DEVELOPMENTAL CHANGES OF AUDIOVISUAL INTEGRATION: A CROSS SECTIONAL STUDY

Preeti Sahu

Registration No: 10AUD023

A dissertation submitted as a part fulfillment of final year

Master of Science (Audiology)

University of Mysore, Mysore.

ALL INDIA INSTITUTE OF SPEECH AND HEARING,

MANASAGANGOTRI, MYSORE- 570006

May, 2012

CERTIFICATE

This is to certify that this dissertation entitled ***'Developmental changes of audiovisual integration: A cross sectional study'****, is a bonafide work submitted in part fulfillment for the degree of Master of Science (Audiology) of the student Registration No.10AUD023. This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any university for the award of any Diploma or Degree.*

Mysore

May, 2012

Dr. S. R. Savithri

Director

All India institute of Speech and Hearing

Manasagangothri, Mysore -570006

CERTIFICATE

This is to certify that this dissertation entitled ***'Developmental changes of audiovisual integration: A cross sectional study'****, has been carried out under my supervision and guidance. It is also certified that this dissertation has not been submitted earlier to any other University for the award of any Diploma or Degree.*

Mysore

May, 2012

Ms.Geetha C.

Lecturer in Audiology,

Department of Audiology,

All India Institute of Speech and Hearing,

Mansagangothri, Mysore -570006.

DECLARATION

This is to certify that this dissertation entitled ***'Developmental changes of audiovisual integration: A cross sectional study'*** *is the result of my own study under the guidance of Ms.Geetha C., Lecture in Audiology, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier to any other University for the award of any Diploma or Degree.*

Mysore ***Registration No.10AUD023***

May, 2012

Dedicated to

Mummy Papa

'Thank u god for giving such a precious gift.'

ACKNOWLEDGEMENT

"We are what we repeatedly do. Excellence, then, is not an act, but a habit."

Aristotle.

A great number of people have contributed in assorted ways to this dissertation and the making of this dissertation deserved special mention. It is a pleasure to convey my gratitude to them all in my humble acknowledgment.

In the first place I would like to thank ***God Almighty*** *for helping me through the successful completion of this dissertation.*

I would like to extend my prayers and regards to ***late Dr.Vijayalakshmi Basavaraj*** *under whose umbrella we have grown gracefully. Madam may your soul rest in peace.*

I would like to record my gratitude to ***Mrs.Geetha ma'am, my guide, my mentor,*** *for her supervision, advice, and guidance from the very early stage of this dissertation as well as giving me extraordinary experiences throughout the work. Above all and the most needed, she provided me unflinching encouragement and support in various ways. Her constant oasis of ideas and passions in science, which exceptionally inspire and enrich my growth as a student. I am indebted to her more than she knows.*

I would like to thank ***Prof. S .R. Savithri****, Director, AIISH, for permitting me to carry out this research.*

Where would I be without my family? My parents deserve special mention for their inseparable support and prayers. My Father, ***Mr.Khamhan****, in the first place is the person who put the fundament of my learning character, showing me the joy of intellectual pursuit ever since I was a child. My Mother,* ***Mrs.Satarupa****, is the one who sincerely raised me with her caring and gentle love and my sisters (****Deepti*** *and* ***Tripti****), jeeju (****Sunil****) and my brother (****Rahul****), thanks for being supportive and caring.*

My genuine thanks to ***Debadutta sir,*** *who inspired me to do my Masters and completing it successfully. Thank you for being with me in my ups and downs of my life.*

I would like to thank ***Dr. Anjwani sir*** *for his constant support and help. Thank you so much sir.*

I would also acknowledge ***Sandeep sir*** *for his advice, help in each and every step from the beginning and his willingness to share his bright thoughts with me, which was very fruitful for shaping up my ideas in my studies.*

I also would like to thank ***Asha ma'am****,* ***Sujeet sir, Animesh sir, Ajith sir, Santosh sir and Priya ma'am*** *for their support.*

My thanks to ***Vasanthlakshmi Ma'am*** *and* ***Santosh sir*** *for their help in formulating my statistics.*

Thanks to ***Nike sir****,* ***Dhathri ma'am****,* ***Akshay raj sir*** *and of course all my cute seniors in all aspects of my dissertation from the day one.*

I would like to thank ***Wasim sir*** *and* ***Sitaram sir*** *for their timely helps and support.*

Collective and individual acknowledgments are also owed to the ***subjects*** *for my data collection was a great help indeed.*

Friendship is a wine, tastes better when it's old. Many thanks go in particular to my close friends ***Rajkishor (Raju), Roshni (mottu), Seby (Sebee)****,* ***Jiten (Gini).***

Thanks to all my classmates of M.Sc. AUDIO and few friends ***Merry****,* ***Saanu****,* ***Deepashree****,* ***Spoorthi****,* ***Kruthika****,* ***Apoorva****,* ***Sachi****,* ***Deepika****,* ***Arpita****,* ***Mythri****,* ***Deepthi, and Satbir*** *in particular and all my juniors of M.Sc. Audio & speech, BASLP and interns.*

And finally I would like to convey a very very special thank to ***Reeny & Pari di*** *for their love, care and support.*

Finally, I would like to thank ***everybody*** *who was important to the successful realization of this dissertation, as well as expressing my apology that I could not mention personally one by one.*

Table of contents

List of Tables

List of figures

Chapter 1

Introduction

Speech perception is an audio-visual phenomenon, in both the degraded and no degraded conditions. The influence of vision on speech perception is evident by McGurk effect. This effect demonstrates the occurrence of sensory integration by presenting conflicting auditory and visual information, for example, an auditory / ba/ presented with a visual /ga/, results in the perception of /da/. This suggests that auditory-visual speech is perceived as a 'whole' perceptual unit rather than as separate unimodal features (Green & Kuhl, 1991).

There are different factors that affect the process of audiovisual integration. One of the major factors is the age. Mcgurk and Macdonald (1976), in one of their early studies noted that significantly fewer children than adults show an influence of visual speech on perception. Further, research on developmental changes of speech perception, from infancy to young adulthood has shown a general trend of increasing use of visual information (Massaro, 1984; Robinson 2004) and increasing audiovisual integration over period of time (Mcgurk & Macdonald 1976; Rosenblum, Schmuckler, & Johnson 1997; Wightman, Kistler & Brungart ,2006).

This pattern of results has been replicated in studies done by different researchers (Desjardins, Rogers, & Werker, 1997; Dupont, Aubin, & Menard, 2005; Hockley &

Polka, 1994; Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Sekiyama & Burnham, 2004; Wrightman, Kistler, & Brungart, 2006).thereason for these may be in part ba a result of ongoing peripheral vision development (Massaro, 1984). Thompson, Barron and Laren (1986) also found that the children are poorer lip readers than adults, which results in lesser visual influence rather than attention playing a role.

The age by which the visual influence completes has also been researched. A few studies have observed benefit from visual speech by the pre-teen/teenage years (Conrad, 1977; Dodd, 1977, 1980; Hockley & Polka, 1994), with one report citing an earlier age of 8 years (Sekiyama & Burnham, 2004). Whereas, Iyar (2005) found that that 7 year olds females and 11 year old males are more likely to use auditory cues for speech perception, while the 11 year old females and 15 year old males used both auditory and visual information in poor listening conditions. There are few other studies which have compared the adults with children till the age of 11 years Sekiyama and Burnham (2004).

From the above it can be observed that, though many of the studies have compared participants with age of below 11 years and adults, there could be changes in audiovisual integration even after 11 years of age and there could gender differences, which is demonstrated in the study done by Iyar (2005).

The auditory visual integration as demonstrated by Mc.Gurk effect was also found to be depended on the linguistic experience of the person. It has been shown that Japanese speakers more frequently notice the incompatibility between auditory and visual cues than the English speakers (Sekiyama, 2004).the findings in Japanese speakers could be in relation to the fact that in Japanese, consonant clusters do not exist (Sekiyama & Tohkura, 1993). The developmental increase in visual influence over age could also possibly be related to the experience in articulating speech sounds.

It has been found that preschool children who shows substitution errors in articulation are less influenced by visual cues than those children who can produce consonants correctly during verbal communication (Desjardins, Rogers & Werker, 1997), when compared to non-substituting children, substitute children with articulatory errors were poorer at speech reading, and had a lower degree of visual influence in auditory visual speech perception. Because the integration of auditory and visual stimuli depends on the linguistic and cross linguistic experience (Sekiyama and Burnham, 2004), it is clear that articulatory experience affects speech reading and the degree of visual influence in auditory-visual speech perception.

Need and justification of the study:

1. Age-related differences in auditory-visual speech perception have been found in a number of studies, which indicating that adults are more prone to visual speech influence than children (Massaro, Thompson, Barron, & Laren, 1986; Mcgurk & Macdonald, 1976).

Though most of the studies done in this area have compared results obtained by children till the age of 11 years and adults, there is a discrepancy with respect to the age by which the complete influence of visual cues occur. One of the study has shown benefit of visual speech by 8 years (Sekiyama & Burnham, 2004), whereas according to Iyar (2005) found to be till age of 15 years, which also depends on the gender.

Hence, studying audiovisual integration across age group in a continuation will help in better understanding of developmental changes for audiovisual integration process in brain over period of time in normal hearing participants. This can also useful to evaluate and rehabilitate different pathological conditions.

2. Further, the integration of auditory and visual stimuli depends on the linguistic and cross linguistic experience (Sekiyama and Burnham, 2004). The studies evaluating the audio-visual integration have mainly investigated Western and Japanese population. The phonological system of the Indian language and the articulatory experience of Indian population integration process is very different western language. The conclusions drawn from the evaluating the audio-visual integration in the western languages these studies may not hold good to the Indian population, as the articulation experience is different for Indian population.

In the Indian context, there is a research study done to perceptually evaluate the auditory integration by Hindi speaking normals. This was done by Mishra (1995) on10 Hindi speaking normal individuals in the age range of 18 to 25years. They used stop consonants in CVC context.

The results revealed that 92% of responses obtained for the auditory mode, the remaining 8% having the effect of the visual mode on the auditory. 73% responses were obtained for the visual mode, the remaining 27% revealed the effect of the auditory mode on the visual mode. Thus indicating that the auditory and the visual mode influence each other. In the auditory visual mode, results revealed that when the stimulus was in the back of the oral cavity the subject were likely to perceive the in between stimulus neither auditory nor visual stimulus. Reflects that cross model interaction i.e. audiovisual integration takes place.

There are few reports available in the participants with hearing impairment and cochlear implant. Khan, Salina, Rajashekhar & Dhamani (2008) studied the normal hearing and cochlear implant children to see impact of temporal envelop and fine structure on audiovisual integration. In addition audiovisual integration is also studied by Ahmed, Chavan, Ameena deema (2010) in dyslexic, poor academic performers and normal reading children.

Apart from these, there are no studies to our knowledge which have evaluated the developmental changes of audiovisual integration.

3. With respect to the stimuli for evaluating the audiovisual integration, most of the time, stop consonants were used and when the bilabial was paired with velar consonants, McGurk effect was found to be more intense (Mc.Gurk & Macdonald, 1976, Dupont, Aubin, & Menard, 2005; Hockley & Polka, 1994; Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Grant and Seitz 1998, Sekiyama and Burnham, 2004), which indicates that place of articulation place a role in the integration process.

The voicing feature which could also be perceived through visual modality via observation of lip movement, little number of studies were found in this aspect

The study done by Schwartz, Berthommier, & Savariaux, (2004), compared identification score for voicing in audio only and audiovisual condition and the improvement in perception when visual cue was added with audition. These was reasoned to the gain of temporal cue through the visual modality regarding the lip gesture onset.

As we are interested to see if the audio and video stimulus is not congruent, then how in incongruent condition the voicing is perceived.

Aim of the study:

Hence, the aim of the present study was to assess the developmental changes of audio visual integration.

