National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

The Development of Audiovisual Speech Perception

Neil Spencer Hockley

School of Communication Sciences and Disorders

McGill University, Montreal

July, 1994

A Thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Master of Science.

ISBN 0-612-00028-1

Canada

Abstract

The developmental process of audiovisual speech perception was examined in this experiment using the McGurk paradigm (McGurk & MacDonald, 1976), in which a visual recording of a person saying a particular syllable is synchronized with the auditory presentation of another syllable. Previous studies have shown that audiovisual speech perception in adults and older children is very influenced by the visual speech information but children under five are influenced by the auditory input almost exclusively (McGurk & MacDonald, 1976; Massaro, 1984; and Massaro, Thompson, Barron, & Laren, 1986). In this investigation 46 children aged between 4:7 and 12:4, and 15 adults were presented with conflicting audiovisual syllables made according to the McGurk paradigm. The results indicated that the influence of auditory information decreased with age, while the influence of visual information increased with age. In addition, an adult-like response pattern was observed in only half of the children in the oldest child subject group (10-12 years old) suggesting that the integration of auditory and visual speech information continues to develop beyond the age of twelve.

Résumé

Le processus développemental de la perception audiovisuelle de la parole a été examiné dans cette expérience selon le protocole McGurk (McGurk & MacDonald, 1976) dans lequel un enregistrement visuel d'une personne disant une syllable particulière est synchronisé avec la présentation auditive d'une autre syllabe. Des études antérieures ont démontré que la perception audiovisuelle de la parole chez les adultes et chez les enfants plus agés est très influencée par l'information visuelle de la parole, mais que les enfants de moins de cinq ans sont presqu'exclusivement influencés par l'information auditive (McGurk & MacDonald, 1976; Massaro, 1984; et Massaro, Thompson, Barron, & Laren., 1986) Dans cette expérience, on a présenté des syllabes audiovisuelles conflictuelles, d'après le protocole McGurk, à 46 enfants agés de 4:7 à 12:4 ans, et à 15 adultes. Les résultats démontrent que l'influence de l'information auditive diminue avec l'âge alors que l'influence de l'information visuelle augmente avec l'âge. De plus, des réponses de type adulte ont été observées chez seulement la moitié des enfants du plus vieux groupe (10-12 ans), suggérant ainsi que l'intégration de l'information auditive et visuelle de la parole continue de se développer au delà de douze ans.

Acknowledgments

The British poet John Donne (1572-1631) once wrote that "No man is an island, entire of itself; every man is a piece of the continent". Donne's words are especially true for any research endeavor. I would therefore like to thank a number of people for their invaluable contributions to this thesis.

This work was partially funded by a grant from the Faculty of Medicine, McGill University which was presented to the author and by an NSERC grant #OPG0105397, awarded to Dr. Linda Polka, School of Communication Sciences and Disorders, McGill University.

I am grateful to my supervisor Dr. Linda Polka who provided a lot of advice, support, encouragement, and enthusiasm for this research project. I would also like to thank my thesis committee members Dr. Shari Baum and Dr. Rachel Mayberry who provided some needed comments and suggestions for both the experiment and the completion of this manuscript.

This research would not have been possible without the help and guidance of Dr. Kevin Munhall of the Queen's University Speech Production and Perception laboratory whose help with the stimuli was invaluable. Dr Munhall also introduced me to the phenomenon known as the "McGurk Effect", which has been an interest of mine for the past three years. I would also like to thank Dr. Gloria Waters and Jackie Williams for the use of the video and audio equipment. Jackie Williams along with Dr. Maggie Bruck gave me some much needed advice about statistical software.

I was given a lot of assistance by some students and research assistants from the School of Communication Sciences and Disorders at McGill University. A big thank you is therefore extended to Joanna Fagg for her inspiration, Lagavulin, help with the data collection, and moral support. I am grateful to Ann Sutton and her friends for their help in the recruitment of the child subjects. I would also like to thank Dora Sisto for her help with the collection of the data and for saying "jumping instead of skipping". Miranda Entwistle very diligently transcribed the recordings of the subjects for the reliability measurements; I admire her tenacity. Je voudrais

remercier Claudine Charpentier de son aide pour la traduction de mon résumé. I would like to thank Charlene Chamberlain who gave me some very useful statistical advice. Finally, Margret Orme, provided a lot of moral support and advice; she also survived working with me for almost three years - the two tallest people in the smallest of rooms. I would therefore like to thank her for her patience, kindness, and her wit.

I would like to give a big "thank you" to all of the subjects, especially the children (along with their parents), for their participation in this project, I hope that it was a worth while endeavor.

I would finally like to thank my parents, Dr Bernard and Mrs Jonquil Hockley, for all of their love and support that they have given to me over the years.

Table of Contents

List of Tables....viii

List of Figures....ix

Chapter

1. Introduction....1

2. Audiovisual perception of speech....3

Auditory speech perception with the addition of vision....3

Vision and the degraded auditory signal....4

Vision as a single source of speech information....5

A model of audiovisual speech perception....6

3. The McGurk paradigm....9

Hearing by eye: The McGurk effect....9

The Manner Place Hypothesis....11

Stimulus manipulations that maintain the McGurk effect....13

Stimulus manipulations that reduce the McGurk effect....15

The McGurk effect in Japanese listeners....18

The McGurk effect in second language learners....21

Conclusions....22

4. Age differences and developmental aspects of audiovisual speech perception....23

The infant's use of auditory and visual information....23

The effect of vision on phonological development....26

Audiovisual development in young children.....27

The present investigation.....32

5. Method.....35

Subjects.....35

Stimuli.....36

Equipment.....42

Procedure.....42

Scoring and Analysis.....44

Inter-transcriber reliability.....44

Scoring method: Auditory Only, Visual Only and AV Conditions.....46

Scoring method: AVC Condition.....46

6. Results.....50

AV, Visual Only, and Auditory Only Conditions.....50

AVC Condition.....54

Confusion matrices.....57

Analysis according to an adult-like pattern.....62

Summary of the results.....64

7. Discussion and conclusions.....66

Age effects in unimodal and bimodal speech perception.....66

Age effects in the perception of conflicting visual and auditory speech information.....70

Overall conclusions, limitations, and directions for future research....74

Clinical implications....81

Summary....83

References....85

APPENDIX A....92
Stimuli and frame number from the Johns Hopkins Lip-reading Corpus Volume 1 (Bernstein and Eberhardt, 1986).

APPENDIX B....93
Order of presentation of the Auditory Only and Visual Only stimuli on Tape 1.

APPENDIX C....95
Alignment of the videodisk file with the digitized audio file in msec for the AV stimulus condition.

APPENDIX D....96
Alignment of the videodisk file with the digitized audio file in msec for the AVC stimulus condition.

APPENDIX E....97
Order of presentation of the audiovisual (AV and AVC) stimuli on Tape 2.

APPENDIX F....99
Language Questionnaire presented to the adult subjects.

APPENDIX G....100
Instructions given to the adult and child subjects for the Auditory Only, Visual Only, AV, and AVC Conditions.

APPENDIX H....101
Score sheets for the Auditory Only, Visual Only, AV, and AVC Conditions.

List of Tables

Table1.....10
Exarnple of stimuli and perceptual results from McGurk and MacDonald (1976).

Table 2.....36
Age range for the four child subject groups.

Table 3.....37
Visual productions of the four AVC stimuli paired with the auditory syllable /ba/.

Table 4.....46
Percent agreement between the two transcripts for the four child age groups and the adult controls.

Table 5.....47
Verbal responses defined as Visual Capture (VC) according to Werker et al (1992).

Table 6.....49
Verbal responses defined as Blends according to Werker et al (1992).

Table 7.....59
Proportion of responses to the AVC stimuli /ba/+/va/ and /ba/+/ða/.

Table 8.....61
Proportion of responses to the AVC stimuli /ba/+/da/ and /ba/+/ga/.

List of Figures

Figure 1.....7
A model of audiovisual speech perception by MacLeod and Summerfield (1987).

Figure 2.....51
Mean proportion of correct responses for the AV, Auditory Only, and Visual Only Conditions.

Figure 3.....55
Mean number of responses defined as AC, VC, Blends and Other in the AVC Condition.

Figure 4.....63
Percentage of subjects classified as meeting the adult-like pattern criterion.

Chapter 1

Introduction

Speech communication, in its most natural form, occurs in a context where an individual has visual contact with the person with whom he/she is conversing. Therefore, the study of speech perception should consider how speech information is processed via the auditory channel as well as the visual channel to derive a complete picture of the processes involved. Studies of audiovisual speech perception have clearly shown that information provided visually facilitates speech comprehension in both favorable and unfavorable listening conditions. The way in which listeners make use of both auditory and visual information in speech perception is revealed when visual speech information is placed in conflict with auditory speech information. This can be seen in the landmark studies of McGurk and MacDonald (1976; MacDonald & McGurk, 1978) in which subjects were presented with a visual sequence of a face articulating one phoneme synchronized with the audio sequence corresponding to a different phoneme. These researchers found that when the visual speech information does not correspond with the auditory speech information an entirely new phoneme can be perceived. This effect of changing speech perception by manipulating visual speech patterns is now referred to as the "McGurk effect". The McGurk effect was an important finding because it clearly demonstrated that visual speech information strongly influences speech perception in adults.

In these early studies, a compelling McGurk effect was observed in adult subjects; however, a clear effect was not evident in

young children. This finding suggests that the ability to interpret visual speech information and to integrate auditory and visual speech information is a product of development. There have been numerous studies exploring the McGurk effect, but there has been very little work addressing the developmental aspects of this phenomenon. This study examined the developmental course of audiovisual speech perception.

Background literature relevant to audiovisual speech perception is presented in Chapters 2, 3, and 4. Chapter 2 provides an overview of research findings on audiovisual speech perception, that is, studies in which the auditory and visual information correspond to the same articulation. In Chapter 3, studies conducted within the McGurk paradigm are discussed. Finally, the focus of Chapter 4 is on age and developmental differences in the use of visual information in speech perception.

Chapter 2

Audiovisual perception of speech

This portion of the literature review will describe studies of audiovisual speech perception. This discussion will consist of four sections. The first section summarizes literature showing that visual information can add to the information that is already present in the auditory channel to aid in the perception of speech. The second section describes some studies that show how the use of visual information can improve the perception of speech when auditory information is degraded. In the third section, evidence will be presented to demonstrate that, despite the usefulness of the visual channel to supplement information received through the auditory channel, vision is not adequate as a single source of speech information. The fourth and final section of this chapter will present a simple model of audiovisual speech perception and will explain how different manipulations of auditory and visual information might be used to explore the integration of these two senses.

Auditory speech perception with the addition of vision

In speech perception, redundant information is provided by the fact that spoken language simultaneously manifests itself in the visual and auditory channels (Byman, 1974). Even though the speech information contained in the visual channel can be extraneous when presented with auditory information, this does not imply that the addition of this information is not useful to the listener. Woodward and Barber (1960) found that phonemic distinctions such as /pa/ and /ka/ are most easily perceived in audiovisual situations. They found that these phonemes were not as easily differentiated

when the auditory information was presented without the visual information. Binnie, Montgomery, and Jackson (1974) found that adding visual information to the auditory channel significantly reduced confusions between CV (a consonant with a vowel, e.g. /ba/) stimuli that differed in place of articulation. Reisberg, McLean, and Goldfield (1987) found that the addition of visual information can aid in the comprehension of a clear acoustic signal for individuals with normal hearing ability. They found that the comprehension of spoken text passages improved by 15% when the speaker was seen and heard instead of just being heard. Thus, visual information can be relied upon for clarification of what is contained in the auditory channel.

Vision and the degraded auditory signal

In speech perception, adults often benefit most when visual information is made available in difficult as well as optimal listening settings. For example, Sumby and Pollock (1954) found that viewing the face of a speaker improved the understanding of speech in noise. The amount of noise interference that could be presented to the subjects in the audiovisual condition, as opposed to the auditory only condition, ranged from between 5 to 22 dB to maintain the same accuracy level in a spondee word repetition task. This led to the conclusion that if visual information can supplement auditory information, then perhaps individuals can tolerate higher noise interference levels than when visual information is not provided.

Summerfield (1979) found similar results to Sumby and Pollock (1954) for test messages embedded in irrelevant passages of prose. Likewise MacLeod and Summerfield (1987) found that Speech

Reception Thresholds (SRT's) for sentence materials in noise can be improved by 6 to 15 dB with the addition of visual information. They also found that key words in their sentence stimuli were correctly reported, when presented at an average signal to noise ratio (S/N) of -11.8 dB in auditory alone conditions as compared to a S/N ratio of -22.8 dB in audiovisual conditions. Thus, the presence of visual information enables listeners to recognize words at higher noise interference levels as compared to the noise levels that support speech recognition for a solely auditory input.

Vision as a single source of speech information

Despite the profound impact of visual information on speech perception, it is clear that visual information alone cannot serve as a substitute for the information that is provided acoustically. Plant and Macrae (1987) emphasize that visual information supplements auditory information but is not a substitute for what is provided acoustically. Massaro (1989) states that vision is inadequate as a single source of information about speech. Visible speech articulations alone do not provide the listener with a complete representation of the speech signal. Massaro (1989) uses the example that silent televisions displaying only lip movements are not a feasible means of precise communication in the same way that a faceless telephone is.

The limits of visual speech perception become apparent when one considers that each phoneme does not have a distinct visual correlate. Woodward and Barber (1960) showed that fewer than half of the English phonemes are perceptible when just visual articulatory information is presented. Furthermore, phonemes that have visible

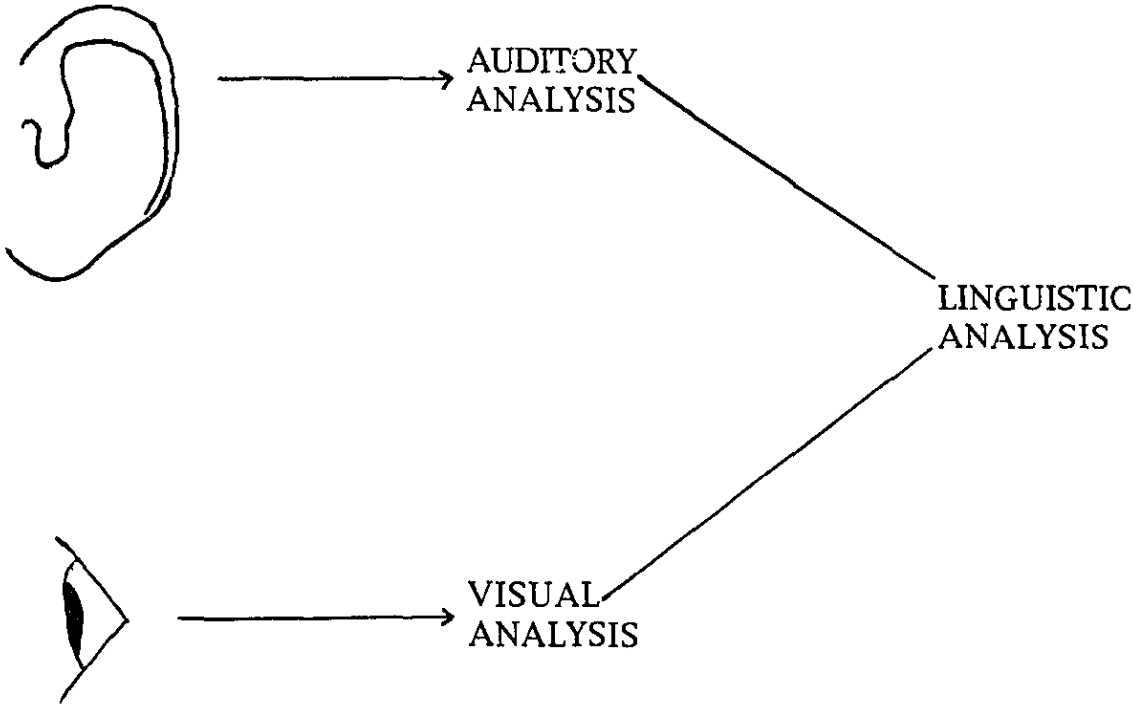
articulatory gestures are often vulnerable to confusions (Fisher, 1968). For example /p/ can be confused with /b/ and /f/ can be confused with /v/. These pairs of phonemes appear to be visually the same, according to the shape of the lips; they can be differentiated acoustically, however, by the presence/absence of voicing. Thus, when the lips have the configuration of /b/ or /p/, if voicing is heard, then /b/ is perceived; likewise, when the lips have the configuration of /f/ or/v/, if voicing is heard, then /v/ is perceived. According to Denes (1963), five of the most frequently occurring consonants in spoken English (/t/,/n/,/s/,/d/, and /l/) are produced in the same place in the mouth and differ only with respect to manner of articulation. Therefore these five frequently occurring English consonants are not discriminable through vision alone.

