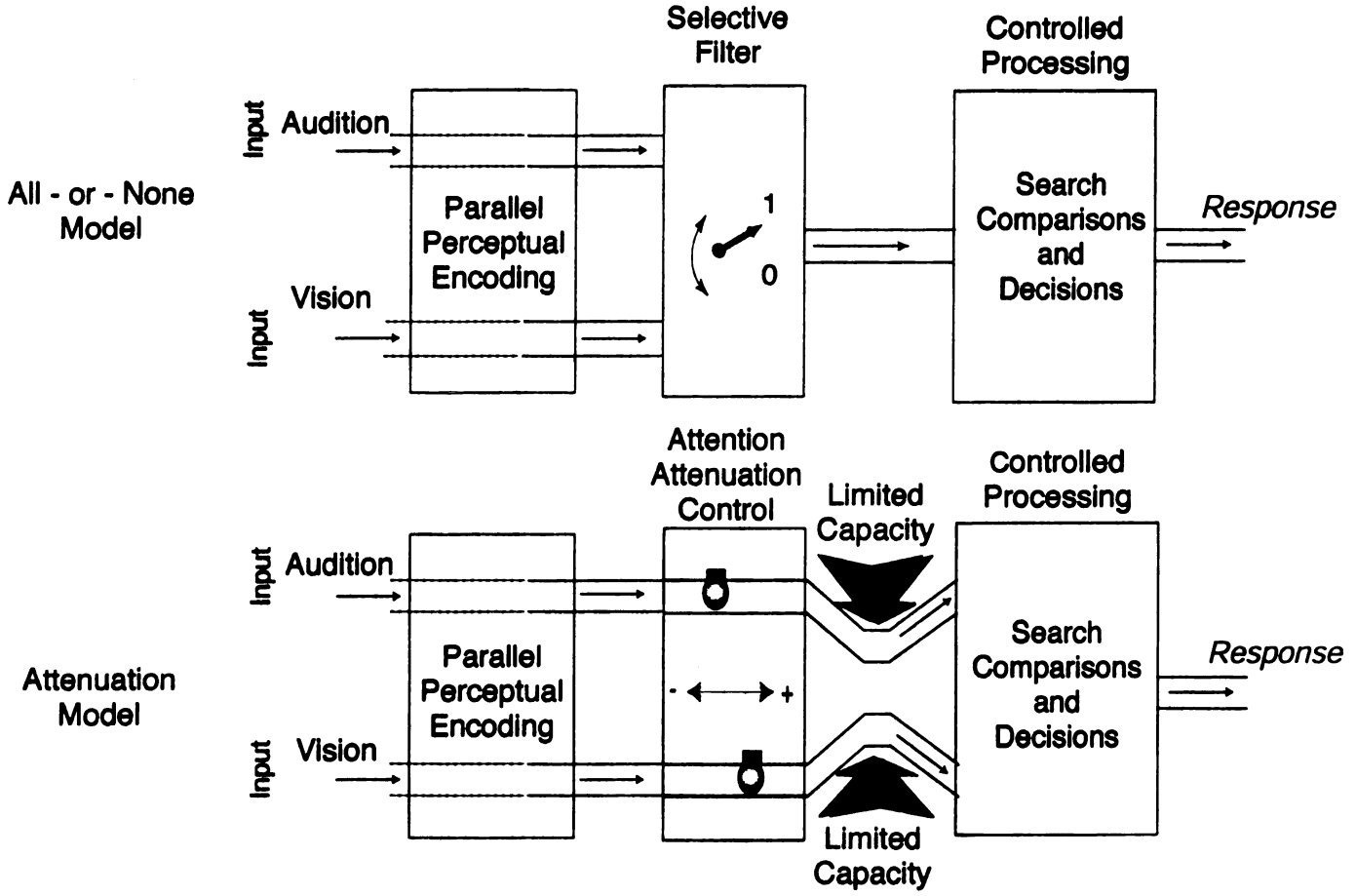

Figure 1. Information processing models.

Figure 2. Dual process neural circuits. This figure corresponds to algebraic relationships that are integrated into a dual process of both excitatory (arrows) and inhibitory (dots) neural input to large (L), medium (M), small (S) opponent cone receptors. The cones, within a neural circuit, encode and convey specific information for red (R), green (G), yellow (Y), blue (B), and [black (Bl) and (Wh)]. From Visual Science by S. Palmer p. 114. Copyright 1999 by the MIT Press. Used by permission from MIT Press.

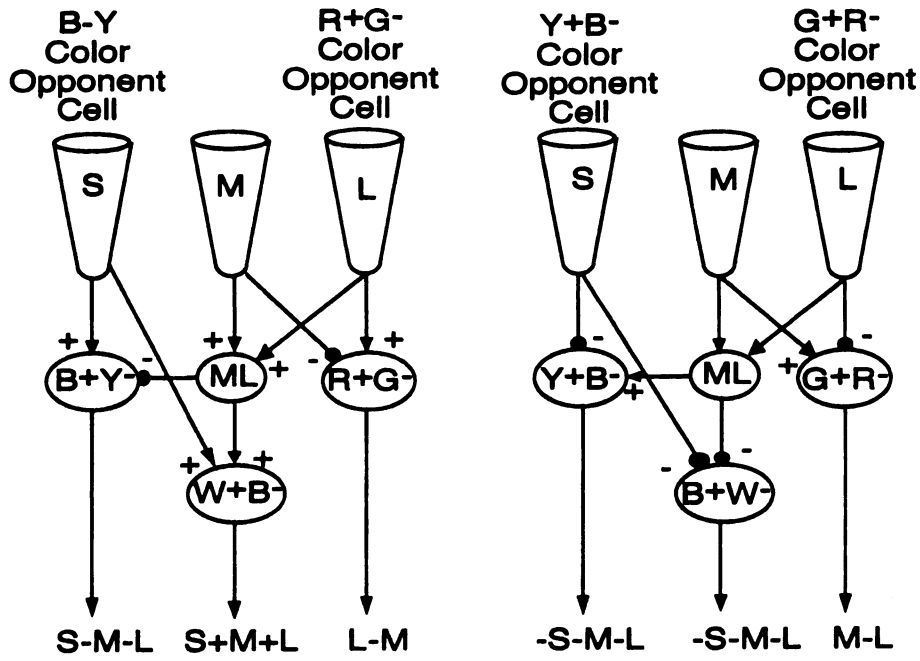

Figure 2. Dual process neural circuits.

Figure 3. Projections from retina to the visual centers. This figure depicts the neural projections from the retina to the visual areas of the thalamus (lateral geniculate nucleus) and midbrain (pretectum and superior colliculus). The projection to the superior colliculus contributes to visually guided eye movements. The projection to the lateral geniculate nucleus and on to the visual cortex processes visual information for perception. From Central Visual Pathway by R. Wurtz & E. Kandel. In E. Kandel, J. Schwartz, & T. Jessel (Eds.) Principles of Neural Science. p. 527. Copyright 2000 by McGraw-Hill. Used by permission by McGraw-Hill.