Objective of the study:

To explore the effect of age on audio visual integration across six age groups (from 6 years to 20 years), as given in the table1, using following conditions.

a) Audio only

b) Visual only

c) Audiovisual congruent

d) Audiovisual incongruent including

I. Audiovisual integration (Mc.Gurk effect)

II. Consonant Voicing effect

\

Table1.1 *Groups and age range in each*.

Serial no.	Group	Age range (in years)
1	I	6 – 7.11
2	II	8 – 9.11
3	III	10 – 11.11
4	IV	12 – 13.11
5	V	14 – 15.11
6	VI	16 – 20

Chapter 2

Review of literature

Earlier, it was thought that the auditory signal was the main channel for speech perception (Desjardins & Werker, 2004). Research done with hearing individuals as well as hearing impaired suggested that the visual information provided by the speaker's mouth facilitated modified auditory speech perception (Desjardins & Werker, 2004; Lachs, Pisoni & Iler Kirk, 2001).

The integration of information from different sensory modalities is clearly beneficial. Multimodal events are detected more accurately and faster than unimodal events (Calvert, 2001; Frens, Vanopstal, & Vanderwilligen, 1995). Human speech is a prime example of this. For example, for individuals with impaired hearing; lip-reading can supplement the auditory signal and enhance its intelligibility (Rosenblum & Saldana, 1996). Visual speech cues are also used by individuals with normal hearing in a noisy environment (MacLeod & Summerfield, 1987) or in recovering a difficult message (Reisberg, McLean, & Goldfield, 1987). McGurk and MacDonald (1976) found that speech perception is strongly influenced by watching the speaker's mouth movements and listening to the auditory signal (Desjardins & Werker, 2004), even in no degrade conditions.

The Mcgurk effect is a perceptual phenomenon where the perceived phoneme is affected by the simultaneous observation of lip movement. This probably reflects the underlying audiovisual integration process, which occurs during speech perception. Grant and Seitz (1998) studied auditory and visual cue integration in 41 listeners with hearing loss and categorized participant responses along a continuum. In their view, optimal cue integration, or equal impact of auditory and visual inputs, occurs when discrepant auditory and visual inputs (e.g., /bi/ & /gi/) result in a clearly fused response such as /di/. By their definition, if a response is not fused, optimal integration does not occur; specifically, one modality is dominant over the other.

There are two typical responses in the Mcgurk effect. The AV combination of 'bilabial' (involving both lips) sounds and 'palatal' (the body of the tongue rose against the hard palate) mouth movement typically results in a 'fusion response', in which a new phoneme different from the originals is perceived. For instance, when an auditory /ba/ is dubbed to the motion picture of mouth pronouncing /ga/, the majority of subjects perceive the phoneme /da/ in adults (Mcgurk & Macdonald 1976). On the other hand, when an auditory /ga/ is paired with the visual articulation of /ba/ syllable, subjects basically perceive either of the phonemes /ba/ or /ga/, and sometimes report the perception of the phoneme /bga/ or /gba/ combination response (Macdonald & Mcgurk 1978).This phenomenon had been widely used to assess audiovisual integration process.

2.1 **Anatomical correlates of audiovisual integration**

Audiovisual integration in speech processing is closely related to the cortical processing of language. While learning about the cortical processing of language, there are several centers in the cortex. We should expect that many of these areas would also be active in audiovisual speech processing. The centers of the brain that are involved in language processing are: inferior frontal lobe, inferior partial lobe and the temporal lobe grossly. Within these lobes of course they have very specific locations. These locations are the inferior frontal lobe (including Broca's area on the inferior frontal gyrus and the precentral gyrus), the inferior parietal lobe (including the supramarginal gyrus and the angular gyrus), and the temporal lobe (including Heschl's gyrus, the anterior and posterior superior temporal planes, the superior temporal gyrus, the middle temporal gyrus, the superior temporal sulcus, and the Insula).

Miller and D' Esposito (2005) concluded that the middle superior temporal sulcus is where audiovisual inputs are first combined. This pathway then progresses along the superior temporal sulcus and the superior temporal gyrus. If the visual input to the superior temporal sulcus matches or supports the auditory input, the superior temporal sulcus provides feedback to the auditory cortex, which strengthens the auditory signal. If the audiovisual stimuli are asynchronous, the superior temporal sulcus recruits the intraparietal sulcus to perform temporal transformations in order to achieve a match between the stimuli. In addition, Broca's area is activated if it is necessary to parse the observed speech into intelligible parts.

The superior temporal sulcus as the major orchestrator and earliest location of audiovisual integration according to Miller and D' Esposito (2005). However, this has recently been disputed, notably by a study that measured even trelated potentials (ERPs) using EEG to resolve the temporal dynamics of audiovisual processing (Bernstein et al 2007). As fMRI does not have the temporal resolution necessary to determine the fine time course of audiovisual processing, EEG was used to test the temporal aspects of the audiovisual processing model.

The major findings of the EEG study were that audiovisual integration occurred first in the dorsolateral prefrontal and inferior frontal cortex, with major activations occurring slightly later in the supramarginal and angular gyri of the inferior parietal lobe and in the intraparietal sulcus. The EEG study showed that the superior temporal sulcus was involved and was activated during the processing of audiovisual information, but it was not the first area to be activated and it was not the primary orchestrator of audiovisual integration.

Indeed, the major areas activated by the perception of audiovisual speech include the superior temporal sulcus, the inferior frontal gyrus (Broca's area), the insula, Heschl's gyrus, and the supramarginal and angular gyri, all areas that are involved in language processing (Miller and Desposito 2005, Bernstein et al 2007, Campbell et al 2001, Calvert et al 1997). This raises the question that actually the audiovisual integration is just a structural representation or the functional integrity between different areas are also important?

There is evidence that crossmodal processing in the cerebral cortex may underlie a phenomenon referred to as "multisensory-integration" (Calvert, 2001), none of the known studies have explicitly talked about it. However, recently, some studies on multisensory integration have focused on the underlying mechanisms. For example, in behavioral studies the main finding has been that reaction times to congruent audio–visual stimuli are typically shorter than to their unimodal counterparts (Frens et al., 1995; Miller, 1982). Incongruent stimuli have the opposite effect, slowing response times (Lewkowicz, 1996; Stein, Meredith, Huneycutt, & McDade, 1989) and producing perceptual anomalies (McGurk & MacDonald, 1976). Some neuro-imaging (Calvert, Hansen, Iversen, & Brammer, 2001; Lewis, Beauchamp, & DeYoe, 2000; Macaluso, Frith, & Driver, 2000) and electro-magnetic (Giard & Peronnet, 1999; Krause, Möttönen, Jensen, Lampinen, & Sams, 2001;Sams & Imada, 1997; Möttönen, Krause, Tiipana, & Sams, 2002) studies also have shed some light on the neural process underlying audio–visual integration.

Most research in humans only demonstrates the existence of the phenomenon rather than reveals the physiological process underlying it (O'Hare, 1991). Finglecurts and Finglecurts (2003), in their study observed the existence of widespread networks of active functional interactions between various cortical brains sites involved in audiovisual speech information integration was investigated. Cortical functional interactions during audiovisual speech integration (Finglecurts et.al, 2003) when the mcgurk subject were compared with non Mcgurk subject the majority of connections typical for the ‘‘McGurk subjects’’ were absent in the ‘‘non-McGurk

Although audio–visual speech integration, is well-established experimentally (Massaro, 1987; Massaro & Cohen, 1996; Rosenblum, Yakel, & Green, 2000), the brain (neural) process that sub serve it remain to be assessed (Giard & Peronnet, 1999).

In animal studies, detailed observation of the behavior of multisensory neurons at the single neuron level has been resulted in four integration rules (Stein & Meredith, 1993). The central rule is that of temporal coincidence, according to which the greatest interaction effects are obtained if inputs are temporarily synchronized (Stein, Meredith, & Wallace, 1994). This rule could also be applied to the large-scale cortical level of the human brain. It is possible to plausibly argue that the cross modal binding in the human brain may be achieved by the synchronized processing of sensory inputs between the unimodal cortical areas (Phillips & Singer, 1997; Salinas & Sejnowsli, 2001), rather than in so-called convergence regions of the cortex.

Indeed, extensive analysis of the lesion studies has found that none of the structures known to receive converging input from more than one sensory system has been shown to be specifically crucial for both the development and display of crossmodal performance (Ettlinger & Wilson, 1990; see also Murray, Malkova, & Goulet, 1998). Instead, synchrony generated intrinsically by functional interactions between distant cortical areas might be the mechanism underlying multisensory integration. (Engel et al., 2001; Ettlinger & Wilson, 1990) e.g. Audiovisual integration.

These indicate that not only the structural, but functional connection within the brain is also important in helping for audiovisual integration process and it's lying at the level of neurons which is a unit cell of brain structure.

2.2 Factors affecting audiovisual integration:

Previous studies of audiovisual speech perception suggest that there are different factors affecting the integration process. Out of them one of the factors is characteristics of the auditory and visual signals, particularly whether they are compromised in some way. The other factors are auditory and visual characteristics of the individual talker, linguistic experience, noise and intensity level, gender, and attention.

2.2.1Characteristics of the auditory and visual signals:

The robustness of the auditory speech signal has been evidenced in a number of studies by measuring speech perception when the speech signal has been degraded in some form. A study by Remez et al. (1981) showed that the reduction of the speech waveform to three sine waves representing the first, second, and third formants in the original signal yields signals with sufficient information to support speech perception for both sentences and isolated syllables. These results show that there exists a degree of redundancy in the auditory signal, and that extensive removal of information from the auditory signal can still result in good intelligibility.

Researchers have extended the examination of the effects of degrading auditory signals to include effects of adding visual stimuli. Summerfield (1987) hypothesized that the addition of visual cues to the auditory signal would result in improved speech perception. He suggested that visual cues could aid in speech perception through the visual stimulus being redundant with auditory stimulus, emphasizing aspects of the Signal and serving as a reinforcer. The second function that visual cues may serve is to complement the auditory cues, filling in information where the auditory signal may be lacking. The third role that visual cues may have is to generate temporal coincidence between the auditory and visual signals to highlight the most significant characteristics of the speech signal.

Grant and Braida (1991) continued to examine the effects of adding visual stimuli to a degraded auditory signal by determining the degree by which perception is improved. The study noted that adding visual cues to auditory speech in noise can improve the signal-to-noise ratio by up to 5 - 18 dB (Sumby and Pollack, 1954) and with each decibel increase resulting in an intelligibility improvement of 5-10 percent (Grant and Braida (1991).

Also a study conducted by Munhall et al. (2004) investigated the nature of the visual image processing during audiovisual speech perception by manipulating the spatial resolution of facial images for a speech-in-noise task. They found that speech signals were most commonly identified with the presentation of the unfiltered visual stimuli while least commonly identified with the presentation of the auditory stimulus only.

In summary the intrinsic characteristics of the sound help them to perceive even if whole auditory information is not available or got distorted. This can also be achieved through the addition of visual information along with the auditory signal as reported by different researchers.

2.2.2. Intensity and noise level:

studies on the effect of sound intensity and noise level on the integration of audiovisual speech had shown that since both sound intensity and noise level are manipulations of the acoustic signal-to-noise ratio (SNR), the visual influence were increased with decreasing SNR, i.e. With decreasing sound intensity and increasing noise level (Tiippana, 2005). Whereas there is contradicting study which reflects that the listening in noise conditions, the presence of visual cues was neither beneficial nor detrimental to the speech perceptual abilities of the cochlear implant and normal hearing participants (Iyar ,2005).

Several possible reasons were attributed for the participants’ poor performance in noise: (Lachs, et al., 2001; Bergeson et al.,2003).

1) Participants were less able to extract the desired target from the background noise due to a low signal to noise ratio,

2) Participants were less able to make use of the semantic information to assist with audio and audiovisual perception and

3) Listener fatigue (although order of test condition was randomized).

The preceding reasons favor that there is increase in influence of visual modality on speech perception in presence of noise due to lack of excssecibility of auditory information.