It has been found that vision provides cues about the place of articulation while audition provides information about the manner of articulation and voicing (Miller & Nicely, 1955; Dodd, 1977, 1980; MacDonald & McGurk, 1978). More specifically, Walden, Prosek, and Worthington (1975) found that visual cues can enhance the transmission of information about duration, place of articulation, frication and nasality but do not enhance certain manner distinctions (e.g. liquid/glide) and voicing information. It is thus clear that even though visual information can complement auditory information, it must not be considered as a pure substitute for the information that is attained via the auditory channel.

A model of audiovisual speech perception

MacLeod and Summerfield (1987) proposed a simple model to explain audiovisual speech perception.

Figure 1. A model of audiovisual speech perception by MacLeod and Summerfield (1987).



This model assumes that there are three processes involved with audiovisual speech perception: auditory analysis, visual analysis, and linguistic analysis. Lip-reading depends upon the process of visual and linguistic analysis while auditory speech perception depends upon the process of auditory and linguistic analysis. These three processes analyze the information contained in speech and occur simultaneously. Therefore, meaning can be derived simultaneously from the articulation movements that are processed visually and the acoustic signals which are processed auditorially. By examining the output of all three of these processes together, and in different combinations, it is possible to explore the nature of audiovisual speech perception. It is for this reason that the

McGurk paradigm is a useful tool to examine how speech is perceived. In the McGurk paradigm it is possible to observe the output or result of linguistic analysis when auditory and visual analyses are manipulated independently of each other. For example, the auditory input can be kept constant while the visual input is manipulated. This permits the separation of the contributions of each channel to the final percept and the determination of the factors influencing the interaction between audition and vision or all three levels of analysis.[1]

---

[1] This model was presented to illustrate the interaction between the auditory and visual channels of information. This model was not presented to draw any conclusions about when and how the analysis of auditory and visual information is accomplished. The focus of this study is on the perception of speech by children, with different manipulations of the auditory and visual channels of information and not the specific mechanisms behind the observed processes.

Chapter 3

The McGurk paradigm

This portion of the literature review focuses on research that has employed the McGurk paradigm in which the auditory and the visual speech signals are manipulated independently from the other. This discussion of the McGurk paradigm consists of seven sections. The first section presents the initial research on the McGurk paradigm. In the second section the Manner/Place hypothesis, developed by MacDonald and McGurk (1978), is described. The third section summarizes a number of studies that show that the McGurk effect is not affected by some stimulus manipulations. The fourth section then describes three studies that show instances where the magnitude of the McGurk effect can be reduced. The fifth section presents studies of the McGurk effect elicited through the presentation of Japanese stimuli to Japanese listeners. The sixth section describes the findings from one study which examined the McGurk effect in second language learners. The seventh section presents some conclusions about the McGurk paradigm studies that have been presented.

Hearing by Eye: The McGurk effect

In the early 1970's McGurk designed some stimuli to see if infants could differentiate between mismatched faces and voices. He dubbed a film of a person uttering nonsense syllables with another sound track consisting of these same nonsense syllables in a different order. Visual articulatory movements were thus paired with different acoustic speech segments. He expected that adults, watching mismatched faces and voices, would detect a "conflict"

between the visual input and the auditory input. He expected adults to notice that something was wrong with the stimuli, that is, that the voices and the faces did not match. However, this was not the case because McGurk found that adults failed to detect such a mismatch. Instead, their responses indicated that they had integrated the visual and auditory speech information that was presented to them, to perceive something new. Examples of McGurk's stimuli and the perceptual results that were reported are shown in Table 1.

Table 1. Example of stimuli and perceptual results from McGurk and MacDonald (1976).

Stimuli			Perceptual Result
"BA" Voice	+	"BA" Lips	"BA"
"GA" Voice	+	"GA" Lips	"GA"
"BA" Voice	+	"GA" Lips	"DA"
"GA" Voice	+	"BA" Lips	"BA" or "BGA"

As shown in Table 1, when /ba/ voice was presented with /ga/ lip movements the listeners perceived /da/. However, when /ga/ voice was presented with /ba/ lip movements, /ba/ or /bga/ was perceived. After McGurk presented these stimuli to some additional subjects and colleagues, he realized that this was evidence of a new perceptual phenomenon: "hearing by eye", or as it is more commonly referred to, the "McGurk Effect" (McGurk & MacDonald, 1976).

After further research (MacDonald & McGurk, 1978; summarized in McGurk, 1988), McGurk made three statements about the phenomenon that he had observed in adults. First, even when subjects are informed that they will be presented with conflicting stimuli they are rarely confused about what they perceive. Second, he surmised that adults perceive a unified stimulus when presented with conflicting auditory and visual information because their years of experience with natural conversation has produced a powerful expectation that lips and voices will say the same thing. Third, McGurk (1988) suggested that a visual bias is observed in audiovisual speech perception, which occurs because adults have an implicit understanding of the constraints placed on the speech production process by the visible articulators. In other words, when we view a talker's face, we are quite reluctant to "hear" speech that is normally articulated at the front of the mouth unless we can see the frontal articulators in the appropriate motion. For example, we are less likely to hear the bilabial stop /b/ unless we can also see the lips closing. We also realize that much of the articulation that goes on in the mouth is not available for visual inspection, therefore we are less influenced by visual information when perceiving sounds that are produced out of view. Thus, some visual articulations are more informative than others. These conclusions provide a coherent interpretation of McGurk's findings in adults.

The Manner Place Hypothesis

MacDonald and McGurk (1978) have argued that the auditory/visual illusion generated within the McGurk paradigm is not a laboratory artifact elicited by peculiar auditory/visual

combinations. Rather, they propose that the McGurk effect is a robust demonstration of the powerful effects of visual information on speech perception. To explain this finding, MacDonald and McGurk (1978) proposed a possible process for audiovisual speech perception where the manner of articulation and the voicing of consonantal utterances is more reliably determined through audition and place is more often determined through vision. Therefore, whether a phoneme is voiced or unvoiced, nasal or nonnasal, stop or continuant is specified by information in the auditory channel, whereas, whether a phoneme is bilabial or velar is specified by information in the visual channel. At some level of processing the visual and auditory information is combined. If a mismatch exists then a "best fit" solution is presented. This can be seen when /da/ is perceived from /ba/ voice and /ga/ lips. The auditory information would convey acoustic features of both /da/ and /ba/ (e.g. rising F1) while the visual information would convey articulatory features of /da/ and /ga/ (e.g. back place of articulation). By responding to the common features shared by both modalities the syllable /da/ is perceived. This theory of using a "best fit" solution was further developed by Massaro who developed a quantitative model (see Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Massaro, 1987; Massaro, 1989; Massaro & Cohen, 1990; Massaro, Cohen, Gesi, & Heredia, 1993 for a complete discussion) using the principles of fuzzy logic to describe the processes of audiovisual speech perception.

Stimulus manipulations that maintain the McGurk effect

Investigations of the effects on speech perception of different auditory and visual syllables presented simultaneously have revealed that the McGurk effect is quite robust. The results of these studies have further clarified the nature of audiovisual speech perception for adult subjects. Some of these studies have shown little or no change in the McGurk effect with various stimulus manipulations. For example, Manuel, Repp, Liberman, and Studdert-Kennedy (1983) found very similar results to McGurk and MacDonald (1976; MacDonald & McGurk, 1978) when they presented adults with an isolated auditory vowel paired with a visually displayed CV syllable. Manuel et al. (1983) wanted to determine if the McGurk effect would still occur when the acoustic signal was completely devoid of consonant manner cues. That is, would the subject still report hearing a consonant when the acoustic signal was just an isolated vowel? Manuel et al. (1983) constructed a four element auditory continuum from /ba/ to /a/ by truncating the natural speech production of /ba/ in three steps to eliminate the first 10, 20, and 50 msec of the syllable. They then paired the acoustic signals with the visual productions of /ba/, /va/, /ɵa/, and /da/. They found that the McGurk effect did in fact remain; therefore, the visual information still had an effect on the perception of the syllable, even when the amount of consonantal information in the auditory channel was severely reduced. The visual channel thus provided the information about the consonant to the listener and he/she thus perceived a CV syllable.

The McGurk paradigm has not only been employed to study isolated speech elements such as CV syllables but also to study the perception of words and continuous speech. Dekle, Fowler, and Funnell (1992) examined the integration of auditory and visual information in the perception of real words. An example of their stimuli is the auditory production of "met" presented with the visual production of "gal". These stimuli commonly produced the McGurk fusion response of "net". Dekle et al. (1992) concluded that the phonetic information that is available in the visual modality can alter the perception of speech for words as well as syllables. McGurk (1988) showed that the McGurk effect can also be produced with continuous speech, as was demonstrated in an example reported on the radio program "Science Now" on the BBC in Britain.

VOICE: "My bab pope me pu brive"
LIPS: "My Dad taught me to drive"
HEARD: "My Dad taught me to drive"

(as cited in McGurk, 1988)

It appears to be the case that the visual channel has an effect on natural continuous speech as well isolated words and syllables, thus illustrating the powerful influence that vision has upon the perception of speech.

The McGurk effect also appears to be quite resistant to some other discrepancies between the auditory and visual channels. Ward (1992) found that a strong visual influence is maintained in adult perception even when there is an asynchrony between the auditory

and visual signals which is as great as 300 msec. The McGurk effect was preserved when the auditory signal preceded or followed the visual signal. Green, Kuhl, Meltzoff, and Stevens (1991) also found that the McGurk effect is unaffected by a mismatch in the gender of the talker producing the auditory stimulus. In this study the subjects viewed video tapes of male voice/female face and female voice/male face as well as video tapes of male voice/male face and female voice/female face. Green et al. (1991) found that there were no significant differences in the magnitude of the McGurk effect between the different tape presentations. This led Green et al. (1991) to conclude that the process of audiovisual integration is not sensitive to a gross difference such as gender. Finally the McGurk effect is also resistant to modality changes. Fowler and Dekle (1991) found that McGurk-like effects could be produced with the Tadoma method of speech perception. This is a method used by deaf/blind individuals in which the sense of touch is used to gather speech information from the talker's mouth and neck. Fowler and Dekle (1991) found, that for normal subjects, when the tactile stimulus was in conflict with the auditory stimulus, a new percept was generated. Together these studies show that, for adults, when unambiguous visual (or tactile) speech information is available, it becomes integrated with speech information that is available via the auditory channel.

Stimulus manipulations that reduce the McGurk effect

Although the McGurk effect has been elicited in several ways, three studies have shown the McGurk effect to be reduced under certain stimulus conditions. For example, Green, Kuhl, and Meltzoff

(1988) found that the magnitude of the McGurk effect is not the same across different vowel environments. These researchers investigated the number of McGurk effect responses (e.g. /d(vowel)/) to the visual /g(vowel)/, auditory /b(vowel)/ stimuli with the vowels /a/, /i/, and /u/. The strongest effects were found for the /i/ vowel, moderate effects were found for the /a/ vowel, and almost nonexistent effects were found for the /u/ vowel.

Green et al. (1988) state that during the articulation of /i/ and /a/ the mouth and lips are either spread or open which provides the subject with a clear view of the oral cavity. During the articulation of /u/ however, the lips are narrowed, protruded and rounded so that the subject cannot see the interior cavity. The information in the visual channel is thus reduced in this /u/ vowel environment. Thus, these results are consistent with the hypothesis that the visual speech signal will have a strong influence when it provides unambiguous phonetic information. Given that the various features of place of articulation, release burst, and the direction and extent of formant transitions vary significantly as a function of vowel space, Green et al. (1988) proposed that there might be some type of perceptual weighting system that is modality specific and that varies depending on the vowel context. That is, a particular syllable might have a particular code composed of different auditory and visual elements.

In another study, Kuhl, Green, and Meltzoff (1988) found that the influence of the visual channel on the McGurk effect decreased as the signal intensity level was reduced. These results appear counterintuitive, at first, but Kuhl et al. (1988) suggest that this

effect can be attributed to the subjects' knowledge of what happens when a person raises his/her voice. When a person talks loudly their articulatory movements become more pronounced. According to Kuhl et al. (1988) exaggerated articulatory movements are more informative to a perceiver when attempting to decipher a message. When a person talks quietly his/her articulatory movements are less pronounced and are therefore not as informative. As a result when presented with quiet speech, the perceiver needs to pay more attention to the auditory channel because the visual information is less accessible. Therefore Kuhl et al. (1988) believe that loud speech may increase the attention paid to the visual modality and thus increase the effects of visual information on the perception of speech.

Green (1994) performed a study which examined the effect that inverting the talker's face would have on the McGurk effect. Two groups of subjects were presented with a videotape of stimuli that were designed to produce a typical McGurk effect. The first group was presented with the stimuli on a monitor in an upright position while the second group was presented with the stimuli on a monitor in the inverted position. Green (1994) found that the second group experienced a significantly weaker McGurk effect than did the first group. It was thus concluded that inverting the face and subsequently the articulatory movements affects the integration of auditory and visual speech information.

These three studies show how the McGurk effect can be changed when certain stimulus parameters are modulated. Overall the findings of Green et al. (1988), Kuhl et al. (1988), and Green (1994) show that the McGurk effect is reduced in stimulus conditions

which decrease the phonetic information that is available via the visual articulation.

The McGurk effect observed for Japanese listeners

The McGurk effect has been examined extensively in English with English speaking listeners. Recently, three interesting studies have been conducted on the McGurk effect with Japanese listeners. Sekiyama and Tohkura (1991) found a weaker McGurk effect for Japanese listeners with Japanese stimuli than was found by other researchers who presented English stimuli to English listeners. Sekiyama and Tohkura (1991) suggest that this finding may be due to structural differences between Japanese and English. For example, the Japanese phonological system does not permit any consonant clusters at a systematic phonemic level. Therefore, the perceived combination "bda" for the auditory /da/ combined with the visual /ba/, which is characteristically seen in English subjects, was rarely seen in this study with Japanese subjects. They also suggest that there may be cultural differences between native Japanese and native English speakers and that this issue should be pursued in a follow-up study.

Sekiyama and Tohkura (1993) investigated the language and cultural aspects of the findings observed in the 1991 study. Sekiyama and Tohkura (1993) tested one group of Japanese subjects with Japanese syllables, another group of Japanese subjects with English syllables, a group of American subjects with Japanese syllables, and another group of American subjects with English syllables. It was found that more visually biased responses were produced by the American subjects with the Japanese stimuli than

the Japanese subjects with the Japanese stimuli. The Japanese subjects thus produced weak McGurk effects when presented with stimuli from their native language, whereas the American subjects produced strong effects in a non-native language. No significant differences were found however between the Japanese subjects and the American subjects for the English stimuli. The Japanese subjects thus produced more McGurk effects for stimuli from a non-native language, but these effects were only comparable to those seen in American subjects with a native language. Therefore, it was found in general that the McGurk effect occurred more frequently for the non-native languages in both the American and Japanese groups. Sekiyama and Tohkura (1993) concluded that the American subjects were more easily influenced by the visual speech information than were the Japanese.

Sekiyama and Tohkura (1993) surmised that the results obtained were not due to the stimuli used, but to the different ways that the American and the Japanese subjects use visual speech information. Specifically, the Japanese do not look at the face of the person they are listening to out of politeness. Sekiyama and Tohkura (1993) also point out that Japanese has a simpler phonological structure, for example, Japanese does not contain the consonants /v/, /ɵ/, and /r/ . This simpler phonological structure may permit native Japanese speakers to discriminate one Japanese syllable from another without the additional help that vision provides. English has a more complicated phonological structure and therefore vision, as was demonstrated in Chapter 2, becomes more important to the native English listener. It would appear, from the evidence

presented in this study, that the visual channel does not provide as much information to the Japanese listener as it does to the English listener in the perception of his/her native language.

Sekiyama and Tohkura (1991) indicated that the intelligibility of the auditory information may be a factor in explaining their findings. These researchers found that when the intelligibility of the auditory stimulus alone was 100% the McGurk effect was absent or very weak. If the intelligibility was less than 100% the McGurk effect could be induced. This indicates that when the information cannot be found or is difficult to perceive through the auditory channel, greater weight is assigned to the information available in the visual channel. These findings led the researchers to question the role of auditory intelligibility in some of the earlier McGurk paradigm studies. Sekiyama and Tohkura (1991) suggest that further research should be performed to examine the effect of auditory intelligibility.

In another study, Sekiyama and Nishino (1994) performed an experiment to examine the issue of intelligibility. In this study the Japanese subjects were presented with Japanese syllables at four different sound intensity levels (65, 58, 51, and 44 dB) combined with four white noise conditions (56, 48, 40, and 32 dB). It was found that when the noise level was less than the speech level the McGurk effect was not as strong. This supported the evidence presented by Sekiyama and Tohkura (1991) which suggested that the McGurk effect is stronger when auditory intelligibility is poor.

