The McGurk illusion: A study of bimodal speech perception

Manisha Thakkar-Willms
The University of Montana

THE McGURK ILLUSION: A STUDY OF BIMODAL SPEECH PERCEPTION

By

Manisha Thakkar-Willms

B.A., University of British Columbia, 1986

Presented in partial fulfillment of the requirements

for the degree of

Master of Arts

University of Montana

1989

Approved by

Chairman, Board of Examiners

Dean, Graduate School

Date

Willms, Manisha, M.A., December 1989 Communication Sciences and Disorders

The McGurk Illusion: A Study of Bimodal Speech Perception (88 pp.)

Director: Michael K. Wynne, Ph.D.

This study examined perceptual responses to the McGurk illusion which involves the simultaneous presentation of conflicting auditory and visual stimuli. The strength of the McGurk illusion using the consonants /b/ and /g/ in combination with the vowels /a/ and /i/ was examined across four subject age groups; 3-5, 11-13, 21-25, 55-65. The results indicated significant differences and interactions across all main effects including age, vowel, and condition. Contrary to hypotheses presented in the literature, descriptive analysis of the data indicated that the percepts of young children are affected by the auditory/visual conflict of certain stimulus combinations. Further descriptive analysis provided intriguing trends in terms of response types within and across subjects. Although the manner in which the illusion was manifested varied across age groups, individuals tended to be either markedly affected or not affected by the auditory/visual conflict. The previously unexplored area of the McGurk illusion in the senior population yielded particularly striking results. Unlike the three younger age groups, the senior subjects were found to be strongly and affected by all four experimental condition of the McGurk illusion.

Several theories regarding the origin of developmental differences were discussed in light of both the past and present findings. These theories included differences in lip reading skills, physiological capabilities of intermodal perception, and attention strategies. Attentional strategies were particularly emphasized as a novel but potentially productive train of thought for further study. Descriptive analyses of the data was also performed for relationships involving the variables of gender and responses type distribution. Suggestions and discussion regarding the need for further research and possible areas of study within intermodal speech perception were presented.

Acknowledgements

I would like to express gratitude to my thesis committee members Dr. Barbara Bain, Dr. Wes Shellen, and especially Dr. Michael Wynne for their wisdom and encouragement. Gratitude is also extended to Mr. Greg MacDonald for his assistance in the preparation of stimuli and to Mr. Richard Keeney for a "custom made" data base system.

For running subjects, checking reliability, proof reading, maintaining perspective, and supporting me always, my deepest appreciation is extended to my partner in life, Michael Willms.

This work is dedicated to the memory of my mother, Hansa Thakkar (1933-1989), who was, is, and will always be my inspiration to succeed.

Manisha Thakkar-Willms

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

The nature of speech perception has been investigated by researchers of many disciplines, yet it remains relatively undefined even after decades of study. Over the years, numerous theories have been put forth and heralded as valid. Unfortunately, these theories have eventually fallen to disrepute through the presentation of contradictory evidence. Such was the evidence discovered in 1976 by McGurk and MacDonald in their study "Hearing lips and seeing voices." Their findings, which cannot be explained by any current theory of speech perception, are the focus of this research.

Prior to the above mentioned study, the perception of speech had generally been regarded as an exclusively auditory event. As discussed by MacDonald & McGurk (1978), the role of vision in speech perception was traditionally considered in only two ways. First, a number of studies had examined vision as an alternative mode to hearing for language input. These papers focused on speech/lip reading skills of hearing-impaired children and adults. Second, vision has been considered as having a compensatory or complementary role to the auditory perception of speech under conditions of noise. The role is compensatory in that visual cues are used when auditory cues are more difficult to perceive. In summary, vision had only been regarded in either an independent or additive manner. The findings of McGurk and MacDonald (1976) questioned these assumptions and forced models of speech perception to account for percepts based on the active integration of visual and auditory stimuli.

McGurk and MacDonald (1976) found that when the auditory information for certain consonant-vowel (CV) syllables was presented simultaneously with the visual information for different CV syllables, the resulting auditory percepts consistently reflected neither of the true presentations in a pure manner. Within phoneme combinations involving /p/, /b/, /k/, and /g/, the strongest effect occurred with the auditory presentation of /ba/ and the visual presentation of /ga/. This combination resulted in the speech percept /da/, thereby suggesting an interactive relationship between hearing and seeing. Since the publication of this study, the relationship between vision and audition in the perception of speech has been referred to by various names including "The McGurk Illusion", "bimodal speech perception", and "intermodal speech perception".

In addition to revealing the interactive effect between vision and audition, McGurk and MacDonald (1976) also studied the strength of this effect across three age groups; preschool children, school age children, and adults. Table 1 presents the overall results of responses to stimuli containing conflicting auditory and visual components. Included in the experiment were conflicting presentations of /ba/ and /ga/ and conflicting presentations of /pa/ and /ka/. The results suggested that adults were more heavily influenced by visual input during speech perception than were the children of either group. Unfortunately, the authors did not propose an explanation for this trend.

Several studies, which are discussed in the following section, have examined the "McGurk Illusion" through variations of the McGurk and MacDonald (1976) study. The original study, however, did not adequately describe subject criteria and methodology. For example, the authors provided very little information regarding the subjects other than general age ranges. A specific description of subject age distribution was not presented. This is particularly disturbing for the 54 adult subjects tested. The age range (18-40) is so broad that a weighting of subjects at any one portion of the continuum may have affected the results.

TABLE 1 Overall responses for conflicting auditory and visual stimuli across age groups from McGurk & MacDonald (1976).

	Auditory only (% correct)*	Aud + Visual (% incorrect)*
preschool (3-5 yrs)	91%	59%
school age (7-8 yrs)	97%	52%
adult (18-40 yrs)	99%	92%

* correct responses were regarded as ones which reflected the auditory component of the presentation.

Second, the subjects were not distributed evenly in terms of gender. The adult population was said to be "predominantly" male, while the two younger populations had an "approximately equal" number of girls and boys. Third, the subject selection process was not addressed. For instance, the inclusion or exclusion of persons with handicaps was not discussed. In addition, and pertaining more to the overall study, the dubbing procedure for developing the stimuli was not adequately described. This information is particularly important in terms of defining the temporal parameters of auditory and visual stimuli to ensure simultaneous presentation for each experimental CV syllable. Differences in the onset of the auditory and visual components of a stimulus item would lead to several unresolved questions regarding the interpretation of the data. For example, the younger children may have been more sensitive to timing differences between the auditory and visual components, thus making them less likely to be affected by the illusion.

Finally, the procedures used to elicit responses were not described. This is of particular concern for the two younger subject groups. For example, did the examiners ascertain whether each child was attending to the monitor during the presentation of stimulus items? It is conceivable that poor visual attention or an averted eye gaze could lead to enhanced influence of the auditory signal.

Several studies within the past decade have presented variations of the initial research of McGurk and MacDonald. The above mentioned problems of confirming eye gaze and stimuli development description with the McGurk and MacDonald study have been addressed by this more current research in speech and hearing science. Other problems, however, have remained unaddressed. In addition to the findings gained from these studies, researchers from a variety of disciplines outside of speech science have completed work that is indirectly related to the McGurk illusion. The integration of information from child development, speech and hearing science, experimental psychology, and psychophysics, has led to the further development of speech science theories and to hypotheses pertaining specifically to the phenomena of intermodal speech perception.

A review of the current literature indicates that researchers appear to be accepting the role of vision as influencing the perception of speech in normal listeners.. Several studies have shown that the perception of an auditory stimulus can be altered by the presentation of a conflicting visual stimulus (Green & Miller, 1985; MacDonald & McGurk, 1978; Massaro & Cohen, 1983; Massaro, Thompson, Barron and Laren, 1986; McGurk & MacDonald, 1976; Summerfield, 1979; Summerfield & McGrath, 1984). Indeed, many authors are now using the more appropriate term "audio-visual speech perception" in reference to visual information being a functional aspect of speech perception. The central issue at present, and therefore, the direction of this research, focuses not on whether visual cues affect the auditory percept, but rather on the degree to which vision affects the perception of speech and the nature of this interaction.

Models of Speech Perception

Prior to 1976, the theories of speech perception for normal listeners operated primarily on the assumption that listeners depended on audition in situations not involving background noise. In other words, the perception of speech under normal listening conditions was thought to rely almost completely on auditory cues, with visual cues playing only a minor or nonexistent role. The first attempt to account for the significant influence of vision can be traced to the development of the manner-place theory, as proposed by MacDonald and McGurk in 1978. They hypothesized that the manner of articulation for consonants was detected primarily by the auditory modality whereas the place of articulation was detected by the visual modality. Although the theory explained some of the responses obtained for conflicting presentations, it failed to account for others. For example, the theory would predict that a conflict involving a nonlabial auditory stimulus /ga/ and labial visual stimuli /ba/ would result in the percept /ba/ since this is the place of articulation that is detected visually. In actuality, subjects most often reported the percept /b'ga/.

Summerfield (1988) provided an excellent overview of the strengths and shortcomings of the current theories of speech perception. Summerfield's discussion supported the hypothesis that visual cues strongly influenced the perception of speech. The author summarized the limitations of the current theories by outlining the primary questions which remained unanswered:

> "...the fundamental task remains of identifying the parameters of the acoustical and optical signals that are of relevance to an observer, and of characterizing them in ways that permit their mutual relationship to be registered."

In his analysis of the current theories, Summerfield presented four situations in which vision influences the auditory perception of speech and contended that an adequate theory must account for speech perception in each of these conditions. The conditions included speech perception in noise, speech perception in hearing-impaired individuals, speech perception of a larynx-frequency pulse train, and finally, speech perception with conflicting auditory and visual information. This paper is, of course, most interested in the final situation.

In 1986, Massaro and his colleagues argued that his earlier findings (Massaro,1984) provided evidence supporting the theory that speech perception occurs as a continuous analysis of components rather than a holistic or categorical process of analysis. In the 1984 study, Massaro independently varied auditory and visual information in a speech perception task. Synthetic speech sounds ranging from /ba/ to /da/ were combined with the visual presentations of /ba/ and /da/. The researchers found that /da/ perceptions increased as the auditory stimulus became more /da/-like and that the effects of the visual variable were found to increase during instances of more spectrally ambiguous auditory stimuli.

According to the continuous analysis theory, the analysis of components relies on a "fuzzy logical model of perception" through which the least ambiguous percept or source of information has the greatest impact on what is perceived. Individuals receive information regarding the degree to which certain events represent particular speech sound categories. Massaro et al. (1986) provided the example of a subject attempting to decide between the percept of /ba/ and /da/ when the auditory and visual modalities conflict. According to the model, the auditory source and the visual source would provide independent evidence with respect to the possibility of the percept of /ba/ versus /da/. In other words, visual and auditory cues typical of productions of /ba/ and /da/ would be weighted against one another. Examples of this evidence are the visual cues of closed lips for /ba/ and slightly open lips for /da/ and the auditory cues of rising second and third formants for /ba/ and slightly falling second and third formants for /da/. Different cues or types of evidence would carry different weightings in terms of the degree to which they affect the percept. According to this theory, then, the percept is determined by the phonetic alternative for which there is the most evidence.

Summerfield (1988) presented a somewhat similar position by stating that the integration of the two modalities during bimodal speech perception occurs prior to phonetic or lexical categorization of the speech segment. Summerfield cited the study by Foster (1982) who reported that variations in the synchrony of audio-visual presentations systematically alter the likelihood of different perceptual judgements. Foster maintained that this finding was better explained by a pre-categorical rather than a post-categorical integration of information from the two modalities.

Green and Miller (1985) also presented evidence supporting the view that integration of modalities precedes phonetic categorization. These authors manipulated visually specified information to demonstrate that both visual and auditory cues affect the judgement of a voiced versus voiceless consonant. The study demonstrated that manipulations of the visual rate information of syllables influenced not only the perceived rate of audio-visual syllables, but also the perception of voicing. The authors thus proposed that the integration of audio and visual information is necessitated prior to phonetic categorization.

Summerfield (1988) summarized the nature of bimodal speech perception by presenting two general principles or assumptions:

> "1) observers know the audio-visual structure of the phonemes of their language, and
>
> 2) when an audio-visual stimulus is presented, be it natural or artificial, observers perceive the phoneme, or sequence of phonemes, whose audio-visual representation is most like that which is presented."

Summerfield combined these assumptions to predict the perceptual outcome of conflicting auditory and visual information. He described the most likely resulting percept as follows:

> "...the consonant that is most easily confused auditorily with the acoustical consonant and which is most visually compatible with the visible consonant".

Conflicts arise when the visible consonant is not a member of the viseme group to which the auditory consonant belongs. Summerfield discussed three possible perceptual outcomes of such conflicts. First, the observer may recognize the conflict and simply perceive the auditory signal. MacDonald and McGurk (1978) provided evidence supporting this outcome as 83% of their subjects perceived /ga/ when presented with a visual /ba/ and an auditory /ga/. Second, observers may recognize the conflict but yet continue to be influenced by the visual signal such that they "hear" a consonant different than what is auditorily presented. Sixty-four percent of the subjects in the MacDonald and McGurk study illustrated this outcome when they perceived /da/ in the presence of visual /ga/ and auditory /ba/. Finally, an observer may deal with a conflict through an auditory percept involving both the visual and the auditory components of the presentation. MacDonald and McGurk demonstrated this outcome as 17% of their subjects perceived /bga/ when presented with visual /ba/ and auditory /ga/.

As demonstrated through the recent research of such investigators as Massaro, Summerfield, and Green, speech scientists are beginning to formulate a theory of speech perception that attempts to account for the phenomena of the "McGurk illusion". When complete, this theory will most probably involve the integration of modality information prior to phonetic categorization and will address the relative weightings of various visual and auditory features in contributing to the "best fit" solution. The developmental studies done to date indicate that the weighting of cues will most likely vary across subject age groups. Research endeavors could therefore logically begin with examinations and clear descriptions of any existing developmental trends. Subject groups should be well defined in terms of variables such as language exposure, hearing sensitivity, and exclusion of those with handicapping conditions. Subject groups must also be more narrowly divided into age ranges to allow for adequate examination of developmental trends.

Infant Studies

Infant studies play an important role in the endeavor to discover the origin and nature of intermodal perception. As will be illustrated, they can also serve to support or refute intermodal perception theories developed around the findings regarding other age groups.

Two major theories have dominated infant intermodal perception literature. The traditional integration theory maintains that infants gradually develop the ability to integrate information from different modalities through the use of association cues. In contrast, the invariant-detection theory argues that the integration ability is present at birth and endows the infant with the capacity to discriminate between matching and conflicting inputs. The primary point of contention between these theories is also shared by the debates regarding many aspects of language, as it focuses on the innateness of ability. Unfortunately, intermodal speech perception in infants is rather poorly understood as only a handful of studies have examined the skills of infants in this area. Several of these studies have focused on the interactions between modalities other than those between vision and audition.

Bahrick and Watson (1985) examined the perception of proprioceptive-visual contingencies in five-month old infants. Measuring preferential fixation, the authors determined that the five month old infants were able to discriminate between a contingent and non-contingent display of their own live leg movement. This discrimination ability, however, was not found in three month old infants as they did not show a significant degree of preferential fixation to the contingent display.

Bahrick (1988) extended this finding by attempting to teach three-month old infants to integrate two modalities. Although speech was not used, this study did examine the integration of auditory and visual signals. "The intermodal learning method" used by Bahrick, consisted of familiarizing each infant to one of four stimulus conditions in which the appropriateness, in terms of matching auditory and visual signals and synchrony of the auditory signal were varied with respect to the visual display. Infants were then tested for their intermodal perceptual abilities during instances in which the auditory signal matched only one of the visual displays on a split screen. Abilities were judged on the basis of preferential fixation. "Intermodal learning" was evidenced only by the group of infants who had been familiarized with the auditory signals which were both appropriate and synchronous to the visual signal. Preferential fixation to appropriate auditory and visual matches was not found to be significant in groups that were familiarized in one of the other three conditions . Bahrick suggested her findings supported the invariant-detection theory of infant intermodal perception.

Walker-Andrews (1986) studied the integration of visual facial displays and vocalizations. Through the use of visual displays of the upper two-thirds of angry or happy facial expressions and auditory vocal expressions of anger or happiness, seven-month old infants were found to fixate significantly more on facial expressions which matched the presented auditory signal. This ability to discriminate matching versus conflicting signals was not found to be significant in five-month old infants.

In a study which directly addressed intermodal speech perception, Dodd (1979) presented three and four-month old infants with matching or conflicting visual and auditory signals during the reading of nursery rhymes. Dodd found that the infants attended less to visual displays which were asynchronous with the presented auditory speech signal than they did to synchronous displays. This finding argues that young infants have the ability to process both visual and auditory cues during a speech perception task even though the speech may not be meaningful to them.

Kuhl and Meltzoff (1982) provided a similar but more detailed study regarding infant intermodal speech perception by manipulating the synchrony of audio and visual vowel presentations. Using the split-screen method common to the studies discussed above, Kuhl and Meltzoff found that 18 - 20 week old infants fixated significantly more to corresponding rather than noncorresponding auditory and visual speech signals. The authors then conducted a second experiment (Kuhl and Meltzoff, 1984) to examine the relative importance of the temporal versus spectral characteristics of the auditory signal. They removed the spectral characteristics of formant frequencies and used auditory signals containing only the temporal characteristics of amplitude and duration. The infant subjects did not show significant preferential fixation during the presentation of these stimuli. Kuhl and Meltzoff concluded that some type of spectral information is necessary for successful intermodal speech perception.

Many issues regarding infant intermodal speech perception remain unresolved. The controversy regarding the innateness of such skills is particularly intriguing. The studies discussed above show that young infants demonstrate the ability of intermodal speech perception. As will be discussed in the next section, this finding seems in conflict with theories suggesting that young children are less affected by bimodal illusions due to poor lip reading capabilities.

Developmental Trends

As discussed earlier and as evidenced by the McGurk illusion, adults can be so affected by visual information that their auditory percept of a speech signal is actually altered. Based on the consistency with which infants have been shown to attend to and utilize lip movement information, it could be logically hypothesized that even young children would have mastered the task of intermodal speech perception. Several studies, however, have shown that children are significantly less affected by the McGurk illusion than are adults.

McGurk and MacDonald (1976) studied the McGurk illusion across three age groups: preschool children, school age children, and an adult group. They found that while the illusion affected the adult group in 92% of the trials, the effect was present in only 59% of trials for the preschool group and 52% for the school age group. Although the authors did not attempt to explain this finding, poor attention to the visual screen may have been a factor. Based on the public description of their procedures, McGurk and MacDonald did not seem to have controlled for their subjects' attention to the visual signal. Poor attention to this signal would have reduced the effect of the illusion by causing the auditory signal to be the dominant percept.

Massaro, Thompson, Barron and Laren (1986) conducted a study quite similar to that of McGurk and MacDonald (1976). These authors, however, did attempt to control for attentional factors by monitoring the gaze of young subjects during testing. The results of the Massaro et al. study supported those of McGurk and MacDonald as they found children to be less influenced by the presence of conflicting auditory and visual information. Massaro et al. contended that this difference between adults and children was due to stronger lip-reading abilities of adults.

Massaro et al. (1986) did not, however, report any data addressing whether the differences in the strength of the illusion were a result of inter- or intra-subject variation. In terms of establishing and describing a developmental trend, it would be essential to examine whether some of the children were being consistently affected while some were not at all affected or whether all of the children were only partially affected. If the results indicated the prior condition, it would be informative to separate and possibly correlate subject descriptors to those being consistently affected and to those not being affected.

Massaro and his colleagues have consistently argued that the developmental trends of the McGurk illusion can be explained by poorer lip reading skills in young children. This explanation, however, is in conflict with the findings of the infants studies discussed in the previous section. The results of Bahrick (1988) and Dodd (1979), in particular, respectively show infants to be quite capable of intermodal perception and recognition of lip reading cues. It thus becomes questionable that children between the ages of 3 and 5 are less affected by the McGurk illusion due to poorer lip reading skills.

None of the studies discussed have addressed the possibility that the apparent developmental trends are related to physiological or cognitive capabilities for modality integration across subjects. In other words, could the decreased effect of the McGurk illusion on younger subjects be a reflection of physiological immaturity in terms of the ability to integrate information from the visual and auditory modality? Due to difficulties in controlling for environmental factors and in separating physiology and cognition, this possibility becomes very hard to test. Such explanations should, however, receive ongoing consideration as new findings emerge.

In addition, the literature does not report findings regarding the effects of the McGurk illusion in the senior adult population. The Working Group on Speech Understanding and Aging (1988) describe in detail how the process of aging is accompanied by progressive neural degeneration. In terms of the auditory mechanism, they explain that this degeneration results in decreased auditory processing capabilities which make the auditory speech signal more difficult to perceive accurately. As such, it may be possible that, even in older persons with essentially normal visual acuity and hearing sensitivity, the visual signal could be relied upon significantly enough to result in strong effects of the McGurk illusion.

An alternate hypothesis might be one which focuses on the physiological ability to integrate two modalities. Could this ability be adversely affected by the beginning of neurological degeneration? A hypothesis based on this premise may predict that the effect of the McGurk illusion would decrease with age as a result of decreased processing capabilities.

Further discussion of these intriguing questions, however, must wait for the in-depth examination and possible establishment of a developmental trend which spans the age ranges. The importance of such developmental information to the understanding of the McGurk illusion and of the speech perception process overall was clearly stated by Massaro (1988) as follows:

> "The framework guiding our research is also ideal for the study of developmental changes in processing audible and visible speech. Any theory of language processing must eventually confront the acquisition of the processes involved in this skill. Thus, developmental studies are central to evaluating theories of the processes responsible for observed differences and similarities in language processing with age."