The addition of visual cues to an unintelligible auditory signal can enhance overall speech perception significantly. Sumby and Pollack (1954) demonstrated that the addition of visual cues could improve the perception of speech by the equivalent of a 5 to 18 dB increase in the signal-to-noise (S/N) ratio, that is, an effective change of up to 60% in word recognition was observed depending on the materials used.

Subsequently, Erber (1969) found that word recognition scores for young adults improved from 20% in the auditory-only condition to 80% in the auditory-plus-visual condition at a 10 dB S/N ratio. Middelweerd and Plomp (1984) and Summerfield (1979) showed similar results for sentence materials. Seeing the face of a speaker can significantly embellish a degraded or noisy auditory speech signal so that it functionally raises the signal-to-noise ratio by as much as 22 dB (Sumby & Pollack, 1954; Rosenblum, Johnson, & Saldana, year).

Also, the rehabilitation strategies that help individuals compensate for an impoverished auditory signal often encourage speech reading, i.e. the use of visual cues to supplement the auditory component (Alpiner & McCarthy, 1987).

2.2.3 Auditory and visual characteristics of the individual talker

Individuals also vary in their auditory articulation, with some talkers' speech being more intelligible than others. It is known that some talkers have better speech intelligibility than others and that there is a large variance in the articulation used by talkers.

When in an environment that makes communication difficult, many talkers engage in "clear speech" to improve intelligibility (Chen, 1980; Picheny et al., 1985; Uchanski et al., 1992; Peyton et al., 1994). When compared to conversational speech, a number of researchers have found clear speech to be marked by a slower speaking rate, increased temporal modulation, greater range of voice fundamental frequency, expanded vowel space, and more stimulus energy in high frequencies.

Gagne et. al. (2002) further investigated the use of clear speech by focusing on its benefits in auditory, visual, and audiovisual presentations. They suggested that articulation in clear speech production aids in perception of auditory, visual, and audiovisual conditions. Though results supported their theory, not all talkers produced clear speech benefits for all conditions. Additionally, the amount of benefit varied across talkers and across iterations within individual talkers. The question is whether a "good" talker is one that provides more benefit auditorily and/or visually or if a good talker is one that provides more ambiguity to allow for greater integration.

Jackson's year study focused on talker characteristics, highlighting the negative effects of visual features, such as facial hair, thin or thick lips etc. on their intelligibility.

In summary, large variance in articulation was found to present among individuals in addition the facial appearance of articulators like thickness of lips, facial hair also contribute for the perception of speech. Along with this the clear production of speech in adverse environmental condition helps in better speech perception.

2.2.4 Individual listener characteristics

Individual characteristics such as age gender and hearing sensitivity has been found to affect the integration ,the details about the effect of age is discussied I details in later in this chapter.

In one the study Grant and Seitz (1998) examined whether individual listeners integrate auditory and visual cues with varying degrees of efficiency and found that there are significant levels of difference between listeners. One difference was that older subjects tended to be less efficient at integration than younger subjects. In a study done by Clark (2005), there was variability in the degree to which subjects exhibited the McGurk effect. When subjects were viewing themselves as talkers, half of the subjects showed a reduced McGurk effect, though none of the subjects exhibited particularly strong McGurk effects to these talkers.

2.2.5 Gender effect:

Gender is also found to be a factor, which affects the audiovisual integration process (Johnson, Hicks, Goldberg & Myslobodsky, 1988; Watson, Qui, Chamberlain & Li, 1996; cited in Desjardins & Werker, 2004; Irwin et al 2006).). Females are found to show a stronger McGurk effect than males (referense). Women show significantly greater visual influence on auditory speech than men did for brief visual stimuli, but no difference is apparent for full stimuli (Irwin et al 2006). Female listeners have been found to have superior lip reading abilities and were better at integrating audiovisual speech information than the male listeners (Johnson et. al. 1988; Watson et. al 1996; Desjardins & Werker, 2004).

Cameron (2004) investigated the audiovisual integration of young children (aged 6 - 9 years) and found that the presence of visual cues did not enhance speech perception scores for male participants. Young males (aged 6 - 9 years) are more likely to attend to auditory cues, where as young females are more likely to benefit from the addition of visual cues when performing a speech perception task (Cameron, 2004). The visual cues seemed to distract the male participants instead of assisting them.

Interestingly, in the study done by Iyar (2005) where across different age group the gender effect was contradicted. There were differences found between male and females studied shows the audiovisual integration effect was only seen in young males (aged 7 and 11) and young females (aged 7) in poor listening conditions, whilst females (aged 11 and 15 year old males) benefited more from the presence of visual cues (Cameroon, 2004; Sloutsky & Napolitano, 2003)

The preference for the auditory modality was investigated in young children Sloutsky and Napolitano (2004). Their findings indicate that when auditory and visual stimuli are presented separately four year olds are likely to process both stimuli. However, when both stimuli are presented simultaneously four year olds are more likely to process auditory stimuli than visual. When the scores of the young children were compared to that of adults the researchers found that the adults were more likely to respond to visual stimuli when auditory and visual information was presented simultaneously (Sloutsky & Napolitano, 2004). This is consistent with; McGurk and MacDonald's (1976) findings.

2.2.6 *Linguistic experience:*

Sekiyama & Tohkura (1993) showed that, as a whole, the Mcgurk effect was weaker in Japanese subjects than in Americans. Studies have also shown that Japanese listeners do not Japanese listeners are more able. This result could be in relation to the fact that in Japanese, consonant clusters do not exist. In noisy environments where speech is unintelligible, however, people of all languages resort to using visual stimuli and are then equally subject to the McGurk effect. The McGurk effect works with speech perceivers of every language for which it has been tested. Studies done in Indian language i. e. Hindi, in adults age ranging from 18-25 years found that the alveolar sound were better identified followed by bilabial and velar stop sound.

Sekiyama and Burnham (2004) tested Japanese- and English-speaking 6-, 8-, 11-yearolds and adults using the McGurk paradigm in three experimental conditions (auditory-only [AO], auditory-visual [AV], and visual-only [VO]). They found that for English speakers there was minimal use of visual speech information at 6 years, but a significant increase from age 6 to 8, which remained stable at 11 years through to adulthood. The use of visual speech information by the Japanese-speaking 6-yearolds was equivalent to that for the English-speaking 6-yearolds; however, the influence of visual speech remained at this level across all Japanese age groups. These results pose the question of what causes the increase in visual speech influence in English-speaking children aged between 6 and 8 years. As regards the differences across languages, several studies (Sekiyama & Tohkura 1991, 1993; Sekiyama 1997) suggested differential occurrences of the Mcgurk effect across various languages (e.g. Japanese, American and Chinese).

In summary across language there is differences observed which is influencing the audiovisual integration process.

2.2.7 Effect of attention

The investigate were done to see whether integration of audiovisual speech occurs automatically so that information from heard speech and seen articulatory movements of the talking face are combined without any voluntary effort (Tiippana et al. 2001) and the results revealed that integration of audiovisual speech is not entirely automatic and as per the fuzzy logical model of perception (FLMP) the good fit of the model implies that attention influences unimodal information processing before integration across modalities takes place. According to this integration process is quite automatic that occurs as soon as a talking face is seen together with speech sounds. But when the attention was directed towards lip movement during conversation i.e. merely fixating lip movements that are incongruent with the auditory signal can deteriorate speech recognition performance which reflects that the directed attention can affect the speech perception which is not natural (Driver, 1996).

Whereas the above finding was argued against by Massaro (1998) finding that by instructing subjects to respond according to either auditory or visual information only, the responses to audiovisual speech stimuli were biased towards the respective modalities, particularly when the other modality gave ambiguous information. Although both behavioral and neurophysiologic studies have converged on an appropriate pre-attentive conceptualization of audiovisual integration (Mcgurk et al, 1976, Bernstein et al, 2004).

So, one way to examine feature integration, that is, integral processing of information, is to use a selective attention task (Gamer, 1974: Lockhead & Pomerantz, 1991).

In summary there are different factors intrinsic to the individual as well as extrinsic factors which influence the audiovisual integration process. Out of these the age and linguistic effect was found to be more effective in influencing integration effect on audiovisual information. Others like attention, intensity and noise level, individual characteristics of talker, stimulus used etc., have impact on integration process but found to be less effective compare to former ones.

2.3 Audiovisual integration across age:

2.3.1 Infants

Although there have been no direct tests of the McGurk effect in prelinguistic infants, there is a good agreement of evidence that infants are sensitive to audiovisual correspondences in speech (Aronson & Rosenbloom, 1971; Dodd, 1979; Kuhl & Meltzoff, 1982, 1984; Kuhl, Williams, & Meltzoff, 1991; Spelke & Owsley, 1979).

Early research showed that infants are sensitive to spatial correspondences in audio and visual speech (e.g., Aronson & Rosenbloom, 1971; Spelke & Owsley, 1979).

Aronson and Rosenbloom (1971) observed that 1- to 2-month-olds became visibly distressed when a mother's voice was displaced away from her face. Research has also shown that infants are sensitive to temporal synchronyin audiovisual speech (Dodd, 1979; Pickens et al., 1994; Spelke & Cortelyou, 1980; Walker, 1982).

Using a gaze preference procedure, Dodd (1979) demonstrated that 3- to 4-month-olds attended longer to audiovisual speech that was in synchrony than to speech presented out of synchrony by 400 msec. She interpreted this as evidence that infants are aware of some congruence between lip movements and speech sounds. Legerstee (1990) has found that 3- to 4-month-old infants will only imitate audio visually compatible vowels (/a/-/u/) and not stimuli that are dubbed to be incompatible. She interprets this finding as evidence that multimodal information is useful for speech acquisition.

More recent research suggests that infants are sensitive to phonetic correspondences in audiovisual speech. Kuhl and Meltzoff (1982, 1984) used a preferential gaze procedure to test whether 4-month-old infants were sensitive to audiovisual correspondences for the vowels /i/ and /a/. They found that, for both vowels, infants looked longer at the face that matched the vowel presented auditorily. Additional research has replicated these findings with /i/ and /u/ vowels (Kuhl & Meltzoff, 1988) and with disyllables such as /mama/ and /lulu/ (MacKain, Studdert found with the use of an operant choice sucking procedure (Walton & Bower, 1993). This study showed that infants ranging in age from 1 to 14 months perform more sucks to audiovisual compatible than to audiovisual incompatible vowels.

McGurk effect would be evident in pre linguistic infants. A study done by Rosenblum et.al (1997) For which, 5-month-old infants were tested under an infant-control habituation of looking time procedure (see, e.g., Best, McRoberts, & Sithole, 1988; Horowitz, 1975; Horowitz, Paden, Bhana, & Self, 1972). This procedure tests the degree to which infants generalize to various test stimuli after habituation to an initial stimulus. For the first experiment, this procedure was used to test infant discrimination of audio /va/-visual */va/* from audio /ba/-visual /va/ and from audio /da/-visual */va/*. Previous research had shown that an audio /ba/-visual */va/* is "heard" as */va/* up to 98% of the time with adult observers (Rosenblum & Saldana, 1992, 1996; Saldana & Rosenblum, 1993, 1994). In contrast, there is evidence that an audio /da/-visual /va/ rarely displays a visual influence and is heard by adults as /da/ over 88% of the time (Repp, Manuel, Liberman, & Studdert Kennedy, 1983).

2.3.2 Children:

In contrast to the infant and adult literatures, the child literature emphasizes that visual speech has less influence on speech perception by children. In their initial research, Mcgurk and MacDonald (1976) noted that significantly fewer children than adults show an influence of visual speech on perception. In response to one type of Mcgurk stimulus (auditory /ba/ - visual /ga/), the percentage of individuals who reported hearing /ba/ (auditory capture) was 40-60% of children, but only 10% of adults. The same results were found in others studty also (Desjardins, Rogers, & Werker, 1997; Dupont, Aubin, & Menard, 2005; Hockley & Polka, 1994; Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Sekiyama & Burnham, 2004; Wrightman, Kistler, & Brungart, 2006).

These findings in children found to be agreeing with the general observation of a bias toward the auditory modality in young children (Sloutsky & Napolitano, 2003).