In general, Sekiyama and Tohkura (1993) suggest that the American listeners automatically integrate visual cues with auditory

cues whereas Japanese listeners only utilize visual information in difficult listening situations. The results from these three experiments (Sekiyama & Tohkura, 1991, 1993; Sekiyama & Nishino, 1994) suggest that both cultural and intelligibility factors should be taken into account in investigations of the McGurk effect.

The McGurk effect in second language learners

Werker, McGurk and Frost, (1992) investigated the effects of second language development on the elicitation of the McGurk effect. They found that French Canadians who were studying English, and English Canadians had differing perceptions of conflicting English auditory/visual syllables. These researchers used a variety of visual CV stimuli that are common in both English and French: /ba/, /va/, /da/, /ʒa/, and /ga/ and an interdental fricative /ðа/ which is present in English but not in French. Visual versions of these CV stimuli were paired with the auditory stimulus /ba/. For the interdental stimulus /ðа/ some of the Francophones reported hearing a /da/ or /ta/. This indicated an assimilation of the visual information to the nearest place of articulation in the French language. The English subjects reported hearing a /ðа/ or a /bða/ which is consistent with the presence of these syllables in the English language. Thus, both the English and French subjects showed that they were integrating the auditory and visual information. However, their percepts were quite different. Thus, it appears that the listener assimilates speech information obtained through both the visual and auditory channels, to the phonology of their native language. For French subjects who had the greatest proficiency in English their response to the /ba/voice and /ðа/ lips stimulus had shifted to the

English pattern. The results from this study suggest that experience with a language facilitates the visual as well as the auditory perception of that language and the integration of the two sources of information.

Conclusions

The studies that have been presented in this Chapter have revealed that the McGurk paradigm is a very useful tool to explore how individuals use visual and auditory information. The McGurk effect appears to be resistant to some stimulus manipulations but not to others. For example, the McGurk effect is resistant to asynchrony between the auditory and visual channels, modifications of speaker gender, and manipulations of modality. On the other hand, the McGurk effect can be reduced by different auditory vowel environments, intensity levels of the auditory stimulus, and spatial orientations of the visual stimulus in the vertical plane. These studies reveal not only the strengths but the limitations of this experimental paradigm and the elicited perceptual effects.

Chapter 4

Age differences and developmental aspects of audiovisual speech perception

This Chapter of the literature review examines age differences along with other developmental aspects of audiovisual speech perception and also outlines the groundwork for the rationale behind the current investigation. This Chapter consists of four sections. The first section examines some precursors to the development of auditory and visual integration of speech. The second section demonstrates the important role which vision plays in phonological development. In the third section, a number of studies that explore how young children utilize auditory and visual information in their perception of speech are presented. The fourth and final section of this Chapter presents the rationale for the current investigation.

The infant's use of auditory and visual information

Developmental research has shown that infants are aware of some correspondence between the auditory and visual channels of information. These studies have been performed with both speech and non-speech stimulus materials. Humphrey, Tees, and Werker (1979) presented four month old infants with light emitting diodes (LED's) and tones. The visual and auditory stimuli were both in and out of synchrony with each other. It was found that the infants looked longer at the visual display that was synchronized with the tones. Dodd (1979) presented nursery rhymes spoken with both asynchronous and synchronous (by 400 msec) acoustic signals and lip movements to 16 week old infants and found that infants preferred to attend to the stimuli that were in synchrony. From

these two studies it appears that infants have a tendency to be interested in auditory and visual perceptual events that are temporally coordinated. This is important for speech perception because the speech signal is simultaneously manifested in the auditory and visual domains.

The studies by Humphrey et al. (1979) and Dodd (1979) show that infants can detect the correspondence between auditory and visual events for non-speech material and connected discourse, but can infants detect this correspondence in speech elements at the phonetic level? Kuhl and Meltzoff (1982) found that 18-20 week old infants can detect the correspondence between auditory and visual speech information for the vowels /i/ and/a/. Infants in their study preferred to look, when given the choice, at films where the auditory information referred to the same vowel as the visual information. This finding shows that infants' preference for the coordination of lips and voices exists for isolated elements of speech. Kuhl and Meltzoff (1982) proposed that infants can relate specific articulatory postures to the corresponding speech sounds. In a later study Kuhl and Meltzoff (1984) showed that infants respond to the spectral information as well as the temporal characteristics of speech sounds. This suggests that infants do not rely on the simple auditory features of timing and amplitude to link visual and auditory speech events. Thus, the perceptual mechanisms that the infants are using appear to be more complex than was previously thought.

In a study, which reported similar findings to those of Kuhl and Meltzoff (1982, 1984), Walton and Bower (1993) used an operant habituation/dishabituation nonnutritive sucking technique. In this

procedure the infants controlled (by varying their sucking rate) the presentation of different audiovisual representations of /u/ and /a/ where the visual signal either matched or did not match the auditory signal. When the audiovisual stimulus shown to the infant changed, both the auditory and the visual signals were changed simultaneously. If the infant sucked on the pacifier with an interval between sucks of less than one second (habituation) then the audiovisual combination that was first presented would be shown repeatedly to the infant. If, however, the interval between sucks was greater than one second (dishabituation), then the next audiovisual combination was shown to the infant. If at anytime the infant stopped sucking, a neutral face with no sound was presented to the infant. As was found in the previously discussed studies, the infants dishabituated and therefore preferred to look at the stimulus of the face which matched the voice.

In a continuation of this research, Walton and Bower (1993) visually presented the infant with a phoneme and auditorially presented the correct articulation, an incorrect articulation, and a "possible" articulation. The stimuli consisted of the visually presented /u/ phoneme paired with the auditory presentations of /u/ (correct), /i/ (impossible), and the French phoneme /y/ (possible). Walton and Bower (1993) found that the infants preferred to look at the visual /u/ paired with the correct auditory articulation, /u/, or the possible auditory articulation, /y/. These results offer further evidence to suggest that infants are aware of some relationship between visual and acoustic speech patterns.

The findings of these studies by Kuhl & Meltzoff (1982, 1984) and Walton & Bower (1993) clearly show that infants detect some correspondence between auditory and visual speech events. Specifically, infants prefer visual and auditory events which correspond to each other. It is important to recognize, however, that these studies do not inform us about the specific elements of speech that the infants are perceiving when exposed to the audiovisual information. In contrast to these studies, investigations of audiovisual speech perception using such techniques as the McGurk paradigm examine identification as opposed to preference responses thereby supplying data that explicitly addresses what the subjects perceive under different conditions. Therefore, the conceptual link between infant preference studies and the findings of studies which tested older children using the McGurk paradigm and other identification methods (described below) is not yet clear.

The effect of vision on phonological development

Studies performed with young children have illustrated the importance of visual information in the development of speech perception and production. For example, Dodd (1987) found that young children aged 19 to 36 months can lip-read familiar words. In this study the children were presented with ten sets of three pictures (e.g. dog, door, bed) and were asked to point to specific target items. This was accomplished by the experimenter saying "Show me the_____" and then the target word was silently mouthed to the child. This experiment showed that these young children were able to use the information obtained from lip-reading

Another study examined the pattern of speech production by young sighted and blind children. In a study of the phonological development of three blind children in their second year of life and three sighted controls, Mills (1987) found that the visibility of phonemes can partially determine the development of their production. For example, poorly visible phonemes showed relatively weak accuracy levels for production in both the blind and the sighted control groups. However, phonemes that were easy to see were produced accurately by the sighted group only. This suggests that access to visual speech information can influence the development of speech production.

Audiovisual development in young children

McGurk and MacDonald (1976) found that there were differences between children and adults with regards to the alteration of percepts by conflicting auditory and visual information. In this investigation a correct response was defined as the accurate identification of a syllable presented to the auditory modality. Specifically, it was found that for preschool children (three to four years), primary school children (seven to eight years) and adults (18 to 54 years) the error rate for the repetition of auditory syllables was 9%, 3% and 1% respectively (an accuracy rate of 91%, 97%, and 99%). The differences between the age groups were not significant. However, when the auditory signal was in conflict with the visual signal the error rate for the repetition of the auditory syllables was 59%, 52%, and 92% respectively. In other words, the correct identification of the syllables presented auditorially was 41% for the preschool children, 48% for the primary school children, and 8% for

the adults. This indicates that the visual input had a significant effect on speech perception for the adults while it had a much weaker effect on the preschool and primary school children. It can thus be concluded from these results that the auditory modality plays a larger role in a child's perception of speech in audiovisual situations.

Massaro (1984) found similar results to those obtained by McGurk and MacDonald (1976). He studied 11 children aged 4:9 to 6:9 and 11 adults. The presented stimuli consisted of five synthesized syllable tokens in a continuum from /ba/ to /da/ combined with the visual tokens of /ba/, /da/, and no articulation. The subjects were asked to press two buttons: one for /ba/ and another for /da/. Massaro (1984) found that the children showed about half the degree of visual influence evident in the adults. In other words, the children were less likely to choose the button that corresponded to the syllable that was presented visually.

In a second experiment Massaro (1984) examined whether the results of the first experiment could be explained as a function of attention by including a condition where the visual stimulus consisted of a still face. Could the influence of the visual source be related to the attention paid to that source? Thus, Massaro asked eight children who had participated in the first experiment to report if they saw the speaker's lips move during the presentation of a syllable. The children were able to accurately indicate whether the speaker's lips moved in this experiment. Furthermore, the demonstrated visual influence was similar to that found in the first experiment. When the auditory continuum was presented with the

speaker not moving his mouth there were no significant differences found between the adults and the children. Thus, the face of the speaker without any articulatory movements did not affect the percept interpreted by the subjects. That is, the presence of a visual stimulus does not necessarily affect the reception of auditory information. Therefore, the results of this experiment cannot be explained as a failure to pay attention to the visual stimulus. Massaro (1984) felt that children can detect the correspondence between visual and auditory information from speech but the auditory component has a larger influence in the perception and utilization of speech categories.

Massaro, Thompson, Barron, and Laren (1986) employed similar stimuli to those used in Massaro (1984) but added another condition of lip-reading only. The results for the audiovisual materials were similar to those found by Massaro (1984). It was found, in the visual only condition, that the adult subjects were much better lip readers than the children. Massaro et al. (1986) also extended the Massaro (1984) findings by testing younger children. The children in this younger group were aged 2:5 to 5:3. The children again showed evidence of a decreased visual influence in audiovisual situations. Therefore both Massaro (1984) and Massaro et al. (1986) demonstrated a decreased influence of the visual channel on the perception of speech in audiovisual situations for young children. Moreover, this effect cannot be explained by a lack of attention paid to the visual stimulus by the child subjects.

The studies of audiovisual integration by McGurk and Macdonald (1976), Massaro (1984), and Massaro et al. (1986)

exemplify the emphasis paid to the auditory modality during speech perception development. These studies show that, unlike adults, the perception of speech in children under the age of eight is not strongly influenced by input from the visual modality. The influence of the visual channel becomes greater with development as indicated by findings for adults. Despite the evidence presented by Dodd (1987) and Mills (1987), that visually perceived information can be used by the young child, it appears that young children are not as easily affected by discrepancies between the auditory and visual information. Thus, there appears to be reduced reliance by children on the visual information contained within the speech signal when children are presented with both auditory and visual sources. This reduced reliance on visual information is what accounts for the results obtained in earlier studies.

The studies on audiovisual speech perception that have been examined so far (McGurk & Macdonald, 1976; Massaro, 1984; Massaro et al., 1986) have provided us with a lot of information about capacities that children have to perceive speech with different manipulations of auditory and visual stimuli. There are however a number of questions that need to be addressed.

All of the studies on audiovisual integration in children that have been discussed above report group data. For example, Massaro (1984) presented the children as one group and the adults as another without any differentiation according to age. Thus, it is not evident from these studies whether there are any individual differences in the pattern of audiovisual speech integration and how tightly these perceptual changes are tied to chronological age differences. For

example, will most five year olds be less influenced by visual information than most seven year olds when tested in the McGurk paradigm?

Previous experiments on audiovisual speech perception have not directly examined age-related differences in normally developing children. There have been some previous studies however, that have examined the use of auditory and visual information in pre-school and school aged children. It is difficult to compare these results to those obtained in the studies discussed previously (McGurk & Macdonald, 1976; Massaro, 1984; Massaro et al., 1986) due to the use of disordered populations and the wide variety of manipulations of the auditory and visual stimulus materials (e.g. Craig, 1964; Conrad, 1977; Dodd, 1980; Green, Green, & Holmes, 1980, lip-reading in hearing and deaf children; Dodd, 1977, audiovisual perception in noise by hearing children; Erber, 1972; Fagg, 1992, lip reading supplemented with minimal acoustic information in hearing and deaf children). It is therefore difficult to ascertain the presence of any developmental trends from these studies due to the many differences among them. Thus, there is a need to examine the use of auditory and visual information in normal children of different ages.

The studies by McGurk and MacDonald (1976), Massaro (1984), and Massaro et al. (1986) did not examine the ability to utilize auditory and visual speech information in children older than eight years of age. Evans (1965) found that in congenitally and prelingually deafened children, lip-reading scores increased rapidly between the ages of eight and eleven, then leveled off, and only improved slightly by the age of 15. Evans thus believed that lip-

reading maturity is gained by eleven years of age. Some amount of caution should be exercised when comparing profoundly deaf and hearing individuals, due to differences in the use of, and reliance on the visual sense modality and also to the wide variation in language proficiency exhibited between these populations. Nevertheless, this study indicates that there is a need to examine the use of visual information by older school age children. By examining older as well as younger children the developmental process of audiovisual perception can be examined in detail.

As was mentioned throughout this Chapter, it appears that adults and children use auditory and visual information in different ways. It is not evident however, at what age children show adult-like responses to audiovisual speech information. In this regard, it would be useful to define an adult-like response pattern in the McGurk paradigm to provide an index for determining the age at which a child's use of the auditory and visual modalities becomes similar to that of an adult.

The present investigation

The purpose of this investigation was to examine age-related differences in audiovisual speech perception. Two main questions were addressed. The first was how do children respond to different manipulations of auditory and visual information at different ages? The second was at what age(s) do adult-like responses become evident in the McGurk paradigm? To address these questions analyses of both group and individual performance were conducted.

In this investigation children aged between 4 and 12 years were tested, along with adults, with different presentations of

auditory and visual information. The stimuli were CV syllables similar to the stimuli employed by Werker et al. (1992). The syllables were presented to the subjects in four different conditions: Auditory Only, Visual Only, Audiovisual Same Syllables (AV), and Audiovisual Conflicting Syllables (AVC). In the first three conditions, the syllables /ba/, /va/, /ða/, /da/, and /ga/ were presented. In the fourth condition the auditory production of /ba/ was paired with the visual articulations of /va/, /ða/, /da/, and /ga/, respectively.

The Auditory Only and Visual Only Conditions were used to determine the subject's perceptual ability when information was presented to one sensory modality. The AV Condition was used to examine the subjects' perceptions of corresponding audiovisual syllables. It was hypothesized that performance on the Auditory Only Condition would exceed performance on the Visual Only Condition for all age groups (Woodward & Barber, 1960; Fisher, 1968). It was also hypothesized that the performance on the AV Condition would exceed performance on the Visual Only and Auditory Only Conditions for all subject groups (Woodward & Barber, 1960; Binnie et al., 1974; Reisberg et al., 1987). It was further hypothesized that the ability to use visual information develops in early childhood. Thus it is expected that performance in the Visual Only and AV conditions will improve as subject age increases. This is consistent with the findings of McGurk and MacDonald (1976), Massaro (1984), and Massaro et al. (1986).

The AVC condition provides further information about the nature of perceptual processing which occurs in response to a bimodal speech stimulus. In this condition the visual and auditory

productions do not match. The perceptual responses to the AVC stimuli may potentially reveal a bias for one modality over the other or they may reveal an integration of the information available across the visual and auditory channels.

It was hypothesized that before the age of seven, visual influence on speech perception will be weak, and likewise the integration of auditory and visual information will not occur, as was found by McGurk and MacDonald (1976), Massaro (1984), and Massaro et al. (1986). The younger children will thus experience a weaker visual influence than the adults and will therefore show a bias towards the auditory information. It was further hypothesized that by the age of nine, visual information will have a greater influence on speech perception. Thus the children between seven and nine will show some integration responses at this age, though not as profound as the effects seen in the adults (McGurk & MacDonald, 1976), and more of a bias towards the visual information. It was also hypothesized that the eleven year old children will show more of an influence from the visual component of speech, thus supporting the findings of Evans (1965). These children will thus show a response pattern similar to that of the adults. The results from this experiment will provide further clarification of the developmental process of audiovisual speech perception.