<u>Contributing Variables</u>

Since publication of the classic study by McGurk and MacDonald (1976), certain combinations of auditory and visual cues have been found to create a stronger illusion than those of other combinations. The strongest illusion has consistently been found to occur with the combination of the visual cues of /ga/ and the auditory cues of /ba/. The effects of the illusion when employing vowels other than /a/ have not been presented in the literature.

The rate of speech has also been shown to affect intermodal speech perception. Green (1987) demonstrated that listeners utilize cues of speaking rate from both the auditory and visual modalities during the perception of speech. In addition, Kuhl, Green, and Meltzoff (1988) suggested that the level of the auditory signal affected the degree to which visual information was utilized during speech perception. Kuhl et al. specifically varied the level of the auditory signal during presentations of visual /ga/ and auditory /ba/. Their findings were somewhat unexpected in that, as the auditory signal of /ba/ became louder, the percept of /ba/ actually decreased. In other words, the increased auditory signal resulted in an increase of the influence of the visual signal. Kuhl et al. tentatively hypothesized that louder speech may automatically cause listeners to attend more to the visual signal. This hypothesis remains to be answered. The finding, however, clearly points to a need for careful consideration and documentation of intensity levels during experimentation with the McGurk effect.

<u>Research Question</u>

The McGurk illusion has been consistently replicated and shown to be fairly robust in the adult population. The discovery of this illusion has added a previously unknown dimension to the study of speech perception. In order to better understand and account for the phenomena, its developmental nature, its variations across populations, and the variations across phonetic presentations must be studied further.

The purpose of the present study is to replicate and extend the pioneer study by McGurk and MacDonald (1976). In particular, this study seeks to examine the developmental trend of the strength of the McGurk illusion across four age groups and along two stimulus variables. The age groups will represent four points along the developmental continuum: preschoolers (3-5 yrs), junior high school students (11-13 yrs), adults (21-25 yrs), and seniors (55-65 yrs). The stimuli will be composed of the consonants /b/ and /g/ varied with the vowels /i/ and /a/.

The hypothesized is that the strength of the McGurk illusion, and therefore, the capabilities of bimodal speech perception, depend partially on the processing and integrative capacities of the individual. Supportive results then, would show the strength of the illusion to increase with age up to the adult group. As the effect of the McGurk illusion has not been documented in normal hearing groups of senior adults, this study poses the research question of examing the any such effect and does not present a hypothesis regarding the outcome of this age group.

In addition, this study aimed to more clearly define the distribution of responses across subjects. For example, if the illusion is present in 60% of the responses of the 3-5 year group, the results will provide information regarding whether "correct" responses were evenly distributed across children or whether the illusion was consistently strong in approximately 60% of children and consistently weak in approximately 40% of children.

Finally, this study examined the possible differential effects of varying the vowel component of experimental CV syllables. Although previous research has examined differences in the illusion due to varying consonant components (McGurk & MacDonald,1976; MacDonald & McGurk,1978; Summerfield,1979), the effect of changing the vowel context has not been adequately addressed. This study focussed on differences between presentations involving the traditionally used vowel /a/ and the less frequently used vowel /i/. The presentations of auditory /b/ and visual /g/ together with /i/ or /a/ were regarded as the experimental conditions since this consonant combination has been identified in the literature to be the strongest exemplar of the illusion (MacDonald and McGurk, 1978, Summerfield, 1979).

CHAPTER II: METHODS

Subjects

Four age groups comprised of fifteen subjects each were used to test for a developmental trend. The youngest group consisted of fifteen local preschool children ranging in age from 3 to 5 years old. The second subject group was made up of fifteen 11 to 13 year old children. Both of these younger groups were recruited either from local summer recreation programs at the university or from the university community. The third subject group was comprised of fifteen young adults ranging in age from 21 to 25 years. These subjects were recruited from among the university student population. The fourth subject group consisted of fifteen senior adults between the ages of 55 and 65. These individuals were recruited from churches or from families among the university community. Sixty of sixty-three individuals completed the study. Two persons were disqualified due to inattentiveness to the monitor, and one person was disqualified for failure to respond correctly to the control items. This left a total of sixty subjects.

All subjects had normal hearing sensitivity (thresholds no greater than 25 dB HL across 1000 to 4000 Hz), and were judged by the examiner to have demonstrated the ability to clearly articulate and discriminate the phonemes /b/, /d/, and /g/. All subjects were reported to have normal vision, although some subjects required corrective lenses. This information was gathered from parents for the two younger groups and from the subjects themselves for the two adult groups. Persons with handicapping conditions were excluded from subject groups as were those for whom normal language exposure or cognitive status were judged to be questionable by the examiners. These judgements were made on the basis of informal case history information regarding language environment and current eduational/employment status. Only native English speakers were included in the study. A description of the subjects' age and gender characteristics is presented in Appendix A.

Stimuli

The stimuli consisted of audio-visual VCV syllables which were created by the pairing of auditory and visual tokens in four possible combinations for each of two vowels used. There were a total of 40 stimulus items which were presented by a female speaker. Twenty presentations involved the vowel /a/ and the remaining 20 involved the vowel /i/. Each vowel was presented in the context of either the consonant /b/ or /g/. See Table 2 for the combinations used for each vowel.

Each stimulus combination was presented five times during each experimental trial, resulting in a total of 40 presentations to each subject. The presentations were randomized within each 20 stimulus item set involving one vowel. The forty stimulus items were presented at 15 second intervals. A blank screen with no auditory signal was presented during the intervals between stimulus presentations.

A training sequence was utilized prior to the experimental set for each subject. The presentation procedure for the four training items was consistent with that of the experimental trials. The training sets served to test the subjects' understanding of the task and their accurate perception of items involving nonconflicting auditory and visual components. Persons who failed to respond correctly to all four training items were excluded from the study. Table 3 presents the items used in the training sequence.

TABLE 2. Stimulus items used during experimental trials.

A. Vowel Context /a/

	Auditory Stimuli	Visual Stimuli
1.	/aba/	/aba/
2.	/aga/	/aga/
3.	/aba/	/aga/
4.	/aga/	/aba/

B. Vowel Context /i/

	Auditory Stimuli	Visual Stimuli
1.	/ibi/	/ibi/
2.	/igi/	/igi/
3.	/ibi/	/igi/
4.	/igi/	/ibi/

TABLE 3. Items included in the training sequence.

Auditory Component	Visual Component
/aba/	/aba/
/aga/	/aga/
/ibi/	/ibi/
/igi/	/igi/

The reliability of the videotapes was confirmed through a pilot study involving eight subjects ranging from 24 to 36 years of age. These subjects responded to presentations involving consistent auditory and visual informations with 100% accuracy. The illusion (i.e. the percept of /d/ in the presence of auditory /b/ and visual /g/) was effective in 97.5% of occurrences.

Instrumentation

Video tapes were prepared in the Department of Radio and Television at the University of Montana. The stimuli were created by videotaping a female speaker who produced the VCV syllables while directly facing the video camera. Videotaping was completed with a Sony Video-Cassette Recorder VD-5800. Conflicting audio and visual stimuli were dubbed using a Toscam M-1B Line Mixer and a Sony RM440 Automatic Editing Control Unit.

The stimuli were presented using one of the following combinations: a Panasonic CTJ-2053R-1 19" monitor and a Hitachi VT-TU98A HiFi Video Cassette Recorder or a Sharp Linytron 20" monitor and a Mitsubishi HS-413UR HiFi Video Cassette Recorder. The auditory stimuli were presented at a level peaking at 60 dB SPL. This intensity level was initially set using a Bruel and Kjaer Type 2203 Precision Sound Level Meter with a Bruel and Kjaer Type 4144 one inch microphone. The level was then monitored by either a 1565-A Survey Sound Level Meter or a Bruel and Kjaer Type 2215 Precision Sound Level Meter on a daily basis.

Subjects were video taped using an RCA Solid State Image Sensor VHS Camcorder or a Sanyo VC 1600x video camera switched through a closed-circuit television system to a Panasonic PV-2700 Video Tape Recorder and a Sanyo Model 4155 black and white television monitor.

The hearing screenings were administered through the use of one of two Maico MA40 portable audiometers with TDH-498 earphones and MX-41/AR cushions calibrated to ANSI (1969) standards for pure tone audiometers.

PROCEDURES

During testing, all subjects were seated in a quiet room at a table facing a TV monitor. The subjects were approximately four feet from the monitor which was placed at approximately eye level.

All subjects were instructed to report to the examiner "what the woman on the videotape said" for each trial. The oral responses of all subjects were recorded on blank data sheets (an example data sheet is presented in Appendix B) following each stimuli presentation. The responses were recorded using the International Phonetic Alphabet. No other information regarding the study was provided to the subjects.

The responses of the preschool children were reinforced with tokens and a verbal "good" presented immediately after every third trial. The children then traded the tokens for a toy prize at the end of their participation. The subjects in the three older groups were verbally reinforced with "good" after every third trial.

A video camera, as described in instrumentation, was positioned to the side of the television monitor and was used to videotape all subjects during the presentations. The results of any persons who were found to be inattentive to the screen during a presentation trial were excluded from the study. The videotapes were also used to assess the reliability of the examiner's perception and recording of subject responses. Intra-examiner reliability was assessed on 100% of the subject data. Inter-examiner reliability was checked on a point by point basis on 25% of the subjects.

All subject data was entered into a computer program which compiled both individual subject and group totals for the various response types within each of the four conditions for the two vowels (see Appendix C). Responses were scored as either correct or incorrect depending on whether or not they represented the auditory presentation of the trial. All responses which did not purely represent the auditory component of the stimuli were counted as incorrect. All data was also checked manually for entry or computational errors.

CHAPTER III: RESULTS

A 4 x 2 x 4 analysis of variance (ANOVA) with two repeated measures was performed on the compiled data for each group. The statistical analyses was perforned on an IBM-AT based microcomputer system using a GB-STAT (Version 1.5) statistics software package. Responses within the experimental groups which were incorrect but did not conform to the type of responses typical to past descriptions of the McGurk illusion were grouped for descriptive presentation. Descriptive analyses were also completed on the responses with respect to gender and vowel differences within age groups.

The results from the 4 x 2 x 4 ANOVA are presented in Table 4. The ANOVA results indicated significant differences across all three main effects (age, vowel environment, auditory-visual condition) and significant interactions between all main effects. The mean number of correct responses and the corresponding standard deviations are presented in Table 5 for all age groups, vowel environments, and auditory-visual conditions. Histograms displaying the percentages of correct responses to the control conditions are illustrated in Figures 1 and 2. The three older age groups showed high (>97%) accuracy across control conditions. Scores for the youngest group were slightly lower, and should thus be considered in the interpretation of their experimental trial results.

The histograms presented in Figures 3 through 6 display the mean number of correct and incorrect responses across the four age groups. The histograms presented in Figures 3 and 4 illustrate the distribution of correct and incorrect results for the presentation of auditory /b/ and visual /g/ across vowels. This combination of auditory /b/ and visual /g/ with either vowel was referred to as the "forward illusion". It is evident that while

Table 4: Three way Analysis of Variance Summary Results

Source	SS	DF	MS	F-Ratio	Probability
Age	52.84177	3	17.61392	8.1463	0.0001
Sub W. Grp	121.0832	56	2.1622		
Vowel	11.4083	1	11.4083	11.8564	0.0011
Age X Vowel	8.70813	3	2.90271	3.0167	0.0374
Vowel X Sub W. Grp	53.88343	56	0.9622		
Condition	1185.492	3	395.1638	197.0446	<0.0001
Age X Condition	94.59148	9	10.51017	5.2408	<0.0001
Condition X Sub W. Grp	336.9163	168	2.00545		
Vowel X Condition	21.09176	3	7.03059	9.9998	<0.0001
Age X Vowel X Cond.	32.79199	9	3.64355	5.1823	<0.0001
Total	2036.925	479			

Table 5. Mean Number of Correct Responses and Standard Deviations Across Age Groups.

Auditory	Visual	Age Group	Mean	SD
/gi/	/gi/	3-5	4.5333	0.9155
		11-13	4.8667	0.3519
		21-25	5.0000	0.0000
		55-65	4.8667	0.3519
/ga/	/ga/	3-5	4.9333	0.2582
		11-13	5.0000	0.0000
		21-25	4.9333	0.2582
		55-65	4.8667	0.3519
/bi/	/bi/	3-5	4.4667	0.7432
		11-13	5.0000	0.0000
		21-25	5.0000	0.0000
		55-65	4.9333	0.2582
/bi/	/gi/	3-5	2.6000	1.7647
		11-13	2.7333	2.0517
		21-25	1.0667	1.5337
		55-65	0.2000	0.4140
/ba/	/ga/	3-5	0.9333	1.3345
		11-13	0.8667	0.9155
		21-25	0.6000	1.2421
		55-65	0.2667	0.4577

Table 5 (cont'd):

/gi/	/bi/	3-5	2.8667	1.6847
		11-13	3.4667	1.8848
		21-25	3.3333	2.1931
		55-65	1.3333	1.6330
/ga/	/ba/	3-5	3.8667	1.4573
		11-13	2.2000	2.1112
		21-25	2.6667	2.1269
		55-65	1.0000	1.6903

Figure 1. Percentage of Correct Responses to Auditory /b/ and Visual /b/.

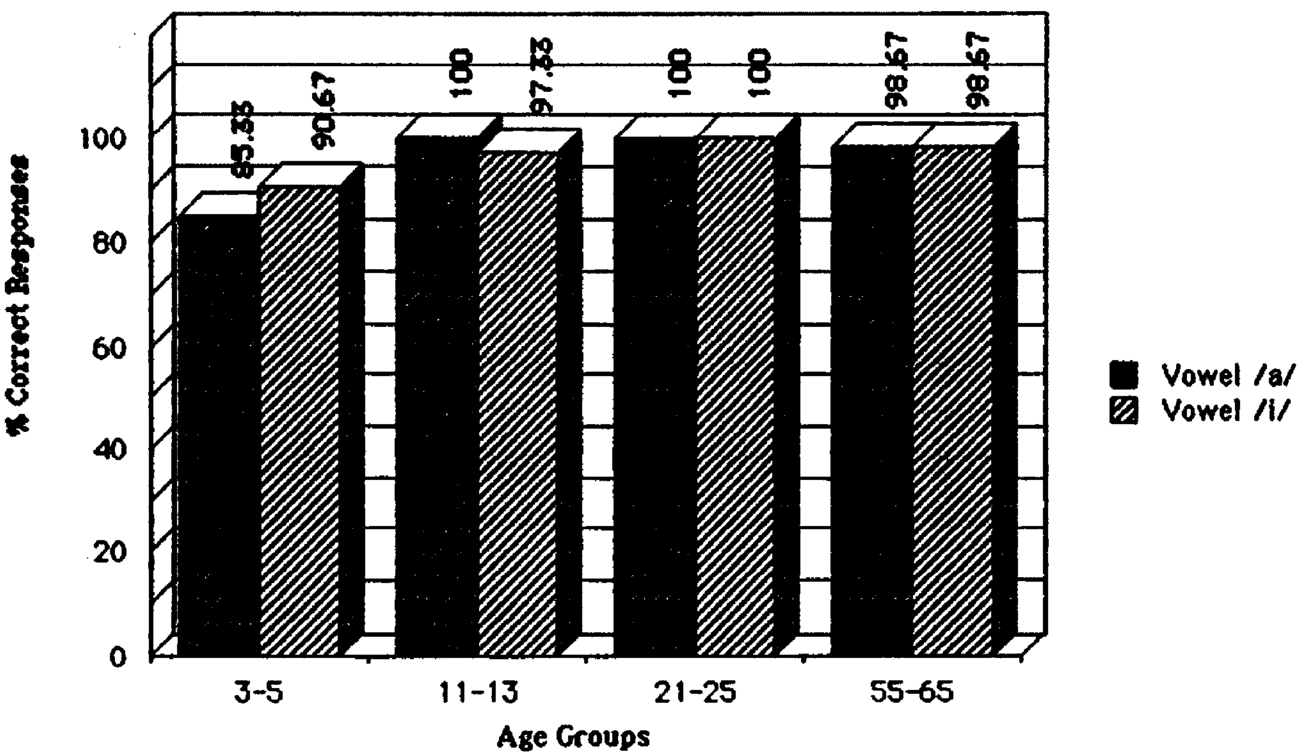


Figure 2. Percentage of Correct Responses to Auditory /g/ and Visual /g/.

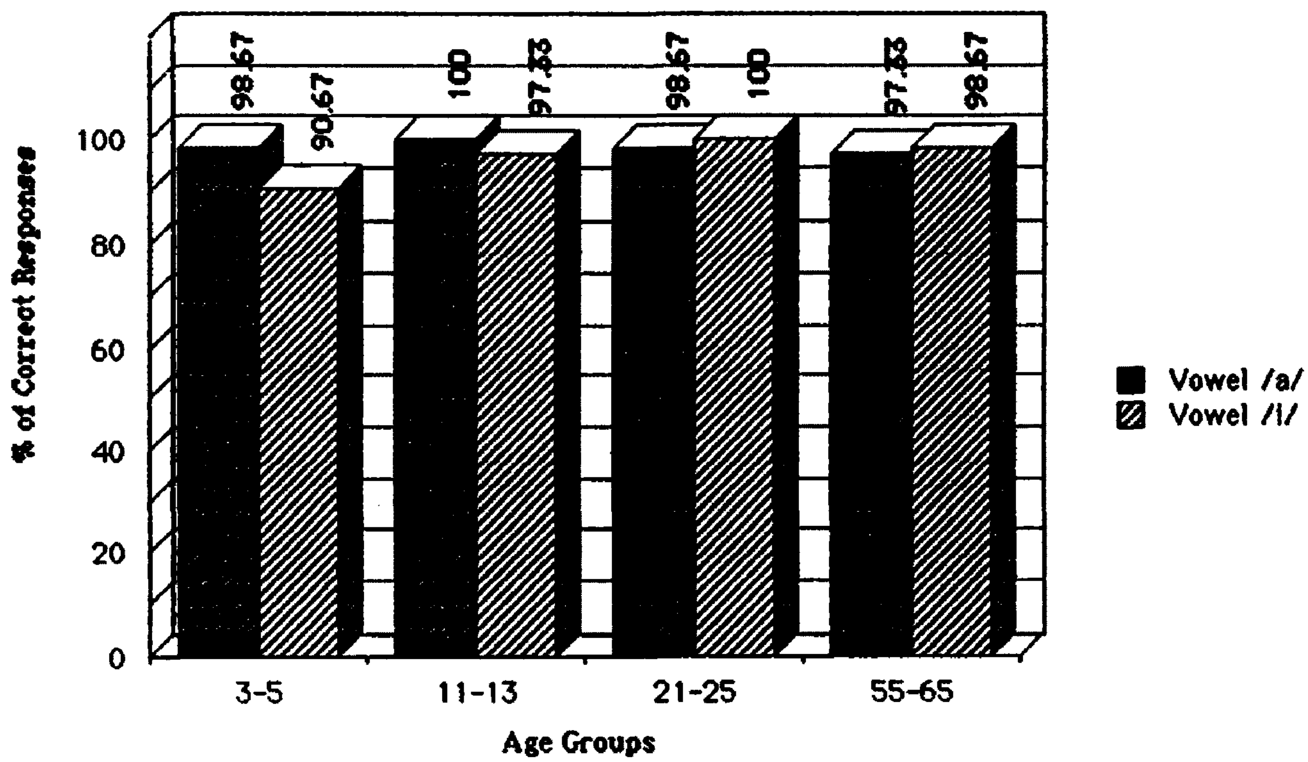


Figure 3. Mean Number of Incorrect Responses for Auditory /ba/ and Visual /ga/.