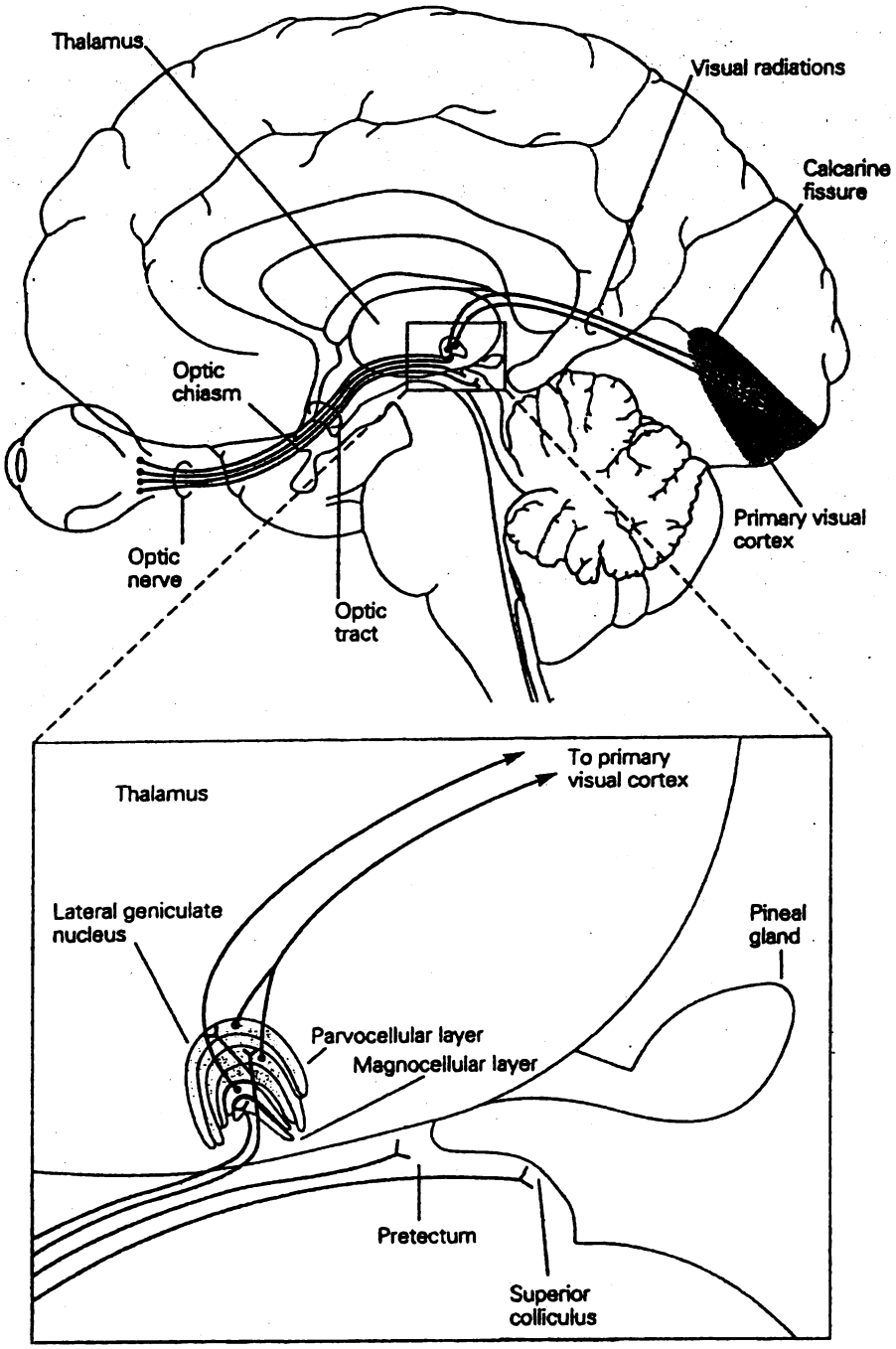

Figure 3. Projections from retina to the visual centers.

Figure 4. Experiment room setup. In the depiction the subject-to-monitor distance was maintained at 1.5m within an arc of ±30° relative to 0° azimuth. The monitor height was 0.76 m (2.5 ft). From 0° azimuth, an inter-subject separation of 0.73 m (2.6 ft) corresponded to 15°. There were two partitions with a height of 0.91 m (3 ft) placed to the left and right of position "B" to visually isolate the subjects from each other.

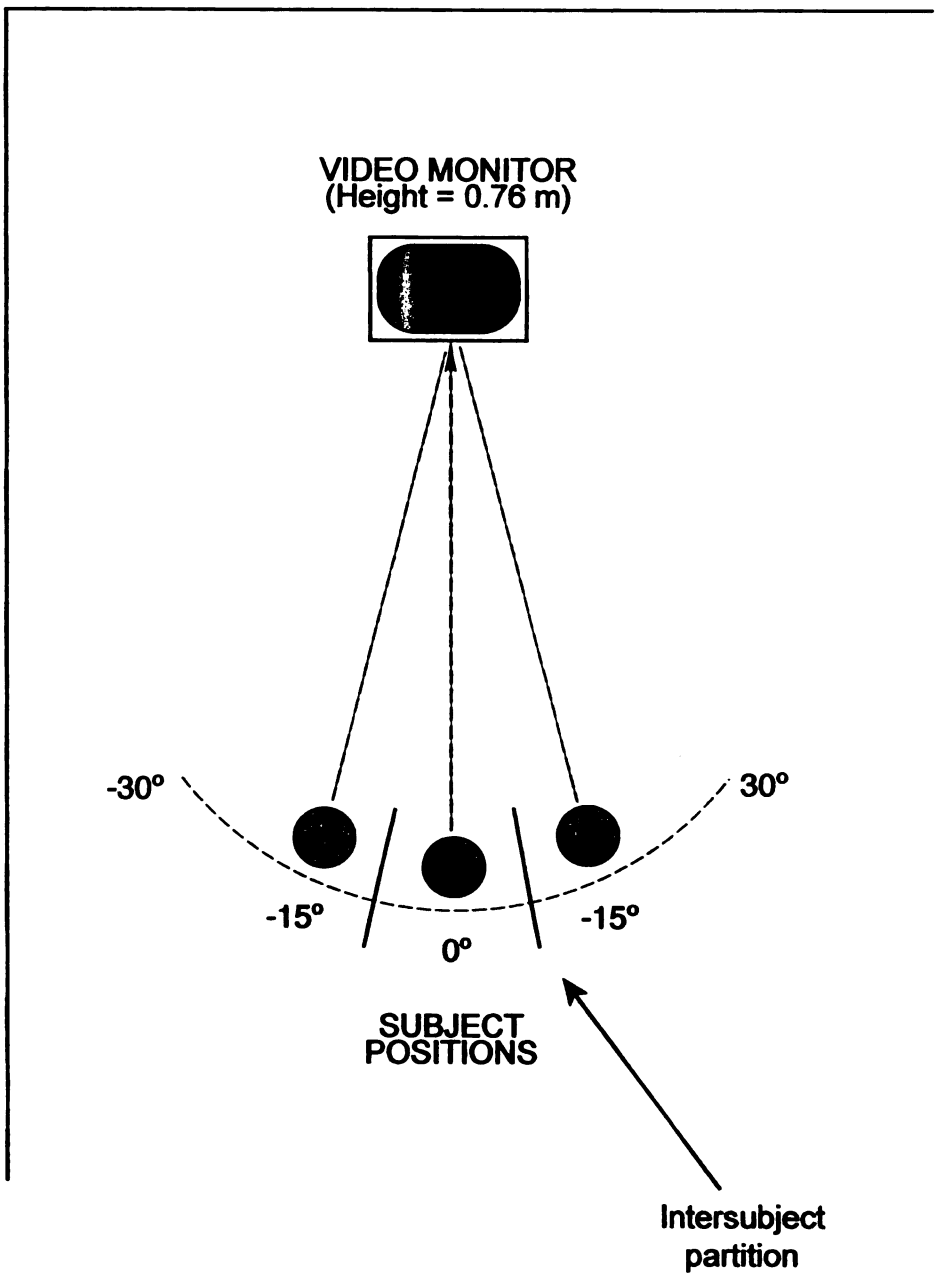

Figure 4. Experiment room setup.

Figure 5. The talker's face in the R-G color condition. Cosmetic lip-color was brush applied to the talker's lips to create an opponent color effect. This image shows the talker minutes before the video recording.

Figure 6. The talker's face in the G-R color condition. Cosmetic lip-color was brush applied to the talker's lips to create an opponent color effect. This image shows the talker minutes before the video recording.

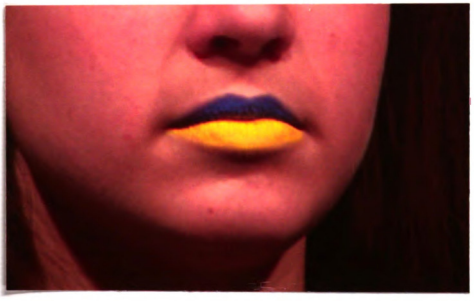

Figure 7. The talker's face in the B-Y color condition. Cosmetic lip-color was brush applied to the talker's lips to create an opponent color effect. This image shows the talker minutes before the video recording.

Figure 8. The talker's face in the Y-B color condition. Cosmetic lip-color was brush applied to the talker's lips to create an opponent color effect. This image shows the talker minutes before the video recording.