Chapter 5

Method

Subjects

The subjects for this experiment consisted of 15 adults and 46 school aged children who had no documented speech or hearing difficulties. The adult and child subjects were all native English speakers. All subjects had normal or corrected to normal binocular vision and were naive to the experimental theory. Both the adults and the children were paid for their participation in this experiment.

The 15 adult subjects were all recruited from the McGill University community.[2] None of the adult subjects reported fluency in a second language. This subject group was comprised of seven males and eight females ranging in age from 18 to 28 with a mean of 21.5 years (SD=3.02). None of the adult subjects demonstrated any articulation problems during conversation with the experimenter.

The 46 child subjects were recruited from the suburban Montreal area.[3] For every child, English was the language used in the home and was the child's preferred language. Some of the children were being regularly exposed to other languages, but none were fluent speakers of any languages other than English. Children were placed into four groups centered around the ages of five, seven, nine, and eleven years. The age range of the child groups can be seen

---

[2] 16 adult subjects participated in this research but one subject was removed due to poor inter-transcriber reliability (<80%) which is discussed in the Scoring and Analysis Section of Chapter 5.

[3] 50 child subjects participated in this research but four subjects were removed for the following reasons: two subjects did not complete the AV and AVC experimental conditions, one subject had a first language which was not English, and another subject had poor inter-transcriber reliability (<80%) which is discussed in the Scoring and Analysis Section of Chapter 5.

in Table 2. The children's articulatory abilities were assessed using the Goldman Fristoe Test of Articulation (1986). All children demonstrated articulation skills within a normal range for their age. Furthermore, in the articulation assessment, every child demonstrated the ability to produce each of the target phonemes presented in the experimental task.

Table 2. Age range for the four child subject groups.

Group	Age Range	Mean Age	Number of Males	Number of Females	Total
five	4:7-5:11	5:0 (SD=0.40)	5	5	10
seven	6:5-7:11	7:3 (SD=0.48)	4	9	13
nine	8:1-9:10	9:2 (SD=0.65)	6	7	13
eleven	10:7-12:4	11:4 (SD=0.57)	6	4	10

Stimuli

The syllables that were used for this experiment were productions of /ba/, /va/, /ɵa/, /da/, and /ga/ presented in four different conditions.[4] The Auditory Only Condition consisted of auditory productions of each syllable. The Visual Only Condition consisted of the visual productions of each syllable (face only). The Audiovisual Same Syllable (AV) Condition contained the visual and

[4] The unvoiced syllable /ɵa/ was used instead of the voiced syllable /ð a/ throughout this study due to a technical error in the production of the stimuli These syllables have the same place of articulation and appear to be visually the same.

auditory presentations of each syllable together. The Audiovisual Conflicting Syllable (AVC) Condition consisted of the visual production of the syllables paired with the auditory production of /ba/ as can be seen in Table 3.

Table 3. Visual productions of the four AVC stimuli paired with the auditory syllable /ba/.

	Visual			
	front----------			----------back
	/va/	/ða/	/da/	/ga/
Auditory	/ba/	/ba/	/ba/	/ba/

Sekiyama and Tohkura (1991) proposed that auditory intelligibility is a factor which can modulate the degree to which an individual will show the McGurk effect. That is, they have suggested that a definite McGurk-like response will be observed when an auditory syllable that is not perfectly intelligible is combined with a conflicting visual syllable. In previous studies (e.g. McGurk & MacDonald, 1976; MacDonald & McGurk, 1978; Mannuel et al., 1983; Massaro, 1984; Massaro et al., 1986) stimuli were developed in the laboratory specifically for each study. Thus, there is likely to be considerable variation in the quality of both the auditory and the visual materials presented in these experiments. Given the potential influence of auditory intelligibility, there is a need to use standard materials to facilitate comparisons in this line of research.

The stimuli for this experiment were developed from the first videodisk in the two volume Johns Hopkins Lip-reading Corpus (Bernstein & Eberhardt,1986).[5] These videodisks contain stimuli ranging from CV syllables to sentences, spoken by a female and a male. These stimuli have been used extensively in past research (Bernstein, Eberhardt, & Demorest, 1989; Bernstein, Demorest, Coulter, & O'Connel, 1991; Demorest, Bernstein, & Eberhardt, 1987; Demorest & Bernstein, 1992; Eberhardt, Demorest, & Goldstein, 1990) and are rapidly becoming a set of standard materials for research in multimodal speech perception, including studies using the McGurk paradigm (Ward, 1992).

The videodisk recording was prepared at Johns Hopkins University using a female actress and a male singer who spoke the same material. Both of the talkers spoke General American English. The original video recordings were made using a Sony BVU 110 video recorder and a Hitachi Z31 camera. The talkers were seated in front of a dark background and the lighting was provided by two face-level direct 600W floodlights positioned 35 degrees on either side of the midline. Two 600W fill lights were positioned above the talkers. The talker's mouth was slightly below the center of the video screen and the face filled most of the screen area. A teleprompter was utilized so that the talkers did not have to take their eyes off the screen (Demorest & Bernstein, 1992). For this experiment the male talker was used due to the fact that in past studies he was found to be easier to speech-read than the female

---

[5] The stimuli were recorded from the Johns Hopkins Lip-reading Corpus in the laboratory of Dr K. G. Munhall at Queen's University in Kingston, Ontario.

talker (Demorest et al., 1987). The frame numbers of the stimuli taken from the videodisk for the present study can be found in Appendix A.

The video files were played on a Pioneer Laservision LD V8000 videodisk player. The audio files were first digitized (22kHz sampling rate, 12 bit resolution) from the videodisk with the program CSPEECH 3.1 (Milenkovic, 1990) running on a Zenith 386/25 computer with a Data Translation 2821 analogue/digital board and a Frequency Devices 901 low-pass filter (9 kHz). All of the recordings were made with a Panasonic professional VHS video tape recorder model AG-6300. All of the audio recordings produced with the videotape recorder were performed by routing the output from the computer through the filter directly into the channel 1 audio input of the videotape recorder (input level set at "2"; Dolby B noise reduction). The audio and the video files were synchronized, timed, and organized into blocks using the program LD (1.0) developed by Broekhaven (1991). A 3000 msec gap of black screen (for the Visual Only and the audiovisual conditions) and silence (for the Auditory Only and the audiovisual conditions) was inserted between each production for all of the stimulus conditions.

The first stimulus tape created for this study consisted of the Auditory Only and Visual Only Conditions. The Auditory Only Condition was created by recording onto the videotape the syllables: /ba/, /va/, /ɵa/, /da/, and /ga/ from the digitized files in the computer. Each digitized auditory syllable was recorded four times in a random order to give a total of 20 stimuli (5 practice and 15 test trials). During testing the subjects viewed a blank screen with no

visual information. The Visual Only Condition was created in a similar fashion to the auditory stimuli. Each visually produced syllable was recorded four times in a random order to give a total of 20 stimuli (5 practice and 15 test trials). During testing, the volume setting of the video monitor was turned as low as possible making the audio signal inaccessible. Thus the first tape contained 20 audio only stimuli (five places of articulation repeated four times) and 20 visual only stimuli (five places of articulation repeated four times). The order of presentation of each syllable for these two conditions can be found in Appendix B.

The second stimulus tape consisted of the two audiovisual conditions: the Audiovisual Same Syllable (AV) Condition and the Audiovisual Conflicting Syllable (AVC) Condition. The AV Condition consisted of the audiovisual productions of /ba/, /va/, /ða/, /da/, and /ga/. The AVC Condition consisted of the visual productions of /va/, /ða/, /da/ and /ga/ paired with the audio production of /ba/. In both of the audiovisual conditions, a digitized audio file was synchronized with a selected video production; natural auditory productions, that is, the original audiovisual productions from the videodisk, were not used in the AV Condition. Thus the AV and AVC stimuli were generated in the same manner.

To create each AV and AVC stimulus it was necessary to synchronize the digitized audio file with the video file so that as "natural" a production as possible was produced. To accomplish this it was necessary to digitize both the audio output of the computer and the audio output of the videodisk so that both waveforms could be viewed, timed and synchronized. A Toshiba 3200 laptop

computer with the CSpeech program was used with channel one connected to the audio output of the Zenith computer fed through the filter and channel two connected to the audio output of the videodisk player. For the AV productions the two audio files from the computer and the videodisk were aligned so that the beginning of each production occurred at the same instant in time. In order for this to occur, the audio file from the computer had to be delayed by a certain number of milliseconds as shown in Appendix C. This was also the case for the AVC productions of /va/, /da/, and /ga/ paired with /ba/ auditory. The audio production of/ba/ was delayed by a certain amount of time so that the videodisk file was aligned with the digitized audio file. It must be noted that there was some prevoicing on the /ba/ audio file. The alignment of the video with the auditory signal was made with respect to the burst, not the prevoicing. For the production of /θa/ the video file was delayed by 32 msec so that the sound file /ba/ was aligned (in time) one third of the way through the frication noise of the digitized audio file, to preserve the natural appearance of the utterance. The duration of time by which the files were delayed for the AVC stimuli can be found in Appendix D.

The aligned audio/video files were then recorded onto the second video tape. Eighty-seven trials were recorded: nine practice trials followed by 13 blocks of six test trials (78 trials). For the practice trials each AV and AVC syllable was presented once. Each test trial block contained the four AVC stimuli and two AV stimuli in a random sequence. There were a total of 56 AVC stimuli (four practice and 52 test trials) and 31 AV stimuli (five practice and 26

test trials). The block structure and the order in which the stimuli were presented can be found in Appendix E.

Equipment

The stimuli were presented to the subjects using a Panasonic AG6300 VHS video cassette recorder with a Panasonic NVA810 wired remote control connected to a Panasonic AG500R 10 inch (0.25 metre) Video Monitor Player. The monitor was placed on a table and was positioned on a support so that it was 0.65 metres from the floor. The subject was seated 1.17 metres from the screen. The image of the speaker on the screen produced a visual angle of 7.36 degrees.

Sound level measurements were recorded with a Brüel and Kjaer sound level meter model 2204 using the A weighting scale and the fast response settings. The ambient background noise level in the testing room was 34 dBA and the sound presentation levels of the auditory stimuli (at approximately ear level) were 70 dBA for /ða/, 72 dBA for /θa/, 73 dBA for /va/, 70 dBA for /ga/ and 74 dBA for /ba/ (volume setting of "6").

The subjects provided verbal responses which were recorded on a Marantz PMD 201 cassette recorder with a Realistic Highball-7 microphone mounted on a floor stand 0.90 metres from the floor.

Procedure

Both the adult and the child subjects were presented with the same stimuli and the responses were recorded in a similar fashion. For the children, an assistant interacted with the subject while the author presented the stimuli and scored the responses. The assistant was present so that the child was not distracted by the operation of

the video and audio equipment. During the presentation of the experimental trials the assistant was able to stop the video tape at any time if it was felt that the child needed a break, was not looking at the screen, or that he/she needed some extra encouragement to complete the experimental task.

In the beginning of each session each child subject completed the Goldman Fristoe Test of Articulation (1986). The parents were interviewed regarding the child's language experience. Each adult subject was also given a short questionnaire about their language experience. This questionnaire can be found in Appendix F.

Each subject was asked to sit comfortably in front of the video monitor. Subjects were simply instructed, for each condition, to report what the man on T.V. "was saying" (for exact instructions see Appendix G). An attempt was made to not bias the subjects towards any one sensory modality by using phases such as "repeat what the man is saying" as opposed to "repeat what you hear (or see)". The subject was asked to respond in the Auditory Only Condition while looking at the blank video monitor. The subject was also told that if he/she was unsure of what the man said that he/she should make a guess. The verbal response method was selected for all of the experimental conditions because it is imperative that the subject's gaze is directed towards the visual stimulus at all times. Such response recording methods as writing or button pushing would require the subject to shift his/her visual fixation and thus his/her attention from the video monitor. The experimenter (adult subjects) or the assistant (child subjects) sat next to the subject at all times to

monitor the subjects' gaze and to prompt the subject if he/she looked away from the video monitor.

Every subject was presented the four conditions in the following order: Auditory Only, Visual Only, and Audiovisual (AV and AVC). A short break was provided between conditions. The responses that the subject gave in each condition were transcribed on line by the experimenter on a sheet that can be found in Appendix H. For the Auditory Only and Visual Only Conditions five practice trials were presented, the instructions were repeated, and then the 15 test trials were presented. For the audiovisual conditions (AV and AVC) nine practice trials were presented, the instructions were again repeated, and the 78 test trials were presented.

The subjects were provided with short breaks at various times during the testing in the AV and AVC Conditions due to the large number of stimuli. For the adults the experimenter always stopped the tape after 36 trials to give the subject a break and at other times as requested by the subject. For the children a break was given after 24 trials, and after 48 trials or as indicated by the child, the experimenter, or the assistant. None of the stimuli were repeated to the subject if he/she failed to respond to a particular item.

Scoring and Analysis

Inter-transcriber reliability. The audio tapes of the subjects were transcribed by an independent transcriber who was familiar with phonetic transcription techniques but naive to the experimental methodology. The reliability measures were obtained for the experimental trials only; the practice trials were not included. Each

response was considered as "in agreement" when the identical syllable or two syllables with the same place of articulation was recorded by both transcribers. For example, if both of the transcribers reported /va/ then this was considered to be "in agreement" but if one transcriber reported /va/ and the other transcriber reported /fa/ this was also scored as "in agreement", since both syllables have the same place of articulation. For a small number of trials the utterances on the audio tape were missing (due to technical error), or were unintelligible, or very difficult to hear (due to tape hiss or a soft voiced subject); these items were not included in the reliability assessment.

The percentage of the responses that were in agreement between the two transcribers was calculated for each subject. As shown in Table 4, the average reliability in each subject age group was above 90% and no subjects had reliability below 80%. These reliability measurements were comparable to those obtained in other studies (Werker et al., 1992). For the analysis of the results the experimenter's on line transcriptions were utilized, even if a discrepancy between the two transcriptions was found.

Table 4. Percent agreement between the two transcripts for the four child age groups and the adult controls.

Group	N	Mean Percent Agreement	Range
five	10	93.11 (SD=5.84)	80.00-99.07
seven	13	94.45 (SD=3.37)	85.85-98.98
nine	13	95.74 (SD=4.02)	85.44-100.00
eleven	10	93.52 (SD=4.46)	85.19-98.15
adult	15	92.93 (SD=4.14)	82.29-98.15

Scoring method: Auditory Only, Visual Only, and AV Conditions. For the Auditory only, Visual only, and Audiovisual Same Syllable (AV) Conditions the responses were scored as either being correct or incorrect with regards to place of articulation. Thus, if the reported syllables were different from the syllables presented to the subject, but the place of articulation was the same, then these were considered to be correct. For example, if the syllable /va/ was presented and the syllable /fa/ was reported then this was considered to be a correct response due to the fact that these syllables both have a labiodental place of articulation. In this manner, a proportion of correct responses was recorded for each subject for each condition.

Scoring method: AVC condition. For the Audiovisual Conflicting Syllable (AVC) Condition scoring of the subject's responses was performed according to the response categories used by Werker et al.

(1992). It must be remembered that the subjects were presented with the auditory production of /ba/ paired with the visual productions of /va/, /ɵa/, /da/, and /ga/. The first response category was Auditory Capture (AC) which was assigned when the subject's response was a bilabial. This response category was defined as such because the auditory signal provides information about the bilabial place of articulation. Thus, the subject's response was unaffected by the visual stimuli. The second response category was Visual Capture (VC) which was assigned when the subject's response corresponded to the place of articulation of the visual stimuli. For example, VC for the visual stimulus of /va/ referred to the responses of /va/ or /fa/, as these articulations appear to be the same syllable on the lips. The responses defined as visual capture are shown in Table 5.

Table 5. Verbal responses defined as Visual Capture (VC) according to Werker et al. (1992).

Visual Stimulus	Accepted Verbal Response (with respect to place of articulation)
/va/	/va/ or /fa/ labiodental
/ɵa/	/ð/a/ or /ɵa/ interdental
/da/	/da/, /ta/, /sa/, /la/ or /na/ alveolar
/ga/	/ga/ or /ka/ velar

The third response category was a Blend. This Blend category was assigned when the subject gave a response which reflects some new

combination of the auditory and the visual information. The Blend response does not correspond exactly to either place of articulation of the auditory or the visual syllable, rather the Blend percept has a place of articulation that falls between that of the visual and the auditory information, indicating an integration of the information from both sources. For example, if the subject was presented with /da/ visual and /ba/ auditory a Blend would be defined as a response that corresponds to a place between the bilabial /ba/ and the alveolar /da/ such as the labiodental /va/. Blend responses could only be recorded for the visual syllables /θa/, /da/, and /ga/ paired with the /ba/ auditory stimulus. Blends are restricted to these syllables because, in English, there is no place of articulation falling between that of /va/ (labiodental) and /ba/ (bilabial) therefore only AC (bilabial) and VC (labiodental) responses can be obtained from this audiovisual stimulus pair. The possible responses which are defined as blends can be seen in Table 6. Blends and VC responses provide evidence for visual influence on phonetic perception.