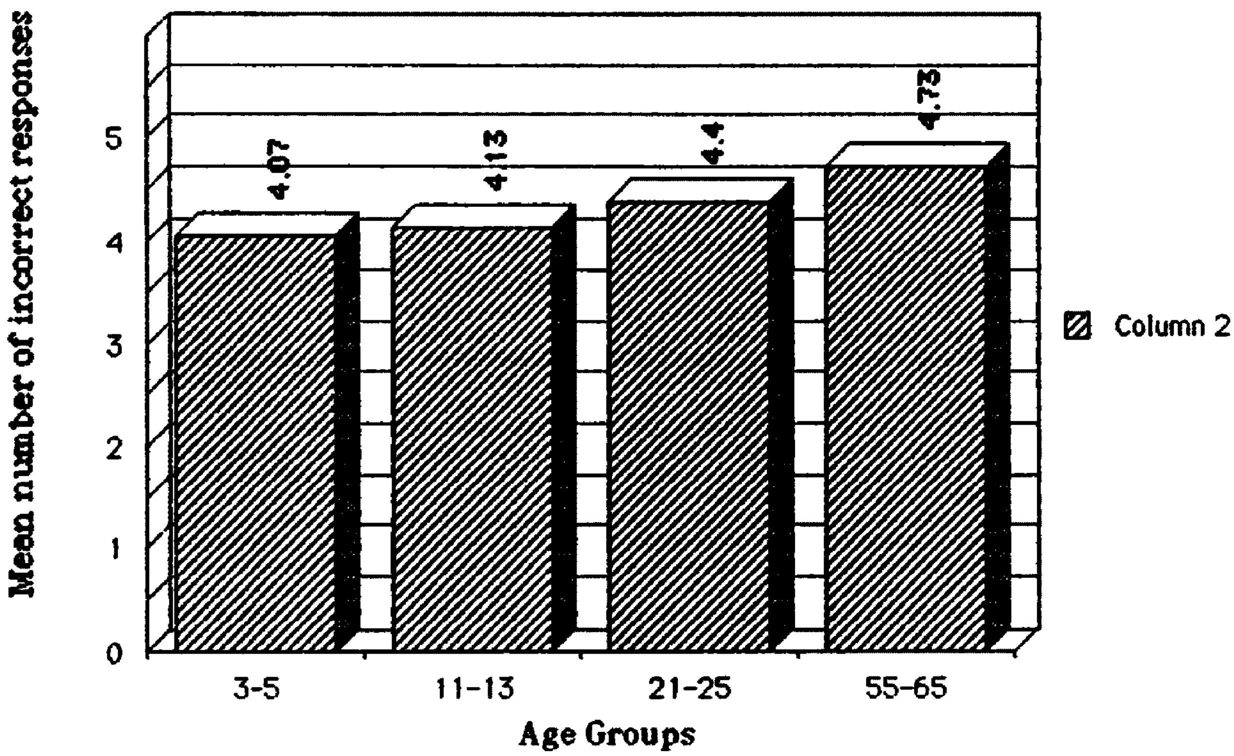


Figure 4. Mean Number of Incorrect Responses for Auditory /bi/ and Visual /gi/

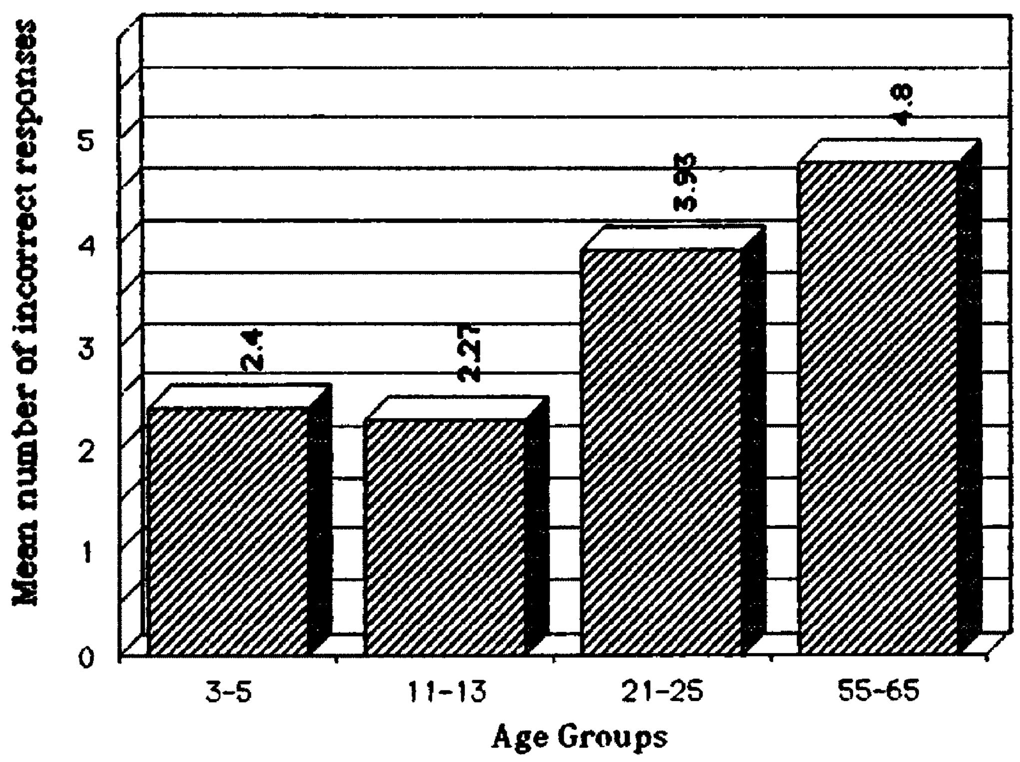


Figure 5. Mean Number of Incorrect Responses for Auditory /ga/ and Visual /ba/

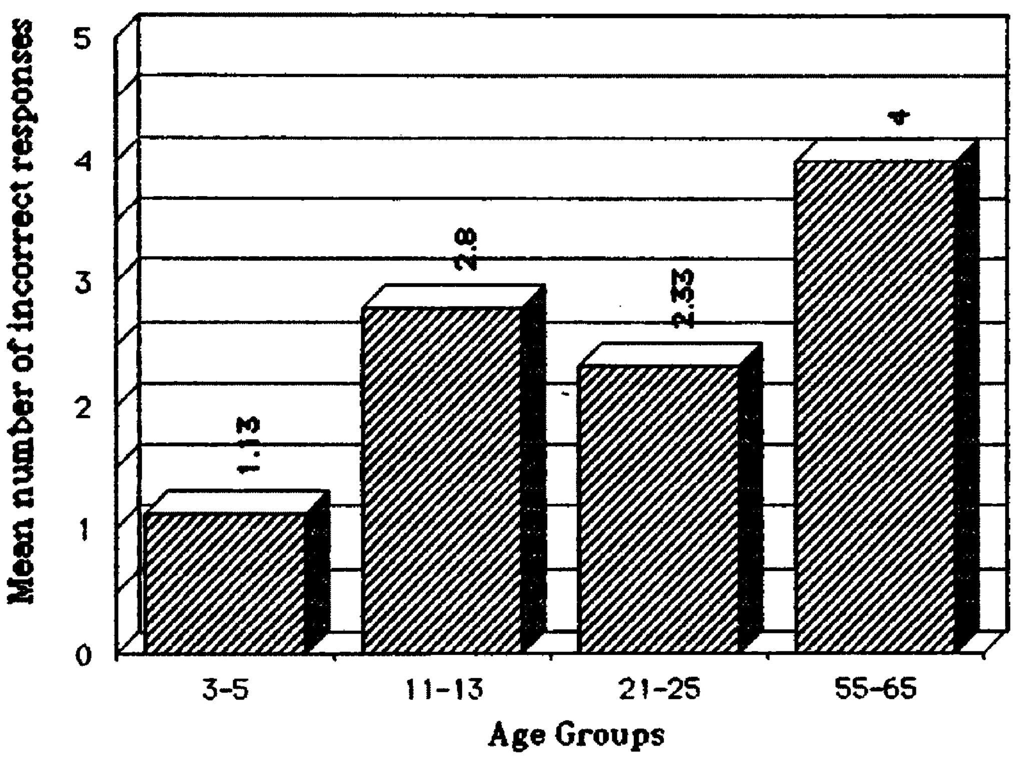
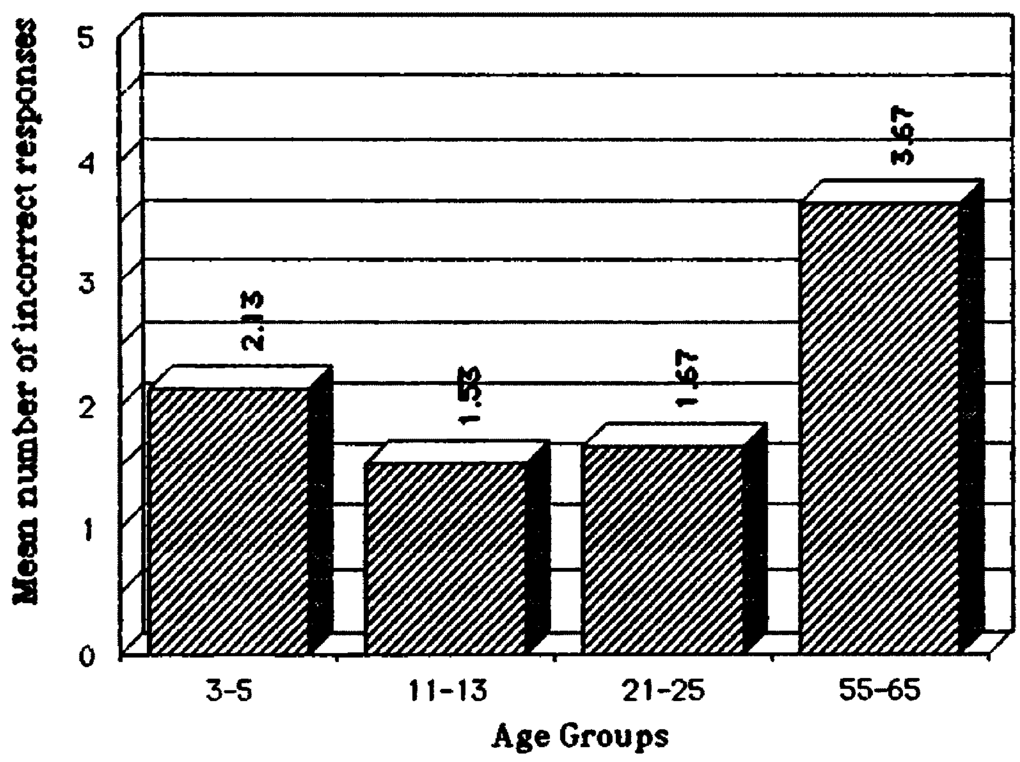


Figure 6. Mean Number of Incorrect Responses for Auditory /gi/ and Visual /bi/.



auditory /ba/ and visual /ga/ produced little variability, auditory /bi/ and visual /gi/ showed greater differences across age groups. Figures 5 and 6 illustrate the distribution of correct and incorrect responses for the presentations involving auditory /g/ and visual /b/ across vowels. The combination of auditory /g/ and visual /b/ with either vowel was referred to as the "backward illusion". This combination revealed variability across both vowels.

In regards to the main research question investigating the developmental trend in the stregth of the McGurk illusion, the results of the ANOVA indicated a significant difference ($p < 0.05$) across age groups as a main effect. Several types of responses were recorded for each of the four experimental conditions. Figures 7 through 22 display a breakdown of response types across age groups for both the forward and backward illusion for each vowel type. "Auditory responses" and "visual responses" are those which purely represented a single component of the experimental stimulus. The number of responses consisting of the phoneme /d/ are separated, as are those consisting of the phoneme /ð/. "Combination" responses, for example /ibgi/, are those consisting of both the consonant phonemes represented by the auditory and visual presentation. Finally, the "Other" category includes responses which could not be included in any of the predominant categories.

It is of note that the responses from the 3-5 year old age group for the forward /a/ illusion were characterized by a tendency to provide only the visual component of the experimental condition. The perception of /ga/ being reported for 49.33% of the trials involving this experimental condition was a pattern quite different than those obtained for the other age groups.

Figure 7.

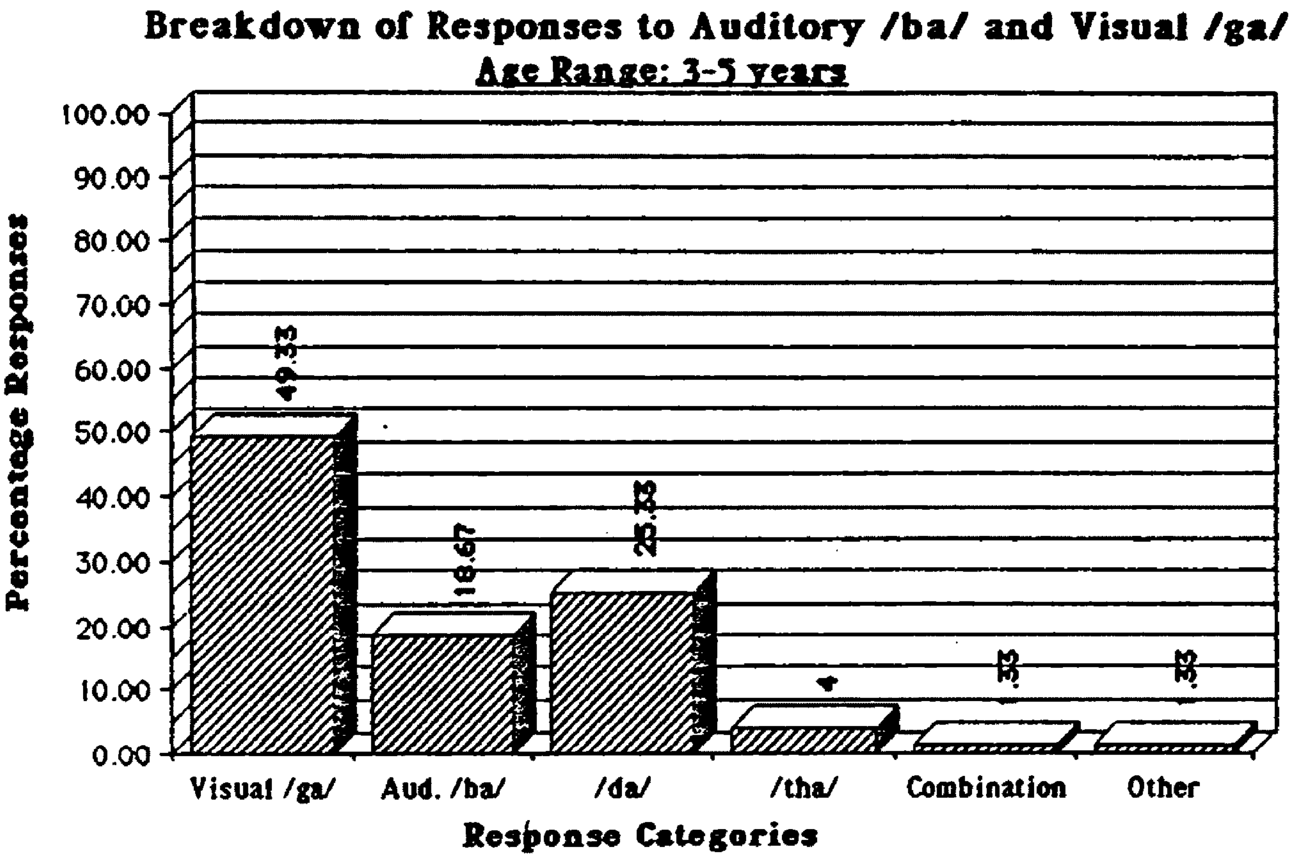

Breakdown of Responses to Auditory /ba/ and Visual /ga/
Age Range: 3-5 years
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Percentage Responses
49.33
18.67
25.33
4
.33
.33
Visual /ga/
Aud. /ba/
/da/
/tha/
Combination
Other
Response Categories

Figure 8.

Breakdown of Responses to Auditory /ba/ and Visual /ga/

Age Range: 11-13 years

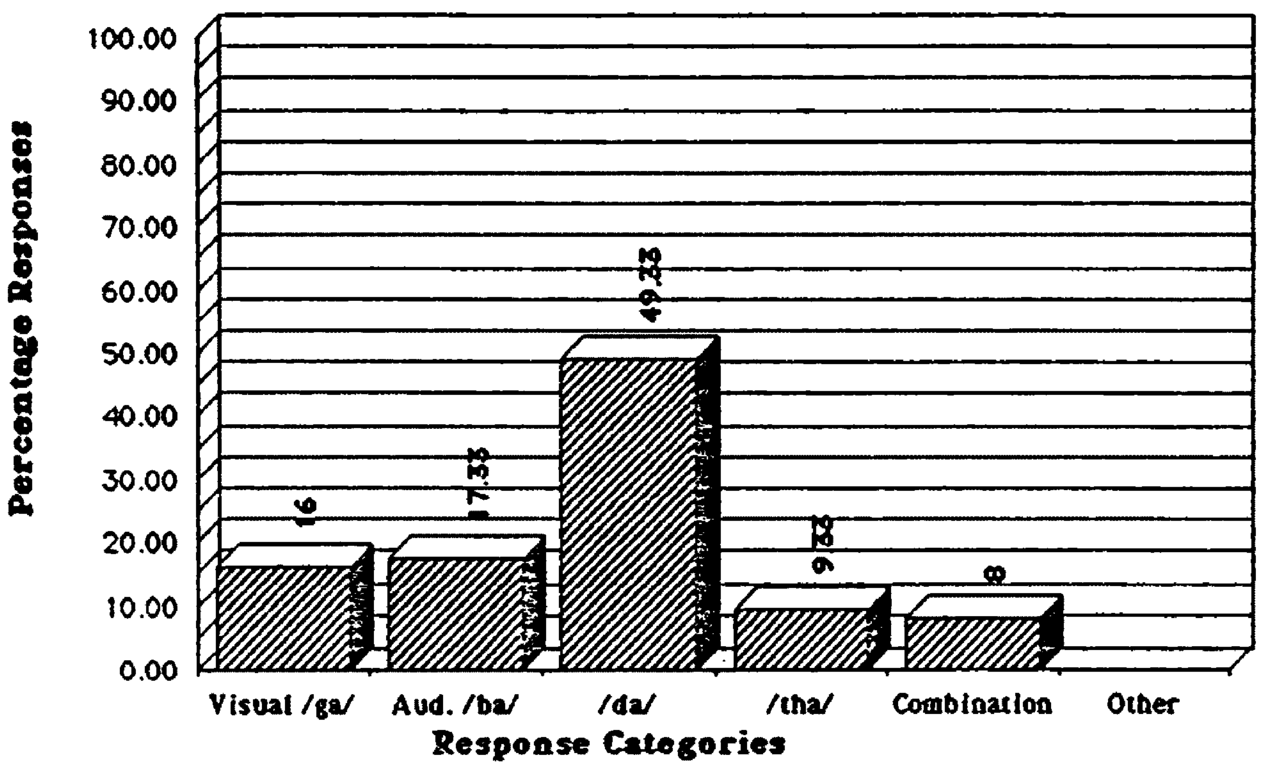


Figure 9.

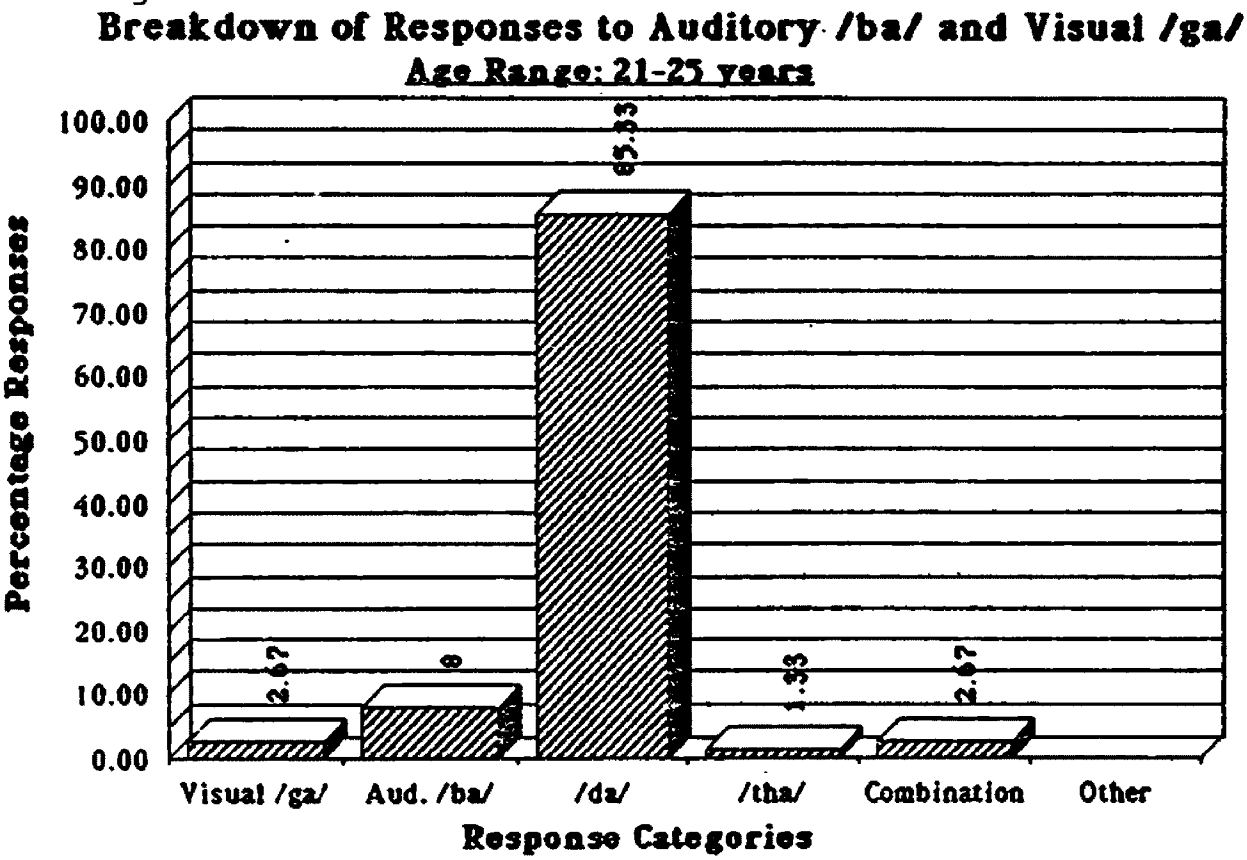

Breakdown of Responses to Auditory /ba/ and Visual /ga/
Age Range: 21-25 years
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Percentage Responses
2.67
8
85.33
1.33
2.67
Visual /ga/
Aud. /ba/
/da/
/tha/
Combination
Other
Response Categories

Figure 10.