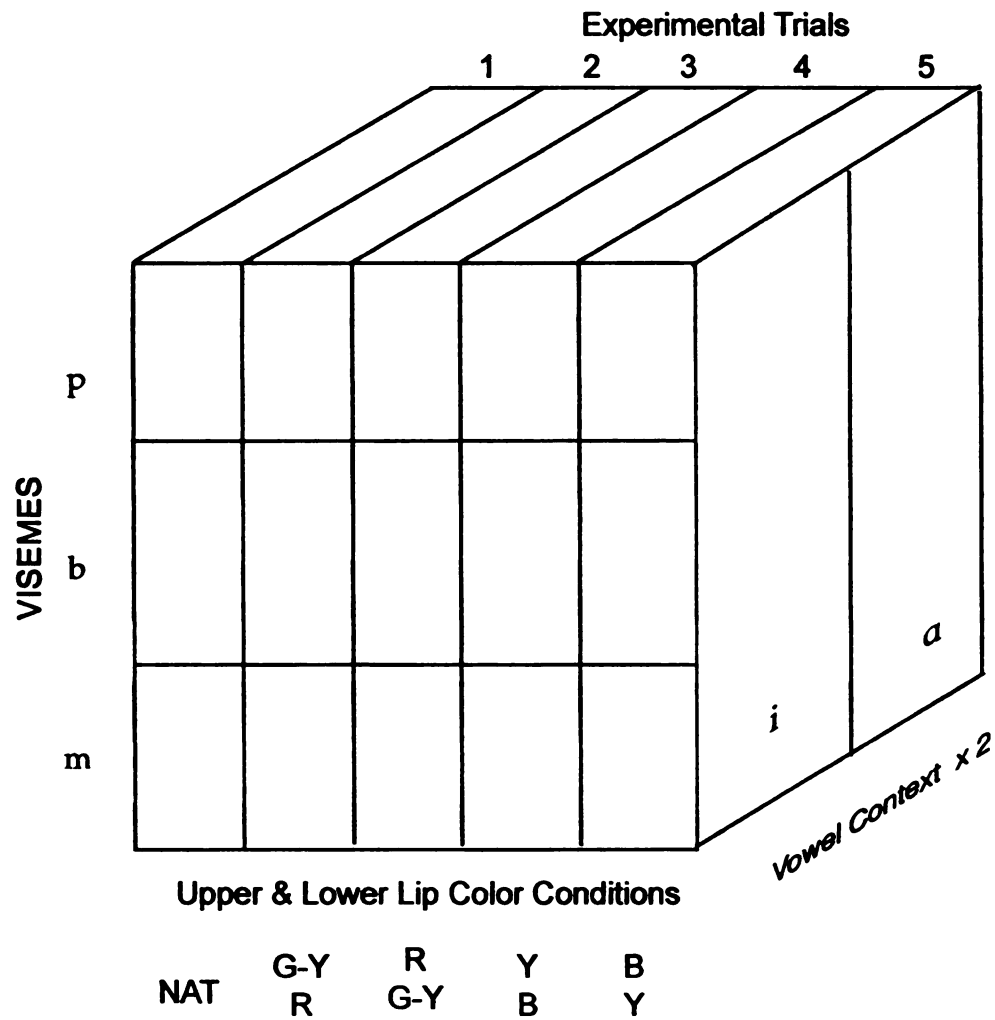

Figure 9. Statistical model for outcome by relevant predictors. This item depicts the design of the experiment for logistic regression analysis. The model included the following relevant predictors for analysis: (1) color, (2) phonemes, (3) vowels, and (4) trials.

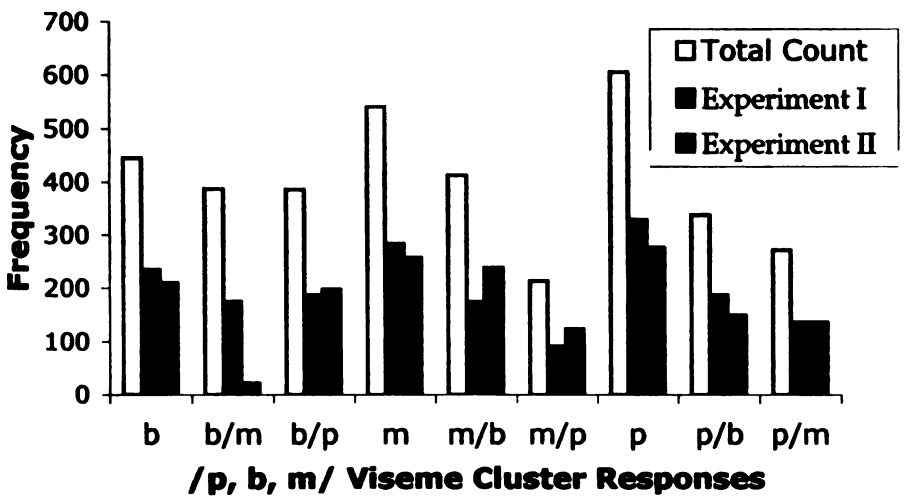

Figure 10. Distribution of visual recognition by experiment. This item depicts the frequency of correct responses and six possible substitution responses for the /p, b, m/ viseme-cluster in Experiments I and II. The frequency distribution indicated a maximum count of 329 for /p/ and a minimum count of 91 for the m/p substitution. *n = 3,600

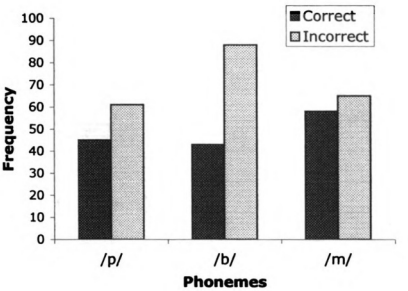

Figure 11. The frequency distribution of R-G in the /i/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

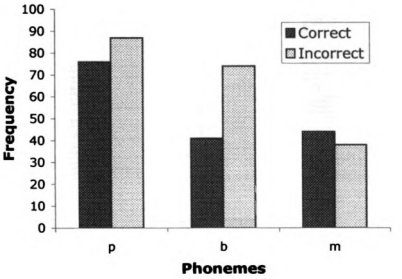

Figure 12. The frequency distribution of G-R in the /i/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

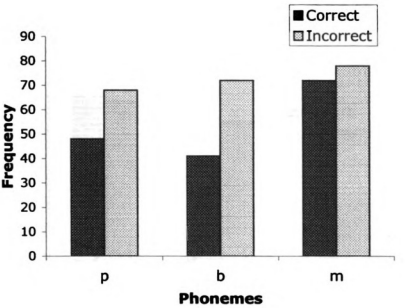

Figure 13. The frequency distribution of B-Y in the /i/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

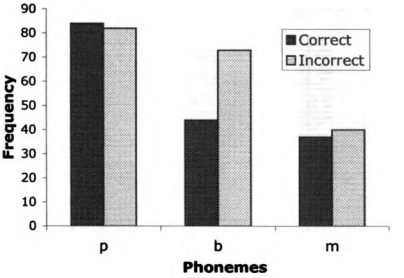

Figure 14. The frequency distribution of Y-B in the /i/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

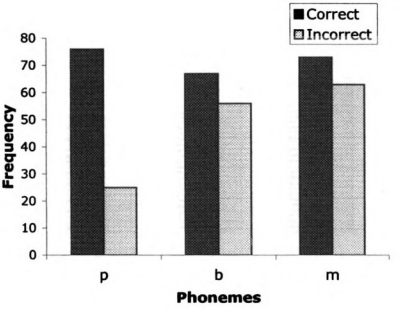

Figure 15. The frequency distribution of NAT in the /i/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

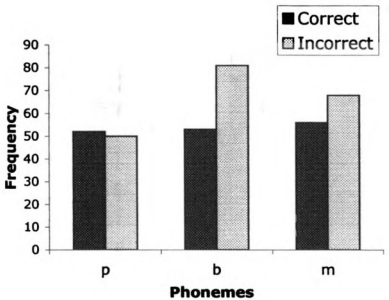

Figure 16. The frequency distribution of R-G in the /a/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

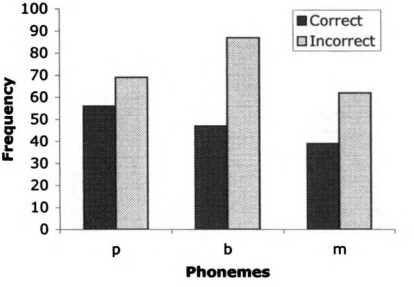

Figure 17. The frequency distribution of G-R in the /ɑ/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