Findings that children show fewer illusory responses than adults (from 10% to approximately 60% of adult levels, depending on the study) suggest that exposure and experience play a role in the development of bimodal speech perception (Massaro, 1998). Although experience appears to increase the strength of the McGurk effect, it is also possible that the less robust effect in children partially reflects other constraints that come with child participants. As in Desjardins et al. (1997) study he reported that 10 of 26 normally developing children (M = 4.3 [years;months]) initially enrolled in their study were unable or unwilling to complete their experimental task despite an attempt to tailor the methods to the children's abilities. Other studies have simply lacked child-friendly procedures. Young children can find it difficult to work with syllables (e.g., /ba/, /da/) that are devoid of meaning. Tasks that require identification of minimal pair CV or CVCV syllables can require metalinguistic capability, a skill that also emerges with age.

Finally, obtaining responses to enough items to provide a reliable estimate of the effect can require sustained attention. If the child's attention to each item is not monitored carefully, then poor attention to the visual information would mimic a reduced McGurk effect or auditory capture types of responses. Any of these factors could affect the measured strength of the visual impact on auditory processing in children.

The age and gender effect was studied by Iyar (2005). The study aimed to explore the audiovisual speech perception abilities in children aged 7, 11 and 15 years and was hypothesized that there would be gender differences in audiovisual speech perception abilities that would differ across age groups results for the children with normal hearing for the three age groups, separately for males and females. The results suggested that normal hearing males and females may use different types of sensory information when perceiving speech. Younger males aged 7 and 11 appeared to prefer auditory cues as they did not benefit from the addition of visual cues, whilst females aged 11 and older males aged 15 did gain slightly from the addition of visual cues in poor listening conditions. Younger females aged 7, like their male counterparts, did not benefit from visual cues in poor listening conditions. It is difficult to see whether visual cues enhanced speech perception in quiet conditions since the scores for these conditions are generally high and there may be a ceiling effect. Critical age of 5- 15 years was found in audiovisual integration process in children by Schorr et. al. (2005).

2.3.3 Adults:

There are several research studies Many studies to see the amount of audiovisual integration taking place in adults. In most of the study's results are consistent with the idea that performance is dominated by visual input in adults than children (Mcgurk & Macdonald 1976; Rosenblum, Schmuckler, & Johnson 1997; Wightman, Kistler & Brungart ,2006; Desjardins, Rogers, & Werker, 1997; Dupont, Aubin, & Menard, 2005; Hockley & Polka, 1994; Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Sekiyama & Burnham, 2004; Wrightman, Kistler, & Brungart, 2006).

Numerous investigations have found a decrease in auditory only speech recognition abilities with increasing age (e.g. Humes & Roberts, 1990; Jerger, Jerger, Oliver, & Pirozzolo, 1989). Correct identification of speech by vision alone (speechreading) also diminishes with age.

Middelweerd and Plomp (1984) report the average benefit of speechreading as expressed in speech in noise ratio for the 50% correct identification of sentence materials was 4.0 dB for older adults in comparison to 4.6 dB for young adults. Other investigators have noted that percent correct identification of key words in sentences in the visual only condition decreases with age (Shoop & Binnie, 1979; Walden, Busaco, & Montgomery, 1993).

As increase in age will lead to more chances of hearing impairment due to presbycusis, so the investigation of auditory plus-visual speech perception in older adults may provide further support for peripheral or central explanations of changes in speech perception that occur with aging. The use of conflicting auditory and visual speech stimuli might provide a test of integration performance for older adults.(reference)

If older adults were selected with normal or near-normal visual and auditory sensory systems, it is likely that their information extraction capabilities would be similar to those of younger adults. Consequently, if the integration performance of the older adults were poorer than that of young adults, changes in central processing mechanisms may be suspect. Conversely, if their integration performance were equal to that of young adults, then the contribution of the peripheral systems is likely to be the most important factor underlying integration.(reference)

Grant and Seitz (1998) studied auditory and visual cue integration in 41 listeners with hearing losses and categorized participant responses along a continuum. In their view, optimal cue integration, or equal impact of auditory and visual inputs, occurs when discrepant auditory and visual inputs (e.g., /bi/ /gi/) result in a clearly fused response such as /di/. By their definition, if a response is not fused, optimal integration does not occur; specifically, one modality is dominant over the other. Thus, even among young adults, optimal cue integration (grant & seitz, 1998), as evidenced by complete fusion of information across discrepant auditory and visual stimuli, is not consistent.

However the participants studied by grant and seitz (1998) ranged in age from 41 to 76 yr, with a mean age of 66 yr, and had mild-to-severe sensori neural losses with sloping configurations. Older adults may be at a disadvantage for processing visual, as well as auditory, speech information.

Degree of auditory-visual fusion has not been examined in older adults with normal or near-normal hearing. Shoop and Binnie (1979) compared the visual recognition abilities of middle-aged and elderly adults with normal hearing using the CID everyday sentences. They found that as age increased, percent correct identification of key words in the visual only (VO) condition decreased.

Evaluation of audio visual integration in terms of response time it was found that there were significant differences between the young adults and older adults and throughout the literature on aging, older adults show a decline in the speed of response (Bashore, Ridderinkhof, & Vander Molen, 1998). When comparing young and older adults on RT tasks, older adults will almost always display slower results (Bashore et al., 1998). Researchers have found consistent differences among the performance of young and older adults on simple, choice, and complex RT tasks (Cerrela, Poon, & Williams, 1989). Older adults show increases in both the amount of time to make a decision and the amount of time to execute a motor response (Salthouse, 1991)

So when the young and older adults are compared for audiovisual integration, there can be different stages might taking place among them for it. Cavanaugh (1997) described a two-stage process of perceptual integration in young and older adults: the extraction or the encoding of important stimulus characteristics and the integration or combination of those characteristics in a meaningful way. Aging might affect these two stages differentially.

Further, the limited access to auditory information occurs because changes in the peripheral and/or central auditory system contribute to age related hearing loss (CHABA, 1988). Older adults may be at a disadvantage for processing auditory and visual information for speech recognition in the unimodal and bimodal conditions. Deficits in the peripheral mechanism may limit the amount of useful information that can be extracted from the input stimulus whereas changes in the central system may affect areas including information processing, memory, and retrieval (Humes & Christopherson, 1991). Whereas Plude and Hoyer (1985) suggested that the ability to extract information remains unaffected by age but that older adults are less successful at integrating information than younger adults.

Furthermore, the literature suggests that older adults may be less successful at perceptual integration, that is, combining information across two or more sensory modalities (Plude & Doussard-Roosevelt, 1989). Studies of selective attention tasks show that older adults show greater interference for non-target items than younger adults (Plude & Hoyer, 1985; Rabbitt, 1965).).

Results from investigations of auditory-visual integration of speech stimuli among older adults have been inconclusive (Cienkowski, 1999; Grant et al., 1998). Cienkowski (1999) found that older adults were not consistently successful at integrating information across sensory modalities. However, hearing loss was not well controlled for in that study and Grant et al. (1998) have shown that hearing levels may affect the integration of auditory and visual syllables in hearing-impaired adults.

In summary, as individual's age, there are changes in the auditory system that result in a decrease in speech recognition abilities (Committee on Hearing and Bioacoustics and Biomechanics, 1988). Loss of sensitivity in the peripheral auditory system, specifically poorer hearing thresholds, contributes substantially to this drop in performance (Humes & Roberts, 1990; Kalikow, Steven, & Elliot, 1977). Some investigators have suggested that changes in cognitive (Pichora-Fuller, Schneider, & Daneman, 1995) or central auditory processing abilities (Jerger, Jerger, Oliver, & Pirozzolo, 1989) play a limited role in the decreased recognition of speech as well. The combination of visual cues from lipreading with the auditory speech signal adds another sensory source to the original signal; it also requires that the listener integrate information from each source to maximize performance (Grant & Seitz, 1998). Consequently, the investigation of auditory plus-visual speech perception in older adults may provide further support for peripheral or central explanations of changes in speech perception that occur with aging.

2.4. Summary

In summary despite the fact that adults show a stronger McGurk effect than children, many children preschool-age and older do, indeed, integrate speech when presented auditorily and visually. Even infants can show evidence of a McGurk effect by 6 months of age (Burnham, 1998; Rosenblum, Schmuckler, & Johnson, 1997).

In the initial McGurk study (McGurk & MacDonald, 1976), three groups of participants were tested: adults, 7- to 8-year-olds, and 3- to 4-year-olds and adults. McGurk and MacDonald (1976) in his study indicated that the responses from the two groups of 3-4 and 7-8 years children did not differ reliably from each other. The adults (n = 54) were influenced more by the visual signal than were the older and younger children (ns = 28 and 21, respectively). And found that only 2% of the adult responses matched the auditory signal, there was a 36% auditory response rate for the older children and a 19% auditory response rate for the younger children.

The McGurk effect is robust in adults for particular combinations of auditory and visual signals as reported by many authors (Green, Kuhl, Meltzoff, & Stevens, 1991; Massaro, 1987; Rosenblum & Saldana, 1996). In contrast, visual articulations influence speech perception to a lesser degree in children as compared with adults (Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; McGurk & MacDonald, 1976).

However, AV benefit remained the same across groups. Visual cues have a broad influence on perceived auditory stimuli, specifically when these visual cues are the lip movements associated with the auditory stimulus of speech. The perception of speech is improved by watching the speaker’s lips as they speak, especially in noisy situations, and visual information can help listeners determine the location of sounds or speech (Calvert et al 1997, Macaluso et al 2004).

Sloutsky and Napolitano (2004) investigated the preference for the auditory modality in young children. Their findings indicate that when auditory and visual stimuli are presented separately four year olds are likely to process both stimuli. However, when both stimuli are presented simultaneously four year olds are more likely to process auditory stimuli than visual.

When the scores of the young children were compared to that of adults the researchers found that the adults were more likely to respond to visual stimuli when auditory and visual information was presented simultaneously (Sloutsky & Napolitano, 2004), whereby the visual stimulus dominates when conflicting auditory and visual cues are presented to adults. Sloutsky and Napolitano (2004) suggested three possible reasons to account for their findings: 1) the auditory system matures earlier than the visual system, 2) the auditory system is also functionally more important for language acquisition than the visual system and this advantage might decrease after the child has mastered language acquisition and 3) the dominance of the auditory system stems from different attentional demands for processing visual and auditory stimuli.

Typically, a sound disappears after a relatively short period of time, whereas a corresponding visual scene may be present for a much longer time, so therefore it seems more adaptive to allocate attention to sound before allocating attention to visual scene (Sloutsky & Napolitano, 2004; Sekiyama, Kanno, Muira & Sugita, 2003).These findings suggests that there is increase in audiovisual integration with age for childhood to adulthood which start declining again towards the older age range (Walden, Busacco, and Montgomery , 1993. McGurk & MacDonald, 1976).

Different reasons were associated in order to explain the age related change in audiovisual integration. Among that Developmental improvement has been attributed to experience in producing speech, changes in the emphasis and perceptual weight given to visual speech cues, and age-related advances in speech reading skills and/or linguistic skills, perhaps consequent on educational training (Desjardins et al., 1997; green, 1998; Massaro, et al., 1986; Sekiyama & Burnham, 2004) again the decrement in perception of integrated information with increase in age attributed to reasons like loss of sensitivity in the peripheral auditory system, i.e. poorer hearing thresholds (Humes & Roberts, 1990; Kalikow, Steven, & Elliot, 1977), changes in cognitive (Pichora-Fuller, Schneider, & Daneman, 1995) or central auditory processing abilities (Jerger, Jerger, Oliver, & Pirozzolo, 1989).

Children's visual speech perception improves with increasing age, but the time course of developmental change is not well understood. A few studies have observed benefit from visual speech by the pre-teen/teenage years (Conrad, 1977; Dodd, 1977, 1980; Hockley & polka, 1994), with one report citing an earlier age of 8 years (Sekiyama & Burnham, 2004). The researchers also found that young children aged 3-5 and 7-8 years were less influenced by the visual signals, particularly when the audio and visual information are mismatched (Desjardins & Werker, 2004).