Breakdown of Responses to Auditory /ba/ and Visual /ga/

Age Range: 55-65 years

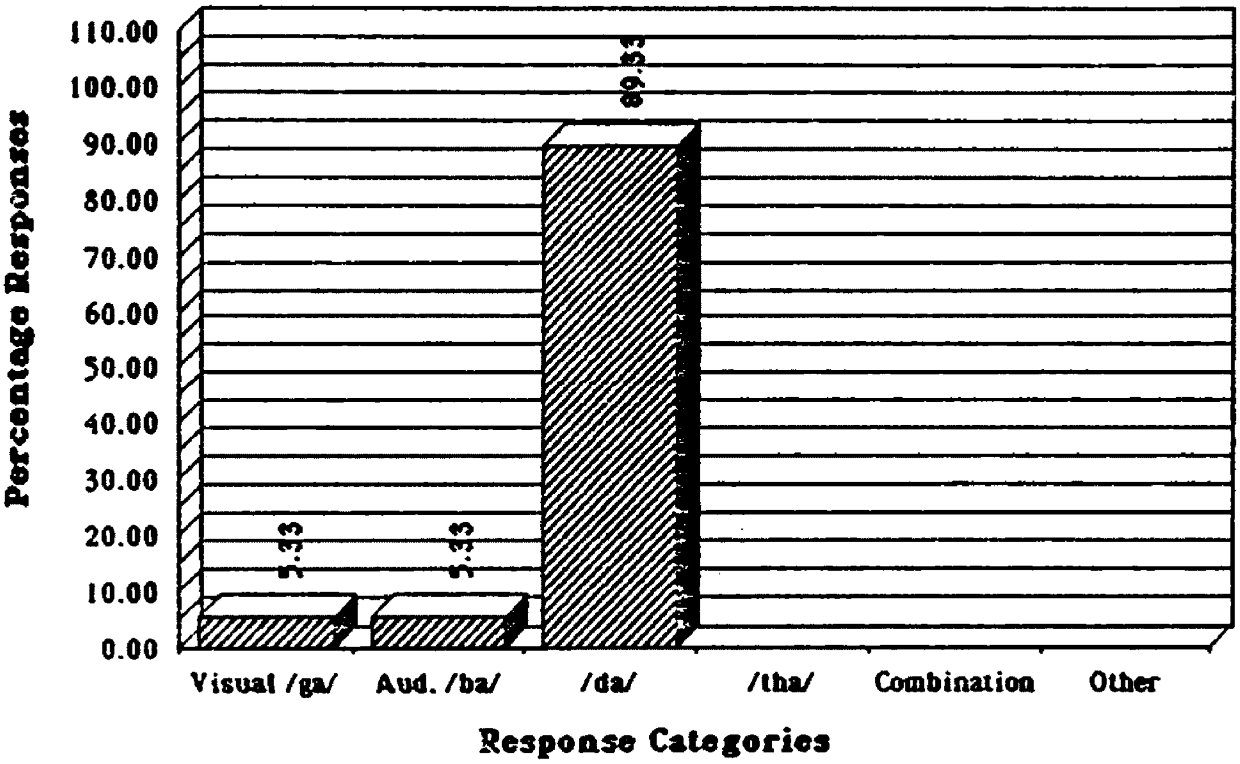


Figure 11.

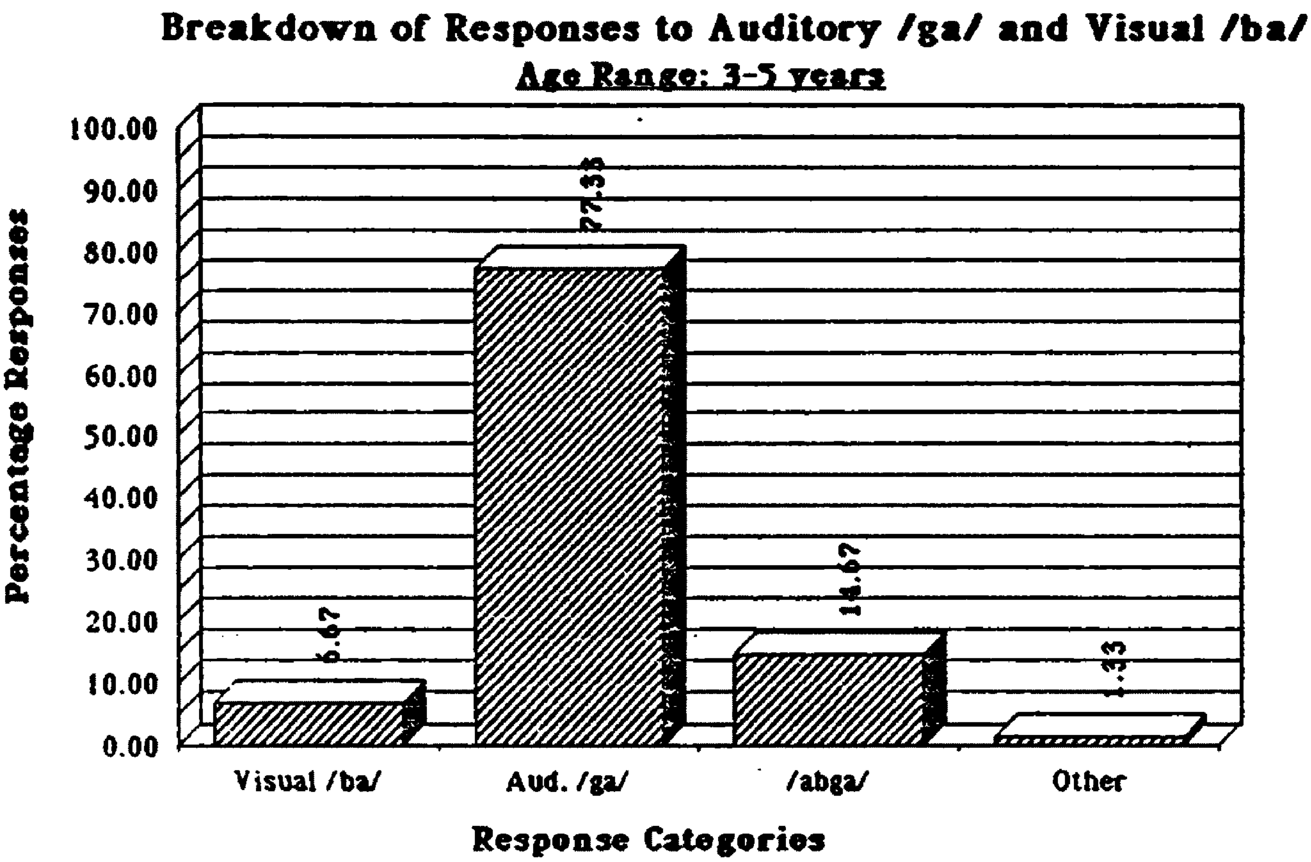

Breakdown of Responses to Auditory /ga/ and Visual /ba/
Age Range: 3-5 years
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Percentage Responses
6.67
77.33
14.67
1.33
Visual /ba/
Aud. /ga/
/abga/
Other
Response Categories

Figure 12.

Breakdown of Responses to Auditory /ga/ and Visual /ba/

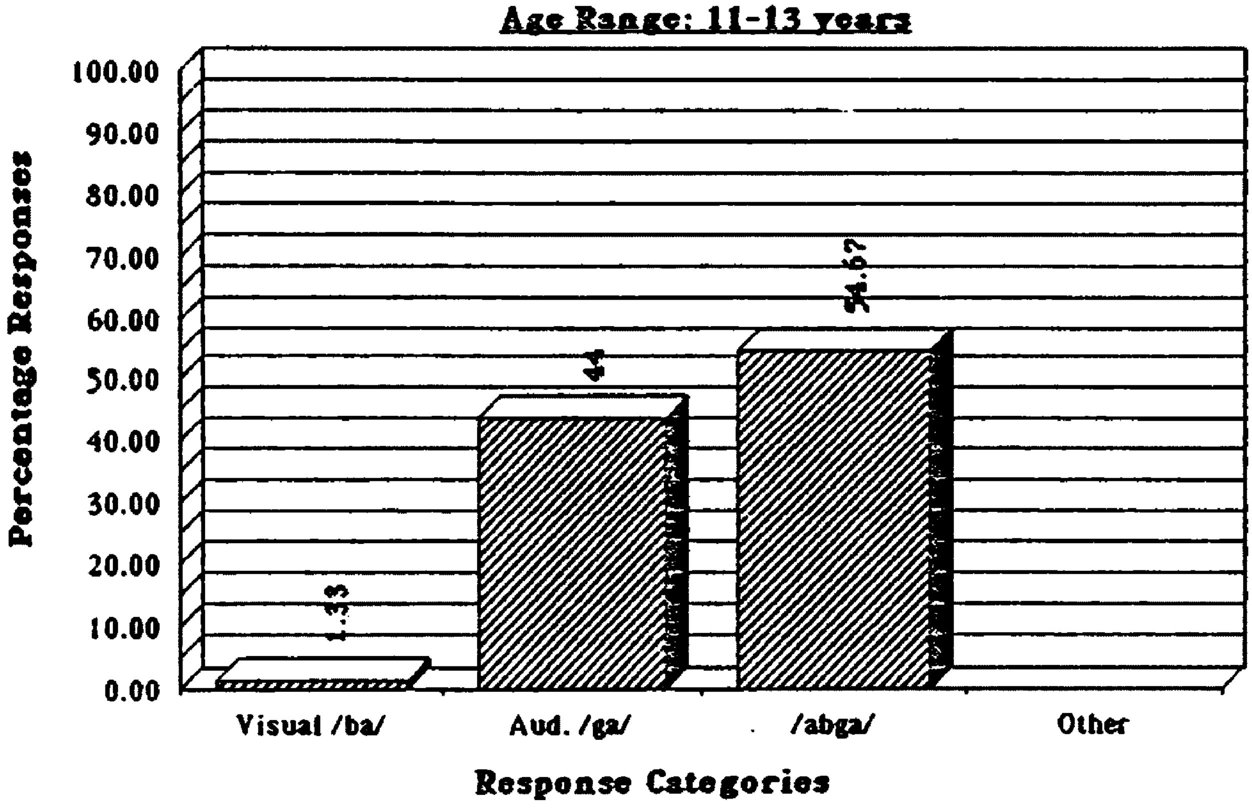


Figure 13.

Breakdown of Responses to Auditory /ga/ and Visual /ba/

Age Range: 21-25 years

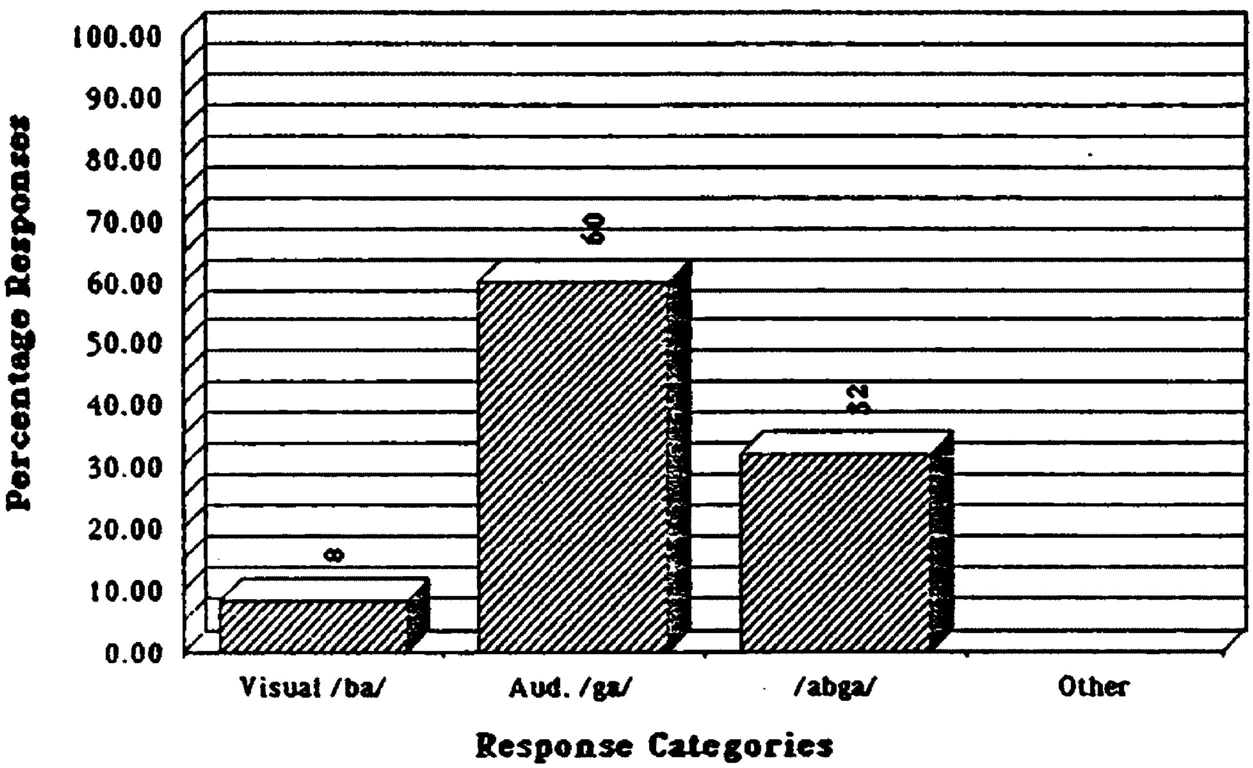


Figure 14.

Breakdown of Responses to Auditory /ga/ and Visual /ba/

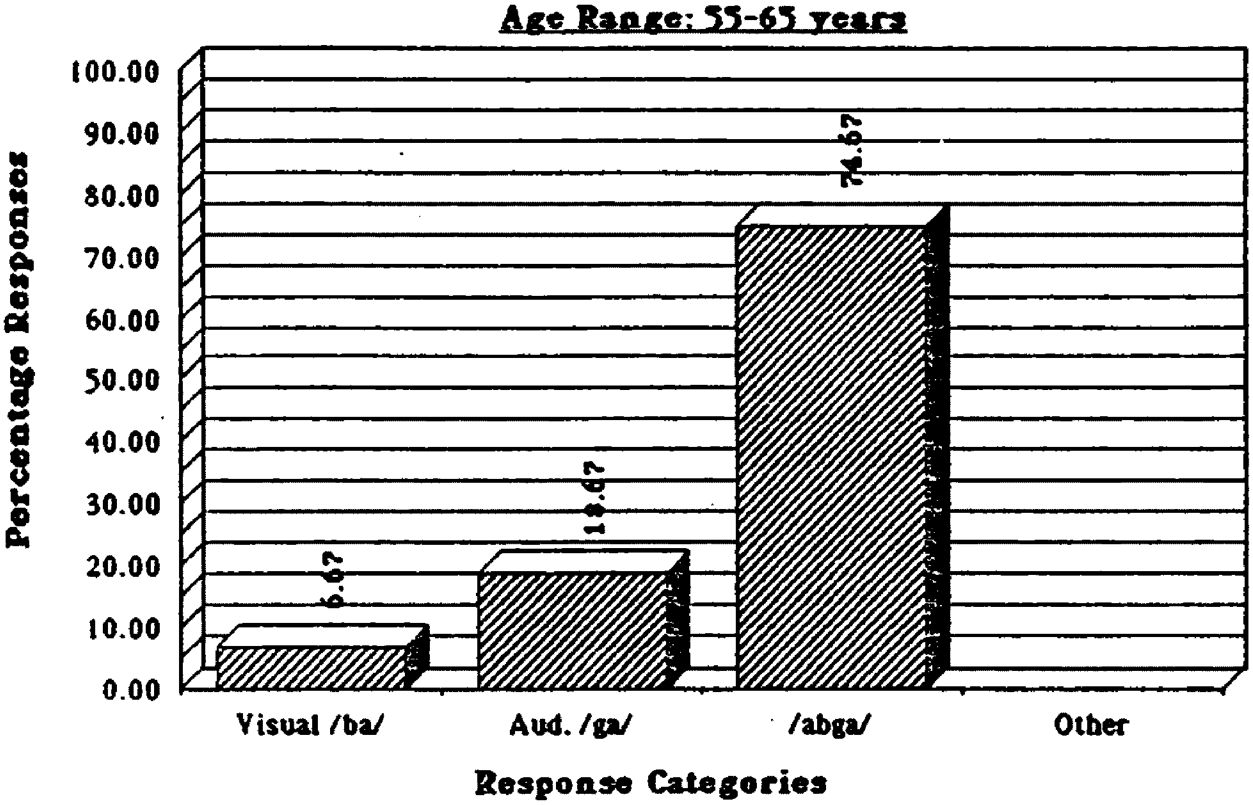


Figure 15.

Breakdown of Responses to Auditory /bi/ and Visual /gi/

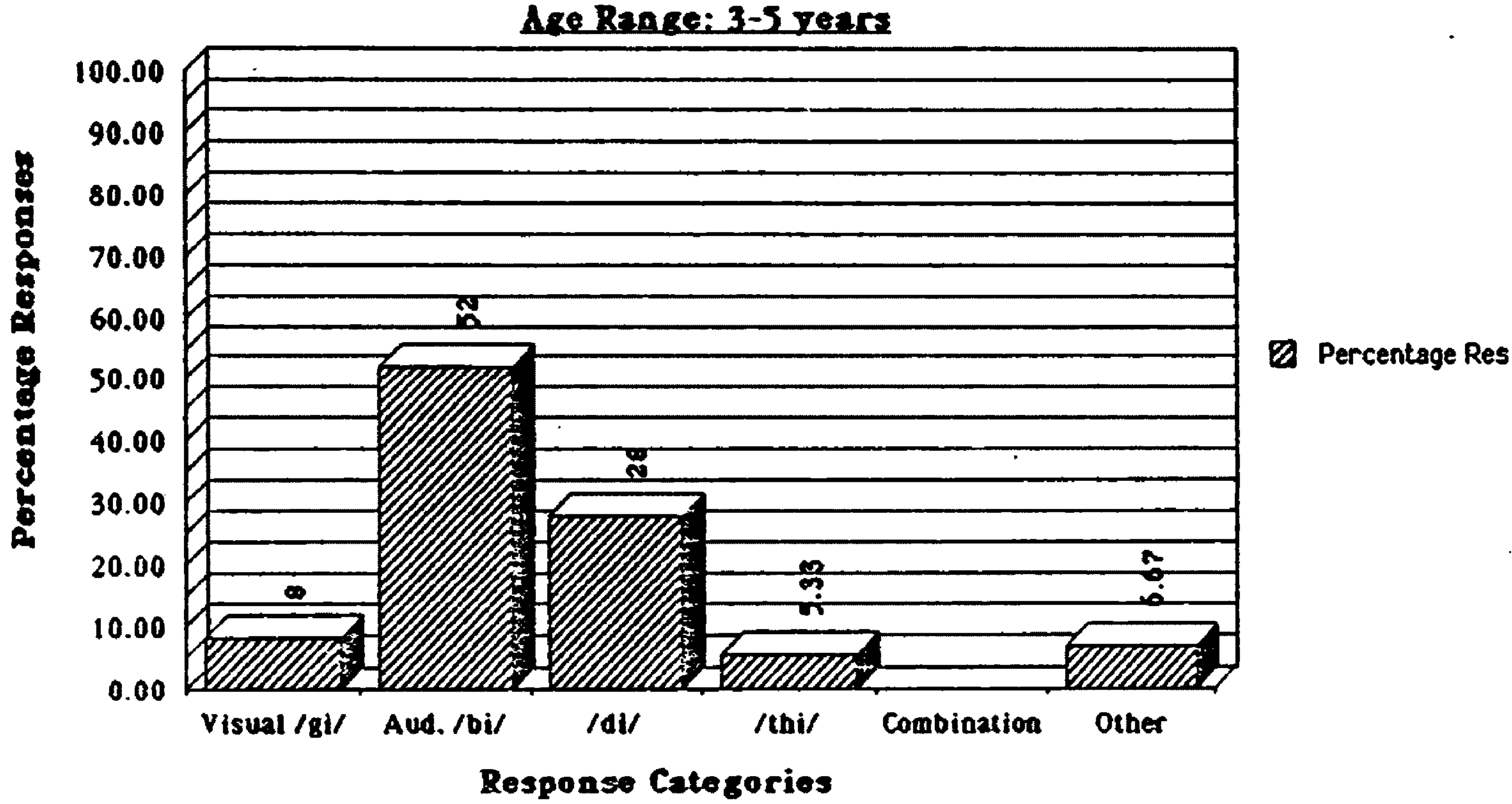


Figure 16.