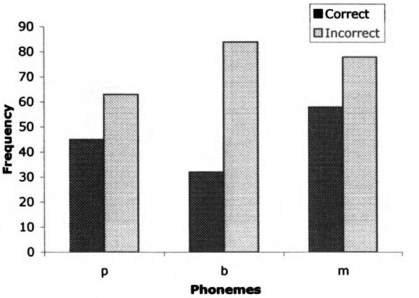

Figure 18. The frequency distribution of B-Y in the /ɑ/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

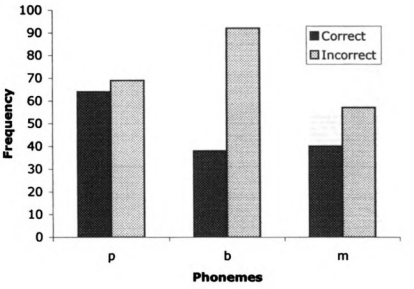

Figure 19. The frequency distribution of Y-B in the /ɑ/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

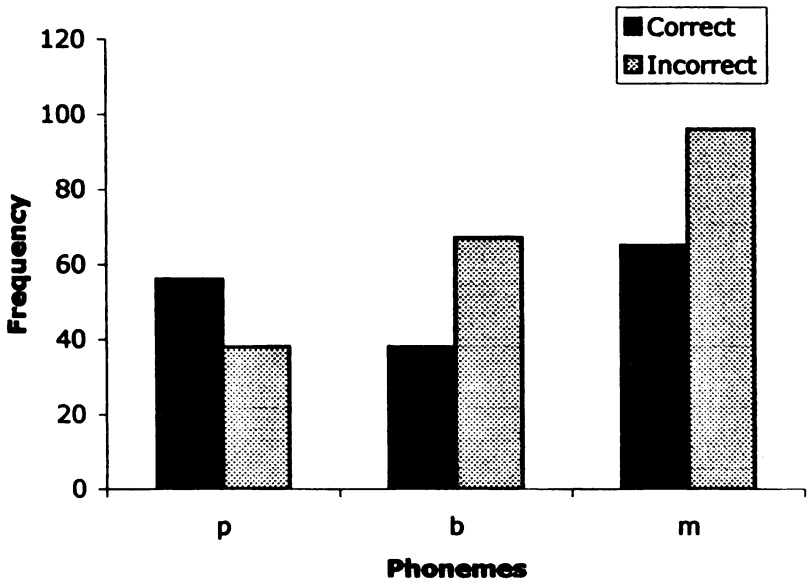

Figure 20. The frequency distribution of NAT in the /ɑ/ context is shown. The dark bars correspond to the correct responses and the light bars correspond to the incorrect responses. *n = 360

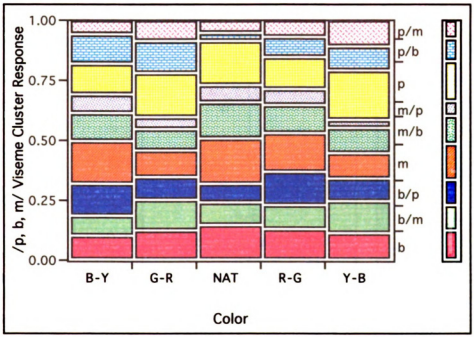

Figure 21. Mosaic plot of visemes by color. This plot that depicts viseme responses by color condition. The "x-axis " represents the proportions of the relative sizes of the color conditions. The proportions along the right y-axis represent the relative sizes of the responses. The scale along the left y-axis indicates the response probability. *n = 3,600

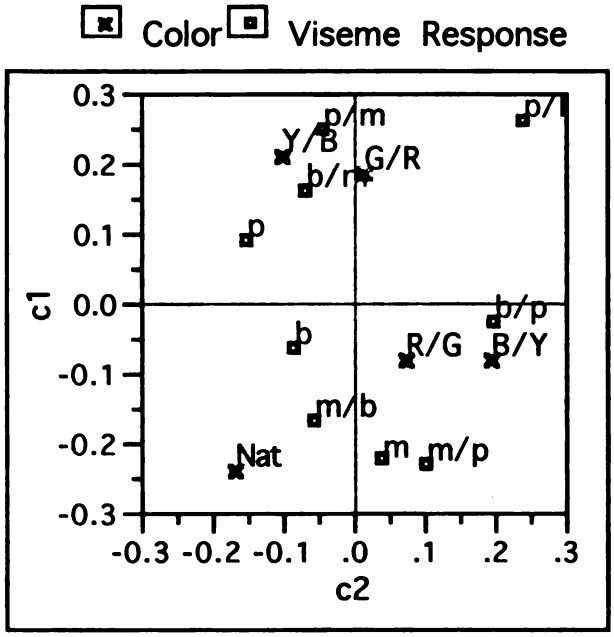

Figure 22. Correspondence analysis of visemes by color. The analysis involves a converted matrix of nonnegative data into the coordinates of c1 and c2, to show which rows and columns have similar patterns. The higher response values are negative in c1 and positive in c2. The moderate response levels are negative in c2 and neutral (0) in c1. *n = 3,600.

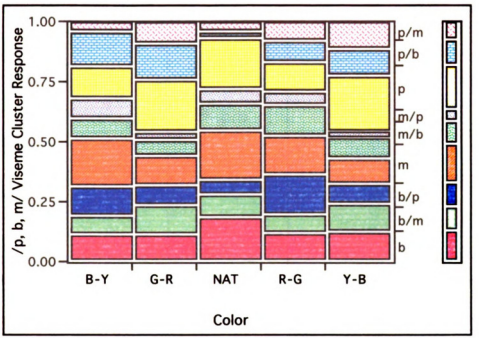

Figure 23. Experiment I mosaic plot. The plot depicts the relative proportion of frequency distribution and counts for /p, b, m/ viseme-cluster. The "x-axis " represents the proportions of the relative sizes of the color conditions. The proportions along the right y-axis represent the relative sizes of the responses. The scale along the left y-axis indicates the response probability. *n = 1,800

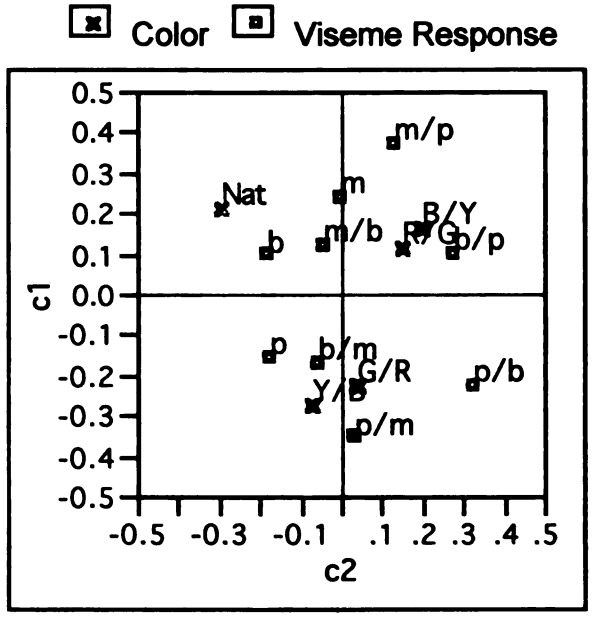

Figure 24. Experiment I. correspondence analysis. This plot for depicts visemes by color. The analysis involves a converted matrix of nonnegative data into the coordinates of c1 and c2, to show which rows and columns have similar patterns. The higher response values are negative in c1 and positive in c2. The moderate response levels are negative in c2 and neutral (0) in c1. n = 1,800

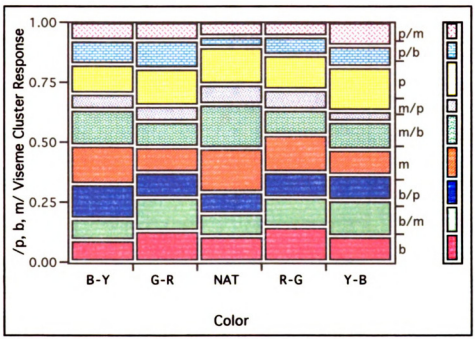

Figure 25. Experiment II mosaic plot. The plot depicts the relative proportion of frequency distribution and counts for /p, b, m/ viseme-cluster. The "x-axis " represents the proportions of the relative sizes of the color conditions. The right y-axis represents the relative sizes of the responses. The left y-axis indicates the response probability. *n = 1,800