Chapter 3

Methodology

The present study aimed at investigating the developmental changes of audiovisual integration from childhood to young adults. The study was carried out in two phases. In the first phase, selection of participants was done and in the second phase, the actual experiment to evaluate Audiovisual Integration was carried out.

Prior to the first phase, a written consent was taken from the caregivers or the participants after briefing about the study, its objectives, method and duration of testing.

3.1 Phase I

3.1.1 Selection of participants:

In the first phase, selection of participants was done using a checklist and various tests. They are as follows: Screening Checklist for Auditory Processing (SCAP), Tumbling-'E' chart test for checking the visual acuity, routine audiological evaluation, Speech in Noise Test (SPIN) and Gap Detection Test (GDT).

Initially, a total of 200 participants were selected only based on the age. The participants' age ranged between 6 to 20 yrs. There were 170 children and 30 adults.

Further selection was based on a series of screening and diagnostic tests. Depending on the age group the tests administered for selection varied. Out of the 200 participants selected, only 150 met the inclusion criteria which are given later in this chapter. Thus, these 150 participants were enrolled for the second phase.

First phase of the study included administration of screening checklist for CAPD and ‘Tumbling – E’ chart for screening of visual acuity. Routine audiological evaluation was done for all the subjects, which included pure tone audiometry and tympanometry.

3.1.1a screening checklist for central auditory processing (SCAP)

This check list was developed by Yathiraj and Mascarenhas (2002, 2004). It is used to screen the school going children for any symptoms of (C)APD. The check list (enclosed in appendix A) comprised of 12 questions related to the symptoms of deficits in auditory processing abilities, auditory memory and other miscellaneous symptoms (enclosed in appendix A). This checklist was administered on 170 participants from different schools. A teacher who had taught the children, at least, for one year, was asked fill the checklist for each of the child.

The check list was scored on a two point rating scale. Each answer marked ‘yes’ was scored ‘1’ and each ‘no’ was scored as ‘0’. The total score ranged from 0-12. The participants with scores less than 6 are considered to have passed the checklist as scores of less than 6 indicates no processing problems. Out of 170 subjects, 140 participants scored less than 6 and were taken up for the visual acuity test in the selection procedure.

3.1.1b Vision acuity test:

The 140 children, who passed SCAP and 30 adults, were tested for vision acuity by using tumbling 'E' chart. This was created by Taylor (1976) and was later on used by different research (Karp, 1988) as standard visual acuity test in their study. For children, this test was carried out in a quiet and distraction free room in the school, and for adults it was carried out in a quiet and well illuminated audiometric room.

This chart has an English alphabet 'E' which varying in degree of size. There are a total of 11 lines. The chart was kept a distance of 6 feet. Participants were instructed to read out the alphabets starting from the top to the bottom of the chart, line by line. If the participant was able to read all the alphabets in all the lines, then, the vision was considered to be normal.

All 170 participants (140 children and 30 adults) passed the visual acuity test. Hence, all of them underwent a routine audiological evaluation.

3.1.1c Routine audiological evaluation:

Routine audiological evaluation was carried out in an acoustically treated room. Air conduction and bone conduction thresholds were established using modified Hughsen and Westlake (Carhart & Jerger, 1959) procedure. Speech audiometry was also carried out on all the participants.

A calibrated two-channel Madsen (Orbiter-922) audiometer with TDH-39 headphones was used to establish air conduction pure tone thresholds and speech audiometry. B-71 bone vibrator was used to establish bone conduction thresholds. Hearing was considered normal if puretone thresholds were within 15 dB HL (ANSI, 1989) bilaterally at all octave frequencies from 250 Hz to 8 kHz.

Tympanometry and Acoustic Reflex Thresholds were established using a calibrated Grason Stadler-Tympstar middle ear analyzer. Presence of 'A' or 'As' type of tympanogram with reflexes present in both the ears within 100 dB HL at 500 Hz, 1 kHz and 2 kHz was considered as normal.

Out of 170, 163 participants showed normal findings in routine evaluation. Seven participants were found to have abnormal type of tympanogram, 'C', 'Cs' or 'B', with reflexes absent and had poor hearing thresholds. These participants were excluded from the study.

For the participants who passed routine hearing test, Speech in noise and Gap Detection Test were administered except for 15 children. These 15 children were below 7 years of age. They could not be administered SPIN as they had difficulty understanding instructions for SPIN. Whereas, GDT has been standardized only for children above seven years of age.

Thus, for these 15 children, between 6- 7 years of age, only the results of SCAP and routine audiological evaluation were considered for inclusion. The remaining 148 (163-15) subjects were considered for SPIN and GDT testing.

3.1.1d Speech in noise test:

Speech in noise test (SPIN) was carried out for all the 148 subjects. Kannada PB word list, developed by Vandana and Yathiraj (1998) was used for this. Speech identification scores in quiet and in the presence of ipsilateral noise were found out. The level of test stimulus was 40 dB SL (ref. SRT). For testing speech perception in the presence of noise, speech noise was presented ipsilaterally at 0 dB SNR.

Scoring was done by calculating percentage correct responses. The difference between percentage correct response in quiet and that of in noise were then calculated. If speech identification scores in the presence of ipsilateral noise were poorer by 40% or more when compared to that of quite condition, it was considered as abnormal. All the 148 participants obtained normal SPIN scores.

3.1.1e Gap Detection test:

Gap detection test (GDT), developed by Shivprakash and Manjula (2003) were also administered on these 148 participants, to further confirm normal auditory processing abilities.

This test consists of 56 trials, including 8 catch trials and broad band noise as stimulus. This broad band noise contained gaps. The duration of gap varied from 1 ms to 20 ms. The signal was presented monaurally at 40 dB SL (ref. PTA). The participants were required to indicate as to which set of noise bursts in a triad contained a gap. The minimum gap that the participant could detect was considered as gap detection threshold. This was compared with the norms given.

Out of the 148 participants, 13 of them obtained abnormal gap detection threshold. Hence, the remaining 135 participants, along with the 15 children below 7 yrs of age were included in the second phase of the study.

Hence, there were a total of 150 participants for next phase. Summary of the details of the selection participants are provided in the table 3.1

Table 3.1: *No of participants passed and failed different tests in selection criterion*

Test administered	**Total no. of participants participated**	**No. of participants passed in test**	**No. of participant failed in test**
SCAP	170	140	30
Visual Acuity	140children 30adults	170	00
Hearing evaluation, tympanometry and Speech recognition threshold	170 (140 children ,30 adult)	163	7
SIPN & GDT	148	135	13

These 150 participants were then divided into 6 groups based on the age (in years), with 25 participants each details of different groups and age are given in the table 3.2

Table 3.2: *Details on different groups, the mean age and SD.*

Group number	Age range (in years)	Mean age (in years)	Standard deviation (SD)
I	6 -7.11	6.92	0.38
II	8 -9.11	8.88	0.52
II	10 -11.11	10.70	1.95
IV	12 -13.11	13.01	0.66
V	14 -15.11	14.93	0.51
VI	16 -20	18.60	1.11

3.2 Phase II: Measurement of audiovisual integration:

In the second phase, the actual experiment of measurement of audiovisual integration was carried out using the following four conditions.

1. Audio only (AO),
2. Visual only (VO),
3. Audiovisual congruent (AV+) and finally,
4. Audiovisual incongruent (AV-).

AO and VO were unimodal conditions, and audiovisual congruent and incongruent were bimodal conditions. In audiovisual congruent condition, the sounds presented through auditory and visual mode were the same, wherein, the audiovisual incongruent condition had two different syllables in auditory and visual mode. They differed either in place of articulation of voicing. Following steps were followed for carrying out the experiment.

- *Stimulus preparation*
- *Sitting arrangement*
- *Stimulus presentation*
- *Response and scoring*
- *Statistical analysis*

3.2.1 Stimulus preparation:

This stage includes the following steps:

- *Stimulus recording*
- *Editing*
- *Goodness to fit test*

3.2.1a Stimulus recording

The stimuli used were six monosyllables /pa/, /ba/, /ta/, /da/, /ka/ and /ga/. All the sounds were stop consonants with the vowel /a/. These CV monosyllables were uttered by a 26 years old female native Kannada speaker. The recording was done in a sound proof audiometric room.

Audio and video recording of these syllables were done. The speaker was seated at a distance of six feet from the camera with the head and neck held erect. The auditory and visual stimulus was recorded simultaneously using a National m-7 movie camera with an inbuilt microphone. A TV zoom lens with the power of 1:12 was used, and 1 KW halogen light was used to illuminate the room.

The subject was asked to practice uttering the syllables before recording. During the recording, she was instructed to utter the monosyllables, four times each with a pause in between each of them. These were video-recorded on the video track of the cassette national VHS (spe-180). The cassette version of the video was then converted in to avi movie file, using the software '**Any video converter'** for easy editing of the stimulus.

3.2.1b Editing:

The avi files made in previous section were edited further using **Virtual dub (version 1.0)**. Based on the clarity of the video and the naturalness of the articulatory movements of the speakers, three best recordings out of the four recordings were selected for further editing.

These stimuli were separated in to audio and video files and were digitized using **Adobe audition (Version 3)**. Video digitizing was done at 29.97 frames/s in 640 × 480 pixels, and audio digitizing was done at 44 kHz in 32 bits. All syllables were normalized in order to avoid the effect of intensity difference between different syllables.

Using these edited audio and video files, stimulus for AO, VO, and AV (+/-) conditions were generated. In AV condition, the audio and video stimuli were synchronized. As mentioned earlier for AV+ condition, the monosyllables in audio as well as video were similar, where as in the AV- condition, monosyllables in the two modalities (audio and video) were different. For AV- , /pa/-/ka/ pair and/ba/-/ga/ pair were made.

3.2.1c Goodness to fit test for quality check.

The prepared stimuli were presented to 10 adults of age range 16 to 30 years. The subjects were asked to rate in a 3 point rating scale which included

1. The quality of it as either clear,

2. Distorted but can be identified

3. Distorted and cannot be identified.

Out of three utterances, the utterance which was rated as clear was taken up for making final stimulus. Further, the stimulus prepared for assessing McGurk effect, that is, Aud /pa – Vis /ka/ and Aud /ba/ – Vis /ga/ were also presented to those subjects. This **Pilot study** revealed that the Mc Gurk effect was weaker for the stimulus pair Aud /ba/ – Vis /ga/ than the other pair. Hence, only Aud /pa – Vis /ka/ was included in the study. The details are given in table 3.3

Table 3.3. *Details of the stimulus prepared for different conditions*.

Serial number	Audio only (AO)	Visual only (VO)	Audiovisual (congruent) AV +	Audiovisual (incongruent) AV- For Mcgurk effect	Audiovisual (incongruent) AV- For voicing
Modality	Auditory	Visual	Auditory- Visual	Auditory - Visual	Auditory Visual
1.	Aud/ba/	Vis/ba/	Aud/ba/ + vis/ba/	Aud/pa/ + vis/ka/	Aud/pa/+vis/ba/
2.	Aud/da/	Vis/da/	Aud/da/ + vis/da/		Aud/ta/+vis/da/
3.	Aud /ga/	Vis/ga/	Aud/ga/ + vis/ga/		Aud/ka/+vis/ga/
4.	Aud/pa/	Vis/pa/	Aud/pa/ + vis/pa/		Aud/ba/+vis/pa/
5.	Aud/ta/	Vis/ta/	Aud/ta/ + vis/ta/		Aud/da/+vis/ta/
6.	Aud/ka/	Vis/ka/	Aud/ka/ + vis/ka/		Aud/ga/+vis/ka/

Note: Aud : auditory signal, Vis: visual signal

This stimulus was copied to a compact disk. The inter stimulus interval was two seconds in order to avoid the higher continuous attention requirement for the task and to make them feel comfortable

3.2.2 Sitting arrangement:

Each participant was tested individually in a double room situation. Participants were seated in a sound treated room comfortably in a chair. A laptop was placed 1 meter away from the participant on the table, in an appropriate height, so as to facilitate easy and normal access of the laptop screen (video) for the participants. A Martin Audio loud speaker was placed at 0 degree azimuth at the ear level of the participants.