Table 6. Verbal responses defined as Blends according to Werker et al. (1992).

Visual Stimulus	Accepted Verbal Response for a blend (a combination of the auditory and visual information)
/θa/	/va/ or /fa/
/da/	/va/ or /fa/; /ða/ or /θa/
/ga/	/ʃa/, /ʒa/, /ça/, or /ja/; /da/, /ta/, /la/, /na/, or /sa/; /ða/ or /θa/; /va/ or /fa/.

The fourth and final response category is the classification of Other. This category was assigned when the subject's response did not correspond with a syllable that was between the place of articulation of the auditory and visual stimuli. For example, if the subject was presented with /da/ visual and /ba/ auditory, the Other response category was assigned if the response corresponded to a place of articulation which did not fall between the place of articulation of the bilabial /ba/ and the alveolar /da/, such as /ga/. The response category of Other was also assigned when a unique combination of the auditory and visual stimuli was reported by the subject such as /bga/ for /da/ visual and /ba/ auditory stimuli or if no response was given by the subject for a particular item.

Chapter 6

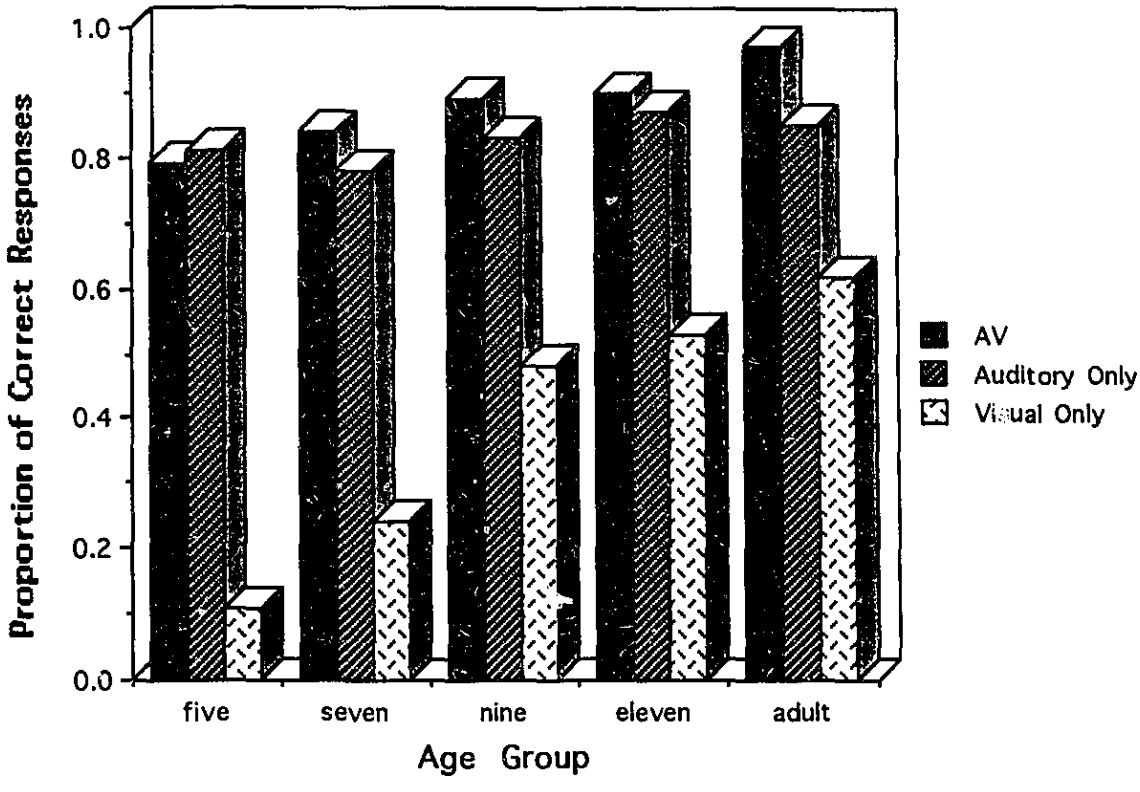
Results

The results of this investigation are organized into five sections. The first section presents the results from the Audiovisual Same Syllable (AV), Visual Only, and Auditory Only Conditions. This will then be followed by a second section which presents the results from the Audiovisual Conflicting Syllable (AVC) Condition. The third section provides a more detailed summary of data from the AVC Condition by presenting a confusion matrix for each of the AVC stimuli and the responses made by the subjects. The fourth section provides an analysis of the percentage of subjects that attained an adult-like criteria in the AVC (McGurk paradigm) Condition. The fifth and final section summarizes the results obtained from this investigation.

AV, Visual Only, and Auditory Only Conditions

The mean proportion of correct responses for the AV, Auditory Only, and Visual Only Conditions across the four child age groups and the adults is displayed in Figure 2. These data were analyzed in a mixed model ANOVA with Age Group as a between subject factor and Modality (AV, Auditory Only, Visual Only) as a within-subjects factor.

Figure 2. Mean proportion of correct responses for the AV, Auditory Only, and Visual Only Conditions.



The main effect of Modality was highly significant ($F(2,112)= 427.31$ $p<.0001$). The main effect of Age Group ($F(4,56) =28.14$ $p<.001$), was also significant. However, as expected, the Age Group X Modality interaction ($F(8,112)=12.28$ $p<.001$) was significant indicating that performance differs across the age groups.

To further probe this interaction, the simple effects of Modality were analyzed separately for each Age Group. The ANOVA analyzing performance in the five year old group showed a highly significant effect for Modality ($F(2,18)=165.15$ $p<.001$). Tukey pairwise comparisons revealed that performance on the AV and Auditory

Only Conditions was significantly better than performance on the Visual Only Condition ($p<.01$). However, a significant difference was not found between the performance on the AV Condition and the Auditory Only Condition. The same pattern of effects and Tukey results ($p < .01$) was found in the one-way ANOVAs conducted for the seven year old group, nine year old group and for the eleven year old group ($F(2,24)=189.538$ $p<.001$ for the seven year olds; $F(2,24)=64.89$ $p<.001$ for the nine year olds; and $F(2,18)=39.74$ $p<.001$ for the eleven year olds). For the adult group a significant effect of Modality was also observed ($F(2,28)=41.51$ $p<.001$). Subsequent Tukey pairwise comparisons revealed that there were significant differences for performance on the AV and Auditory Only Conditions when compared with the Visual Only Condition ($p<.01$) and also for the AV Condition compared with the Auditory Only Condition ($p<.05$).

These findings reveal that, in every age group, bimodal perception of speech was significantly better than the perception of speech via visual information alone. However, bimodal perception exceeded performance with auditory information alone only for the adults. The addition of visual speech information to auditory speech information, therefore, does not appear to significantly improve the accuracy of perception for the child subject groups. However, there is a fairly consistent trend for bimodal to be better than auditory only speech perception and auditory to be better than visual speech perception in the seven, nine, and eleven year old groups. The pattern of performance in which AV>Auditory Only>Visual Only is seen in 11 of the 15 adults, 7 of the 10 children in the eleven year

old group, 12 of the 13 children in the nine year old group, 10 of the 13 children in the seven year old group, and 6 of the 10 children in the five year old group.

As can be seen in Figure 2, there are differences in the performance in each condition as the subject age increases. In the AV and Visual Only Conditions a general developmental trend can be seen; that is, as subject age increases so does the performance on these conditions. To evaluate the pattern of age related change in the AV, Visual Only, and Auditory Only Conditions, simple effects of Age were also analyzed separately for each condition.

The analysis of performance in the AV Condition revealed a significant Age Group effect ($\underline{F}(4,56) = 12.05$ $p<.001$). Tukey-Kramer pairwise comparisons for unequal cell frequencies are consistent with an age-related increase in performance. Specifically, the five year olds were significantly less accurate than the nine year olds ($p<.05$), eleven year olds ($p<.01$) , and the adults ($p < .01$). The adults were significantly more accurate in their responses in the AV Condition than the seven year olds ($p<.01$) and the nine year olds ($p<.05$). There were no significant differences found between the eleven year old group and the adult control group, indicating that there were no significant changes in accuracy for the AV Condition beyond the nine year old group.

Analysis of performance in the Visual Only Condition also revealed a significant Age Group effect ($\underline{F}(4,56) = 23.31$ $p<.001$). Tukey-Kramer pairwise comparisons ($p <.01$) were consistent with an age-related increase in performance in the Visual Only Condition. Specifically, the nine year olds, eleven year olds and the adults were

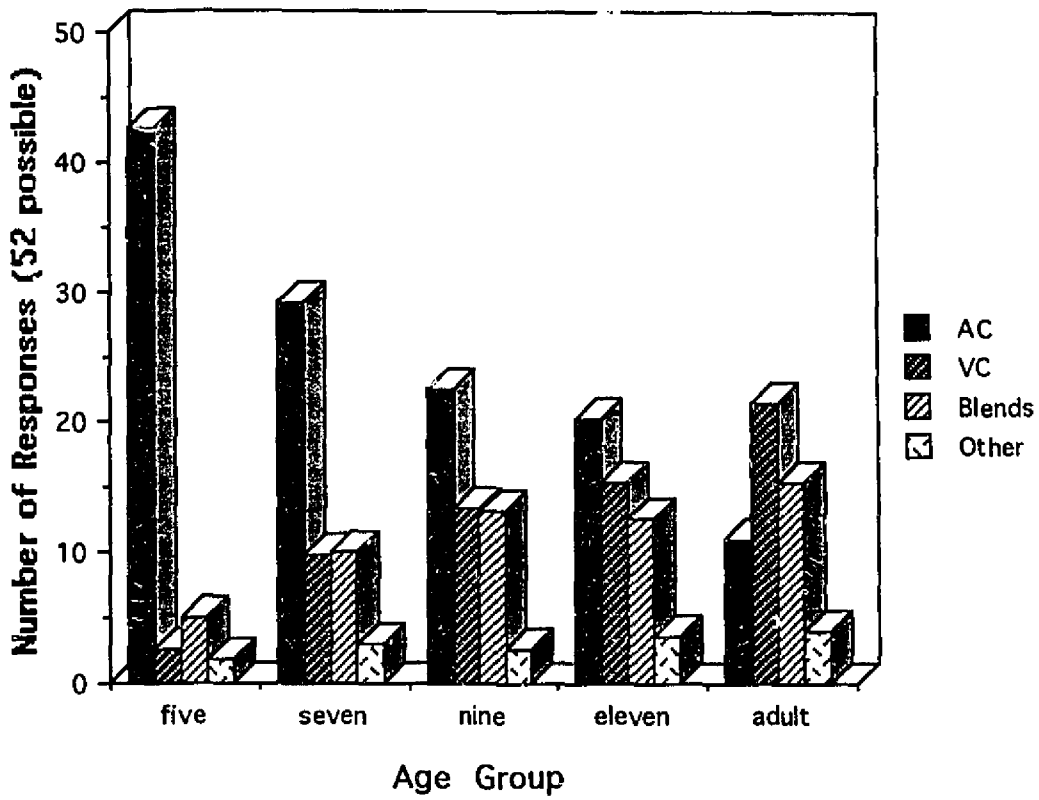
significantly more accurate than the five year olds. The nine year olds, eleven year olds and the adult controls were also significantly more accurate than the seven year olds. No significant differences were found between the nine year old, eleven year old and the adult control groups, indicating that there were no significant changes in lip-reading ability among these older age groups.

Analysis of performance in the Auditory Only Condition revealed a significant Age Group effect ($F(4,56) = 3.02$ $p<.05$). Tukey-Kramer pairwise comparisons revealed that eleven year olds were significantly more accurate than the seven year olds ($p<.05$); no other age group differences were observed. Thus, there was little evidence for age-related differences in performance in the Auditory Only Condition.

AVC Condition

As was discussed in the Scoring and Analysis section of the Methods (Chapter 5), the responses to the AVC condition were classified according to a specific set of criteria. This system produced the four possible classifications of Auditory Capture (AC), Visual Capture (VC), Blends, and Other. Figure 3 shows the number of each type of response made to the 52 AVC stimuli across the five age groups.

Figure 3. Mean number of responses defined as AC, VC, Blends, and Other in the AVC Condition.



To examine the way in which subjects in the different age groups responded to the AVC stimuli, separate one way ANOVAs examining the Age Group effect were conducted for each type of response. The ANOVA conducted on the AC responses will be presented first, followed by ANOVAs for VC, Blends and Other responses. The data are presented in this manner due to the fact that the four response categories are not independent of each other. These four ANOVAs will be interpreted as one set of results.

Analysis of AC responses demonstrated a significant Age Group effect ($F(4,56)=10.91$ $p<.001$). Results of Tukey-Kramer pairwise

comparisons ($p < .01$) are consistent with an age-related decrease in AC responses. Specifically, both the five year olds and the seven year olds had significantly more AC responses than did the adult controls. The five year olds also made significantly more AC responses than both the nine year olds and the eleven year olds. There were no significant differences found between the adult group, and the nine and eleven year old group, indicating that the number of AC responses did not decrease significantly across these groups.

Analysis of VC responses also revealed a significant Age Group effect ($F(4,56)=14.30$ $p<.001$). Results of Tukey-Kramer pairwise comparisons ($p < .01$) are consistent with an age-related increase in VC responses. Specifically, the nine year olds, the eleven year olds and the adults made significantly more VC responses than did the five year olds. Also, the adult control group made significantly more VC responses than the seven year old group and the nine year old group ($p<.05$). However, there was no significant difference between the seven year olds when compared with the nine or eleven year old groups. Likewise, there was no significant difference in the number of VC responses between the adults and the eleven year olds.

Analysis of Blend responses also revealed a significant Age Group effect ($F(4,56)=3.62$ $p<.05$). Tukey-Kramer pairwise comparisons ($p<.01$) revealed that adults made significantly more Blend responses than did the five year olds. No other significant differences were found. Thus, the number of Blend responses did not differ significantly across the three older child groups and the adults.

Analysis of "Other" responses failed to show a significant Age Group effect.

Overall, it can be seen that as the age of the subjects increased the number of AC responses decreased while the number of VC responses increased. For the Blend responses it appears that only the adults made significantly more of these responses than the five year olds.

Confusion Matrices

Confusion matrices were constructed to examine the extent to which general patterns of age-related differences are shown for specific AVC syllables. Each matrix presents the proportion of responses that correspond to each possible place of articulation; the scoring category for each place of articulation is also noted (i.e., AC, VC, Blend, Other).

Table 7 presents confusion matrices for the /ba/+/va/ and the /ba/+/ɵa/ stimuli, both of which provide a frontal visual place of articulation. For these two AVC stimuli it can be seen that the proportion of AC responses (bilabials) clearly decreased as the subject age increased. The proportion of VC responses can also be seen to increase for both the /ba/+/va/ (labiodental) and the /ba/+/ɵa/ (interdental) AVC stimuli. As was mentioned in the Scoring and Analysis section of Chapter 5 there was no Blend response defined for the /ba/+/va/ AVC stimulus due to the fact that there are no syllables found in English between the places of articulation of the visual /va/ and auditory /ba/. For the /ba/+/ɵa/ AVC stimulus however, there are a number of Blend responses (labiodental) for each age group. It is interesting to note that the nine year old group

gave the most Blend responses. It was expected that the adults would produce more Blend responses indicating a fusion of information from the visual and auditory channels. However, it appears that, for adults, the visual information dominates the perception of the /ba/+/ɵa/ stimulus to produce more VC responses than Blends. This result is not surprising in that the visual /ɵa/ provides clear evidence for the interdental place of articulation. There is also a small number of Other (alveolar) responses to the /ba/+/ɵa/ AVC stimulus; this occurs in all five of the subject groups.

Table 7. Proportion of responses to the AVC stimuli /ba/+/va/ and /ba/+/ða/.