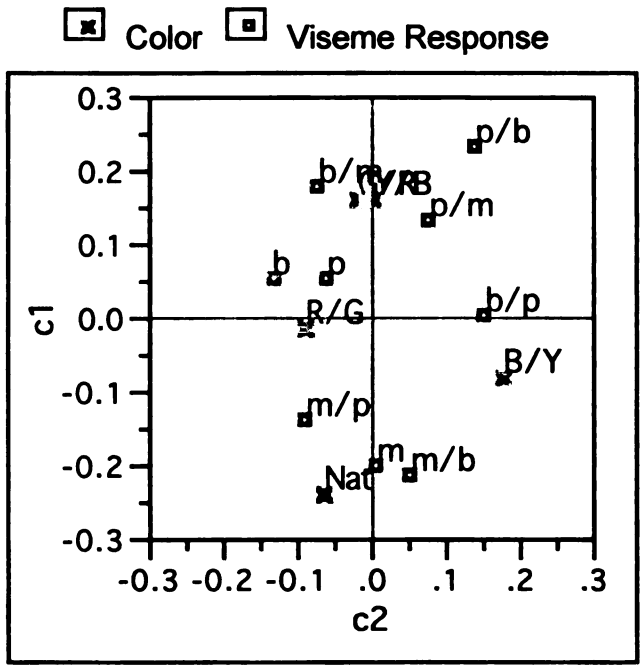

Figure 26. Experiment II. correspondence analysis. The analysis involves a converted matrix of nonnegative data into the coordinates of c1 and c2, to show which rows and columns have similar patterns. The higher response values are negative in c1 and positive in c2. The moderate response levels are negative in c2 and neutral (0) in c1. n = 1,800

APPENDIX C

FORMS

Form 1. Subject Consent Form

Note. This form was used to inform the subjects about the study, their, rights, protection of privacy and compensation. The form was approved by the Michigan State University University Committee on Research Involving Human Subject (UCRIHS).

Form 1. Subject Consent Form

Speechreading Research

This investigation will examine the influence of lip-color upon the visual recognition of the consonants /p, b, m/. Subjects will view a video monitor of an actor speaking vowel-consonant-vowel disyllables (for example, apa) in five lip-color combinations. One experiment will use the vowel /i/; a second experiment will use the vowel /a/. Subjects will attend two sessions of 40-60 minutes each. The recorded responses of six subjects will be compiled for data analysis. The investigators are interested in determining whether lip-color will influence the accurate perception of the consonants /p/, / b/, and /m/.

1. Data gathered in this study will be treated in strict confidence. Your name will not be used in any research findings and reports. In this study, you will be identified by a code system known only by the investigators. After data analysis has been completed, your identification by that code system will be destroyed. Your privacy will be protected to the maximum extent allowable by law. On request and within these restrictions, results may be available to you.

2. If you agree to be a subject in this experiment, you will be financially compensated at a rate of $8 per hour covering two sessions of approximately 60 minutes each. Participation in this study will not result in any cost to you nor will the study provide you with any direct benefit.

3. Participation in this study is voluntary. You may choose not to participate at all. You may also discontinue as a subject at any time without penalty or loss of benefits to which you are otherwise entitled.

4. If you have any questions or concerns regarding your participation in the study, you may contact Dr. Leo V. Deal, Principal Investigator (517) 332.8228 email ideal@msu.edu or Mr. Willard C. Hooks, Jr., Doctoral Candidate, (815) 753.6332 (work) or 630) 208.9852 (home), email hookswil@msu.edu. If you have an questions or concerns regarding your rights as study participant, please contact Ashir Kumar, M.D., University Committee on Research Involving Human Subject (UCRIHS) by phone (517) 355-2180, email: UCRIHS@msu.edu.

Your signature below indicates that you voluntary agree to participate in this study.

Signature _____

Date _____

Form-2. Color Deficiency Screening Form

Note: Potential subjects were required to view "pseudo-isochromatic" plates (Beck, 1965) with the illumination of a 40-Watt incandescent bulb.

Form 2. Color Deficiency Screening Form

Score Sheet

Subject Number _____ Date_____

Plate Number	Normal Response	Subject's Response	Plate Number	Normal Response	Subject's Response
1	12		9	56	
2	6		10	27	
3	42		11	89	
4	56		12	86	
5	57		13	15	
6	75		14	74	
7	5		15	47	
8	3				

Form 3. Stereopsis Screening Form

Note: Potential subjects were required to view a Verhoeff Stereoptor® (American Optical Company, Southbridge, MA) at eye level from a distance of 60.96 cm (24 in). Responses were recorded for the indicate verbally either to the "left," "middle," or "right."

Form 3. Stereopsis Screening Form

SUBJECT NUMBER _____

DATE: _____

1.	Middle	Far	Near
2.	Left	Far	Near
3.	Left	Far	Near
4.	Left	Far	Near

1.	Left	Far	Near
2.	Right	Far	Near
3.	Right	Far	Near
4.	Middle	Far	Near

Form 4. Hearing Screening

Note. Potential subjects were screened for normal-hearing sensitivity with a GSI-20 audiometer. The threshold for passing was 20 dB HL (ANSI) S3.6-1989.

Form 4. Hearing Screening

Hearing Screening

SUBJECT NUMBER _____

DATE: _____

Hearing Level	500 Hz	1000 Hz	2000 Hz	4000 Hz
R / L				
Result Pass / Fail				

PTA R _____dB HL

PTA L _____dB HL

GSI-17 Audiometer

Form 5. Dyslexia Questionnaire

Note. As a component to the subject screening process, this questionnaire included selected items that correspond, in some instances to signs and/or symptoms of dyslexia. This investigator included subjects with no more than three confirming responses to questions 1, 3, 6, 8, 13, 17, and 20.

Initial Dyslexia Screening Test (2002) Available http://www.brad.ac.uk/ qmmaps/dis/ dyslexia.htm University Of Bradford: Author

Form 5. Dyslexia Questionnaire

SUBJECT NUMBER _____

DATE: _____

AT WORK OR COLLEGE

1. Do you get confused about following instructions, for example with a new procedure at work or a new routine at college?
YES ☐ NO ☐

2. Are there any particular tasks that you find difficult in your work/studies?
YES ☐ NO ☐ If so, please describe them:

_____

_____

3. Do you experience frustration at your inability to cope with certain tasks at work or college?
YES ☐ NO ☐

4. Do you have difficulty remembering telephone numbers?
YES ☐ NO ☐

5. Can you easily remember what you hear for example lyrics, conversations, lectures?
YES ☐ NO ☐

6. Do you find it hard to concentrate for long periods?
YES ☐ NO ☐

7. Can you easily remember what you hear and see for example lyrics, conversations, lectures?
YES ☐ NO ☐

LEFT AND RIGHT

8. Do you ever confuse left and right?
YES ☐ NO ☐

9. Are you left-handed?
YES ☐ NO ☐ NOT CLEARLY LEFT OR RIGHT-HANDED ☐

10. Is anyone else in the (blood-related) family left-handed?
YES ☐ NO ☐ If so, who?

11. With which foot do you naturally choose to kick a ball?
RIGHT ☐ LEFT ☐

12. Take a piece if paper, roll it into a tube, and look through it like using a telescope. Which eye did you naturally put it to?
RIGHT ☐ LEFT ☐

13. Do you hesitate before writing the 'b' or the 'd' or other letters or numbers because you have to think which way round they go? If so, please
YES ☐ NO ☐

MEDICAL

14. Is there anything unusual in your medical history?
YES NO If so, what?

_____

_____

15. What was your weight at birth - if you know? _____
do not know ☐

16. Do you have any eyesight problems?
YES ☐ NO ☐ If so, please describe:

17. Do you ever find that you do not hear what people say ?
YES ☐ NO ☐ If so, please describe. _____

18. Did you suffer from repeated ear infections, or been to hospital to have tubes/grommets inserted in the ear, at any time in your childhood as far as you know?
YES ☐ NO ☐

19. Can you easily read a numerical clock face?
YES ☐ NO ☐

20. Do find it hard to remember instructions?
YES ☐ NO ☐

Form 6. Subject Response Form and Data Sheet

Note. The subjects marked their perceived selection for each phoneme in a three-choice, write-down multiple-choice format. Each experimental condition test session generated 90 observations in six trials. There was a separate response form for viewing each experimental condition four times.