3.2.3 Stimulus presentation:

Stimuli were presented at 40 dB SL (reference to speech recognition threshold). In all conditions, the Task was to repeat the syllable presented in auditory and/or visual modality. The responses were open set responses. In each condition, the task was preceded by a practice phase if necessary and a short break was given, if necessary. This was most often necessary for the children in the Group I (age range 6-7.11 yrs).

The stimuli in the unimodal conditions were presented first i. e, AO and VO. After 3 hrs of break testing was done in the bimodal conditions (AV+, AV-). A gap of 3 hrs was given between the two conditions to avoid the memorization of the stimulus by the subjects. The responses were noted by an experimenter (who remained behind the participants). After each condition participates were instructed .The experiment lasted about 15 minutes for each condition (AO, VO, AV+ and AV-).

3.3.4 Response elicitation and scoring:

The participants were asked to give an oral response of, what he/she had perceived. As previously explained each stimulus was presented three times, if 2 responses out of the 3 presentation were similar, that was taken as final response, regardless whether it was repeated correctly or not. The number of CV monosyllable responded correctly for individual participants in different condition, were noted.

3.3.5 Statistical analysis:

The SPSS software (version 16) and SSP were used for statistical analysis. . Later Two way MANOVA and Test of equality of proportions were used for analysis of the data obtained through scoring.

Chapter 4

Result and discussion

The aim of the present study was to evaluate the developmental trend of audiovisual integration across six age groups. The results are discussed under the following four headings.

I. Auditory only (AO)

II. Visual only (VO)

III. Audiovisual congruent (AV+)

IV. Audiovisual incongruent (AV-)

   a. McGurk effect

   b. Perception of Voicing

I. Audio only condition:

In this condition, signal was presented only through auditory mode and the participants were made to listen and speak what they heard. This auditory only perception was assessed for all the six sounds. The number of CV syllables perceived correctly by each individual was calculated for each group.

From the table 4.1, it can be observed that the auditory only performance is not very different across different age groups, though the younger groups have performed a little poorer than the older groups.

Table 4.1*Mean and SD of number of correctly identified syllable for AO condition across different age groups for both the groups.*

Group	**Age range (in years)**	**N**	**Mean for male**	**SD**	**Mean for female**	**SD**	**Total mean**	**SD**
I	6 – 7.11	25	5.33	1.56	6.00	0.00	5.68	1.10
II	8 – 9.11	25	5.8	0.577	5.54	0.87	5.68	0.74
III	10 –11.11	25	5.50	0.90	5.23	1.01	5.36	0.95
IV	12 – 13.11	25	6.0	0.00	6.0	0.00	6.00	0.00
V	14 – 15.11	25	5.58	1.44	6.0	0.00	5.80	1.00
VI	16 – 20	25	6.0	0.00	6.00	0.00	6.00	0.00

Note: Here 'N' depicts the number of participant in that group and 'SD' shows standard deviation.

In order to see, male and female listeners if there is a statistically significant difference across groups and between both the genders, two way MANOVA was done.

The analysis revealed no significant difference among different age groups [F (5,138) = 2.349, p>0.05] and between both male and female participants [F (1,138) = 0.462, P>0.05] in this AO condition.

The above results show that participants of all age groups, that is, from 6 to 20 years of age, were able to perform equally when only auditory information was delivered. These were expected results as all the individuals had normal hearing sensitivity.

The above findings support the results in audio only condition reveals that both children and young adults are found to be equivalent in speech sound perception and discrimination through audition. In speech sound discrimination experiment with preschool children approximately 2 years old and older have indicated that they can make cetain phonological distinctions. Barton (1976) has cautioned, however that speech sound discrimination experiment have not demonstrated that children beginning to speak all the sounds can perceive all of them relevantly.

II. Visual only condition

These were the responses of the participants for the unimodal visual only task. Here, the participants were given only visual stimulus (6 in numbers) and were instructed to give an oral response of what they saw. For each individual, the number of correctly identified CV syllables was counted. The mean and standard deviation of the scores for all the groups and for both the genders are given in table 4.2.

Table 4.2 *Mean and standard deviation of correctly identified syllables for VO condition across different age groups*.

Group	Age range (in years)	N	Mean for male	SD	Mean for female	SD	Total mean	SD
I	6 – 7.11	25	4.92	0.96	3.84	0.61	4.36	1.65
II	8 – 9.11	25	3.08	1.44	3.08	1.48	3.44	1.47
III	10 -11.11	25	3.08	1.78	2.92	1.25	3.00	1.50
IV	12 – 13.11	25	2.07	1.04	2.50	0.90	2.28	0.97
V	14 – 15.11	25	2.58	1.93	2.31	1.49	2.44	1.68
VI	16 – 20	25	3.33	0.77	3.23	1.57	3.84	1.06

Note: Here ‘N’ represents the no. of participants, ‘SD’ represents, standard deviation.

It can be observed from the table 4.2 that the Group I (6 – 7.11 years) performed higher than all the other groups. Group VI got the second highest score followed by Group II. Groups III (10 – 11.11 years), IV (12 – 13.11 years) and V (14 – 15.11 years) performed poorer than the other three groups.

Two way MANOVA was used to determine if there was a statically significant difference across groups. Results of two way MANOVA revealed a significant difference across groups ($F(5,138) = 8.251$, $p < 0.05$).

In order to see which of the groups had the difference, Bonferroni pairwise comparison test was carried out. The results revealed that there was a significant difference between Group I and Groups III, IV and V. Further, the performance of group IV and V was significantly different from group VI. However, there was no statistically significant difference between Group I and VI; Group III and VI; and Group II and all the other groups.

In summary, for VO condition, individuals between 6- 7.11 yrs of age and 16-20 yrs of age had better performance. Individuals with 10 to 15.11 performed with poorer s than other groups.

These results are not in agreement with other studies (Mc.Gurk and MacDonald, 1976; Desjardins, Rogers, & Werker, 1997; Dupont, Aubin, & Menard, 2005; Hockley & Polka, 1994; Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Sekiyama & Burnham, 2004; Wrightman, Kistler, & Brungart, 2006), where it has been found that the visual influence is lesser in the younger age and hence, younger children are poorer speech readers when compared to adults.

In the present study, the better performance of the youngest age group, i.e., 6-7.11 yrs could be because of the practice session given before the actual test. For group I and II, as this age group was not able to understand the instruction easily, practice sessions prior to the actual testing were given. They were repeatedly instructed to pay attention which might have lead to better performance. Further, the equivalent response in age range of 8 to 15.11 years may be because of the the low interest and low attention during the testing.

.probably during the testing , for this age group attention to each item was not monitored carefully where as it was done for the younger group.adults does not require such monitoring, hence the performance as good.

Comparison between the performance of the female and male is seen and the interaction between the group and gender interaction were analyzed. From the table 4.2, it can be observed that between male and female there is not much of difference is seen, results of two way MANOVA revealed no significantly significant differences for gender [F (1,138) = .175, p> 0.05]. In addition there was no interaction between the performance of groups and the gender [F(5, 138) = 1.763,p> 0.05].

III. Audiovisual congruent condition

These were the responses for the congruent stimuli pairs, where the stimuli in auditory and the visual modality were same in terms of place of articulation and voicing The participants responded to the stimulus as a whole, regardless of any specific modality.

Analysis of the sounds perceived correctly by each individual for all the groups was done and the mean and SD are given in the table 4.3. It is shown that all the groups performed equally and there was 100% response.

These findings can be attributed to the better access of auditory and visual information when presented simultaneously. It is also observed that the mean value of this condition is higher than that of the AO condition. This is a clear cut indication of increase in performance because of addition of visual information.

This is in agreement with the results of previous studies, which revealed that the addition of the visual cues to an auditory signal facilitates/ enhances overall perception of speech sound significantly (Sumby & Pollack 1954; Erber, 1969).

Table 4.3 *Mean and standard deviation of correctly identified syllables across diffent groups in audiovisual congruent condition*.

Group	Age range (in years)	N	Mean for male	SD	Mean for female	SD	Total mean	SD
I	6 – 7.11	25	4.92	.96	3.84	0.61	6.0	0.0
II	8 – 9.11	25	3.08	1.44	3.08	1.48	6.0	0.0
III	10 –11.11	25	3.08	1.78	2.92	1.25	6.0	0.0
IV	12 –13.11	25	2.07	1.04	2.50	0.90	6.0	0.0
V	14 –15.11	25	2.58	1.93	2.31	1.49	6.0	0.0
VI	16 – 20	25	3.33	0.77	3.23	1.57	6.0	0.0

Note: Here 'N' represents the no. of participants, 'SD' represents, standard deviation .

IV. Audiovisual incongruent condition.

a. *McGurk effect*

For studying the presence of McGurk effect, auditory /pa/ was dubbed onto the face movements for /ka/ was presented. McGurk effect was said to be present, if there was a fused response, that is, /ta/ for auditory /pa/ and visual /ka/ stimuli. The Mcgurk effect was said to be absent, if the participants perceived either auditory or visual stimulus, i.e., /p/ or /k/ respectively, instead of /ta/. If the participants' perceived /pa/ his/her perception was said to be auditory biased. If he/she perceived /ka/, his/her perception was said to be visual biased.

The total number of participants who perceived /t/ and the total number of participants who perceived either /p/ (auditory biased) or /k/ (visual biased) were calculated. In the table 4.4, the details of these responses across different groups are given.

Table 4.4: *details about the fused and nonfused responses for Aud /pa/ and Vis /ka/ pair*

Group	Total no. of participants perceived /ta/	Total no. of participants perceived /pa/	Total no. of participants perceived /ka/
I	17	4	4
II	7	14	4
III	21	2	2
IV	20	3	2
V	20	3	2
VI	21	0	2

Note: Here 'Aud' represents auditory mode and 'Vis' represents visual mode.

From the table 4.4, it can be observed that the performance of group II is considerably lower than the other groups. It can also be observed that the performance of all the other groups did not differ fromeach other except for the Group I, which has relatively poorer performance when compared to Groups III, IV, V and VI.

Table 4.5 *Details of the pairwise comparison of McGurk effect*

stimulus pa-ka	**Pair wise comparison**				
group	**I-II**	**II-III**	**II-IV**	**II-V**	**II-VI**
Z value	2.8307	3.9886	3.6888	3.6888	3.9886
P value	.004**	.000007**	.00002**	.00002**	.000007**

Note: Here ** depicts the significant difference between groups at the level of p< 0.01.

Test of equality of proportion using Smith's statistical package (SSP) software was performed in order to find out if there is a significant difference in across groups. Only the pairs of groups between which there was significant different found were given in the table 4.5.

Table 4.5 shows that there is a statistically significant difference in the performance of group II when compared to all the other groups. That is, the children in the age range of 8-9.11 years, performed significantly poorer than that of children in the age range of 6-7.11 yrs and that of 9 yrs older children.

Further, there was no significant difference between all the other groups (p>0.05), though performance of the youngest group was a little poorer when compared to Group III, IV, V and VI. That is the performance of children in the age group of 6-7.11 yrs and of children who were of 9 yrs and older did not differ significantly.

This is contradicting to the results obtained in the studies (Mcgurk & Macdonald 1976; Rosenblum, Schmuckler, & Johnson 1997; Wightman, Kistler & Brungart ,2006, Desjardins, Rogers, & Werker, 1997; Dupont, Aubin, & Menard, 2005; Hockley & Polka, 1994; Massaro, 1984; Massaro, Thompson, Barron, & Laren, 1986; Sekiyama & Burnham, 2004; Wrightman, Kistler, & Brungart, 2006), where, they have found poorer integration for children below 11yrswhen compared to adults. The reason for this is not clear.

Though this may be because, the innate differences of articulatory experience of the Indian children when compared to western children. The support for this may be taken from the finding that the phonological structure of the language influenses the audiovisual integration (Sekiyama &Burnham, 2008).may be the children takeb in the present study fully in the developed phonological skills.but this was not assessed in the present study.