Auditory = ba Visual=va	Bilabial b,p,m	Labidntl v,f	Interdntl ð , θ	Alveolar d,t,n,s,l	Palatal ʃ,ʒ,ʧ,j	Velar k,g	Undefined According to Place
Response Category	Aud Capture	Vis Capture	Other	Other	Other	Other	Other
four/five	0.81	0.17	0.02	0.00	0.00	0.00	0.01
six/seven	0.41	0.48	0.06	0.03	0.00	0.00	0.02
eight/nine	0.26	0.65	0.08	0.00	0.00	0.00	0.01
ten/eleven	0.25	0.70	0.04	0.00	0.00	0.00	0.02
adult	0.07	0.74	0.11	0.07	0.00	0.00	0.02

Auditory = ba Visual=ða	Bilabial b,p,m	Labidntl v,f	Interdntl ð , θ	Alveolar d,t,n,s,l	Palatal ʃ,ʒ,ʧ,j	Velar k,g	Undefined According to Place
Response Category	Aud Capture	Blend	Vis Capture	Other	Other	Other	Other
four/five	0.82	0.13	0.02	0.01	0.00	0.00	0.03
six/seven	0.51	0.23	0.16	0.10	0.00	0.00	0.00
eight/nine	0.35	0.30	0.27	0.04	0.00	0.01	0.03
ten/eleven	0.32	0.20	0.27	0.18	0.00	0.00	0.03
adult	0.13	0.09	0.69	0.07	0.00	0.01	0.02

Table 8 presents confusion matrices for the AVC stimuli /ba/+/da/ and /ba/+/ga/ both of which have a more back place of articulation. For these two AVC stimuli the proportion of AC responses (bilabials) clearly decreased with increasing subject age. The proportion of VC responses can also be seen to increase for the /ba/+/da/ (alveolar) stimulus. It is interesting to note however, that seven and nine year olds give the same proportion of VC responses. Proportion of VC responses are also the same for the eleven year olds and the adults, indicating that the VC responses are not changing as regularly with age as was observed for the /ba/+/va/ and /ba/+/ɵa/ AVC stimuli. In fact, almost no VC responses are made for the /ba/+/ga/ AVC stimulus in any age group. This finding is not surprising given that the place of articulation of /ga/ is far back in the oral cavity so therefore there is not much visual information available to support the judgment of a velar response. In general there was an increase in Blend responses (labiodental and interdental) for the /ba/+/da/ AVC stimulus as subject age increased. This is also the case for the /ba/+/ga/ AVC stimulus which indicates that there was a general increase in Blend responses (labiodental, interdental and alveolar) as the subject age increases. The Blend response /da/ (alveolar) to the /ba/+/ga/ AVC stimulus is the classic demonstration of the McGurk effect that is presented widely throughout the audiovisual speech perception literature.

Table 8. Proportion of responses to the AVC stimuli /ba/+/da/ and /ba/+/ga/ stimuli

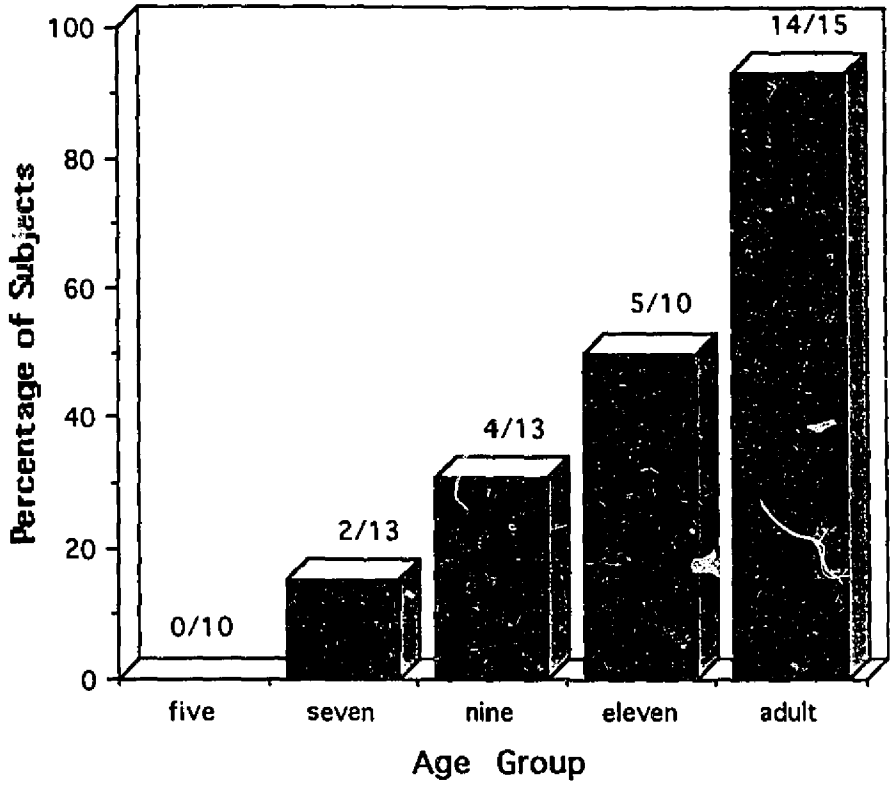
Auditory = ba Visual=da	Bilabial b,p,m	Labidntl v,f	Interdntl ð, θ	Alveolar d,t,n,s,l	Palatal ʃ,ʒ,ʧ,j	Velar k,g	Undefined According to Place
Response Category	Aud Capture	Blend	Blend	Vis Capture	Other	Other	Other
four/five	0.79	0.08	0.06	0.02	0.00	0.00	0.05
six/seven	0.63	0.15	0.09	0.11	0.00	0.01	0.01
eight/nine	0.57	0.17	0.14	0.11	0.00	0.01	0.01
ten/eleven	0.45	0.12	0.22	0.21	0.00	0.01	0.00
adult	0.28	0.13	0.37	0.21	0.00	0.00	0.01

Auditory = ba Visual=ga	Bilabial b,p,m	Labidntl v,f	Interdntl ð, θ	Alveolar d,t,n,s,l	Palatal ʃ,ʒ,ʧ,j	Velar k,g	Undefined According to Place
Response Category	Aud Capture	Blend	Blend	Blend	Blend	Vis Capture	Other
four/five	0.85	0.10	0.02	0.00	0.00	0.00	0.03
six/seven	0.69	0.12	0.06	0.12	0.00	0.01	0.01
eight/nine	0.58	0.16	0.16	0.09	0.00	0.00	0.01
ten/eleven	0.55	0.08	0.17	0.19	0.00	0.01	0.01
adult	0.37	0.06	0.32	0.23	0.00	0.01	0.01

Analysis according to an adult-like pattern

The adult data in the AVC condition were examined to describe a response pattern which characterizes adult performance in the McGurk paradigm for the stimuli used in this study. This description provides an index which can be used to evaluate whether individual child performance conforms to an adult pattern. The criterion for an adult-like response was defined with respect to the pattern of AC, VC, and Blend responses. It was consistently noted in the adult data that the number of VC responses exceeded the number of AC responses, indicating that the influence of vision is quite strong. As well, adults consistently showed some Blend responses (at least 6 were typically observed), also reflecting a substantial effect of visual information in adult speech perception. Thus, for the purpose of this study, these 2 indices jointly defined an adult-like response. That is, an individual was considered to show an adult-like pattern if they had made the same number or more VC than AC responses and at least six Blend responses. Fourteen of the 15 adults tested showed this pattern. There was one adult who did not conform to this pattern. However, his results (41 AC responses, 10 VC responses and 1 Blend) were, in fact, quite deviant when compared with other adults and even when compared with older children. His individual data were most consistent with the pattern observed in the 5 year olds. The percentage of subjects who showed the adult-like pattern in each age group is plotted in Figure 4. It is shown in this graph that there is a general developmental trend observed, that is, as the subject age increases the percentage of subjects who reach the adult-like criterion is also seen to increase.

Figure 4. Percentage of subjects classified as meeting the adult-like pattern criterion.



An Analysis of Proportion (Marascuilo, 1966, 1971) showed that the proportion of subjects demonstrating the adult-like pattern was reliably different across the age groups (X^2 (4)=28.36 $p<.001$). Multiple comparisons revealed that a significantly lower proportion of subjects met the adult-like criteria in each of the child subject groups than in the adult group ($p<.001$ for the five, seven, and nine year olds; $p<.05$ for the eleven year olds). A significant difference was also found between the eleven year olds and the five year olds ($p<.01$), indicating that more subjects showed the adult-like pattern in the eleven year old group than in the five year old group. No

other significant differences were found among the subject groups. Thus, even the oldest child group tested failed to consistently show an adult-like response.

Summary of the results

The results from this experiment have revealed a number of interesting aspects about the development of the audiovisual perception of speech. Performance on the AV and Auditory Only Conditions was significantly better that the performance on the Visual Only Condition for all of the subject groups. For the adult group the performance on the AV Condition was significantly better than the Auditory Only Condition, indicating that the addition of visual cues made substantial contributions to the perception of speech. This advantage of bimodal over auditory perception was seen as a general trend for the child age groups though it was not found to be statistically significant. Thus, even the eleven year old group was not performing in an adult-like fashion.

Performance on the AV, and Visual Only tasks was generally seen to increase as subject age increased. The adults had the highest level of performance in these two conditions. The nine and eleven year old subjects performed in a similar fashion which was better than the five and the seven year olds who also performed in a similar fashion. There was little evidence to support an age related change for the Auditory Only Condition.

For the AVC Condition some interesting effects were found. The number of AC responses decreased with age, while it was found that the number of VC responses increased with age. Thus, the amount of visual bias increased with age while the amount of

auditory bias decreased. As was seen in the AV and Auditory Only Conditions, the older child groups (nine and eleven year olds) performed in a similar fashion and younger child groups (five and seven year olds) also performed in a similar fashion. Specifically, the older child groups had fewer AC responses and more VC responses than the younger child groups. The number of Blend responses, in general, did not change across the age groups, though a significant difference was found between the adults and the five year old group. There were no significant differences found across the age groups for the number of Other responses.

For the adult-like criterion, defined according to the results found in the AVC Condition of this experiment, significantly more adults were found to have reached the adult-like criteria than all of the child subject groups. A general trend was seen that as subject age increased so did the number of subjects who reached the adult-like criteria. The only significant difference within the child groups, however, occurred between the five year old group and the eleven year old group. The results from this analysis suggest that the development of audiovisual speech perception is still progressing beyond the age of 12.

Chapter 7

Discussion and conclusions

In this investigation two main questions were addressed. The first was how do children respond to different manipulations of auditory and visual information at different ages? The second was at what age(s) do adult-like responses become evident within the McGurk paradigm? It is the goal of this final Chapter to examine how the findings from this investigation can address these questions and the previous findings in the literature. This Chapter contains four sections. The first section discusses the findings that emerged from the investigation of unimodal and bimodal perception across the age groups. The second section addresses the findings from the McGurk paradigm in which conflicting auditory and visual speech information was presented. In the third section overall conclusions are presented, limitations of the study are outlined, and some directions for future research are proposed. In the fourth and final section possible clinical implications of this research are considered.

Age effects in unimodal and bimodal speech perception

It was hypothesized that, in every age group, performance would be less accurate when subjects were presented with visual speech information than when they were tested with either auditory or both auditory and visual speech sequences. The present findings are consistent with this hypothesis in that, in every age group, performance in the AV and in the Auditory Only Condition was significantly better than in the Visual Only Condition. Not surprisingly, these findings indicate that vision is a less efficient route by which subjects can obtain phonetic information, as has been

shown repeatedly in the literature (e.g. Woodward and Barber, 1960; Fisher, 1968).

It was also hypothesized that subjects would show more accurate perception when both auditory and visual information were provided than when just the auditory information was present. That is, subjects were also expected to show an advantage in processing a bimodal input over a unimodal auditory input. This hypothesis was also upheld, but only for the adult group, which is consistent with numerous previous studies (e.g. Woodword and Barber, 1960; Binnie et al., 1974; Reisberg, et al., 1987). Unexpectedly, there was no significant difference between performance in the AV and Auditory Only Conditions in any of the child groups. Although there was a trend in this direction for the majority of the children in the seven, nine, and eleven year old groups, the differences were very small. These findings suggest that there may be an advantage of bimodal audiovisual speech perception over unimodal auditory speech perception which emerges with increasing age.

Performance in the bimodal and in each of the unimodal conditions was also analyzed to address several hypotheses regarding age-related differences in absolute performance within each condition. It was hypothesized that the ability to use visual speech information increases with age. Accordingly, it was predicted that there would be a systematic increase in performance in the Visual Only Condition and in the AV Condition. However, no age-related increase was predicted with respect to perception in the Auditory Only Condition. Each of these hypotheses was supported.

As expected, analysis of the results from the AV Condition showed a systematic increase in accuracy of perceptual responses across the age groups. The poorest performance (mean accuracy of 78%) was observed in the youngest age group. The best performance in this condition was observed in the adult group who exhibited a mean accuracy rate of 97% which represents ceiling performance in this task. The largest changes in audiovisual speech processing appear to occur between the ages of five and nine. There do not appear to be any significant developmental changes in audiovisual speech perception occurring in children older that nine but perhaps if a larger number of subjects were tested, additional age differences would have been observed. Overall, these findings show that as age increases subjects improve in the ability to perceive bimodal speech patterns. Given that there was no clear evidence of age-related increases in perception via audition alone (to be discussed below), it can be concluded that these findings reflect an age-related increase in the ability to interpret visual speech information. These results conform to the earlier findings of McGurk and MacDonald (1976), Massaro (1984), and Massaro, et al. (1986).

The findings from the Visual Only Condition also support the hypothesis that the use of visual speech information increases with age in that, as the age of the subject groups increased, there was a clear increase in performance in the Visual Only Condition. Performance in this condition varied from an average of 11% correct in the five year old group to an average of 62% correct for the adult group. It was also found that the older age groups were significantly better lip-readers than the two younger age groups. It appears then,

that lip-reading is approaching the adult level of performance by nine years of age. This finding also supports the findings of Evans (1965), McGurk and MacDonald (1976), Massaro (1984), and Massaro et al. (1986) who found that children under the age of eight are less able to use visual speech information easily. Again, if a larger number of subjects were tested then perhaps more age differences would have become apparent.

As predicted, analysis of performance in the Auditory Only Condition failed to show a systematic improvement in auditory perceptual performance as a function of age. A significant age effect was observed. However, the only reliable difference between the groups was found between the seven and eleven year olds; there was no difference between the seven year old group and the adults. Thus, it can be concluded that the difference observed was not due to an age related process. Rather, mean accuracy in the Auditory Only Condition across all of the age groups centered around 80%, indicating that all of the groups were performing at a high level of accuracy. This result was expected since the auditory modality is the primary source for speech information.

Overall, the comparisons of unimodal and bimodal findings in this experiment indicate that audiovisual and visual perception of speech is improving as subject age increases, with large improvements observed between five and nine years of age. In both conditions further non-significant improvements were also observed in children between 9 and 12 years of age. In contrast, a comparable age related increase in auditory speech perception performance was not observed. This pattern of age related changes across the three

conditions suggests that as the child's ability to use visual information increases there is a concomitant increase in their ability to perceive bimodal audiovisual patterns. However, the developmental pattern seen in the relative performance across the AV, Visual Only, and Auditory Only Conditions suggests that the ability to benefit from bimodal speech materials does not emerge until later in development, as only the adults demonstrate superior performance when bimodal stimuli were presented over performance when unimodal auditory stimuli were presented. The developmental patterns of audiovisual speech perception were further examined in the findings from the AVC condition.

Age effects in the perception of conflicting visual and auditory speech information

The results from the AVC condition provided further information about the nature of perceptual processing which occurs in response to a bimodal speech stimulus. Recall, in this condition the visual and auditory productions do not match and that subjects' responses were scored according to a system developed by Werker et al. (1992), as was discussed in Chapter 5. The first response category was Auditory Capture (AC) which indicates an inherent bias towards the auditory modality in the subjects' responses. The second response category was Visual Capture (VC) which indicates that the subject is responding according to the visual information and thus illustrates a bias towards the visual modality. The third category was a Blend which was assigned when the subject gave a response whose place of articulation was in between the place of the auditory and the visual stimulus (e.g. the response /da/ given for the

/ba/+/ga/ AVC stimulus). A Blend response is an illustration of the integration of auditory and visual place information; as such it also indicates that the subject is utilizing visual speech information. The final response category was that of Other which was assigned if the subject gave no response or a response that could not be classified according to the other three categories.

It was hypothesized that before the age of seven, visual influence on speech perception would be weak, and likewise the integration of auditory and visual information would not occur, as was found by McGurk and MacDonald (1976), Massaro (1984), and Massaro et al. (1986). The younger children would thus show a weaker visual influence than the adults and would demonstrate a bias towards the auditory information. It was further hypothesized that by the age of nine visual information would have a greater influence on speech perception. That is, the children aged between seven and nine would show some integration responses at this age, though not as profound as the effects seen in the adults, (McGurk & MacDonald, 1976) and more of a bias towards the visual information. It was also hypothesized that the eleven year old children would show more of an influence from the visual component of speech, thus supporting the findings of Evans (1965). These children would show a response pattern similar to that of the adults. Accordingly, it was expected that both VC and Blend responses would increase with age, with the largest changes occurring before the age of nine. Likewise it was expected that AC responses would decrease with increasing age.