Form 6. Subject Response Form

Experimental Condition ☐ I ☐ II

Subject ID _____ Date_____ Age _____

For Experimenter Use Only - Write Below This Line.

Check the box that corresponds to your perception of lip movements. Circle your position of observation (A. B. or C).

Trial # 1

Seating Position

A B C

(1) Natural

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(2) Red-Green

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(3) Green-Red

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(4) Blue-Yellow

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(5) Yellow-Blue

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

Check the box that corresponds to your perception of lip movements.

Trial # 2

Seating Position

A **B** **C**

(2) Red-Green

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(3) Green-Red

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(4) Blue-Yellow

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(5) Yellow-Blue

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(1) Natural

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

CHANGE SEATING POSITION

Check the box that corresponds to your perception of lip movements.

Trial # 3

Seating Position

A **B** **C**

(3) Green-Red

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(4) Blue-Yellow

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(5) Yellow-Blue

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(1) Natural

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(2) Red-Green

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

Check the box that corresponds to your perception of lip movements.

Trial # 4

Seating Position

A **B** **C**

(4) Blue-Yellow

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(5) Yellow-Blue

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(1) Natural

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(2) Red-Green

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

(3) Green-Red

☐p ☐b ☐m
☐p ☐b ☐m
☐p ☐b ☐m

Check the box that corresponds to your perception of lip movements.

CHANGE SEATING POSITION

Trial # 5 **Seating Position**

(5) Yellow-Blue

			A	B	C
☐p	☐b	☐m			
☐p	☐b	☐m			
☐p	☐b	☐m			

(1) Natural

☐p	☐b	☐m
☐p	☐b	☐m
☐p	☐b	☐m

(2) Red-Green

☐p	☐b	☐m
☐p	☐b	☐m
☐p	☐b	☐m

(3) Green-Red

☐p	☐b	☐m
☐p	☐b	☐m
☐p	☐b	☐m

(4) Blue-Yellow

☐p	☐b	☐m
☐p	☐b	☐m
☐p	☐b	☐m

Check the box that corresponds to your perception of lip movements.

Trial # 6 **Seating Position**

(1) Natural

			A	B	C
☐p	☐b	☐m			
☐p	☐b	☐m			
☐p	☐b	☐m			

(2) Red-Green

☐p	☐b	☐m
☐p	☐b	☐m
☐p	☐b	☐m

(3) GreenRed

☐p	☐b	☐m
☐p	☐b	☐m
☐p	☐b	☐m

(4) Blue-Yellow

☐p	☐b	☐m
☐p	☐b	☐m
☐p	☐b	☐m

(5) Yellow-Blue

☐p	☐b	☐m
☐p	☐b	☐m
☐p	☐b	☐m

REFERENCES

REFERENCES

Agresti, A. (1990). Categorical data analysis. New York: John Wiley & Sons.

Agresti, A. (1996). An introduction to categorical data analysis. New York: John Wiley & Sons.

Alpiner, J. (1987). Evaluation of adult communication function. In J. G. Alpiner & P. A. McCarthy (Eds.), Rehabilitative audiology: Children and adults. (pp. 15-16, 44-70) Baltimore: Williams & Wilkins.

Annett, M. (1996). Laterality and types of dyslexia. Neuroscience and Biobehavioral reviews. 20, 631-636.

Arnold, P. (1993). The optimization of hearing-impaired children's speechreading. International Journal of Pediatric Otolaryngology, 26, 209-223.

Auer, E. T., & Bernstein, L. E. (1997). Speechreading and the structure of the lexicon: Computationally modeling the effects of reduced distinctiveness on lexical uniqueness. Journal of the Acoustical Society of America, 102, 3704-3711.

Beck Engraving Company. (1965). Psuedo-Isochromatic Plates. NewYork: Author.

Berg, F. S. (1976). Hearing impairment, the hard of hearing child, and educational audiology. Educational audiology: Hearing and speech management, (pp. 1-2). New York: Grune & Stratton.

Berger, K. W. (1972). Speechreading principles and methods. Baltimore: University Park Press.

Binnie, C. A., Montgomery, A. A., & Jackson, P. A. (1977). Auditory and visual contributions to the perception of consonants. Journal of Speech and Hearing Research, 17, 619-630.

Breeuwer, M., & Plomp, R., (1984). Speechreading supplemented with frequency-selective sound pressure information. Journal of the Acoustical Society of America, 76, 686-691.

Broadbent, D. E. (1958). Perception and communication. London: Pergamon.

Bunger, A. M. (1961). Speech reading, Jena method; a textbook with lesson plans in full development for hard of hearing adults and discussion of adaptations for hard of hearing and deaf children. Danville, IL; Interstate.

Burgess, P. R., & Perl E. R. (1973). Cutaneous mechanoreceptors and nociceptors. In: A. Iggo (Ed.) Handbook of sensory physiology. Vol. 2. Somatosensory System. (pp. 29-78) Berlin: Springer-Verlag.

Burnham, D. (1998). Language specificity in the development of auditory-visual speech perception. In R. Campbell, B. Dodd, & D. Burham (Eds.) Hearing by eye II. (pp. 27-60) East Sussex, UK: Psychology Press.

Campbell, R. (1998) Everyday speechreading: understanding seen speech in action. Scandinavian Journal of Psychology, 39, 163-167.

Croner, L. J., & Albright, T. D. (1999). Segmentation by color influence responses to motion-sensitive neurons in the cortical middle temporal (MT) visual area. Journal of Neuroscience, 19, 3935-3951.

Dadson, R. S., & King, J. H. (1952). A determination of the normal threshold of hearing and its relation to the standardization of audiometers. Journal of Laryngology and Otology, 66, 366-378.

Davis, H. (1970). Abnormal hearing and deafness. In H. Davis & S. R. Silverman (Eds.), Hearing and deafness (pp. 83-139). New York: Holt, Rinehart & Wilson.

Davis, J., & Hardick, E. (1981). Rehabilitative audiology for children and adults. New York: Wiley & Sons.

Deland, F. (1931/1968). The story of lip-reading, its genesis and development. Washington, D.C.: The Volta Bureau.

Demorest, M. E., & Bernstein, L. E. (1992). Sources of variability in speechreading sentences: A generalizability analysis. Journal of Speech and Hearing Research, 35, 876-891.

Demorest, M. E., Bernstein, L. E., & DeHaven, G. P. (1992). Generalizability of speechreading performance on nonsense syllables, words, and sentences: subjects with normal hearing. Journal of Speech and Hearing Research, 39, 697-713.

De Valois , R. L., Abramov, I., & Jacobs, G. H. (1966). Analysis of response patterns in LGN cells. Journal of Optical Society of America, 56, 966-977.

De Valois, R. L., & Jacobs, G. H. (1968). Primate color vision. Science, 162, 533-540.

Erber, N. P. (1974a). Effects 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. (1974b). Discussion: Lipreading skills. In R. E. Stark (Ed.), Sensory capabilities of hearing-impaired children (pp. 69-73). Baltimore: University Park Press.

Erber, N. P. (1979). Auditory-visual perception of speech with reduced optical clarity. Journal of Speech and Hearing Research, 22, 212-223.

Erber, N. P. (1983). Speech perception and speech development in hearing-impaired children. In I. Hochberg, H. Levitt, & M. J. Osberger (Eds.) Speech of the Hearing Impaired: research, training, and personnel preparation (pp. 131-145). Baltimore: University ark Press.

Fisher, C. G. (1968). Confusions among visibly perceived consonants. Journal of Speech and Hearing Research, 11, 796-804.

Fitch, R. H., Miller, S., & Tallal, P. (1997). Neurobiology of speech perception. Annual Review of Neuroscience, 20, 331-353.

Food, Drug, & Cosmetic Act (2001). United States of America Code of Federal Regulations, (Title 21 (7) CFR Part 700 to 740). Washington, DC: U.S. Government Printing Office via GPO Access.