Analysis of the number of subjects who perceived /p/ and who perceived /k/, instead of a fused response, was also done. Table 4.4 provides the details of this. It is shown in the table that, only seven of 25 of the participants in Group II perceived /t/ i.e. fused response.

Most of the other participants in this group (14 in number), perceived /p/ and four of them perceived /k/. It is also shown in the table that, in all the other groups, almost equal number of participants perceived /p/ and /k/.

Hence, test of equality of proportions was done only for Group II to check if there was a statistically significant difference between the auditory and visual response. It was found that there was a significant difference ($p<0.01$). This implies that, in this group of 8 – 9.11 years of age, the auditory influence was more and they had significantly poorer audiovisual integration compared to all others.

This is in the agreement with the results with the study done by Erderner & Burnham (2005) also found that there was a drop in ability of decline in the visual speech influence at around 7 yrs of age. Further, at the age of 8 yrs their integration jumped back that of 6 years old's and their reading skill also was found to be increased substantially. The reason attributed was that the reading skill start as an automatic skill at 7 yrs of age, so visual speech perception is not needed as a back support. Because of this they performed poorer.

However, in the present study, the performance drop was between 8 to 9.11 years old children. Hence, we further analyzed the age of seven participants who showed integration in this group. It was noticed that these seven children were of nine yrs or older. All the others in the (who did not show integration) were younger (at around 8.2 yrs) than those seven children. Hence, children at around 8 yrs of age showed poorer integration in contrast to the children (at 7 yrs of age) in Erderner & Burnham's study

The reason for this difference may be that, in the present study, all the children had Kannada language as mother tongue (communicate in Kannada language at home) and English as the medium of instruction at school.

Given this and the complexity of phoneme-grapheme correspondence of English language, the children of 8 yrs of age, in the present study, might have just developed reading as an automatic process. However, this is just a hypothesis, which need to tested and confirmed

Another objective of the study was to analysis for gender effect. The responses of male subjects and female subjects are also given in Table 4.6. It was observed that more number of female was giving fused responses compared to male participants except for the Group I. In the Group I, none of the female subjects showed integration.

Table 4.6 *No. of male and female participants who gave fused responses*

Fused response for Pa-ka stimulus			
Group	**Total no. of participants given fused responses**	**Total no. of feaml given fused responses**	**Total no. of male given fused responses**
I	17	0	8
II	7	5	6
III	21	8	7
IV	20	9	6
V	20	11	5
VI	21	13	6

Note: Here the total no. of participants will not be the sum of male and female participants, The test of equality of proportion was done in order to compare the significance of difference between male and female participants across different age group. It was found that, only for group I, V and VI, there was a significant difference between male and female participants. Female performed better when compared to male participants, except for the group I where male did significantly better than females. In the other groups, there was no significant gender differences found.

Table 4.7 within group gender differences in audiovisual integration

Stimulus **pa – ka**	**Within group gernder comparison for fused responses.**		
Group	**I**	**V**	**VI**
Z value	3.23	2.76	2.92
P value	0.001**	0.005**	0.003**

Note:* * significant difference at the level of $p < 0.01$.

In most of the studies, female subjects are found to show a stronger McGurk effect than males (Iyar et al 2005). This was attributed to female listeners' ability to have superior lip reading abilities and was better at integrating audiovisual speech information than the male listeners (Johnson et. al. 1988; Watson et. al 1996; Desjardins & Werker, 2004; Iyar, 2005) T/his is in consonance with the findings of the present study for children of 14 yrs of age and above. This may be may be because of in Iyar's study the integration was studied in the presence of noise.

However below this age there is no gender difference was seen. For this the reason is not known.

.

b. Perception of Voicing

In this condition, the voiced and voiceless sounds were paired (audio voiceless with voiced video or wise versa). The performance was checked to find out if there is an influence visual stimulus in the perception of voicing when an incongruent auditory stimulus is given. It can be observed from the table 4.8 that all the participants in all the groups gave repeated the audiotory stimuli and there was no influence of vision.

That is, when a voiced stimulus presented through auditory modality was voiced, participants a perceived voiced stimulus and when it was voiceless, participant also perceived a voiceless sound.

Table 4.8 *Mean and standard deviation of the numbersof syllables perceived for voicing perception across age groups*.

Group	Age range (in years)	N	Mean For audio	SD	Mean for video	SD
I	6 – 7.11	25	6.0	0.0	6.0	0.0
II	8 – 9.11	25	6.0	0.0	6.0	0.0
III	10 – 11.11	25	6.0	0.0	6.0	0.0
IV	12 – 13.11	25	6.0	0.0	6.0	0.0
V	14 – 15.11	25	6.0	0.0	6.0	0.0
VI	16 – 20	25	6.0	0.0	6.0	0.0

Note: 'N' represents no. of participants in each group; 'SD' represents standard deviation**.**

Because this was 100% response given by all the groups .thus there was no statistical analysis was done.

The findings of the present study are not in consonance with the previous studies (Schwartz et. al., 2004; Berthommier et. al, 2011). This could be attributed to the stimulus used by the above studies. Those studies used a prevoicing continuum between /p/ and /b/, which was realized by varying the amplitude for a unique prevoicing segment obtained from a /b/. This is quite different from the /pa/-/ba/ combination used .Sounds in present study.

Chapter 6

Summary and conclusion

The present study aimed to evaluate developmental changes in the audiovisual integration process and the integration of voicing cue in the normal hearing Kannada speaking subjects. Six groups age ranging from 6 to 20 years, in difference of 2 years, were made. There were 25 subjects taken in each group. Six monosyllable were recorded and stimuli were prepared for four conditions; Audio only; visual only; audiovisual congruent; and audiovisual incongruent.

Administration of a checklist and various other tests including. SCAP, Tumbling 'E' chart, routine hearing evaluation GDT and SPIN was done. This was done to ensure that all the participants had no auditory processing difficulties and no hearing difficulties were present. After the evaluation, 150 subjects were chosen for the present study. They were divided into six groups based on the age. All the participants were presented the stimuli in all the four condition. The summary of the results are given below.

- The analysis of the results of AO condition was done using two way MANOVA. The results revealed that all the groups performed almost equally, and this is because of the normal hearing sensitivity and all of them had normal speech and language skills.

- In VO condition there was no developmental pattern seen rather beter response were given by 6- 7.11 young age and 16-20 yrs, the children between 8- 15.11 yrs of age found to have poorer responses. The better response by younger group may be because of the repeated instruction and practice session given to the younger group prior to the testing.

- Audiovisual congruent condition resulted in 100% correct response by all the age groups. The performance in this condition was better than the AO condition. This revealed this led that the addition of visual information to auditory stimulus improved the performance, which is in consonances with the other studies.

- There were no gender difference for AO and VO and audiovisual congruent conditions.

- For the McGurk effect which reflects the by the age of AV which was found to be achieved as adults by the age of 6-7.11 yrs. The groups except for 8-9.11yrswho performed poorer. The reason for this may be that all the participants in the group had good phonological skills (which are found to affect audiovisual integration).

  Further the 8 yrs old probably had just developed reading as an automatic skill because with these was a drop in the integration. However, this hypothesis will not be confirmed until tested.

- Female participants were found to be better than the male participants in three of the six groups and for two groups there was no gender difference was observed. Hence, it is important to study both the genders separately.

- The results of voicing perception task revealed that there the entire participant in all the groups repeated the auditory stimulus only and there was no influence vision on voicing perception was being observed. The previous studies reported to be perception of voicing through the temporal cue of lip gesture onset, which was not revealed by the present study may be due to the difference in the stimulus used in present stud.

- Results of audiovisual integration for voicing cue showed no influence in any of the groups.

- Further research is need in order to study the different factors that can affect this auditory visual integration process in which can further help to utilize the information to control it during the use of those perceptual activities in the rehabilitation of the needful population.

The implication of the present study can be:

- The data from the present study will be useful to compare the normal to different clinical population.
- It will add the information to literature regarding the auditory visual integration in Indian population.

- To conclude, the audiovisual integration of the kannada speaking young children were as good as older children and young adults except for the 8 years children.
- There is a need to evaluate the phonological skills and articulatory experience and reading skills. When we study audiovisual integration to have a clear explanation. This again confirms that the vision facilities speech perception in the congruent condition. this study also emphasis on the importance of studying male and female participants separately as ntegration performance differ.

References

Alpiner, J., & McCarthy, P. (1987). *Rehabilitative Audiology: Children and Adults*. Baltimore: Williams & Wilkins.

Alsius, A., Navarra, J., Campbell, R., & Soto-Faraco, S. (2005). Audiovisual integration of speech falters under high attention demand. *Current Biology*, **15**, 839–843.

Bashore. T., Ridderinkhof, R & van der Molen, M. (1998). The decline of cognitive processing speed in old age. *Current Directions in Psychological Science.* 6, 163-169.

Bergeson, T.A., Pisoni, D.B., & Davis, R.A.O. (2003). A longitudinal study of audiovisual speech perception by children with hearing loss who have cochlear implants. *The Volta Review*, 103, 347-370

Burnham, D. (1998). Language specificity in the development of auditory–visual speech perception. In Hearing by eye II (eds R. Campbell, B. Dodd & D. Burnham), ch. 2, pp. 27–60. East Sussex, UK: Psychology Press.

Bernstein, L., Auer, E. T. J., and Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading.

Cerella. I ., Poon. L., & WiJllams, D. (1989). Age and the complexity hypothesis. In J. Binen (Ed.). *Aging in the1980's* (pp. 332 - 340). New York: Erlbaum.

Committee on Hearing and Bioacoustlcs and Biomechanics (CHABA) (1988). Speech understanding and aging. *Journal of the Acoustical Society of America. 83.* 859-893.

Cavanaugh, J.C (1997). *Adult development and aging.* New York: Brooks/Cole. Cerella. I .. Poon. L.. & WiJllams. D. (1989). Age and the complexity hypothesis. In J. Binen (Ed.). *Aging in the 1980's* (pp. 332 - 340). New York: Erlbaum

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

Calvert,G.A. (2001). Crossmodal Processing in Human Brain: Insight from functional neuroimaging studies. Cerebral Cortex,11, 1110-1123

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., and David, A.S. (1997). Activation of auditory cortex during silent lipreading. Science 276, 593–596.

Calvert, G.A., Hansen, P.C., Iversen, S.D., Brammer, M.J. (2001) Detection of multisensory integration sites by application of electrophysiological criteria to the BOLD response.NeuroImage 14:427–438.

Cameron, C. (2004). What are the speechreading abilities of children and adults with normal hearing and children with auditory processing disorders? Unpublished Master of Speech Language Therapy Practice research project, The University of Auckland.

Cienkowski, K. (1999). Auditory-visual speech perception across the lifespan [Doctoral Dissertation. University of Minnesota. 1999l. *Dissertmion Abstracts International*, 60/01. 116.

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.

Desjardins, R., Rogers, J., & Werker, J. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology, 66*, 85-110.

Desjardins, R., & Werker, J. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology, 45*, 187-203.

Desjardins, R. N. & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? Developmental Psychobiology *45*, 187–203.

Driver, J. (1996) “Enhancement of selective listening by illusory mislocation of speech sounds due to lipreading”, *Nature* 381: 6577, 66-68.

Dupont, S., Aubin, J., & Menard, L. (2005). A study of the McGurk effect in 4- and 5-year-old French Canadian children. *ZAS Papers in Linguistics, 40*, 1-17.

Engel, A.K., Fries, P., & Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top-down processing. *Nature Reviews Neuroscience,* **2** , 704-716.

Erber, N. (1969). Interaction of audition and vision in the recognition of oral speech stimuli. *Journal of Speech and Hearing Research*, *12*, 423–425.

Ettlinger, G.,& Wilson,W.A. (1990). Cross-modal performance: behavioral processes, phylogenetic considerations and neural mechanisms. *Behavioral Brain research, 40*, 169-192.