Breakdown of Responses to Auditory /bi/ and Visual /gi/

Age Range: 11-13 years

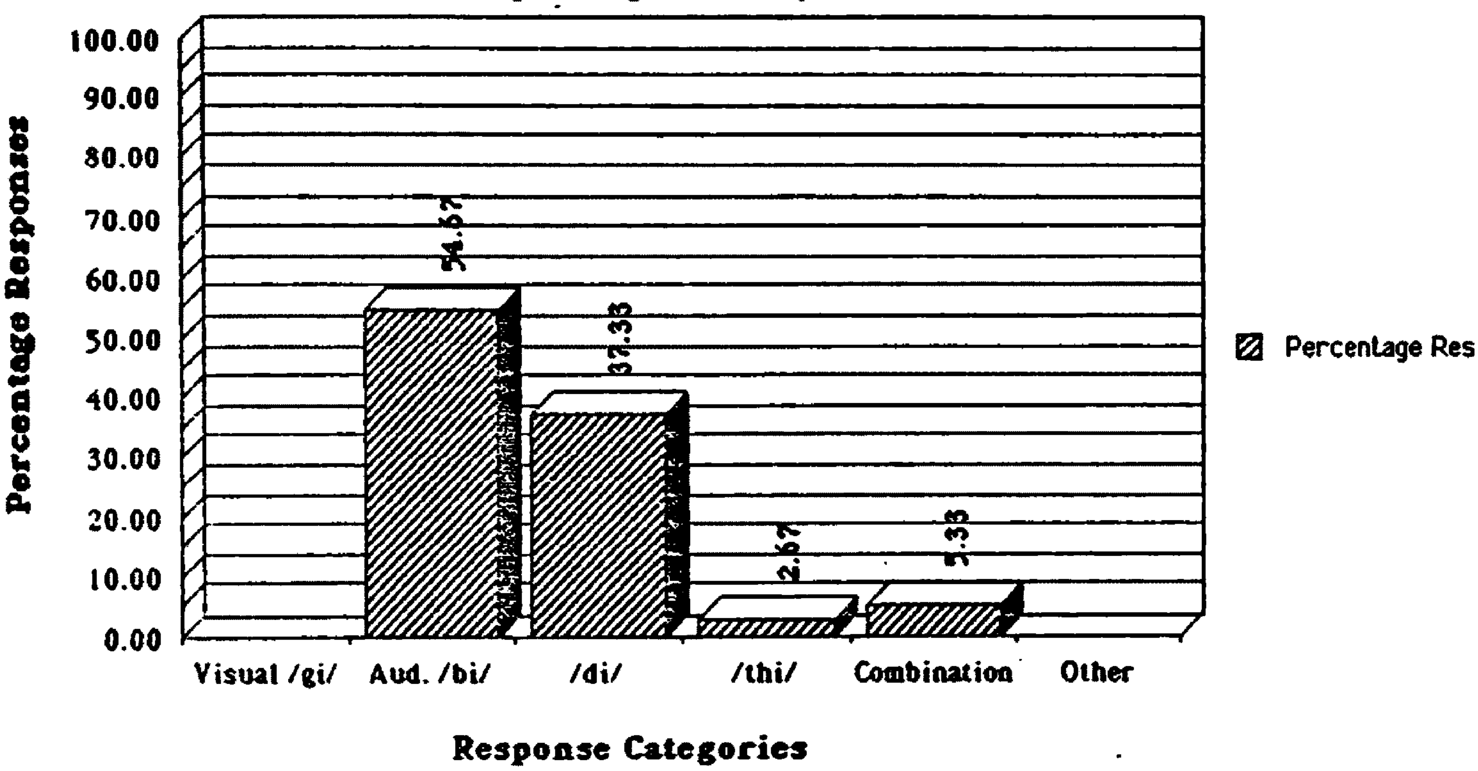


Figure 17.

Breakdown of Responses to Auditory /bi/ and Visual /gi/

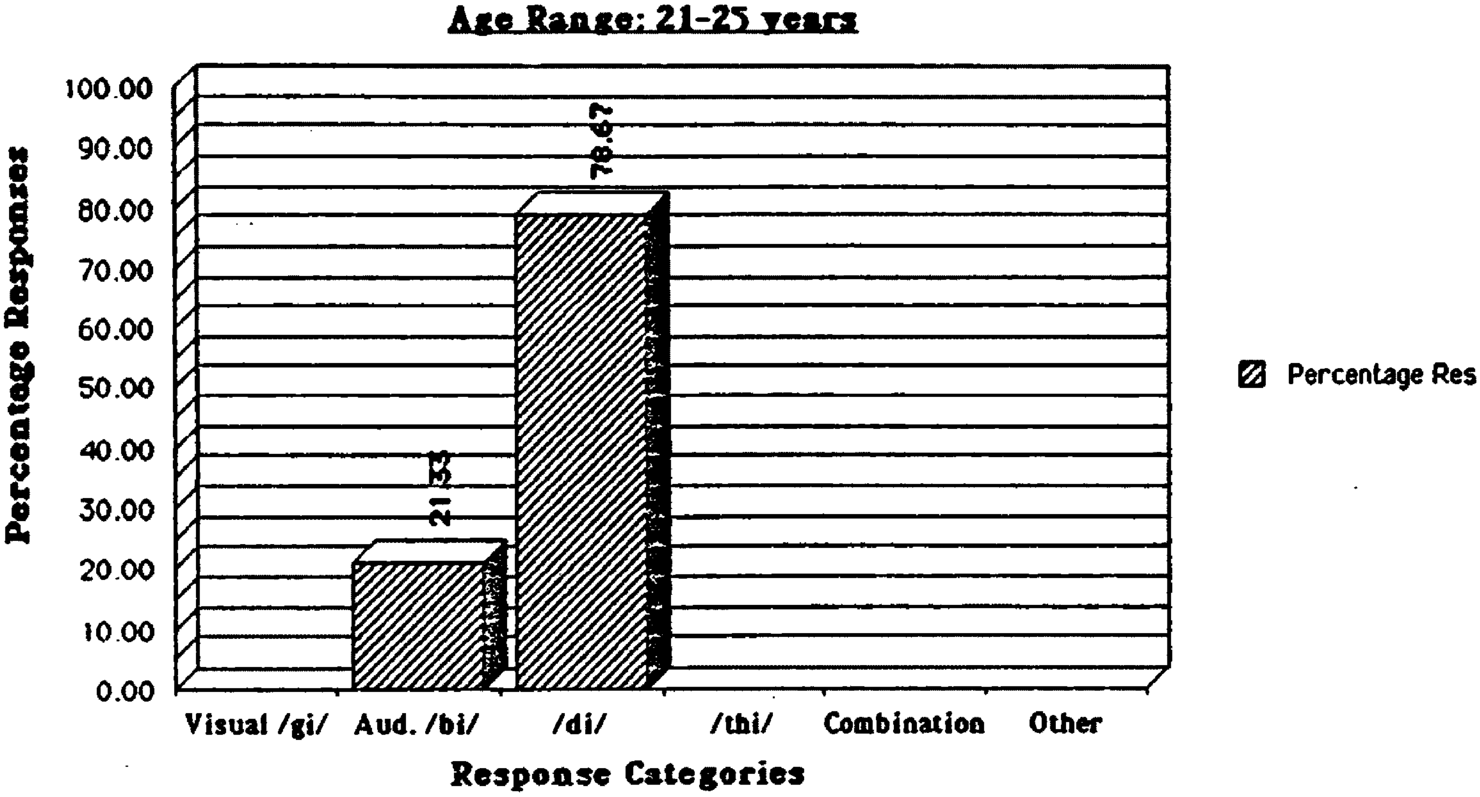


Figure 18.

Breakdown of Responses to Auditory /bi/ and Visual /gi/

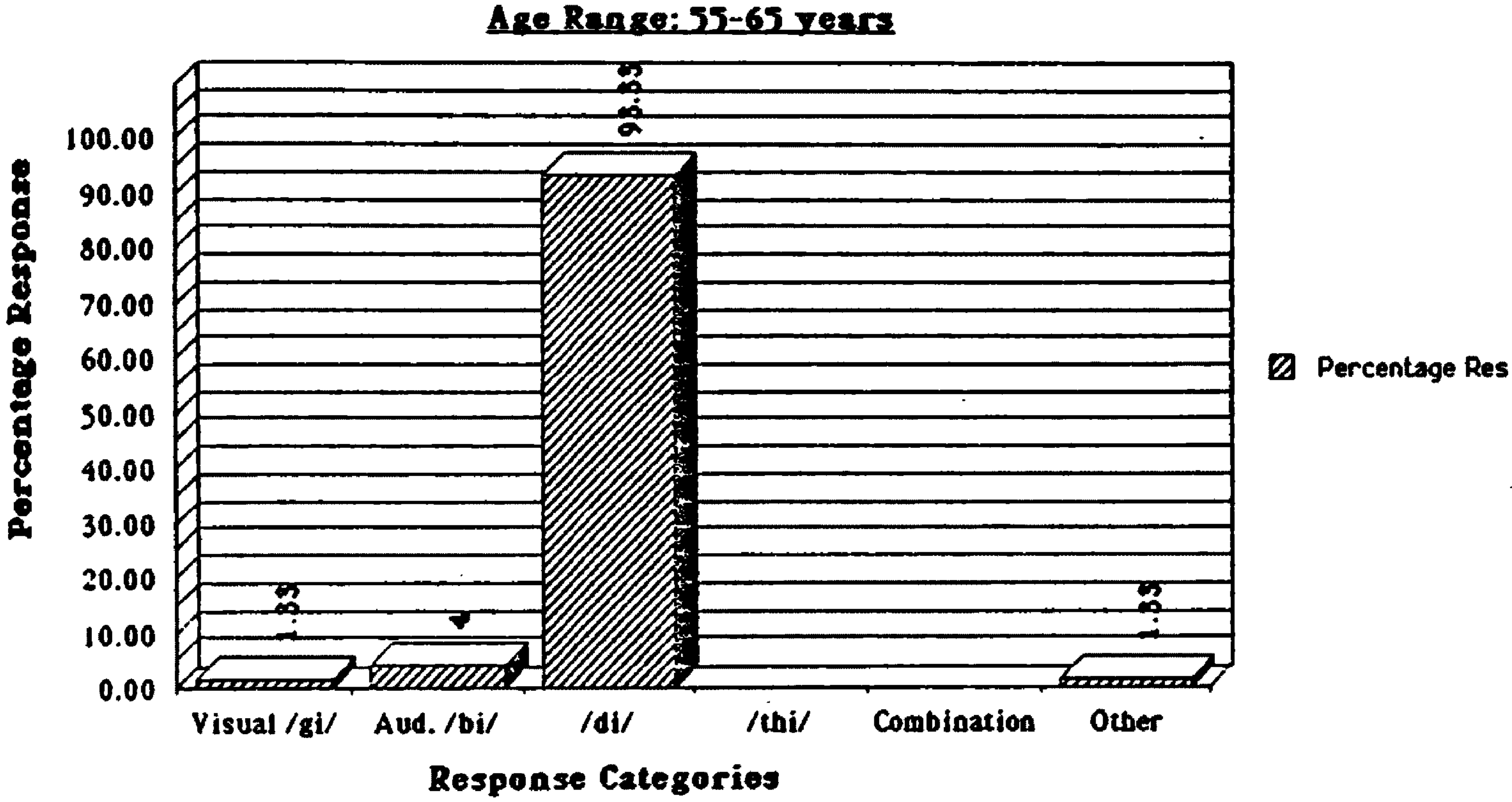


Figure 19.

Breakdown of Responses to Auditory /gi/ and Visual /bi/

Age Range: 3-5 years

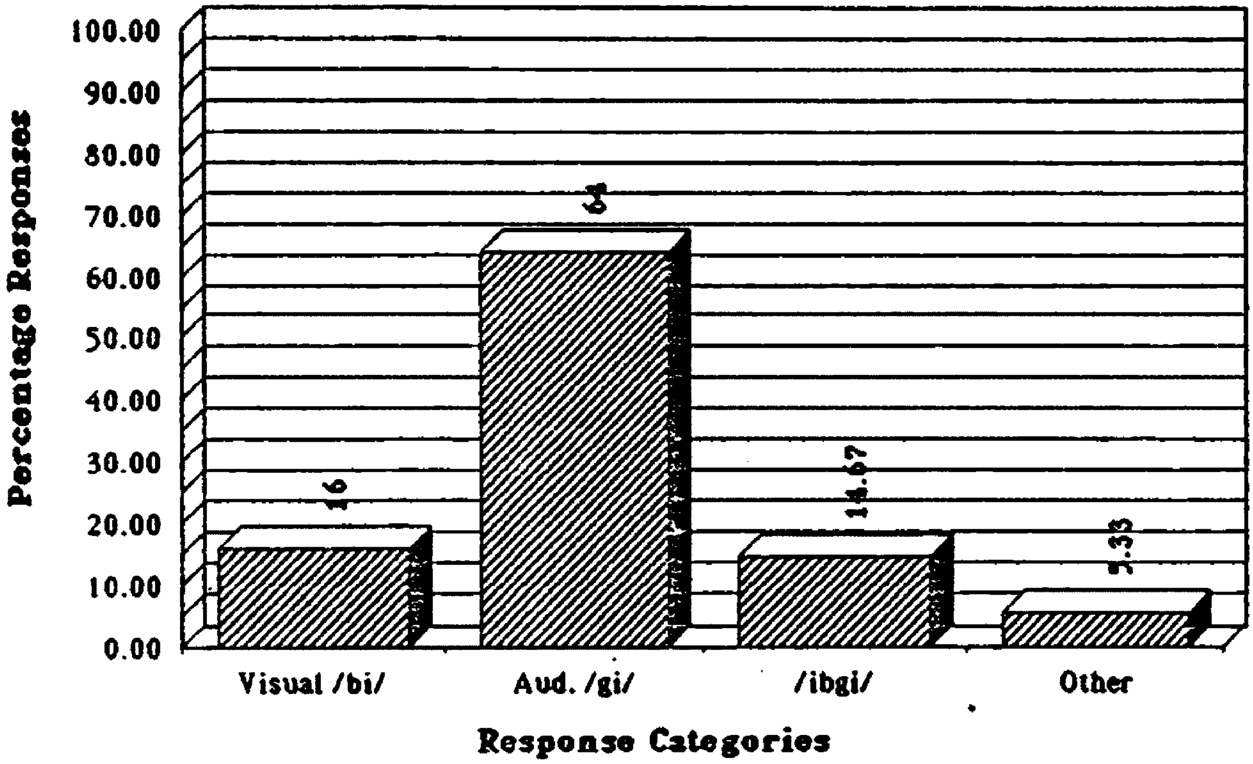


Figure 20.

Breakdown of Responses to Auditory /gi/ and Visual /bi/

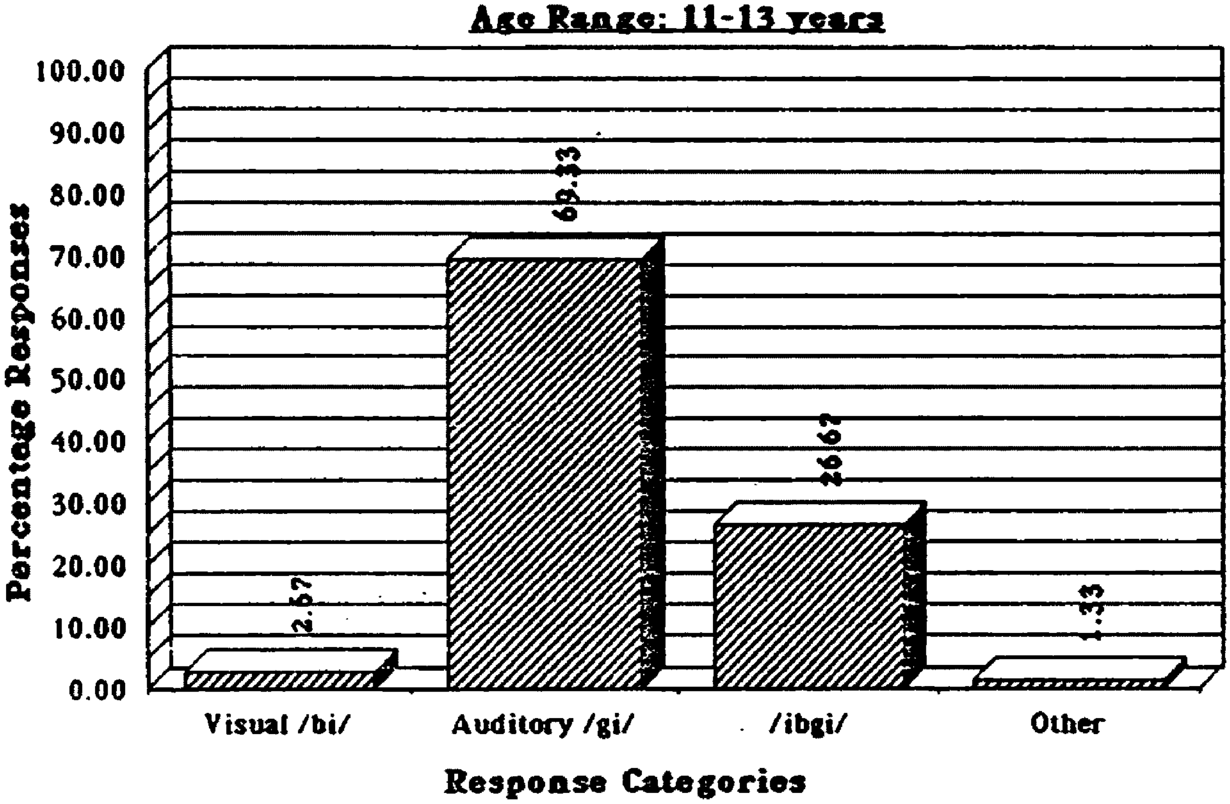


Figure 21.

Breakdown of Responses to Auditory /gi/ and Visual /bi/

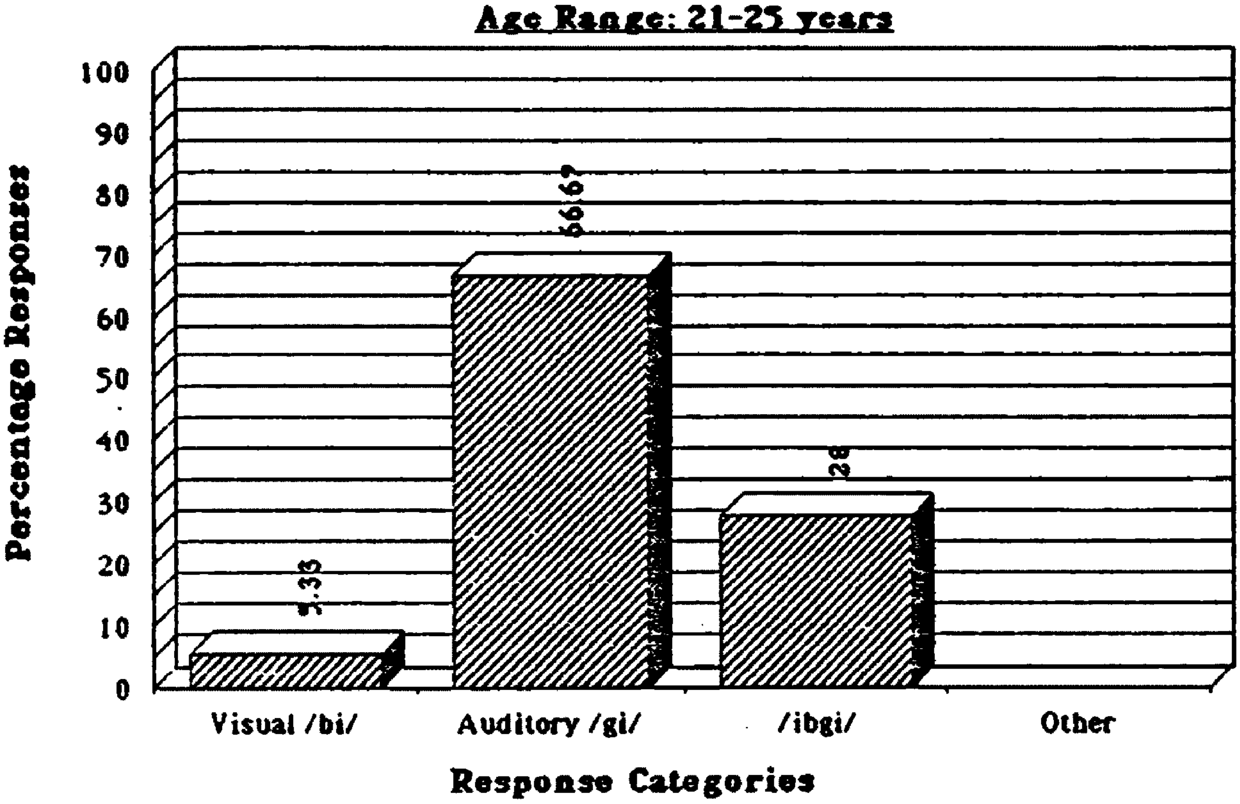
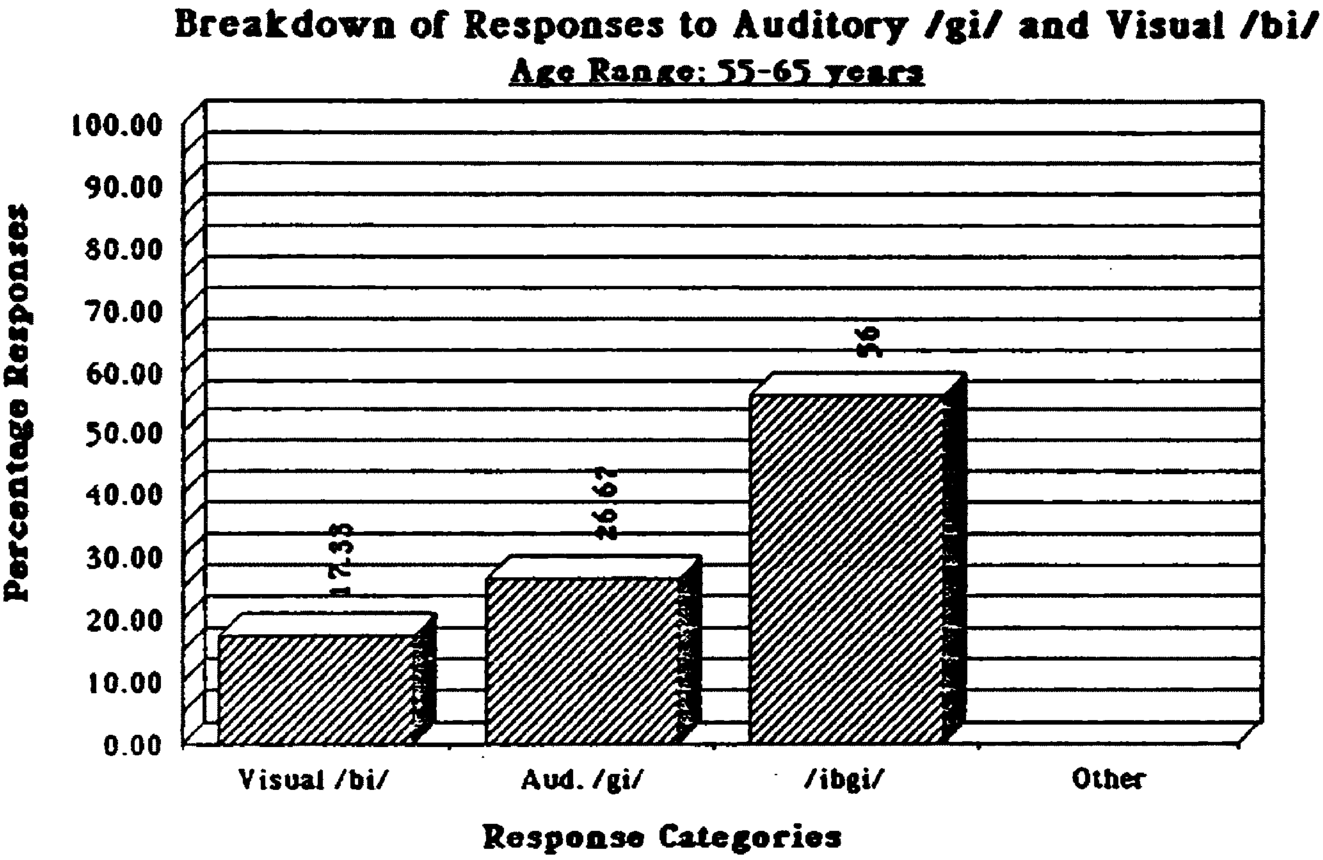


Figure 22.