As expected, it was found that AC responses decreased with age while VC responses increased, indicating that the use of visual

information becomes more prominent with increasing age. It appears that the number of AC responses decreased between the age of five and nine; further decreases were evident in the older age groups but did not reach statistical significance. The number of VC responses was seen to increase between the ages of five and nine, and also between age nine and adulthood. Thus, there is an increased influence of vision on speech perception across a broad span of development. These results are consistent with earlier findings reported by McGurk and MacDonald (1976), Massaro, (1984), and Massaro et al. (1986) in showing that children under the age of eight appear to be more influenced by the auditory channel than the visual channel in their perception of speech.

As expected, Blend responses, which reveal the integration of auditory and visual place information, also changed as a function of age, though not to the same degree as seen for the number of AC and VC responses. Although there was a steady increase in the number of Blend responses, it was found that the only significant difference was between the youngest age group (five year olds) and the oldest age group (adult). This finding again supports the results of McGurk and MacDonald (1976), Massaro (1984), and Massaro et al. (1986) indicating that young children are not as easily influenced by the visual source of information, and are not as adept at integrating auditory and visual speech information compared to adults.

Finally, it is important to note that the number of Other responses was very low and did not differ significantly between subject age groups. This indicates that the majority of subjects were able to complete the entire task and to provide responses that were

typical of those reported in previous studies using the McGurk paradigm.

To further assess the ages at which individual children begin to show a mature response pattern in the McGurk paradigm, the AVC results were also evaluated with respect to an adult response criterion. Recall that an adult-like response pattern was determined from the adult AVC data gathered in this study and was defined as VC≥AC and Blends ≥6. It was predicted, on the basis of previous findings (McGurk & MacDonald, 1976; Massaro, 1984; Massaro et al., 1986) that the five, seven, and nine year old groups would not show the adult-like pattern. However on the basis of the findings of Evans (1965), an adult-like pattern was expected to be evident in the eleven year old group.

Overall, it was found that none of the subjects in the five year old group showed the adult-like pattern, but after this age more subjects were observed to show the adult-like pattern as age increased. As predicted, significantly fewer subjects demonstrated an adult-like pattern in the three youngest age groups (five, seven, and nine year olds). However, the hypothesis that the eleven year old group would show an adult-like pattern was not confirmed. Rather, there was a jump from 50% (5/10) of the subjects in the eleven year old group to 93% (14/15) of the adults reaching the adult-like criteria. This finding suggests that there is further development of audiovisual integration occurring in children older than 12. These results supported the earlier findings of McGurk and MacDonald (1976), Massaro (1984), and Massaro et al. (1986) that adults and young children process audiovisual speech information

differently; the present findings also reveal that audiovisual speech processing in most children is still developing beyond the age of 12 years.

Overall conclusions, limitations and directions for future research

When we examine the findings from all four conditions it is possible to draw some conclusions about the nature of the development of audiovisual speech perception. Age related changes were seen in all of the conditions except for the Auditory Only Condition. Changes were observed in performance on the Visual Only, and AV Conditions which was seen to improve with age, while the number of VC responses (AVC condition), and Blends were seen to increase in conjunction with a decrease in AC responses. This indicates that visual information has more of an influence on speech perception as age increases. A general pattern of age related changes occurred from age five to nine, and additional changes, though not as dramatic, were seen from ages nine to eleven. There were still differences in the responses between the eleven year old group and the adults, especially apparent when an adult-like pattern of responses was defined. The data thus suggest that the development of audiovisual speech perception is still continuing beyond the age of 12.

As with any scientific investigation there are often limitations that are found that should be explored in future follow-up studies. One limitation is that of statistical power. Due to the number of subjects that were recruited for this study it was not possible, in the time frame that was given, to place an equal number of subjects in each group. In the statistical analysis of the data this was taken into

account and the appropriate statistics were chosen. However, the use of unequal cell sizes reduces statistical power and thus some of the age differences might have been missed due to this. Future studies should test more subjects to attain equal cell sizes to increase statistical power and to discover if there are any more developmental differences than were found in this experiment.

It should be mentioned that the results from this study are possibly confounded with the effects of stimulus editing that were employed to create the AV and AVC stimuli. Recall that both the AV and AVC stimuli were created by matching a digitized audio file with a visual articulation sequence from the videodisk. Natural audiovisual productions were not recorded directly from the videodisk. Although there was a lot of effort made to create as natural an utterance as possible, the perceived naturalness of the stimuli was based upon adult judgements. Werker et al. (1992) point out that young children may be more sensitive than adults to the degree to which an articulatory movement matches that which has been previously experienced. Therefore, perhaps the children are detecting an auditory/visual synchrony conflict and thus the children do not use the visual information in the processing of an utterance. Therefore, in future research it would be useful to replicate the present findings with natural unedited AV syllables as well as synthetically created AVC syllables.

Overall the performance in the Auditory Only Condition observed in the current study was poorer than has been observed in past research. For example, McGurk and MacDonald (1976) found that the accuracy for the adult group at repeating the phonemes in

their auditory only condition was 99%. In the present investigation the adult subjects performed at the 85% level. The youngest group that McGurk and Macdonald used was the preschool group (three to four years) which performed at the 91% level, while the youngest group in this study (4:7 to 5:11) performed at the 81% level. Sekiyama and Tohkura (1991) suggest that poorer auditory intelligibility may produce more of a reliance on the visual channel.

To explore this issue further it is necessary to look at the errors made during this condition. For the adult group in this study, out of 34 errors 94% (32/34) of them were confusions between /va/ and /ea/. The remaining 6% of the 34 errors (2 from one subject) were confusions between /ba/ and /va/. Both /va/ and /ea/ have been found to be vulnerable to confusions in perceptual studies (e.g. Miller & Nicely, 1955).

In the Auditory Only Condition some adult subjects commented that they thought that the speaker had an English accent while others thought that he had a strong American accent. The issue of accent may become clearer as specific stimulus materials become standard and are used with subjects with diverse dialects of English. Although these comments did not seem to be related to the accuracy in the Auditory Only Condition, it is possible that the results could be influenced by differences in dialect.

Another reason for the lower auditory intelligibility findings in this study could be due to the fact that the stimuli were presented using standard VHS equipment. The quality of the soundtrack, even with Dolby B noise reduction, was not up to the standards of VHS HiFi, or the videodisk from which the stimuli were copied. This was

a limitation that could not be avoided due to the equipment that was available at the time of running, but in future studies a better recording medium should be employed. The stimuli were also played through a loudspeaker in the video monitor; a separate loudspeaker is likely to provide better audio fidelity. A separate high quality loudspeaker mounted away from the video monitor has been used by Ward (1992) with good results but there is some research currently being undertaken to examine how manipulations of the spatial location of the loudspeaker can affect subjects' perceptions of McGurk stimuli (Munhall, 1994, personal communication). In conclusion, a separate loudspeaker mounted close to the video monitor may improve the auditory quality of the stimuli, but the location of this loudspeaker should be carefully determined in order to not disrupt the McGurk effects.

In general therefore, the intelligibility of the auditory stimuli is an important factor in studies on the McGurk effect, as was discussed by Sekiyama and Tohkura (1991). The use of standard materials for investigations involving the auditory and visual perception of speech is also an important issue to pursue, but caution should be paid to dialectical differences between the subjects and the materials used. Standard materials may still be used but modulation in the effects produced can be due to differences in the equipment used to deliver the stimuli to the subject. In future studies a great deal of attention should be paid to these stimulus factors.

It was found that the five year old subjects were not very influenced by the visual information and that none of them attained the adult-like criteria. It was felt that the results for these subjects

may have been influenced by the fact that they often became disinterested in the experimental task, and were easily distractible. In any experiment with young children it must be taken into account that subject cooperation in the experimental task may influence the ultimate results of the experiment. In this experiment breaks were given to the child along with small tokens (stickers) to try and maintain their interest. Often the parents stayed in the testing room (otherwise they watched through a one way window) to encourage the child in conjunction with the experimenter and the assistant. The conclusions made about the five year old group are thus tentative.

The issue of subject cooperation was very apparent in the Visual Only Condition which appeared to be especially difficult for the younger children. The difficulty of the task was apparent not only in their responses but also by the lack of cooperation seen in some children in the youngest groups. Some of the younger children were unable to complete the task while others who persisted appeared to be guessing, thus the performance in the youngest group was poor but is also a less precise index of their ability to use visual information compared to other groups. These two factors support the general hypothesis that children do not use the visual sense to the same degree as do adults. The children are thus presented with a task that is not only something they find difficult and novel, but is a task that they are in fact unable to do, due to the point at which they are at in their perceptual development. This issue is one which should be kept in mind whenever research of this type is conducted with young children.

As has been discussed in this section there are a number of limitations of this current investigation. Future research should address these concerns. New studies should be designed to examine statistical power, the use of natural versus synthetic pairings of audiovisual stimuli, the issue of auditory intelligibility, and the cooperation of young subjects in difficult experimental tasks. Additional studies should also be undertaken to examine the findings from the current investigation, specifically, that the development of audiovisual speech perception is still progressing beyond the age of 12 and that more differences may be found between children of different ages in the development of audiovisual speech perception by testing a larger number of subjects.

One issue which could not be addressed by the data from this study is that of the importance of the observed developmental pattern in the use of visual information in speech perception, as illustrated by the findings of this investigation and McGurk & MacDonald (1976), Massaro (1984), and Massaro et al. (1986). The question that naturally comes to mind is, why does this visual bias occur and why is it age related? McGurk, Turnure, and Creighton (1977) and Massaro (1984) proposed that initially, the young child's sensory systems are independent and that synthesis gradually takes place. The reliance on the auditory channel by the young child, as was shown in this study and the investigations by McGurk and MacDonald (1976), Massaro (1984), and Massaro et al. (1986), makes sense for speech perception. In learning to speak the young child must monitor both his/her own speech production while receiving and processing the speech of others in his/her environment.

Although the child can monitor both auditory and visual patterns in the speech of others, he/she can typically only monitor his/her own productions via the auditory channel. Furthermore, as discussed in Chapter 2 on the audiovisual perception of speech, speech information is less reliably conveyed by the visual channel alone. Therefore, the young child has to depend on the auditory input to learn how to produce and process the speech sounds and ultimately associate them with meaning without the influence of the visual speech information (Erber,1974; McGurk, 1988).

Perhaps the lack of or reduced use of the visual aspects of speech plays an important role in linguistic and perceptual development by enhancing the attention paid to the auditory modality. Intuitively one might assume that an intellectually mature organism can learn a larger amount of information more efficiently than an immature one. When young children and infants are examined however, this does not seem to be the case. Although infants are intellectually immature they do seem to acquire information very efficiently (Bjorklund and Green, 1992). In fact, according to Turkewitz and Kenny (1982), cognitive immaturity plays an important role in development. The limitations of the cognitive abilities of a young human being are not handicaps to overcome. Rather, these cognitive limitations provide adaptive advantages for perceptual organization. For example, the sensory limitations of many young animals are adaptive features that reduce the amount of information that has to be dealt with and thus create a simple and comprehensible world for the young animal (Turkewitz and Kenny, 1982). Elman (1991), using a neural network model,

found that limitations placed on the memory of a system, or the presentation of information in small pieces to that system, permit the network to learn the elements of a complex grammar. Elman suggested that these processes are similar to that which occurs within the child so that early limitations in a learner are developmentally advantageous to the acquisition of complex systems of information. This reasoning could explain why young children do not appear to use the visual information provided in speech. Their bias toward speech information conveyed via the auditory channel may reflect such an adaptive strategy. From this perspective, it can be hypothesized that the developmental patterns of audiovisual speech perception reflect some global constraints of cognitive development.

McGurk (1988) suggests that the developmental course of audiovisual speech perception generally follows phonological and articulatory development in children. McGurk (1988) feels that the development of language guides the development of audiovisual speech perception. As a child develops, so does his/her language, cognitive capacities, and the ability to use vision in speech perception. In future studies the use of materials which contain lexical and linguistic information (words and sentences, as used by Dekle et al., 1992; & McGurk, 1988) may help to demonstrate the effects of the visual channel on linguistic development.

Clinical implications

By examining the development of audiovisual speech perception in normal children it might be possible to design some tools to examine how hearing impaired children make use of visual

information in the perception of speech. Perhaps the McGurk paradigm may prove useful in examining how hearing impaired children utilize the visual and minimal auditory information that is available to them.

The study of audiovisual integration using the McGurk paradigm may provide findings that permit us to make clearer decisions about the rehabilitative needs of children who have received cochlear implants or who use amplification. It must be remembered that cochlear implants and amplification do not restore "normal hearing" to the recipient of this technology. These devices typically provide an auditory signal which can supplement visual speech information and thus make it possible for the perceiver to benefit from the integration of auditory and visual speech information.

This study along with the work of McGurk and MacDonald (1976), Massaro (1984), and Massaro et al. (1986), has shown that children use auditory and visual speech information in different ways than adults. By examining how hearing children integrate auditory and visual information it is possible to reassess the expectations that we have for recipients of amplification systems and cochlear implants. Tests of audiovisual speech perception are already being employed by cochlear implant teams to assess the speech perception abilities of candidates for this surgery and audiovisual materials are also used after the surgery for training and rehabilitation (e.g. Cook, 1991; Cooper,1991; Faulkner and Read, 1991; Tyler & Champe, 1993; Tyler, Opie, Fryauf-Bertschy, & Gantz, 1992). Audiovisual materials are also used by clinicians to assess the

benefits of amplification for an individual (e.g. Hasselrot, 1974; King, 1991).

The results of this study indicate that the development of audiovisual speech perception is still occurring after the age of 12 in normal children suggesting that rehabilitation efforts designed to facilitate the use of visual and audiovisual information should not be abandoned in children over the age of 12 years. This study also reveals that the integration of auditory and visual speech information is a developmental skill and this should be taken into account when assessments are made of the speech perception performance of a child who has received a cochlear implant or amplification system.

Summary

This investigation has examined the developmental process of auditory and visual information in children aged between 5 and 12 years of age. It has been found in general that there is a large increase in the use of visual information up until the age of nine. It appears, however, that beyond the age of nine there is still some audiovisual development occurring, possibly into adolescence.

The information obtained in this experiment is very important, not only to researchers but to clinicians as well. It was shown in the review of the literature how important the use of visual information is in the perception of speech. Studies such as this one provide more information about how the two channels of auditory and visual information are used separately and how they are combined in the perception of speech by children It is only through the examination of all the senses that specific information can possibly be gathered,

that conclusions can be made about the perception of the source of that information. Speech information is manifested in both the visual and the auditory domains, so it is only through the examination of the integrated percept that we can derive a comprehensive understanding of the perception of speech.

References

Bernstein, L. E., Demorest, M. E., Coulter, D. C., & O'Connel, M. P. (1991). Lip-reading sentences with vibrotactile vocoders: Performance in normal-hearing and hearing-impaired subjects. Journal of the Acoustical Society of America, 90(6), 2971-2984.

Bernstein, L. E. & Eberhardt, S. P. (1986). Johns Hopkins Lip Reading Corpus I-II. [videodisk] Johns Hopkins University, Baltimore, MD.

Bernstein, L. E., Eberhardt, S. P., & Demorest, M. E. (1989). Single-channel vibrotactile supplements to visual perception of intonation and stress. Journal of the Acoustical Society of America, 85(1), 397-405.

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

Bjorklund, D. F. & Green, B. L. (1992). The adaptive nature of cognitive immaturity. American Psychologist, 47(1), 46-54.

Byman, J. (1974). Testing of audiovisual speech perception related to everyday communication. Scandinavian Audiology Supplement, 4, 97-113.

Conrad, R. (1977). Lip-reading by deaf and hearing children. British Journal of Educational Psychology, 47, 60-65.

Cook, B. O. (1991). Testing and rehabilitation of cochlear implant patients at the department of audiology, Södersjukhuset, Stockhom. In H. Cooper (Ed.), Cochlear implants: A practical guide, (pp. 240-250). San Diego, CA: Singular Publishing Group Inc.

Cooper, H. (1991). Training and rehabilitation for cochlear implant users. In H. Cooper (Ed.), Cochlear implants: A practical guide, (pp. 219-238). San Diego, CA: Singular Publishing Group Inc.

Craig, W. N. (1964). Effects of preschool training on the development of reading and lip-reading skills of deaf children. American Annals of the Deaf, 109, 280-296.

Dekle, D. J., Fowler, C. A., & Funnell, M. G. (1992). Audiovisual integration in perception of real words. Perception and Psychophysics, 51(4), 355-362.

Demorest, M. E., Bernstein, L. E., & Eberhardt, S. P. (1987). Reliability of individual differences in lip-reading. The Journal of the Acoustical Society of America Supplement, 82, S24.

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.

Denes, P. B. (1963). On the statistics of spoken English. The Journal of the Acoustic Society of America, 35(6), 892-904.

Dodd, B. (1977). The role of vision in the perception of speech. Perception, 6, 31-40.

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

Dodd, B. (1980). Interaction of auditory and visual information in speech perception. British Journal of Psychology, 71, 541-549.