Friendly, M. (1994). Mosaic displays for multi-way contingency tables. Journal of the American statistical Association, 89, 190-200.

Ginsberg, A. P., & Hendee, W. R. (1993). Quantification of visual capability. In W. R. Hendee & P. N. T. Wells (Eds.) The perception of visual information (pp. 52-72). New York: Springer-Verlag.

Giolas, T. G., Vaughn, G. R., Lightfoot, R. K., & Karp, A. (1989). Overview of aural rehabilitation. In R. L. Schow & M. A. Nerbonne (Eds.), Introduction to aural rehabilitation (3rd ed., pp. 3-29). Austin, TX: Proed.

Goldberg, M. E., Eggers, H. M., & Gouras, P. (1993). The ocular motor system. In: E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), Principles of neural science (3rd ed., pp. 661-678). Elsevier: New York.

Gordon, J. & Ghez, C. (1991). Muscle receptors and spinal reflexes: the stretch reflex. In: E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), Principles of neural science (3rd ed., pp. 564-568). Elsevier: New York.

Greenacre, M. (1988). Correspondence analysis of multivariate categorical data by weighted least-squares. Biometrika, 75, 457-467.

Hardy, M. P. (1970). Speechreading. In H. Davis & S. R. Silverman (Eds.), Hearing and deafness (pp. 335-345). New York: Holt Rinehart and Wilson.

Hartigan, J. A. & Kleiner (1984). A mosaic of television ratings. The American Statistician, 38, 32-35.

Helmholtz, H. von. (1867). Handbuch der physiologischen optic. [Treatise on physiological optics]. In James P. C. Southall (Ed.), Handbuch der physiologischen Optik., Rochester N.Y.: The Optical Society of America.

Hering, E. (1878/1964). Outlines of a theory of the light sense (L. M. Hurvich & D. J. Jameson, Trans.). Cambridge, MA: Harvard University Press.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences of the United States of America, 79, 2554-2558.

Hosmer, D.W., & Lemeshow, S. (1989). Applied logistic regression. New York: John Wiley & Sons.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology, 185, 215-243.

Hurvich, L. M., & Jameson, D. (1957). An opponent process theory of color vision. Psychological Review, 64, 384-404.

Ijsseldijk, R. J. (1992). Speechreading performance under different conditions of video image, repetition, and speech rate. Journal of Speech and Hearing Research, 35, 466-471.

Jacobs, G. H. (1996). Primate photopigments and primate color vision. Proceedings of the National Academy of Science 93, 577-581.

Jeffers, J., & Barley, M. (1971). Speechreading (lipreading). Springfield: Charles C. Thomas.

JMP 5.0 [Computer software]. (2002a). Cary, NC: SAS Institute.

JMP 5.0 [Computer software manual]. (2002b) Cary, NC: SAS Institute.

Johansson, K., Rönnberg, J., &, Lyxell, B. (1991) Contrasting subjective judgment and objective test in the hearing-impaired. Scandinavian Audiology, 20, 91-99.

Jorgensen, A. M. (1962). The measurement of homophenous words. Unpublished master's thesis, Michigan State University, East Lansing.

Kandel, E., R. & Wurtz, R. H. (2000). Constructing the visual image. In: E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), Principles of neural science (4th ed., pp. 492-505). New York: McGraw-Hill.

Kinzie, C. E. (1920). The Kinzie method of speech reading. Volta Review 17, 471-476.

Kricos, P. B., & Lesner, S. A. (1982). Differences in visual intelligibility across talkers. Volta Review, 64, 19-25.

Lambe, E. K. (1999). Dyslexia, gender, and brain imaging, Neuropsychologia 37, 521-523.

Lansing, C. R., & McConkie, G. W. (1999). Attention to facial regions in segmental and prosodic visual speech perception tasks. Journal of Speech-Language-Hearing Research, 42, 526-539.

Leahey, T. H. (1985). Memory and information processing: perceptual and short-term memory processes. Englewood Cliffs: Prentice-Hall.

Lebo, C., & Reddell, R. (1972). The presbycusis component in occupational hearing loss.Laryngoscope 82, 1399-1409.

Lesner, S. (1988). The talker, Volta Review, 90, 88-89.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. Cognition, 21, 1-36.

Livingstone, M. S., & Hubel, D. H. (1987). Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. Journal of Neuroscience, 9, 3416-3468.

Livingstone, M. S., & Hubel, D. H. (1988). Segregation of form, color, movement, and depth: Anatomy, physiology, and perception. Science, 240, 740-749.

Livingstone, M. S., Rosen, G. D., Drislane, F. W., & Galaburda (1991). Physiological and anatomical evidence for a magnocellular defect in developmental dyslexia. Proceedings of the National Academy of Science of the United States of America, 88, 7943-7947.

Long, J., & Luke, J. T. (2001). The new Munsell student color set (2nd ed.). New York: Fairchild.

Losee, R. M. (1999). Information theory beginnings [On-line]. Available: www: http://ils.unc.edu/~ losee/b5/node7.html.

Lyxell, B., & Rönnberg, J. (1991). Visual speech processing: Word-decoding and word-discrimination related to sentence-based speechreading and hearing impairment. Scandinavian Journal of Psychology, 32, 9-17

Lyxell, B., Rönnberg, J., Anderson, J., & Linderoth, E. (1993). Vibrotactile support: initial effects on visual speech perception. Scandinavian Audiology, 22, 179-183.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264, 746-748.

Marassa, L. K., & Lansing, C. (1995). Visual word recognition in two facial motion conditions: Full-face versus lips-plus-mandible. Journal of Speech and Hearing Research, 38, 1387-1394.

Marshall, C. M., Snowling, M. J., & Bailey, P. J. (2001). Rapid auditory processing and phonological ability in normal readers and readers with dyslexia. Journal of Speech and Hearing Research, 44, 925-940.

Massaro, D. W. (1987) Speech perception by ear and eye: A paradigm for psychological enquiry. Hillsdale, NJ: Erlbaum.

Massaro D. W., Cohen, M. M., & Smeele, P. M. (1995). Cross-linguistic comparisons in the integration of visual and auditory speech. Memory and Cognition 23, 113-31.

Menard, S. (2002). Applied logistic regression analysis (2nd ed.). Sage: Thousand Oaks, CA:

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision: two cortical pathways. Trends in Neuroscience, 6, 414-417.

Moores, D. (1996). Educating the deaf—Psychology, principles, practices (4th ed.) Boston: Houghton Mifflin.

Motulsky, H. (1995). Intuitive bisostatistics, Oxford: New York.

Munsell, A. H. (1929). The atlas of the Munsell color system. Baltimore: Munsell Color Company.

Nábêlek, I. V. (1985). Noise Measurement and Engineering Controls, In A. S. Feldman & C. T. Grimes (Eds.) Hearing conservation in industry. (pp. 35-42) Baltimore: Williams and Wilkins.

Nathans, J. (1989). The genes for color vision. Scientific American, 260, 42-49.

National Institute on Deafness and Other Communication Disorders - Hereditary Hearing Impairment Resource Registry (1995). Omaha, Nebraska.

Neitz, J. & Neitz, M. (1994). Color Vision Defects. In A. F. Wright & J. Barrie (Eds.) Molecular genetics of inherited eye disorders. (pp. 217-257). Chur, Switzerland: Harwood Academic Publishers.

Newhall, S. M., Nickerson, D. & Judd, D. B. (1943). Final report of the O.S.A. subcommittee on spacing of the Munsell colors, Journal of the Optical Society of America, 33, 385-397.

Nitchie, E. B. (1905). Lessons in lip-reading for self-instruction. New York, Surdus.

Norton, S. J., Schultz, M. C., Reed, C. M., Braida, L. M. D., Durlach, N. I., Rabinowitz, W. M., & Chomsky, C. (1977). Analytic study of the Tadoma method: Background and preliminary results. Journal of Speech and Hearing Research, 20, 381-393.

FDA Authority over Cosmetics. (1995). Office of Cosmetics Fact Sheet [On-line]. Available: http:// www.cfsan.fda.gov/~dms/cos-206.html

Öhngren, G., Rönnberg, J., & Lyxell, B. (1992). Tactiling: A usable support system for speechreading. British Journal of Audiology, 26, 574-595.