Breakdown of Responses to Auditory /gi/ and Visual /bi/
Age Range: 55-65 years
100.00
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
Percentage Responses
17.33
26.67
56
Visual /bi/
Aud. /gi/
/ibgi/
Other
Response Categories

The ANOVA results also indicated a significant main effect for vowel conditions ($p < 0.05$) and a significant interaction between age group and vowel conditions ($p < 0.05$). Figures 23 and 24 display this interaction by condition in terms of percentage of incorrect responses. Responses for the forward illusion involving /i/ showed a greater developmental trend than did responses for the forward illusion involving /a/. There was a notable lack of difference between the two vowels for the senior group in both the forward and backward illusion conditions.

Comparison of responses across gender within the age groups showed very little difference for the forward illusion. Two major points emerge from gender comparisons involving the backward illusion. First, seniors evidenced a marked gender difference in both vowel conditions with females being affected more than males. Second, males in the 3-5 year old age group were affected considerably more than females for the backward illusion involving auditory /ga/ and visual /ba/.

A comparison of the results obtained in the present study with those obtained by McGurk and MacDonald (1976) indicated overall similarities in terms of the strength of the illusion. These results are illustrated in Figures 25 and 26. As will be demonstrated in the discussion section, the nature of incorrect or illusion response types found within each age group was not consistent across the two studies.

Results of reliability measures indicated acceptable reliability for both intra- and inter-examiner conditions. Intra-examiner reliability was assessed on 100% of the subject data. Point to point reliability calculations based on illusion versus no illusion judgements confirmed 2380 out of the 2400 responses thus yielding a reliability score of 99.17%. Chance reliability for the intra-examiner condition was satisfactory at 56.17%. Inter-examiner

Figure 23. Percentage of Incorrect Responses Across Vowels for Auditory /b/ and Visual /g/.

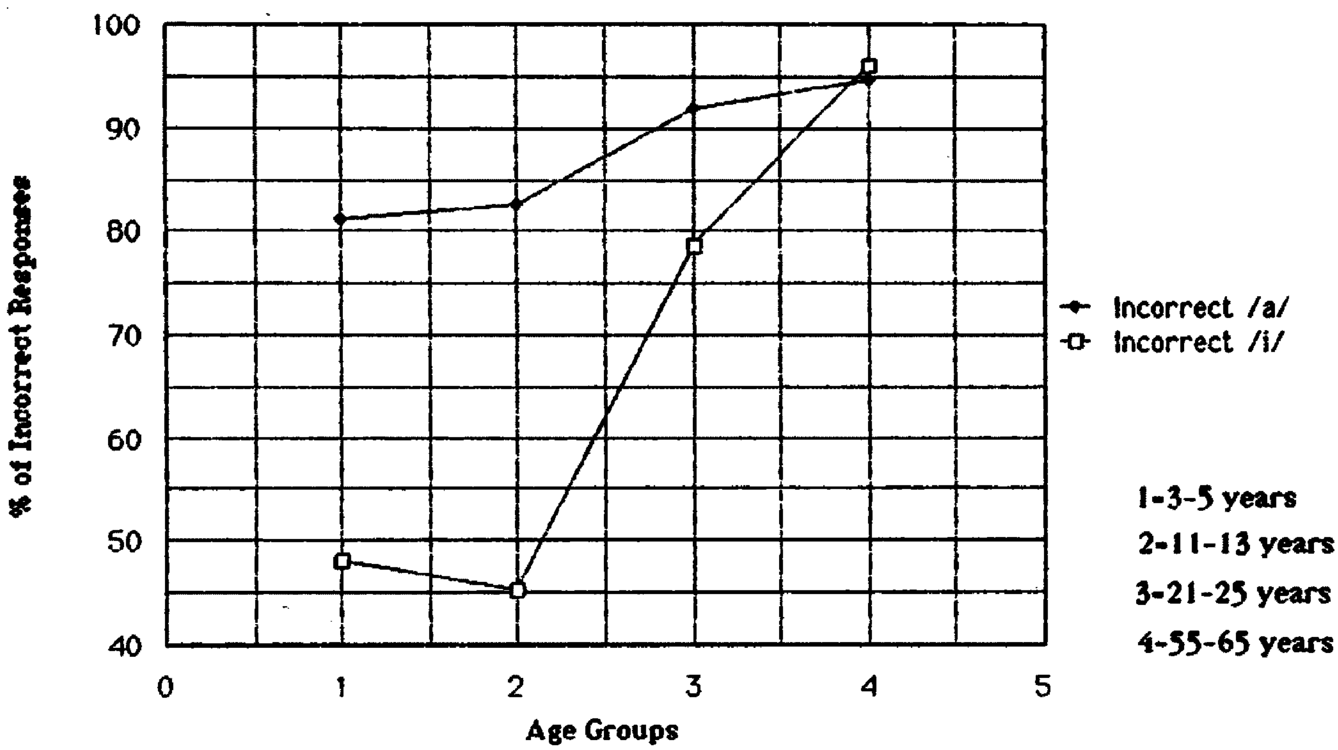


Figure 24. Percentage of Incorrect Responses Across Vowels for Auditory /g/ and Visual /b/

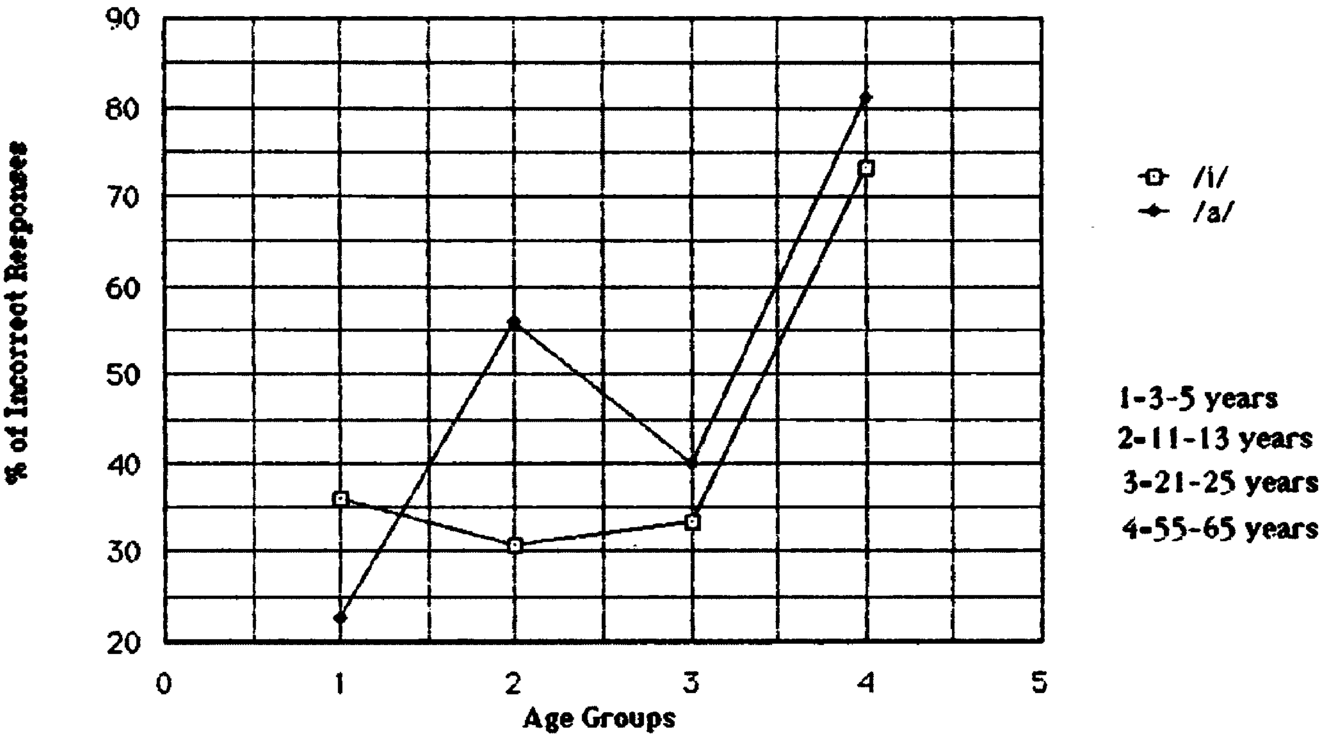


Figure 25. Responses from McGurk and MacDonald (1976) Versus Present Study for Auditory /ba/ and Visual /ga/.

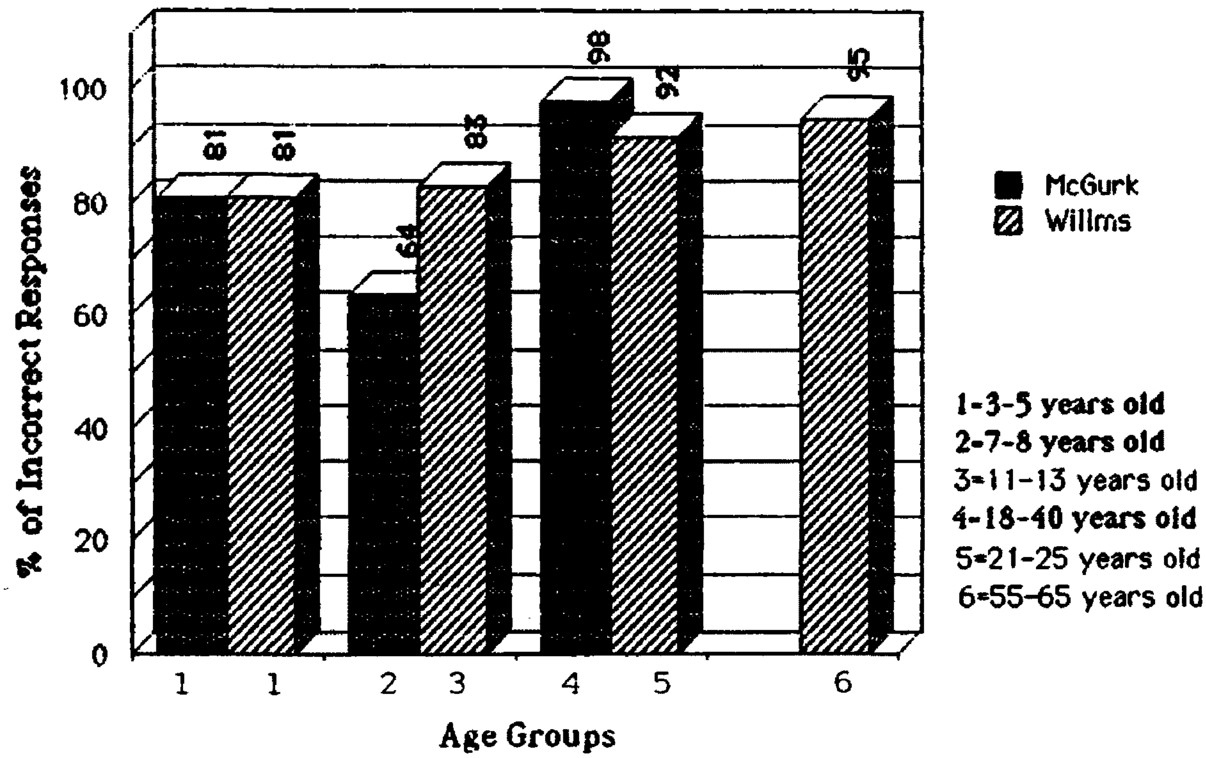
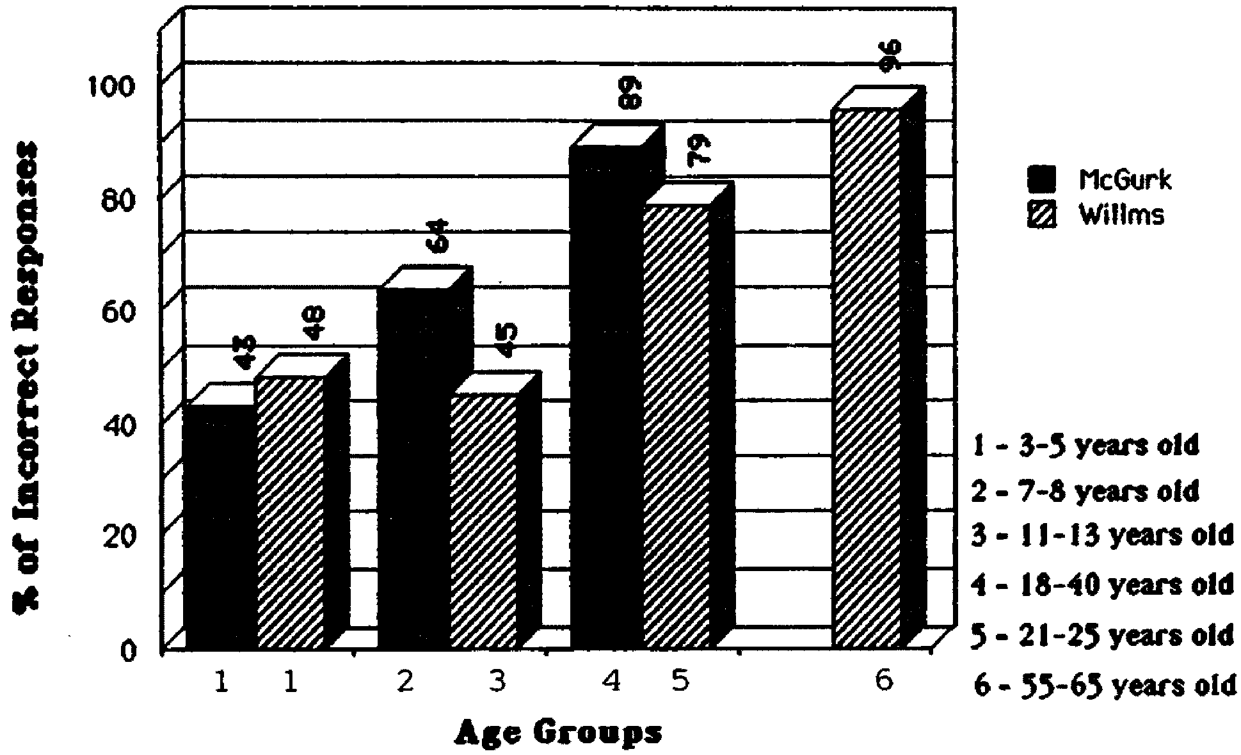


Figure 26. Responses from McGurk and MacDonald (1976) Versus Present Study for Auditory /ga/ and Visual /ba/.



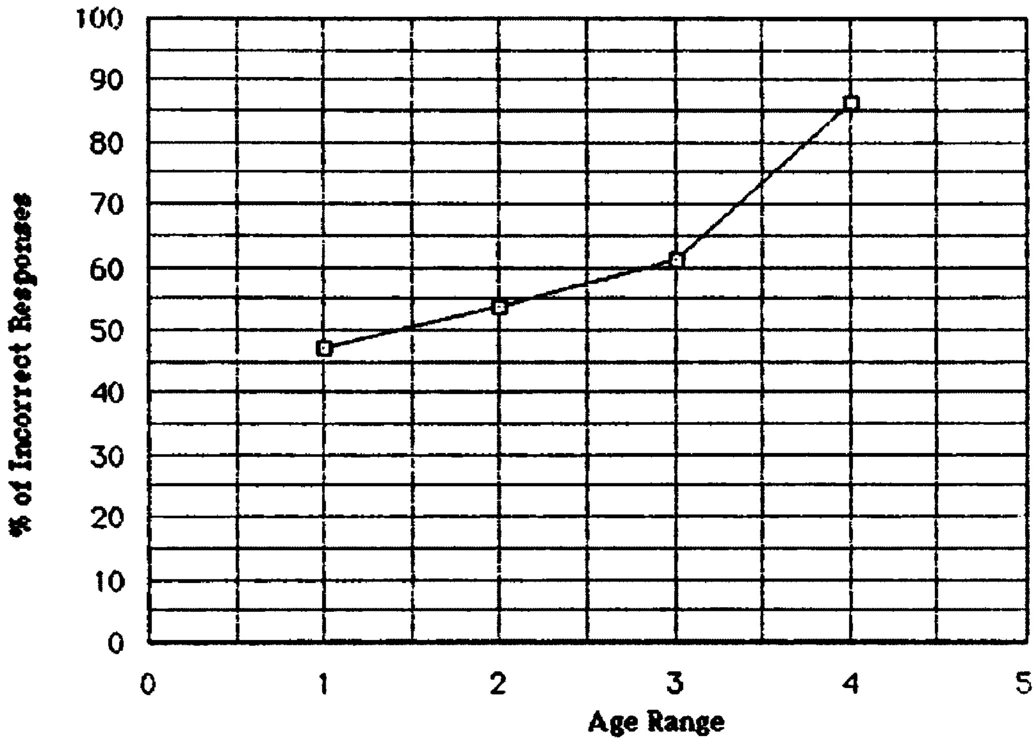
reliability was checked on a point by point basis on 25% of the subjects. 592 out of a possible 600 responses were confirmed, thus yielding a reliability score of 98.67%. Chance reliability for the inter-examiner condition was also satisfactory at 55.89%. It is of note that, although not included as mismatches during reliability, a number of differences during both intra- and inter-examiner reliability occurred within the parameter of an illusion response. During inter-examiner reliability, for example, there were seven differences which involved the phonemes /d/ and / / which both fall within the category of an incorrect or illusion response.

CHAPTER IV: DISCUSSION

The purpose of this study was to examine the trends of the McGurk illusion across a developmental age range and across two vowel conditions. Descriptive information regarding the strength of the illusion in terms of gender, response types, and distribution of responses was extrapolated from the data. The results provided support for both a developmental trend and a difference between vowel contexts. Response types contributing to the strength of the illusion showed patterns which differed from those reported in the literature. Results for the previously unexplored area of illusion strength in the senior population revealed interesting patterns as well as strong evidence of a developmental difference when compared to the other groups.

Results of the ANOVA across experimental conditions suggested a developmental trend in the strength of the McGurk illusion. This overall trend is illustrated in Figure 27 and detailed in Table 6. Close examination, however, showed that the trend was not consistent with the results reported in the literature (McGurk and MacDonald, 1976 & Massaro et al., 1986). Since past research has involved use of the vowel /a/, direct comparisons could only be made with the corresponding portion of the current study. The results for the forward illusion (auditory /b/ and visual /g/) involving the vowel /a/ showed only a slight increase in strength across age groups (Table 7). Of particular interest is the behavior of the 3-5 year old group, who responded incorrectly on 81.33% of the forward /a/ illusion experimental trials. Contrary to what has previously been reported, this figure indicated children to be affected by the visual portion of the conflicting stimulus. In fact, a surprisingly large number of responses

Figure 27. Percentage of Incorrect Responses Across All Four Experimental Conditions



1- 3-5 years old
2= 11-13 years old
3= 21-25 years old
4= 55-65 years old

Table 6. Strength of the McGurk Illusion Across Age Groups for All Experimental Conditions Combined.

Age Range	Overall Percentage of Incorrect Responses
3-5	47.0%
11-13	53.7%
21-25	61.0%
55-65	86.3%

Table 7. Incorrect Responses to Auditory /ba/ and Visual /ga/.

Age Range	Percentage of Incorrect Responses
3-5	81.33%
11-13	82.67%
21-25	92.00%
55-65	94.67%

(49.33%) were representations of the visual component of the stimulus. This observation directly contradicts the following statement made by McGurk and MacDonald (1976) in interpreting their data:

> "...where responses are dominated by a single modality, this tends to be the auditory for children and the visual for adults."

Results for the 11-13 year old age group also indicated a strong influence of the visual stimulus, in that 82.67% of their responses to the forward /a/ illusion were incorrect. This group, however, tended not to report representations of the visual component of the stimulus as only 2.67% of their responses were /ga/. Despite the lack of exclusively visually representative responses in the 11-13 year old group, both groups of children presented evidence of strong intermodal speech perception abilities.

Comparison with other studies

Closer examination of the results obtained by McGurk and MacDonald (1976) is needed before comparing these findings to those of the present study. To begin, the percentages of incorrect responses presented in McGurk and MacDonald's literature review took into account the data for all four of the possible conflicts within the combinations of /ba/ and /ga/ and the combinations of /pa/ and /ka/. When the data were separated for the various stimuli, notable differences in response trends occurred. Appendix D displays these differences and shows that young children were strongly affected by the forward /a/ illusion. The overall developmental trend reported by McGurk and MacDonald was more reflective of the data obtained for the backward /a/ illusion in which young children were affected in only 43% of the trials.