Frens, M.A. Vanopstal, A.J., & Vanderwilligen, R.F.(1995) Spatial and temporal factors determine auditory-visual interaction in human saccadic eye movement perception and psychophysics,*57*,802-816

Giard, M.H., Peronnet, F. (1999). Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci 11*:473–490.

Grant, K. W., and Seitz, P. F. (1998). “Measures of auditory-visual integration in nonsense syllables and sentences,” *J. Acoust. Soc. Am. 104*, 2438– 2449.

Grant, K.W., Braida ,L.D.(1991). Evaluating the articulation index for AV input. *J Acoust Soc Am*; *89*. 2952– 2960. [PubMed: 1918633]

Garner. W, (1974). *Theprocessingofinformationandstructure.* Potomac, MD: Erlbaum.

Green, K. P. & Kuhl, P. K. (1991). Integral processing of visual place and auditory voicing information during phonetic perception. *J. Exp. Psychol*.: Hum. Percept. Perform *17*, 278–288. (doi:10.1037/0096-1523.17.1.278)

Green, K.P., Kuhl, P.K., 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. *Percept Psychophys;50*. 524–536. [PubMed: 1780200]

Green, K. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In R. Campbell, B. Dodd & D. Burnham (Eds.), *Hearing by eye II. Advances in the psychology of speechreading and auditory-visual speech* (pp. *3-25*). Hove, UK: Taylor & Francis

Humes, L., & Roberts, L. (1990). Speech recognition difficulties of the hearing impaired elderly: The contributions of audibility. *Journal of Speech and Hearing Research*, *33,* 726–735.

Humes, L., & Christopherson, L. (1991). Speech identification difficulties of hearing-impaired elderly persons: The contributions of auditory processing deficits. *Journal of Speech and Hearing Research*, *34,* 686–693.

Hockley, N., and Polka, L. (1994) .“A developmental study of audiovisual speech perception using the McGurk paradigm,” *J. Acoust. Soc. Am. 96*. 3309

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

Kalikow, D., Stevens, K., & Elliot, L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *Journal of the Acoustical Society of America*, *61*, 1337–1351.

Kuhl, P. K. & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science 218*, 1138–1141. (doi:10.1126/science.7146899)

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

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception* & *Psychophysics*, SO, 93-107.

Kuhl, P. K., & Meltzoff, A. N. (1988). Speech as an intermodal object of perception. In A. Yonad (Ed.), *Perceptual development in infancy:* The *Minnesota symposia on child psychology* (pp. 235-266). Hillsdale, NJ: Erlbaum

Krause, C.M., Mottonen, R., Jensen, O., Lampinen, J., & Sams, M. (2001). Brain oscillations during audiovisual speech perception – a MEG study. *Paper was presented at the Human Brain Mapping Conference, June 10-14, England.*

Lachs, L., Pisoni, D.B., & Iler Kirk, K. (2001). Use of audiovisual information in speech perception by prelingually deaf children with cochlear implants: A first report. *Ear and Hearing*, *22,* 236-251

Lockhead, G., & Pomerantz, I. (Eds.). (1991). *The perception of structure.* Washington, DC: American Psychological Association.

Lewkowicz,D.J. (1996). Perception of auditory –visual temporal synchrony in human infant. *Journal of Experimental psychology. Human perception and performance 22*,1094-1106.

Lewis, J.W., Beauchamp, M.S., DeYoe, E.A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. Cereb Cortex10:888.

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

MacDonald, J. & McGurk, H. (1978). Visual influences on speech perception processes. *Percept.Psychophys. 24*, 253–257

Massaro, D. W. (1984). “Children’s perception of visual and auditory speech,” *Child Dev. 55*, 1777–1788.

Massaro, D. W. (1998). *Perceiving Talking Faces: From Speech Perception to a Behavioral Principle* (MIT, Cambridge, MA)

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

Macaluso, E., Frith ,C.D., & Driver, J. (2000a). Modulation of human visual cortex by crossmodal spatial attention.*Science 289*:1206–1208.

Macaluso, E., Frith ,C., & Driver, J. (2000b). Selective spatial attention in vision and touch: unimodal and multimodal mechanisms revealed by PET.*J Neurophysiol, 83*,3062–3075.

Miller, L.M., & D'Esposito, M.D. (2005). Perceptual fusion and stimulus coincidence in the crossmodal integration of speech. *Journal of Neuroscience, 25*, 5884–5893.

Millar, J.O. (1982). Divided attension :Evidence for coactivation with redundant signals. *Cognitive psychology*, *14*,247-279.

Middelweerd, M., & Plomp, R. (1984). The effect of speechreading on the speech reception threshold of sentences in noise. *Journal of the Acoustical Society of America*, *82*, 2145–2147.

Murray, E.A., Malkova, I., & Goulet, S. (1998). In A.D. Milner (Ed.). *Comparative neuropsychology* (pp. 51-69), Oxford: Oxford University Press.

Munhall, K., & Vatikiotis-Bateson, E. (1998). The moving face during speech communication. In R. Campbell, B. Dodd & D. Burnham (Eds.), *Hearing by eye II: Advances in the psychology of speechreading and auditory-visual speech* (pp. 123-139). Hove, UK: Psychology Press.

Repp, B.H., Manuel, S.Y., Liberman, A.M., & Studdert-Kennedy, M. (1983). Exploring the "McGurk effect". *Paper Presented at the 24 annual meeting of the Psychonomicth Society in San Diego.*

Rosenblum, L. and Saldana, H. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 22*(2), 318-331.

Rosenblum, L. D., Schmuckler, M. A. & Johnson, J. A. (1997) The McGurk effect in infants. *Percept. Psychophys. 59*, 347–357.

Reisberg, D., McLean, J., and Goldfield, A. (1987). "Easy to hear but hard to understand: A lip-reading advantage with intact auditory stimuli," in *Hearing* *by Eye: The Psychology of Lip-reading*, edited by B. Dodd and R. Campbell Erlbaum, London, pp. 97–113.

Hockley, N., and Polka, L. (1994).“A developmental study of audiovisual speech perception using the McGurk paradigm,” *J. Acoust. Soc. Am. 96,* 3309.

Massaro, D. W. (1984). “Children’s perception of visual and auditory speech,” *Child Dev. 55*, 1777–1788.

Massaro, D. W., Thompson, L. A., Barron, B., and Laren, E. 1986. “Developmental changes in visual and auditory contributions to speech perception,” *J. Exp. Child Psychol. 41*, 93–113.

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) . Hillsdale, NJ: Erlbaum.

Massaro, D.W., & Cohen, M.M. (1996). Perceiving speech from inverted faces. *Perception and Psychophysics, 58* , 1047-1065.

Mottonen, R., Krause, C.M., Tiipana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Brain Research, Cognitive Brain Research, 13* , 417-425.

O'Hare, J.J. (1991). Perceptual integration. *Journal of the Washington Academy of Sciences, 81* , 44-59.

Pichora-Fuller, M. K., Schneider, B., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *Journal of the Acoustical Society of America, 97*, 593–608.

Phillips, W.A., & Singer, W. (1997). In search of common foundations for cortical computation. *Behavioral and Brain Sciences, 20* , 657-683

Plude, D., & Hoyer, W. (1985). Attention and performance:Identifying and localizing age deficits. In N. Charness (Ed.), *Aging and Human Performance* (pp. 47–99). Chichester, England: Wiley.

Plude, D., & Doussard-Roosevelt, J. (1989). Aging. selective attention. and feature integration. *Psychology* & *Aging, 4*, 98- J 05.

Rabbit, P. (1965). An age-decrement in the ability to ignore Irrelevant information. *Journal of Gerollology. 20, 233-238,*

Rosenblum, L.D., Yakel, D.A., & Green, K.P. (2000). Face and mouth inversion affects
on visual and audiovisual speech perception. *Journal of Experimental Psychology. Human Perception and Performance, 26* , 806-819.

Remez, R., Rubin, P., Pisoni, D., and Carrell, T. (1981). Speech perception without traditional speech cues. *Science. 212,* 947–949.

Sams, M., & Imada, T. (1997). Integration of auditory and visual information in the human brain: Neuromagnetic evidence. *Society for Neuroscience. Abstracts, 23* . 1305.

Salinas, E., & Sejnowsli, T.J. (2001). Correlated neuronal activity and the flow of neural
information. *Nature Reviews Neuroscience, 2* , 539-550.

Sekiyama, K., & Burnham, D. (2004). Issues in the development of auditory-visual speech perception: Adults, infants, and children. *Interspeech-2004*, 1137-1140.

Schwartz, J. L., and Cathiard, M. A. (2004). “Modeling audio-visual speechperception. Back on fusion architectures and fusion control,” in Proceedings of the ICSLP’2004, Jeju, Korea, pp. 2017–2020.

Schorr, E.A., Fox, N.A., Van Wassenhove, V., Knudsen, E.L. (2005). Auditory-visual fusion in speech perception in children with cochlear implants. *PNAS*, *102*, 18748-18750.

Sekiyama, K. & Tohkura, Y. 1993 Inter-language differences in the influence of visual cues in speech perception. *J. Phonet. 21*, 427–444.

Shoop, C., & Binnie, C. (1979). The effects of age on the visual perception of speech. *Scandinavian Audiology*, *8*, 3–8.

Salthouse, T. (1991). *Theoretical perspectives 011 aging"* Hlllsdale. NI: Erlbaum.

Stein BE, Meredith MA, Huneycutt WS, McDade L (1989) Behavioural indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. *J Cogn Neurosci* 1:12–24.

Stein, B.E., & Meredith, M.A. (1993). *The merging of the senses.* Cambridge, MA: MIT Press.

Stein, B.E., Meredith, M.A., & Wallace, M.T. (1994). Development and neural basis of multisensory integration. In D.J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 81-105). Hillsdale, NJ: Erlbaum

Sumby, W. H., and Pollack, I. (1954). "Visual contribution to speech intelligibility in noise," *J. Acoust. Soc. Am. 26*, 212–215.

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

Summerfield, Q. (1987). Some preliminaries to a comprehensive account of audio-visual speech perception. In B. Dodd & R. Campbell (Eds.), *Hearing by eye:* The*psychology oflip-reading* (pp. 3-51). London: Erlbaum.

Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: the McGurk effect in Chinese subjects. *Percept. Psychophys, 59*, 73–80.

Sloutsky, V.M., & Napolitano, A.C. (2003). Is a picture worth a thousand words? Preference fro auditory modality in young children. *Journal of Child Development*, *74*, 822-833

Walden, B.E., Busacco, D.A., & Montgomery, A.A. (1993). Benefit from visual cues in AV speech recognition by middle-aged and elderly persons. J Speech Hear Res;36:431–436. [PubMed: 8487533]

Wightman, F., Kistler, D., & Brungart, D. (2006). Informational masking of speech in children: Auditory-visual integration. Journal *of the Acoustical Society of America*, *119*, 3940-3949.

Taylor, H.R., "Applying New Design Principles to the Construction of an Illiterate E Chart." Am J Optom & Physiol Optics 55:348, 1978.

Appendix

.

Screening checklist for central auditory processing (SCAP)

Yathiraj and Mascarenhas (2002)

Please place a tick (√) mark against the choice of answer that is most appropriate.

Sl. No.	Questions	Yes	No
1	Does not listen carefully and does not pay attention (requires repetition of instruction).		
2	Has a short attention span of listening (approx 5-15mins).		
3	Easily distracted by background sound.		
4	Has trouble in recalling what has been heard in the correct order.		
5	Forgets what is said in few minutes.		
6	Has difficulty in differentiating one speech sound from other similar sound.		
7	Has difficulty in understanding verbal instruction and tent to misunderstand what is said which other children of the same age would understand.		
8	Show delayed response to verbal instruction or questions.		
9	Has difficulty in relating what is heard with what is seen.		
10	Poor performance in listening task, but performance improves with visual cues.		
11	Has a pronunciation problem (mispronunciation of words).		
12	Performance is below average in one or more subjects, such as social subjects, I/II language.		