Olsen, W., & Matkin, N. D. (1979). Speech audiometry. In: W. F. Rintlemann (Ed.), Hearing assessment (pp. 133-206). Baltimore: University Park Press.

O'Neill, J. (1954). Contributions of the visual components of oral symbols to speech comprehension. Journal of Speech and Hearing Disorders, 19, 429-439.

O'Neill, J. J., & Oyer, H. J. (1981). Visual communication for the hard of hearing (2nd Ed.) Englewood Cliffs: Prentice-Hall.

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

Oyer, H. J. (1983). Founding philosophy of the Academy of Rehabilitative Audiology. Journal of Academy of Rehabilitative Audiology, 15, 116-130.

Palmer, S. E. (1999). Vision science: Protons to phenomenology. Cambridge, MA: MIT Press.

Palmer, S. E., & Kimchi, R. (1986). The information processing approach to cognition. In: T. J. Knapp & L. C. Robertson (Eds.) Approaches to cognition: contrasts and controversies Hillsdale, NJ: Lawrence Erlbaum Associates (pp. 37-78).

Pichora-Fuller, M. K., & Benguerel, A. (1991). The design of CAST (computer-aided speechreading training). Journal of Speech and Hearing Research, 34, 202-212.

Pickett, J. M., (1999). The acoustics of speech communication: fundamentals, speech perception theory, and technology. Boston: Allyn & Bacon.

Pollack, D. (1970). Educational audiology for the limited-hearing infant. Springfield, IL: Thomas.

Poynton, C. (2002). Frequently asked questions about gamma [On-line]. Available: http://www.inforamp.net/%7Epoynton/.

Preminger, J. E., Lin, H., Payen, M., & Levitt, H. (1998). Selective visual masking in speechreading. Journal of Speech-Language-Hearing Research, 41, 564-575.

Prentiss, S. (1971). How to use vectorscopes oscilloscopes, and sweep-signal generators. Tab Books: Blue Ridge Summit, PA.

Ramachandran, V. S., & Gregory, R. L. (1978). Does color provide input to human motion? Nature 275, 55-56.

Roback, I. M. (1961). Homophenous words. Unpublished master's thesis, Michigan State University, East Lansing.

Rodgers, W. (1995). Analysis of cross-classified data. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics (pp. 169-215). Washington, DC: American Psychological Association.

Rönnberg, J., Andersson, Samuelsson, S., Söderfeldt, B., Lyxell, B., & Risberg, J. (1999). A speechreading expert: The case of mm. Journal of Speech-Language-Hearing Research, 42, 5-20.

Rosenblum, L. D., Johansson, J. A., & Saldaña, H. M. (1996). Point-light displays enhance comprehension of speech in noise. Journal of Speech and Hearing Research, 39, 1159-1170.

Sanders, D. A. (1982). Aural rehabilitation: a management model. Englewood Cliffs: Prentice-Hall.

SAS Institute. (2002a). Convergence criteria [On-line]. Available http://www.pop.upenn.edu/cores/computing/sasdoc/sashtml/stat/chap39/sect9.htm#logisticmgconv.

SAS Institute. (2002b). PROC logistic [On-line]. Available: http://www.luc.edu/infotech/research/saswin/logistic.htm.

Schneider, W., & Shiffrin, R. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. Psychological Review, 84, 1-67.

Schow, R. L., & Nerbonne, M. A. (1996). Overview of aural rehabilitation. In R. L. Schow & M. A. Nerbonne (Eds.), Introduction to aural rehabilitation (3rd ed., pp. 3-29). Needham Heights, MA: Allyn & Bacon.

Schwartz, S. H. (1999). Visual perception: A clinical orientation (2nd ed.). Stamford, CT: Appleton & Lange.

Shannon, C. E., & Weaver, W. (1962). The mathematical theory of communication. Urbana: University of Illinois Press.

Shapley, R., & Perry, V. H. (1986). Cat and monkey retinal ganglion cells and their visual functional roles. Trends in Neuroscience, 9, 229-235.

Shapley, R., Caelli, T., Grossberg, S., Morgan, M., & Rentschler, I. (1990). Computational theories of visual perception. In L. Spillman & J. Werner (Eds.) Visual Perception (pp. 417-448). San Diego: Academic Press

Sharp, P. F., & Philips, R. (1993). Physiologic optics. In W. R. Hendee & P. N. T. Wells (Eds.) The perception of visual information (pp. 17-29).New York: Springer-Verlag.

Shaywitz, S. (1996). Dyslexia. Scientific American 275, 98-105.

Skottun, B., Parke, C., & Lesley, A. (1999). The possible relationship between visual deficits and dyslexia: Examination of a critical assumption. Journal of Learning Disabilities, 32, 2-5.

Smeele, P. M. T., Massaro, D. W., Cohen, M. M., & Sittig, A. C. (1998). Laterality in visual speech perception. Journal of Experimental Psychology, 24, 1232-1242.

Stein, J. F., & Fowler, M. S. (1993). Unstable binocular control in dyslexic children Journal of Research in Reading, 16, 30-45.

Tessier-Lavigne, M. (2000a). Constructing the visual image. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), Principles of Neural Science (4th ed., pp. 492-521). New York: McGraw-Hill.

Tessier-Lavigne, M. (2000b). Visual processing by the retina. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), Principles of Neural Science (4th ed., pp. 507-522). New York: McGraw-Hill.

Tessier-Lavigne, M. (2000c). Color vision. In E. R. Kandel, J. H. Schwartz, & T. M. Jessell (Eds.), Principles of Neural Science (4th ed., pp. 572-579). New York: McGraw-Hill.

Thiele, A., Dobkins, K. R., & Albright, T. D. (1999). The contribution of color to motion processing in macaque middle temporal area. Journal of Neuroscience, 19, pp. 6571-6587.

Treisman, A. M. (1960). Contextual cues in selective listening. Quarterly Journal of Experimental Psychology, 12, 242-248.

Treisman, A. M. (1985). Preattentive processing in vision. Computer Vision, Graphics, and Image Processing, 31, 156-177.

Treisman, A. M. (1969). Strategies and models of selective attention. Psychological Review, 76, 282-299.

Treisman, A. (1986). Features and objects in visual processing. Scientific American 255, 114B-125.

Treisman, A. M., & Davies, A. (1973). Divided attention to ear and eye. In S. Kornblum (Ed.), Attention and Performance (pp. 101-117). New York: Academic Press.

University Of Bradford. (2002). Initial Dyslexia Screening Test [On-line] Available: http://www.brad.ac.uk/qmmaps/dis/ dyslexia.htm

Van Son, N., Bosman, A. J., Lamoré, & Smoorenburg. (1993). Auditory pattern perception in the profoundly hearing-impaired and lipreading of Dutch phonemes. Scandinavian Audiology Suppl., 38,111-123.

Vatikotis-Bateson, E., Eigtsi, I. M., & Yano, S. (1994). Listener eye movement behavior during audiovisual perception. Proceedings of the International Conference on Spoken Language. (pp. 527-530).

Vernon, M., & Andrews, J. (1990). The Psychology Of Deafness. White Plans, NY: Longman Press.

Walden, B. E., Prosek, R. A., & Worthington, O. W. (1974). Predicting audiovisual consonant recognition performance of hearing-impaired adults. Journal of Speech Hearing Research, 17, 270-278.

Wise, C. M. (1957). Detailed view of speech sounds of English speech. Englewood Cliffs: Prentice-Hall.

Wickens, C. D. (1987). Information processing, decision-making, and cognition. In G. Salvendy (Ed.), Handbook of Human Factors (pp. 76-107). New York: Wiley.

Wickens, C. D. (1992). Attention and Perception. In G. Salvendy (Ed.), Engineering Psychology and Human Performance (pp. 72-94; 102-109). New York: Harper Collins.

Wright, R. E. (1995). Logistic regression. In L. G. Grimm & P. R. Yarnold (Eds.), Reading and understanding multivariate statistics (pp. 217-244). Washington, DC: American Psychological Association.

Wyszecki, G. & Stiles, W. S. (1982). Color Science: Concepts and Methods, Quantitative Data and Formulae (2nd ed.). New York: Wiley.

Yoshinaga-Itano, C. (1988). Speechreading instruction for children. Volta Review, 90, 241-255.

Zeki, S. (1980). The representation of colours in the cerebral cortex. Nature, 284, 412-418.