In terms of correct/incorrect responses for auditory /ba/ and visual /ga/, McGurk and MacDonald's results for the 3-5 year old group are actually similar to those of the present study (Fig. 25). Differences exist, however, between the type of incorrect responses given across the two studies. McGurk and MacDonald reported that all incorrect responses for the 3-5 year age group were "fused" or /da/ responses. The current study, however, revealed that for auditory /ba/ and visual /ga/, a large proportion of the responses for the youngest group were representations of the visual component (/ga/) of the stimulus.

The results obtained for the backward illusion were, again, similar to the results of McGurk and MacDonald (Fig. 26) as they showed a stronger developmental trend than the forward illusion results. Comparisons of statistical significance across variables could not be made between the two studies as such information was not reported by McGurk and MacDonald.

The study of McGurk and MacDonald (1976), although having served as a catalyst for the examination of bimodal speech perception, contained several methodological difficulties. The seemingly apparent consistency between the data obtained in the two studies, however, indicate that the strength of the McGurk illusion was sufficient to overcome the design flaws of this previous research.

In comparing the results of the present study to those of Massaro et al. (1986), a close examination of their research design and results is required. Massaro et al. (1986) presented synthetic speech sounds that varied on a continuum from /ba/ to /da/ to two age groups; children 4-10 and adults 16-32. They concluded that the visual component had a greater influence on the perceptual judgements of adults. The authors discussed the possibility of existing differences between the age groups in terms of their ability to process the visual source. Massaro et al. (1986) further argued that the extent of the visual influence was directly related to the lip reading abilities of the different age groups.

Results of the auditory /ba/ and visual /ga/ combination of the current study do not support the findings of Massaro et al. (1986) as young children were found to be affected by the visual component of the illusion. A number of design differences need to be considered. First, the subjects in the Massaro et al. (1986) study were given a closed response set in that they were told to report whether they heard "ba" or "da". As such, any other perceptions of the stimuli were not reported. In addition, the children who participated in the Massaro study responded orally while the adults provided written responses. Perhaps this variation in procedure may have resulted in some difference in the results obtained across the studied age groups.

The results of Massaro et al. (1986) may also have been affected by the intensity level at which their stimuli were presented. Kuhl, Green, and Meltzoff (1988) found that increased intensity of the auditory signal led to a greater influence of the visual component of a conflicting presentation. Kuhl et al. found significant differences between the intensity levels of 45 dB SPL, 58 dB SPL, and 66 dB SPL for their adult subjects. Kuhl et al. (1988) did not investigate the effects of varying the intensity of the auditory stimuli in presentations to children. Massaro et al. (1986) presented all of their auditory stimuli items at 70 dB SPL. Based on the research of Kuhl et al. (1988), this intensity level may cause the adult subjects to be more strongly influenced by the visual source. Perhaps children are not as influenced by the intensity effect and therefore gave fewer visually dominated responses than did adults in the study of Massaro et al. (1986). Clearly, the effects of intensity across age groups in terms of its interaction with the McGurk illusion is an area for further research.

The results of the present study also better correspond with infant speech perception data than do the results and consequent theory of Massaro et al. (1986). A number of researchers have shown that young infants are capable of bimodal speech perception (Dodd, 1979, Kuhl & Meltzoff, 1982). It is thus counterintuitive to argue that children are less affected by a bimodal conflict due solely to poorer lip reading capabilities. Rather, based on the data from these infant studies and the results of the current study, children should be capable of matching the visual and auditory components of a speech signal and would therefore be more heavily influenced by a conflict across these modalities. The term "conflict" is purposely used here since it is, in fact, the conflict of information rather than the visual signal itself which produces the illusion.

The comparisons between the findings of McGurk and MacDonald and Massaro et al. (1986) with the current study indicate a danger in generalizing the effects of the McGurk illusion across various consonant/vowel combinations or across different experimental designs. Until more is known about the parameters of the illusion and the acoustic or spectral factors which affect it, unwarranted generalizations of the research data could be misleading. For instance, it should be made clear that the predominance of visual representative responses among the 3-5 year old group was confined to the illusion involving the auditory /ba/ and the visual /ga/. Why the youngest group was affected in a manner so different from that of the others is intriguing and deserves further investigation. McGurk and MacDonald (1976) defined the term "fused" response as "one where information from the two modalities is transformed into something new with an element not presented in either modality". The 3-5 year old age group in this study gave a much lower number of fused (/da/) responses than did the other three age groups. This is particularly interesting in terms of the 11-13 year old group which responded more like the two adult groups by giving predominantly fused responses to the forward illusion. Although young children are capable of bimodal speech perception and are influenced by visual cues, their processing skills in terms of auditory /ba/ and visual /ga/ appear to differ from older individuals in the manner in which they deal with the modality conflict. Further study into the parameters of this effect and possible corresponding subject characteristics is warranted.

Perception of the McGurk illusion had not previously been investigated in the senior population. The results of the current study indicated the 55-65 year old age group was strongly affected by all combinations of conflicting stimuli across both vowel conditions. In particular, they were the only age group to be affected (> 90% incorrect responses) by the backward illusion. Since all subjects had normal hearing sensitivity, theories expounding this group's reliance and greater attention to visual cues due to changes in their hearing sensitivity are unsatisfactory for explaining the strong effects of the illusion demonstrated by seniors.

In a discussion of their findings, Jerger, Jerger, Oliver, & Pirozzolo (1989) indicated that reduced speech understanding in the elderly was due to a decline in the subjects' central auditory processing skills, rather than in their peripheral auditory sensitivity curves. The authors cite structural changes within the central auditory system as a cause for this decrease in skills. Incorporation of these findings with those of the current study leads to two possible hypotheses to explain the increased strength of the McGurk illusion in seniors. One, visual cues have more influence due to the older subjects' difficulties in central auditory processing; or two, a decrease in the subject's central auditory skills leads to difficulties specific to processing modality conflicts in speech. Further study of the McGurk illusion would, no doubt, provide critical information to the area of speech understanding in the senior population.

Differences between the age groups may also be explained through differing attention strategies. Although the present study was controlled for visual gaze towards the monitor during presentations, the qualitative aspect of the cognitive response of "attending" is difficult to control. Attention and cognition are difficult to separate, and even more troublesome to measure independently in a perceptual task. It is hypothesized, however, that the degree of focus, awareness, and/or cognizance of perceptual stimuli varies within subjects across time and across subjects. Such differences may or may not be related to lip reading and/or physiological variations across the developmental range.

Up to now variation across responses to consonant-vowel combinations in the McGurk illusion have mainly been explained by the perceptual salience or various visual cues. The greater the perceptual salience, the more "attention" paid to the cues. The idea being presented shifts from examining "attention getting" differences across stimuli to examining "attention giving" differences across subjects. Manipulations of subjects' attentional focus through verbal instructions or visual highlighting may be a fruitful avenue of study towards understanding the McGurk illusion.

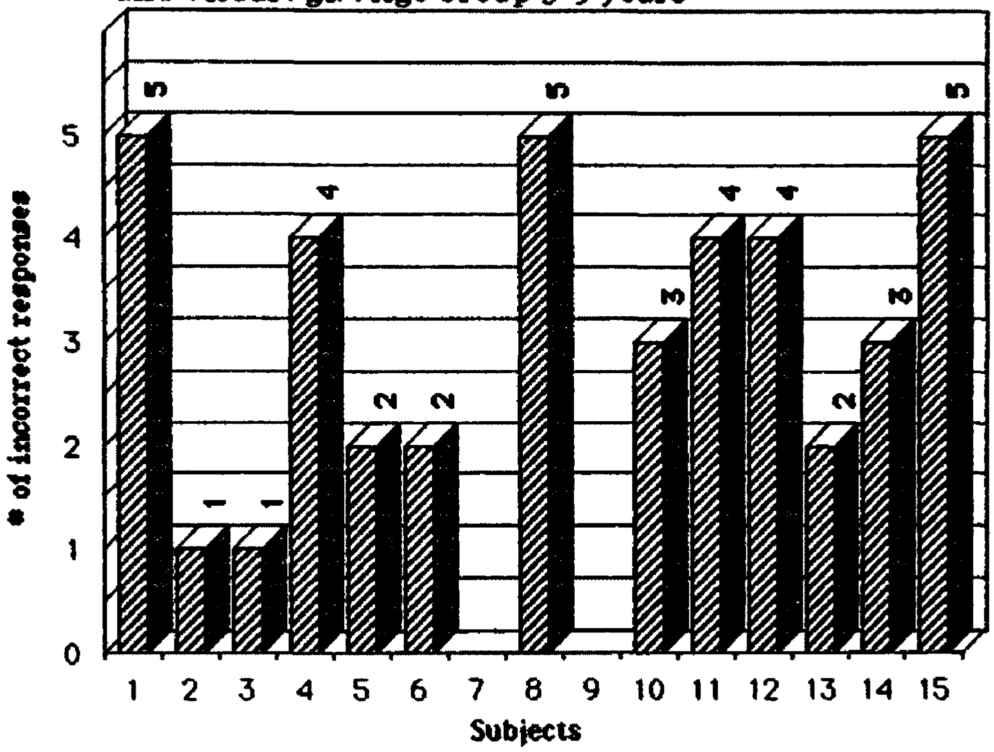
<u>Response Distribution</u>

The data across all conditions and age groups indicated that most subjects were either markedly affected or not affected by the McGurk illusion. This is particularly evident when comparing correct versus "fused" or "combination" responses for the forward and backward illusion respectively. Figures 28 and 29 illustrate the number of incorrect responses across subjects in one age group for selected stimulus conditions. In the case of 3-5 year olds responding to the forward /i/ illusion (Fig. 28), all 5 of the subjects who had either two or three incorrect responses, responded with the visual representation /gi/ or "other" rather than the "fused" /da/ response.

In the case of 21-25 year olds responding to the backward /a/ illusion (Fig. 29), only three of the fifteen subjects had two or three incorrect responses. Of the remaining twelve subjects, five were markedly affected (4-5 incorrect responses) and seven were markedly not affected (0 incorrect responses) by the illusion. This trend was observed throughout the data for the experimental conditions across the age groups. As such, the trend may be explained by the above presented hypothesis of differing attentional strategies across subjects; some subjects attend to evidence of auditory/visual conflict while others don't. Very few subjects attend inconsistently.

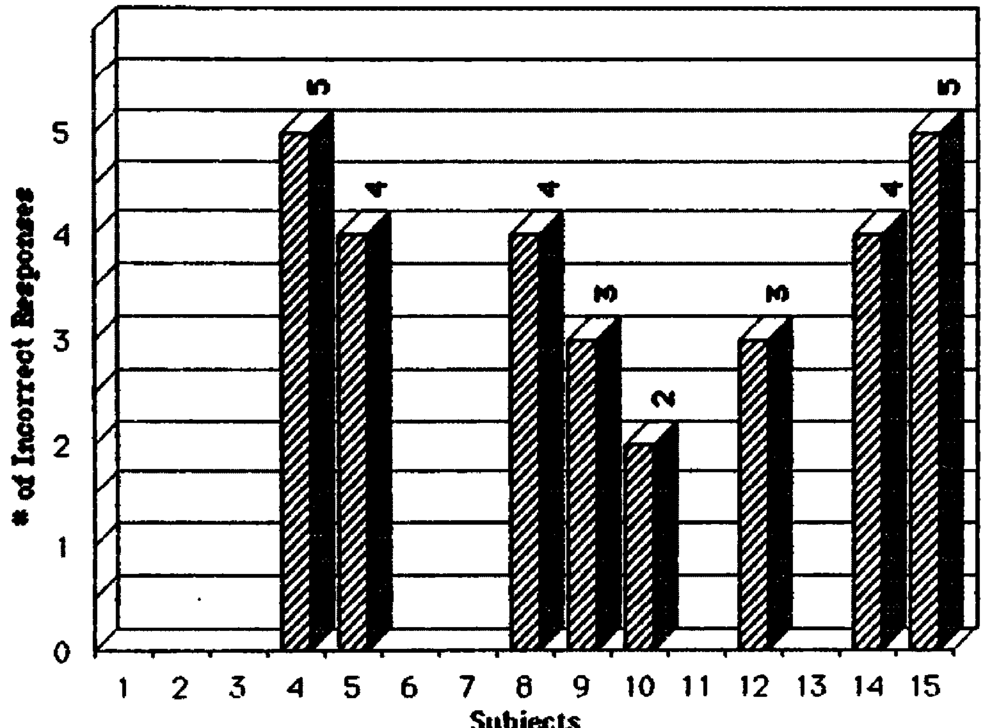
Furthermore, individual subject's response types were consistent within illusion conditions. For example, those with four incorrect responses within a condition most likely responded with four fused or visually representative responses rather than with responses that were evenly distributed across the incorrect response types. This clustering of response types indicates the need for further research which would more carefully define subject variables which possibly influence the strength of the illusion and the distribution of incorrect response types. Careful research in this

Figure 28. Individual Subject Responses to Auditory /bi/ and Visual /gi/: Age Group 3-5 years



*blanks indicate no incorrect responses

Figure 29. Individual Subject Responses to Auditory /ga/ and Visual /ba/: Age Group 21-25 years



*blanks indicate no incorrect responses

area would contribute valuable information to the development of a more complete theory of speech perception.

Vowel Differences

Results of the ANOVA indicated a significant difference in the number of incorrect or illusion type responses given across vowels. This is not surprising since the past literature (McGurk & MacDonald, 1976, MacDonald & McGurk, 1978) has found significant differences between responses depending on the consonant contexts used. Figure 23 and 24 illustrate the vowel differences across age groups for the forward and backward illusion respectively. Looking first at the forward illusion (Fig. 23), it is apparent that the responses to the illusion involving the vowel /i/ showed a stronger developmental trend than did the responses for the vowel /a/. The responses to both vowels did, however, suggest a positive correlation between the subject's age and the increasing strength of illusion. For both vowel conditions, but more so for /i/, a large increment in the number of incorrect responses is seen between the two children's and the two adult groups.

Results across vowels for the backward illusion (Figure 24) show yet another pattern. Incorrect responses to the backward illusion were fairly similar across the vowel conditions for three out of the four age groups. The exception was the 11-13 year old age group whose total number of incorrect responses to the backward illusion involving /a/ was substantially greater than to the backward illusion involving /i/. This higher figure also departed from a steady developmental trend as it was greater than the incorrect responses of the 21-25 year old group for the same condition. As mentioned earlier, a striking characteristic of the responses to the backward illusion was the large number of incorrect responses across vowels for the senior group. This age group was clearly much more affected by the backward illusion than were any of the other three groups. The seniors were also the only age group to be almost equally affected across vowels for both of the auditory/visual experimental conditions.

The differences found between vowel conditions, particularly for the forward illusion, support the earlier discussed danger of generalizing the strength of the McGurk illusion across stimuli conditions. More research examining the vowel differences is needed, not only for different vowels, but also for these vowels across consonant contexts.

With all of the data available the respective influences of the acoustic/spectral information and the visual components of the bimodal stimuli remain difficult to separate. One hypothesis regarding the greater strength of /a/ for the forward illusion centers on the visible differences of mouth configuration for the two vowels and the interaction of these differences with consonant contexts. The production /aga/, for instance, involves much more mouth movement than does the production /igi/. It is possible that the strength of the McGurk illusion is correlated to the amount of mouth movement for the forward illusion condition. If so, the relative lack of difference across vowels for the backward illusion may be partially explained by the greater amount of mouth movement involved in the production of /ibi/ versus /igi/.

It is suggested that future research focusing on vowel differences control for presentation order effects through counterbalancing. The vowel differences found within the current study would have been more reliable if half of the subjects in each age group had been first presented with trials involving the vowel /a/.

Gender Differences

Subjects within the age groups of the current study were distributed approximately equally across gender to allow for comparison in the strength of the illusion across males and females. These comparisons for both the forward and backward illusion are displayed in Figures 30 through 33. The results indicated little difference between males and females on either vowel condition of the forward illusion. There were, however, substantial differences between males and females of the senior group on both vowel conditions of the backwards illusion. The results showed that senior male subjects tended to be much less affected by the modality conflict than were senior female subjects. A substantial difference was also revealed between males and females of the 3-5 year old group for the backwards illusion involving the vowels /a/. In this instance, however, females were much less affected by the conflict of modalities.

Although there was no substantial difference between 3-5 year old males and females in terms of correct versus incorrect responses to the forward /a/ illusion, there appears to be a difference in the type of incorrect responses given. As discussed earlier, a large number (49.33%) of the the 3-5 year old group's responses to the forward /a/ illusion represented the visual component of the stimulus rather than a fused response. When analyzed by gender, it is apparent that the preschool girls responded with visual component representation during 60% of such trials while their male counterparts responded with the visual representation on 33.3% of the trials. Although 33% is still greater than the visual responses of any other age group, there does appear to be gender differences for this particular experimental condition.

Figure 30.

Male versus Female Responses for Auditory /ba/ and Visual /ga/

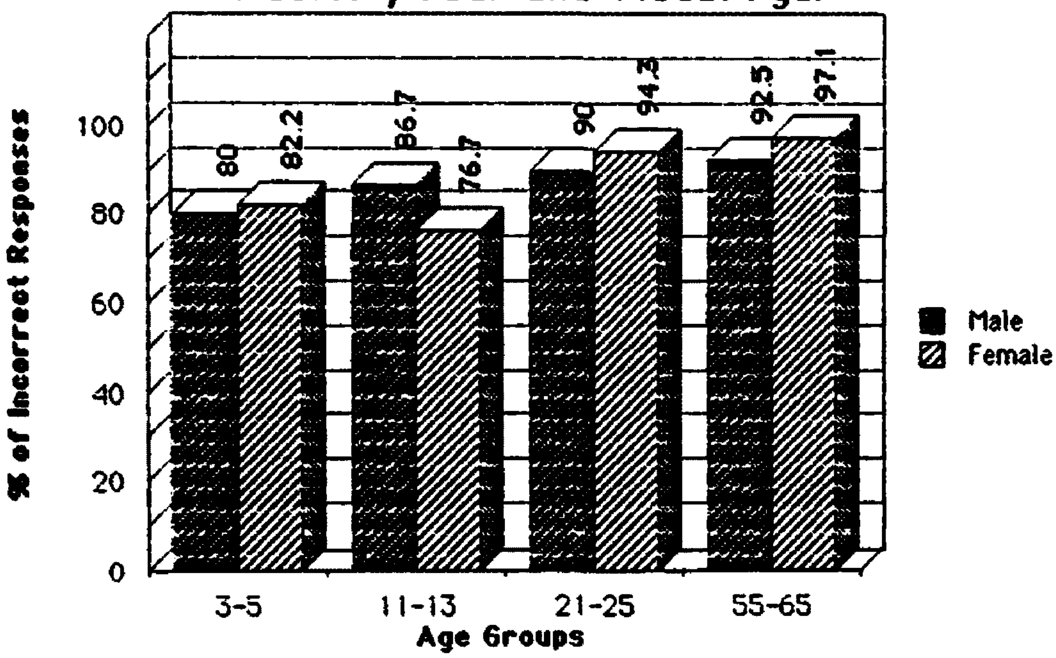


Figure 31.

Male versus Female Responses for Auditory /ga/ and Visual /ba/

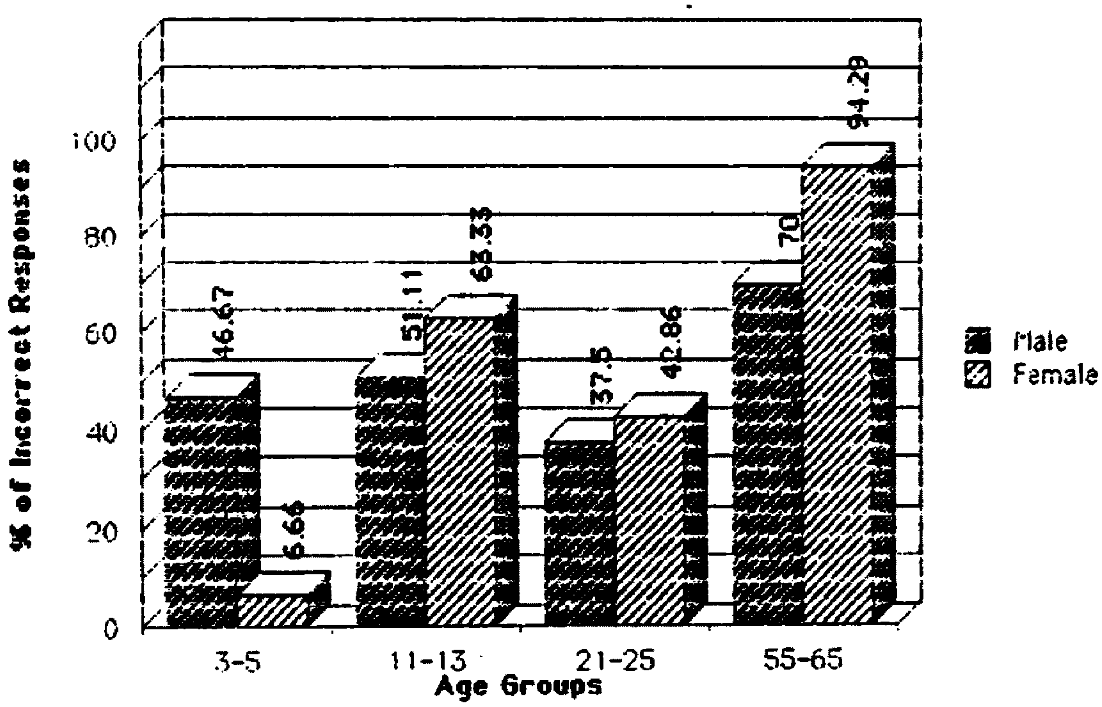


Figure 32.