Dodd, B. (1987). The acquisition of lip-reading skills by normally hearing children. In B. Dodd, & R. Campbell (Eds.), Hearing by eye: The psychology of lip-reading, (pp 163-175). London: Lawrence Erlbaum Associates Ltd.

Eberhardt, S. P., Demorest, M. E., & Goldstein, Jr, M. H. (1990). Speech-reading sentences with single-channel vibrotactile presentation of voice fundamental frequency. Journal of the Acoustical Society of America, 88(3), 1274-1285.

Elman, J. L. (1991). Incremental learning, or the importance of starting small. CRL Technical report, University of California, San Diego.

Erber, N. P. (1972). Speech-envelope cues as an acoustic aid to lip-reading for profoundly deaf children, Journal of the Acoustical Society of America, 51(4-2). 1224-1227.

Erber, N. P. (1974). Visual perception of speech by deaf children. Scandinavian Audiology Supplement, 4, 97-113.

Evans, L. (1965). Psychological factors related to lip reading. The teacher of the deaf, 63, 131-137.

Fagg, J. D. (1992). Voice pitch as an aid to speechreading in young children. Unpublished Masters Thesis. University of Manchester, United Kingdom.

Faulkner, A. & Read, T. (1991). Speech perception and its assessment. In H. Cooper (Ed.), Cochlear implants: A practical guide, (pp. 251-282). San Diego, CA: Singular Publishing Group Inc.

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

Fowler, C. A. & Dekle, D. J. (1991). Listening with eye and hand: cross-modal contributions to speech perception. Journal of Experimental Psychology, 17(3), 816-828.

Goldman, R. & Fristoe, M. (1986). The Goldman Fristoe test of articulation. Pines MN: American Guidance Service Inc.

Green, K. P. (1994). The influence of an inverted face on the McGurk effect. The Journal of the Acoustical Society of America Supplement, 95(5), 3014.

Green, K. P., Kuhl, P. K., & Melzoff, A. N. (1988). Factors affecting the integration of auditory and visual information in speech: The effect of vowel environment. Poster presented at the meetings of the Acoustical Society of America, November 15-18, Honolulu, Hawaii

Green, K. P., Kuhl, P. P., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the McGurk effect. Perception and Psychophysics, 50(6), 524-536.

Green, K. W., Green, W. B., & Holmes, D. W. (1980). Speechreading abilities of young deaf children. American Annals of the Deaf, 906-908.

Hasselrot, M. (1974). Exploration of an audiovisual test procedure with background noise for patients with noise-induced hearing loss using hearing aids. Scandinavian Audiology Supplement, 4, 165-181.

Humphrey, K., Tees, R. C., & Werker, J. (1979). Auditory-visual integration of temporal relations in infants. Canadian Journal of Psychology, 33(4), 347-352.

King, A. (1991). Audiological assessment and hearing aid trials. In H. Cooper (Ed.), Cochlear implants: A practical guide, (pp. 101-108). San Diego, CA: Singular Publishing Group Inc.

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

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

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 16th-20th , Seattle, Washington.

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

MacLeod, A. & Summerfield, Q. (1987). Quantifying the contribution of vision to speech perception in noise. British Journal of Audiology, 21, 131-141.

Manuel, S. Y., Repp, B. H., Liberman, A. M., & Studdert-Kennedy, M. (1983, November). Exploring the "McGurk Effect". Paper presented at the meeting of the Psychonomic Society, San Diego, California.

Marascuilo, L. A. (1966). Large-sample comparisons. Psychological Bulletin, 65(5), 280-290.

Marascuilo, L. A. (1971). Statistical methods for behavioral science research. Montreal: McGraw-Hill Book Company.

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

Massaro, D. W. (1987). Speech perception by ear and eye. In B. Dodd, & R. Campbell (Eds.), Hearing by eye: The psychology of lip-reading, (pp 53-83). London: Lawrence Erlbaum Associates Ltd.

Massaro, D. W. (1989). Multiple Book review of, speech perception by ear and eye: A paradigm for psychological inquiry. Behavioral and Brain Sciences, 12, 741-794.

Massaro, D. W. & Cohen, M. M. (1990). Perception of synthesized audible and visible speech. Psychological Science, 1(1), 55-63.

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.

Massaro, D. W., Cohen, M. M., Gesi, A., Heredia, R., & Tsuzaki, M. (1993). Bimodal perception of speech: an examination across languages. Journal of Phonetics, 21, 445-478.

McGurk, H. (1988). Developmental psychology and the vision of speech. Inaugural Lecture by Professor Harry McGurk 2nd March 1988, Personal communication.

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

McGurk, H., Turnure, C., & Creighton, S. J. (1977). Auditory-visual coordination in neonates. Child Development, 48, 138-143.

Miller, G. A. & Nicely, P. F. (1955). An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 27, 338-352.

Mills, A. E. (1987). The development of phonology in the blind child. In B. Dodd, & R. Campbell (Eds.), Hearing by eye: The psychology of lip-reading, (pp. 145-162). London: Lawrence Erlbaum Associates Ltd.

Plant, G. & Macrae, J. (1987). Testing visual and auditory visual speech perception. In M. Martin (Ed.), Speech Audiometry, (pp. 179-206). London: Whurr Publishers Ltd.

Resiberg, D., McLean, J., & Goldfield, A. (1987). Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli. In B. Dodd & R. Campbell (Eds.), Hearing by eye: The psychology of lip-reading, (pp. 97-113). London: Lawrence Erlbaum Associates Ltd.

Sekiyama, K. & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. Journal of the Acoustical Society of America, 90(4),1797-1805.

Sekiyama, K. & Tohkura, Y. (1993). Inter-language differences in the influence of visual cues in speech perception. Journal of Phonetics, 21, 427-444.

Sekiyama, K. & Nishino, K. (1994). Exploring the Japanese McGurk effect using various S/N ratios and speakers. The Journal of the Acoustic Society of America Supplement, 95(5), 2872.

Sumby, W. H., & Pollock, I. (1954). Visual contribution to speech intelligibility in noise. Journal of the Acoustical Society of America, 26, 212-215.

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

Turkewitz, G. & Kenny, P. A. (1982). Limitations on input as a basis for neural organization and perceptual development: A preliminary theoretical statement. Developmental Psychobiology, 15, 357-368.

Tyler, R. S. & Champe, G. (1993). An audiovisual feature test for young deaf children in English, French, Spanish, and German. In B. Fraysse & O. Deguine (Eds.), Advances in Otorhinolaryngology: Vol 48. Cochlear implants: New perspectives, (203-206), Basel: Karger.

Tyler, R. S., Opie, J. M., Fryauf-Bertschy, H., & Gantz, B. J. (1992). Future directions for cochlear implants. Journal of Speech-Language Pathology and Audiology, 16(2), 151-164.

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

Walton, G. E. & Bower, T. G. R. (1993). Amodal representation of speech in infants. Infant Behavior and Development, 16, 233-243.

Ward, M. (1992). The effect of audio-visual desynchrony on the integration of auditory and visual information in speech perception. Unpublished Honours thesis, Queen's University, Kingston, Ontario.

Werker, J. F., McGurk, H., & Frost, P. E. (1992). La langue at les levres: Cross-language influences on bimodal speech perception. Canadian Journal of Psychology, 46(4), 551-568.

Woodward, G. A. & Barber, C. G. (1960). Phoneme perception in lip-reading. Journal of Speech and Hearing Research, 3, 212-222.

APPENDIX A. Stimuli and frame number from the Johns Hopkins Lip-reading Corpus Volume 1 (Bernstein and Eberhardt, 1986)

STIMULI	INITIAL FRAME NUMBER	FINAL FRAME NUMBER
/ba/	13943	13970
/va/	14090	14119
/ða/	13549	13573
/da/	13171	13204
/ga/	12823	12852

APPENDIX B. Order of presentation of the Auditory Only, and Visual Only stimuli on Tape 1

auditory alone with blank screen = a

Practice

1 /ba/.a
2 /va/.a
3 /ga/.a
4 /ɵa/.a
5 /da/.a

Test

1 /da/.a
2 /ba/.a
3 /da/.a
4 /va/.a
5 /ga/.a
6 /ɵa/.a
7 /ga/.a
8 /ɵa/.a
9 /va/.a
10/ba/.a
11 /ga/.a
12 /ba/.a
13 /ɵa/.a
14 /va/.a
15 /da/.a

Summary

Practice: 5 Auditory Only syllables: 1 repetition of 5 syllables
Test: 15 Auditory Only syllables: 3 repetitions of 5 syllables

N=20

visual alone no sound = v

Practice

1 /ga/.v
2 /ɵa/.v
3 /ba/.v
4 /va/.v
5 /da/.v

Test

1 /da/.v
2 /ga/.v
3 /ɵa/.v
4 /va/.v
5/ba/.v
6 /va/.v
7 /ba/.v
8 /ga/.v
9 /da/.v
10/va/.v
11 /da/.v
12 /ɵa/.v
13 /ga/.v
14 /ba/.v
15/ɵa/.v

Summary

Practice: 5 Visual Only syllables: 1 repetitions of 5 syllables
Test: 15 Visual Only syllables: 3 repetitions of 5 syllables

N=20

APPENDIX C. Alignment of the videodisk file with the digitized audio file in msec for the AV stimulus condition

AUDIOVISUAL STIMULI	DELAY OF AUDIO FILE
/ba/	195 msec
/va/	109 msec
/ɵa/	30 msec
/da/	420 msec
/ga/	162 msec

APPENDIX D. Alignment of the videodisk file with the digitized audio file in msec for the AVC stimulus condition

VISUAL STIMULI	DELAY OF VIDEO FILE	AUDITORY STIMULI	DELAY OF AUDIO FILE
/va/	0 msec	/ba/	133 msec
/ɵa/	32 msec	/ba/	0 msec
/da/	0 msec	/ba/	295 msec
/ga/	0 msec	/ba/	39 msec

APPENDIX E. Order of presentation of the audiovisual (AV and AVC) stimuli on Tape 2

auditory/visual same syllable: av
auditory/visual conflicting syllable: avc

Practice

1 /da/.av
2 /da/.avc
3 /ɵa/.av
4 /va/.av
5 /ga/.av
6 /ɵa/.avc
7 /ga/.avc
8 /va/.avc
9 /ba/.av

Test

one	two	three
1 /ɵa/.avc	1 /ga/.av	1 /da/.avc
2 /ɵa/.av	2 /va/.av	2 /da/.av
3 /ga/.av	3 /ga/.avc	3 /ɵa/.avc
4 /da/.avc	4 /da/.avc	4 /ba/.av
5 /ga/.avc	5 /va/.avc	5 /va/.avc
6 /va/.avc	6 /ɵa/.avc	6 /ga.avc

four	five	six
1 /ba/.av	1 /da/.av	1 /ga/.avc
2 /ɵa/.av	2 /ba/.av	2 /va/.av
3 /ɵa/.avc	3 /ga/.avc	3 /da/.avc
4 /va/.avc	4 /ɵa/.avc	4 /ga/.av
5 /da/.avc	5 /va/.avc	5 /ɵa/.avc
6 /ga/.avc	6 /da/.avc	6 /va/.avc

seven	eight	nine
1 /ɵa/.av	1 /da/.av	1 /ɵa/.avc
2 /da/.avc	2 /va/.avc	2 /da/.av
3 /va/.av	3 /ba/.av	3 /va/.avc
4 /ɵa/.avc	4 /ga/.avc	4 /ga/.avc
5 /ga/.avc	5 /ɵa/.avc	5 /ga/.av
6 /va/.avc	6 /da/.avc	6 /da/.avc

ten	eleven	twelve
1 /ɵa/.av	1 /va/.av	1 /ga/.avc
2 /da/.av	2 /ga/.avc	2 /ba/.av
3 /ga/.avc	3 /ga/.av	3 /da/.avc
4 /va/.avc	4 /va/.avc	4 /ɵa/.avc
5 /ɵa/.avc	5 /ɵa/.avc	5 /va/.av
6 /da/.avc	6 /da/.avc	6 /va/.avc

thirteen
1 /ɵa/.avc
2 /ga/.avc
3 /da/.avc
4 /ga/.av
5 /ɵa/.av
6 /va/.avc

Summary

Practice: 9 AV and AVC syllables: 1 repetition of 9 syllables

Test: 26 AV syllables: 5 repetitions of 5 syllables + 1
52 AVC syllables: 13 repetitions of 4 syllables

N=87

APPENDIX F. Language Questionnaire presented to the adult subjects.

Language Questionnaire

Name:_____

Age:_____ Sex:_____

Birthplace:_____
town/city

_____
province/state/country

Please list the places that you have lived (from birth) and the ages that you were when you lived there:

Parent's birthplace (city, province, country)

Father:_____Mother:_____

_____

Do you speak any other languages other than English fluently? yes no

If yes, then what language (s)?_____

Was English the first language that you learned? yes no

If no, then which language (s)?

_____

Have you ever had any courses in Phonetics (the scientific study of speech sounds, Phonetics is taught at the University level in Speech Science, Linguistics, or Psycholinguistics courses)? yes no

APPENDIX G. Instructions given to the adult and child subjects for the Auditory Only, Visual Only, AV and AVC Conditions.

Auditory Only Condition

"You will hear a man say some sounds like /ba/, /ga/ and /da/. I would like you to look at the T.V., even though you won't see anything, and tell me what the man says. Try and keep your eyes on the T.V at all times. If you are not sure of what he said then you can make a guess. He is going to say five sounds first, for practice."
[after presentation of practice trials the instructions were repeated and then the test trials were presented]

Visual Only Condition

"Now you will see the man but not hear him, he is going to say some more sounds like /ba/, /ga/ and /da/, just like you just heard. I would like you to look at the T.V. and tell me what the man says. Try and keep your eyes on the T.V at all times. If you are not sure of what he said then you can make a guess. He is going to say five sounds first, for practice."
[after presentation of practice trials the instructions were repeated and then the test trials were presented]

AV and AVC Conditions

"Now you will see the man and hear him, (this is easier) he is going to say some more sounds like you just heard. I would like you to look at the T.V. and tell me what the man says. Try and keep your eyes on the T.V at all times. Sometimes the man may sound a little funny but if you are not sure of what he said then you can make a guess. He is going to say nine sounds first, for practice."
[after presentation of practice trials the instructions were repeated and then the test trials were presented]

APPENDIX H. Score sheets for the Auditory Only, Visual Only, AV, and AVC Conditions

auditory alone with blank screen: a

Practice

1 /ba/.a _____
2 /va/.a _____
3 /ga/.a _____
4 /θa/.a _____
5 /da/.a _____

Test

1 _____
2 _____
3 _____
4 _____
5 _____
6 _____
7 _____
8 _____
9 _____
10 _____
11 _____
12 _____
13 _____
14 _____
15 _____

Summary

Practice: 5 Auditory Only syllables: 1 repetition of 5 syllables
Test: 15 Auditory Only syllables: 3 repetitions of 5 syllables

N=20

visual alone no sound: v

Practice

1 /ga/.v _____
2 /ɵa/.v _____
3 /ba/.v _____
4 /va/.v _____
5 /da/.v _____

Test

1 _____
2 _____
3 _____
4 _____
5 _____
6 _____
7 _____
8 _____
9 _____
10 _____
11 _____
12 _____
13 _____
14 _____
15 _____

Summary

Practice: 5 Visual Only syllables: 1 repetitions of 5 syllables
Test: 15 Visual Only syllables: 3 repetitions of 5 syllables

N=20

auditory/visual same syllable: av
auditory/visual conflicting syllable: avc

Practice

1 /da/.av _____
2 /da/.avc _____
3 /ɡa/.av _____
4 /va/.av _____
5 /ga/.av _____
6 /ɡa/.avc _____
7 /ga/.avc _____
8 /va/.avc _____
9 /ba/.av _____

Test

	one	two	three
1	_____	_____	_____
2	_____	_____	_____
3	_____	_____	_____
4	_____	_____	_____
5	_____	_____	_____
6	_____	_____	_____

	four	five	six
1	_____	_____	_____
2	_____	_____	_____
3	_____	_____	_____
4	_____	_____	_____
5	_____	_____	_____
6	_____	_____	_____

	seven	eight	nine
1	_____	_____	_____
2	_____	_____	_____
3	_____	_____	_____
4	_____	_____	_____
5	_____	_____	_____
6	_____	_____	_____

	ten	eleven	twelve
1	_____	_____	_____
2	_____	_____	_____
3	_____	_____	_____
4	_____	_____	_____
5	_____	_____	_____
6	_____	_____	_____

	thirteen
1	_____
2	_____
3	_____
4	_____
5	_____
6	_____

Summary

Practice: 9 AV and AVC syllables: 1 repetition of 9 syllables

Test: 26 AV syllables: 5 repetitions of 5 syllables + 1
52 AVC syllables: 13 repetitions of 4 syllables

N=87