Male versus Female Responses for Auditory /bi/ and Visual /gi/

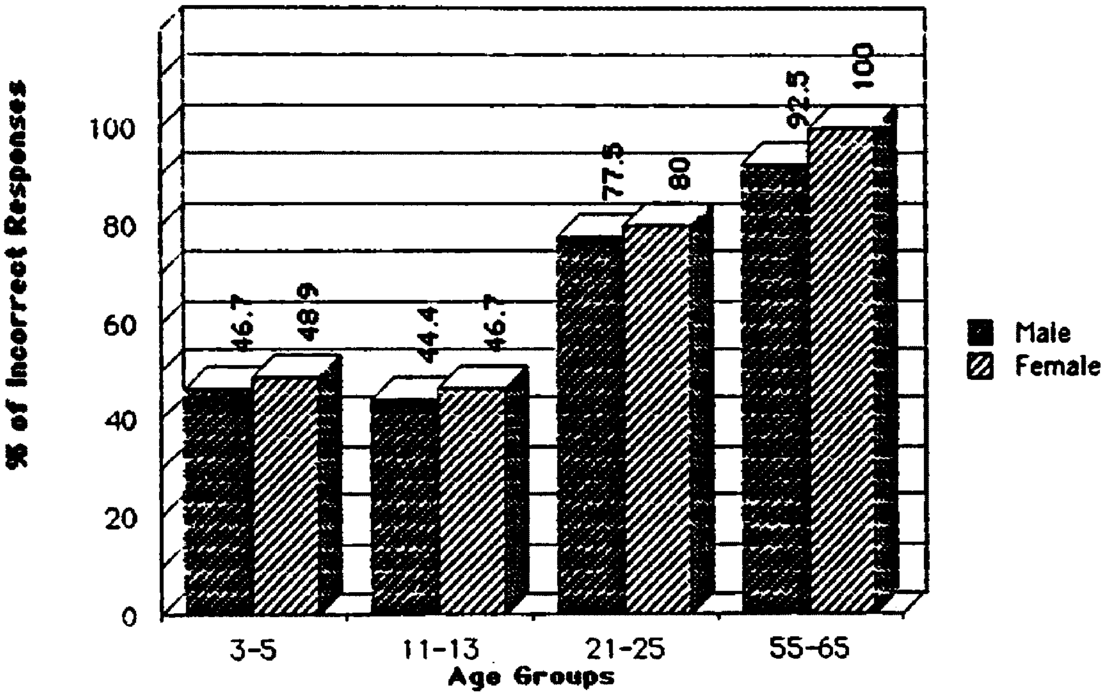
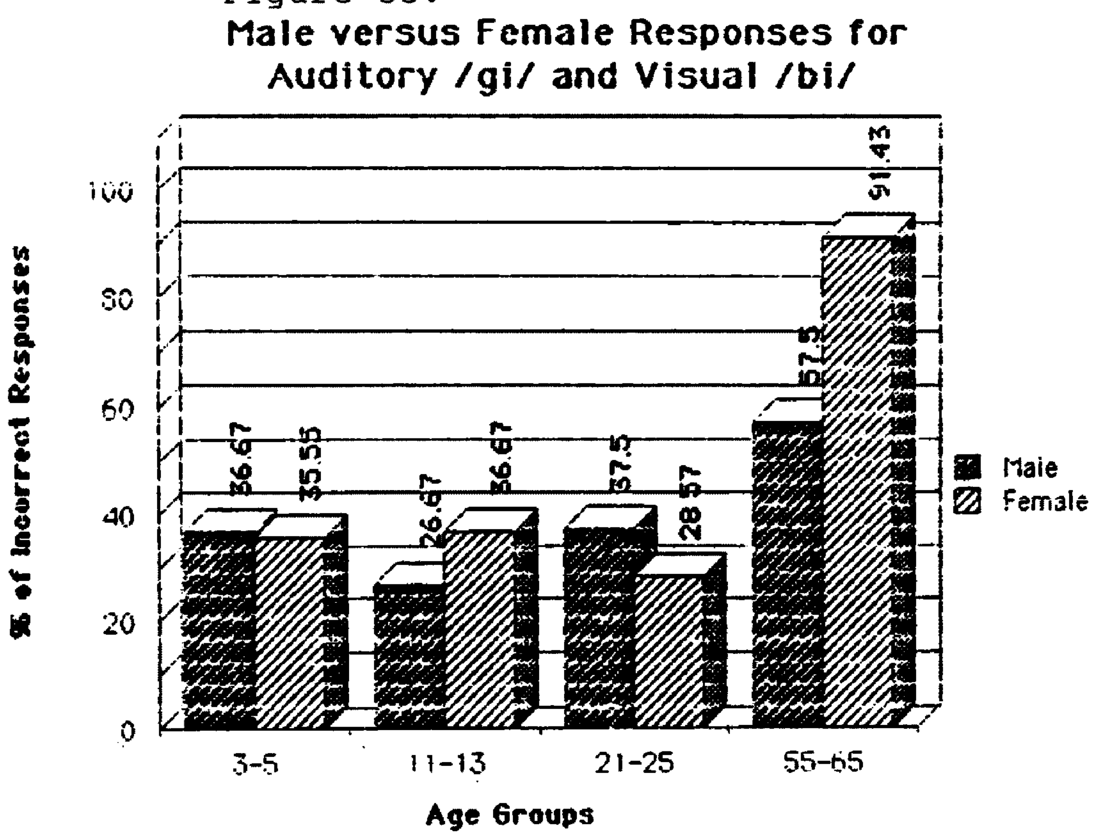


Figure 33.


Male versus Female Responses for
Auditory /gi/ and Visual /bi/
% of Incorrect Responses
100
80
60
40
20
0
36.67
35.55
26.67
36.67
37.5
28.57
57.5
91.43
3-5
11-13
21-25
55-65
Age Groups
Male
Female

Due to the relatively small number of males and females within each subject group, establishing the significance of the differences found across gender is difficult. Additional studies investigating the McGurk illusion, which focus on gender differences, particularly in young children and seniors, are needed to ascertain whether the findings of this study were indeed valid or whether they represent a sampling error due to small number of male and female subjects within each age group.

Summary

The strength of the McGurk illusion using the consonants /b/ and /g/ in combination with the vowels /a/ or /i/ was examined across four subject age groups. The results indicated significant differences and interactions across all main effects including age, vowel, and condition. Contrary to hypotheses presented in the literature, descriptive analysis of the data indicated that the percepts of young children are affected by the auditory/visual conflict of certain stimulus combinations. Further description provided intriguing trends in terms of response types within and across subjects. The manner in which the illusion was manifested varied across age groups and individuals tended to be either markedly affected or not affected by the auditory/visual conflict.

The previously unexplored area of the McGurk illusion in the senior population yielded particularly striking results. Unlike the three younger age groups, the senior subjects were found to be strongly affected by all four experimental conditions of the McGurk illusion. This population was affected approximately equally by the forward and backward illusion and by the two vowel conditions.

Several theories regarding the origin of developmental differences were discussed in light of both the past and present findings. These theories included differences in lip reading skills, physiological capabilities of intermodal perception integration, and attention strategies. Attentional strategies were particularly emphasized as a novel but potentially productive manner of approach for further study.

Descriptive analyses of the data was also performed for relationships involving the variables of gender and response type distribution. Finally, suggestions and discussion regarding the need for further research and possible areas of study within intermodal speech perception were presented.

APPENDIX A. Breakdown of Age and Gender Across Subjects

Grp. A			Grp. B			Grp. C			Grp. D		
Subj.	Gender	Age	Subj.	Gender	Age	Subj.	Gender	Age	Subj.	Gender	Age
1	Male	5;2	1	Male	12;6	1	Female	23;6	1	Male	65;10
2	Male	5;0	2	Male	12;0	2	Female	25;0	2	Male	60;0
3	Male	3;10	3	Male	12;4	3	Male	25;2	3	Female	61;5
4	Female	4;3	4	Female	11;6	4	Female	24;1	4	Male	62;2
5	Female	3;8	5	Female	12;3	5	Male	21;11	5	Female	65;8
6	Female	4;9	6	Male	13;8	6	Male	24;11	6	Female	62;10
7	Female	4;0	7	Male	11;2	7	Female	25;11	7	Male	57;3
8	Female	5;1	8	Female	12;6	8	Female	24;0	8	Female	59;6
9	Female	3;9	9	Female	12;7	9	Female	25;2	9	Male	56;5
10	Male	3;4	10	Female	11;6	10	Male	24;11	10	Female	57;3
11	Male	3;9	11	Male	12;3	11	Male	23;0	11	Male	59;4
12	Female	5;4	12	Female	11;0	12	Female	21;10	12	Female	57;11
13	Female	5;4	13	Male	11;3	13	Male	22;7	13	Male	62;7
14	Female	3;10	14	Female	12;6	14	Male	22;4	14	Male	64;8
15	Male	4;5	15	Male	12;4	15	Male	21;8	15	Male	61;1
	Ratio F - 9 M - 6			Ratio F - 7 M - 8			Ratio F - 7 M - 8			Ratio F - 6 M - 9	

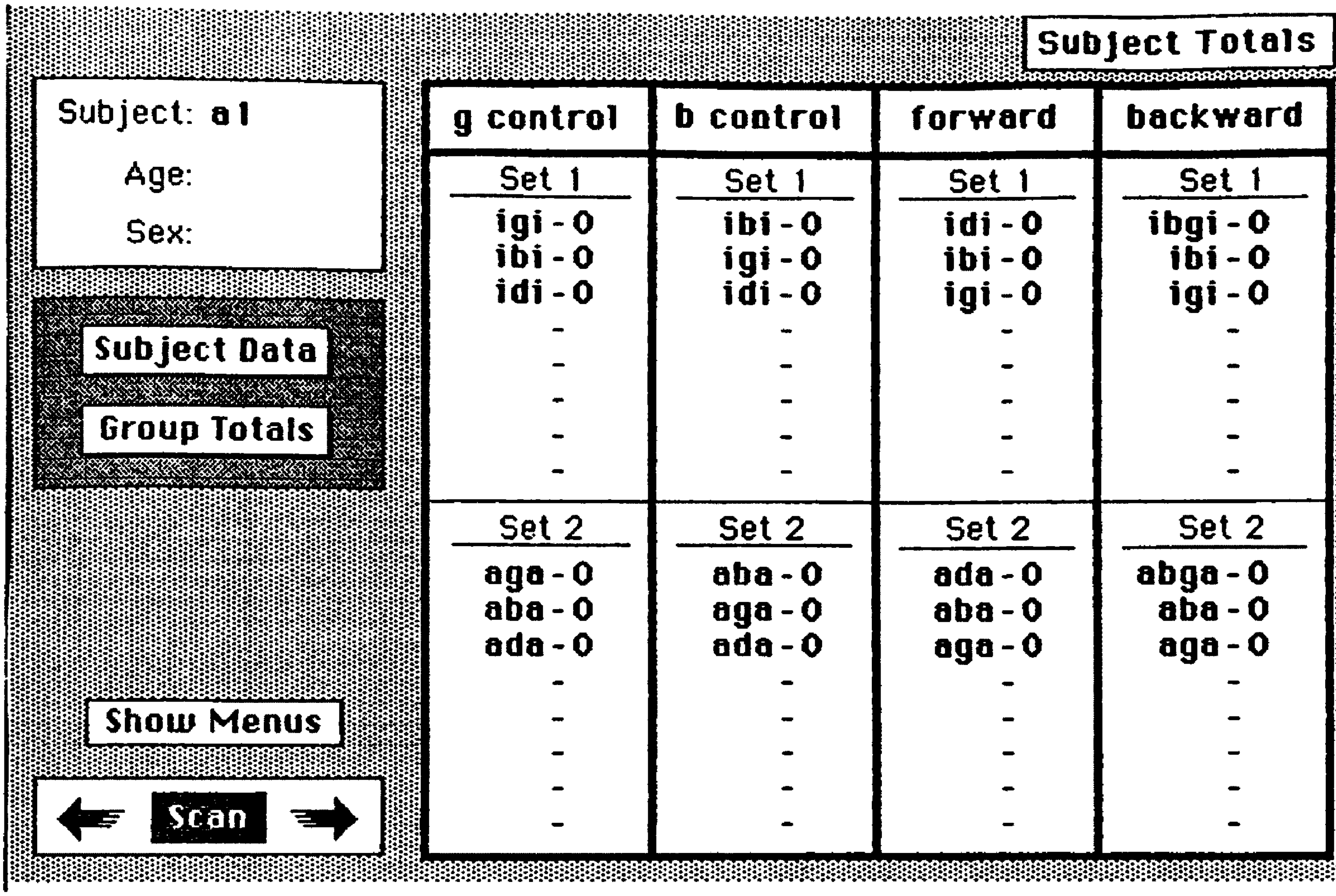
APPENDIX B

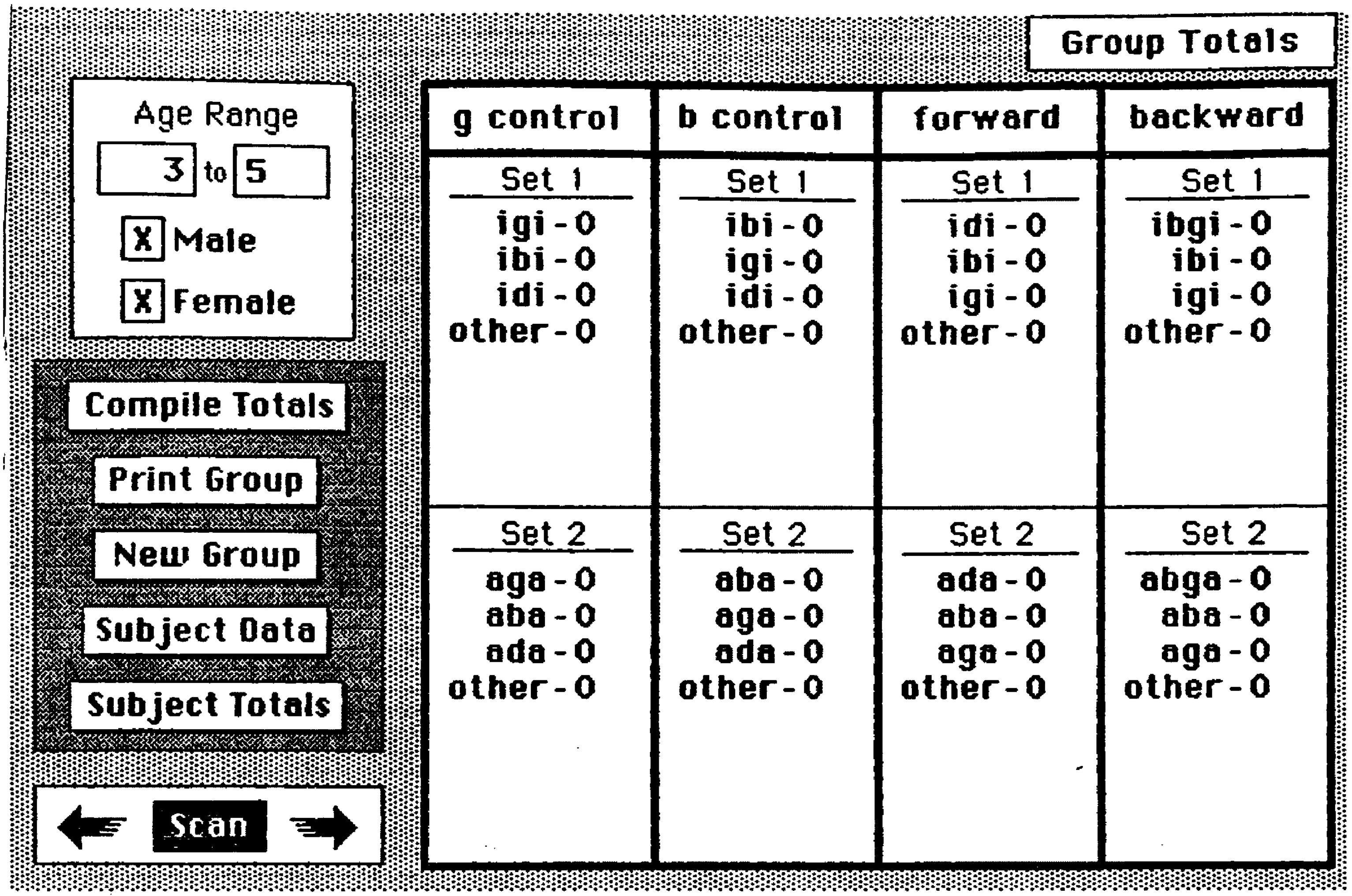
DATA RECORDING SHEET

Subject No.______

TRIAL ITEMS	FIRST SET	SECOND SET
1. __________	1. __________	1. __________
2. __________	2. __________	2. __________
3. __________	3. __________	3. __________
4. __________	4. __________	4. __________
	5. __________	5. __________
	6. __________	6. __________
	7. __________	7. __________
	8. __________	8. __________
	9. __________	9. __________
	10. __________	10. __________
	11. __________	11. __________
	12. __________	12. __________
	13. __________	13. __________
	14. __________	14. __________
	15. __________	15. __________
	16. __________	16. __________
	17. __________	17. __________
	18. __________	18. __________
	19. __________	19. __________
	20. __________	20. __________

APPENDIX C. Data Base Sample for Individual Subject Totals




Group Totals
Age Range
3 to 5
X Male
X Female
Compile Totals
Print Group
New Group
Subject Data
Subject Totals
Scan
g control
b control
forward
backward
Set 1
igi - 0
ibi - 0
idi - 0
other - 0
Set 1
ibi - 0
igi - 0
idi - 0
other - 0
Set 1
idi - 0
ibi - 0
igi - 0
other - 0
Set 1
ibgi - 0
ibi - 0
igi - 0
other - 0
Set 2
aga - 0
aba - 0
ada - 0
other - 0
Set 2
aba - 0
aga - 0
ada - 0
other - 0
Set 2
ada - 0
aba - 0
aga - 0
other - 0
Set 2
abga - 0
aba - 0
aga - 0
other - 0

APPENDIX D. Percentage Results Across Conflicts for the Study of McGurk and MacDonald (1976).

STIMULI Aud	Vis	SUBJECTS	% CORRECT	% INCORRECT
baba	gaga	3-5	19	81
		7-8	36	64
		18-40	2	98
gaga	baba	3-5	57	43
		7-8	36	64
		18-40	11	89
papa	kaka	3-5	24	76
		7-8	50	50
		18-40	6	94
kaka	papa	3-5	62	38
		7-8	68	32
		18-40	13	87

REFERENCES

Bahrick, L.E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relationin audible and visible events. Child Development, 59, 197-209.

Bahrick, L.E., and Watson, J.S. (1985). Detection of intermodal proprioceptive-visual contingency as a potential bases of self-perception in infancy. Developmental Psychology, 21(6), 963-973.

Dodd, B. (1979). Lip reading in infants: attention to speech presented in- and out-of-synchrony. Cognitive Psychology, 11, 478-484.

Dodd, B. (1986). The acquisition of lip-reading skills by normally hearing children. In Campbell R. and Dodd, B. (Eds.) Hearing by Eye, (1986). Erlbaum Associates: London.

Foster, G.A. (1982). The integration of audio-visual speech stimuli as a function of temporal desynchronisation. Proceedings of the Autumn Conference of the Institute of Acoustics (1982), H3.1-H3.5.

Green, K.P. and Miller, J.L. (1985). On the role of visual rate information in phonetic perception. Perception and Psychophysics, 38(3), 269-276.

Jerger, J., Jerger S., Oliver, T. & Pirozzolo, F. (1989). Speech understanding in the elderly. Ear and Hearing, 10(2), 79-89.

Kuhl, P.K., Green, K.P., & Meltzoff, A.N., (1988). Factors affecting the integration of auditory and visual information in speech: the level effect. Paper presented at the meetings of the Acoustical Society of America, May 16, 1988.

Kuhl, P.K. and Meltzoff, A.N. (1982). The bimodal perception of speech in infancy. Science, 218(10), 1138-1141.

Kuhl, P.K. and Meltzoff, A.N. (1984). The intermodal representation of speech in infants. Infant Behaviour and Development, 7, 361-381.

MacDonald, J. and McGurk, H. (1978). Visual influences on speech perception processes. Perception & Psychophysics, 24(3), 253-257.

Massaro, D.W. (1984). Children's perception of visual and auditory speech. Child Development, 55, 1777-1788.

Massaro, D.W. (1986). Speech perception by ear and eye. In Campbell R. and Dodd, B. (Eds.) Hearing by Eye, (1986). Erlbaum Associates: London.

Massaro, D.W. and Cohen, M. M. (1983). Evaluation and integration of visual and auditory information in speech perception. Journal of Experimental Psychology: Human Perception and Performance, 9, 753-771.

Massaro, D.W., Thompson, L.A. Barron, B. & Laren, E. (1986). Developmental changes in visual and auditory contributions to speech perception. Journal of Experimental Child Psychology, 41, 93-113.

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

Summerfield, Q. (1979). Use of visual information for phonetic perception. Phonetica, 36, 314-331.

Summerfield, Q. (1986). Some preliminaries to a comprehensive account of audio-visual speech perception. In Campbell R. and Dodd, B. (Eds.) Hearing by Eye, (1986). Erlbaum Associates: London.

Walker-Andrews, A.S. (1986). Intermodal perception of expressive behaviors: Relation of eye and voice? Developmental Psychology, 22(3), 373-377.

Working Group on Speech Understanding and Aging. (1988). Speech Understanding and Aging. Journal of the Acoustical Society of America, 83(3), 859-894.