Developmental Changes of Audio-Visual Integration: A Cross Sectional Study

[1]Preeti Sahu & [2]Geetha C.

Abstract

The study aimed to evaluate the developmental changes of audio-visual integration using McGurk phenomenon and the integration of voicing cue in normal hearing Kannada speaking participants. One hundred and fifty participants, age ranging from 6 years to 20 years were chosen for the study and were divided into six groups (25 participants in each group) based on age. Auditory and visual recording of six stop consonants in combination with vowel /a/ was done and stimuli were prepared for four conditions: audio only (AO); visual only (VO); audio-visual congruent (AV+); and audio-visual incongruent- for studying McGurk effect and audio-visual integration of voicing cue. Two way MANOVA and test of equality were used for analysis of data. Results revealed that all groups performed equally in AO condition. In VO condition, better responses were observed in 6 to7.11 year olds and 16 to 20 year olds. Audio-visual congruent condition resulted in 100% correct response by all the age groups. In McGurk task, 8 to 9.11 year olds performed poorly. There was no significant difference between the performance of 6 to 7.11 year old children and of 9 years or older. It can be concluded that the developmental changes of audio-visual integration of Indian children are different from the developmental changes reported in western studies. The assessment of phonology, reading and writing skills of the children whilst looking into the audio-visual integration will give better insight about the process. It can also be concluded that it is important to study the female and male participants separately and, also, to have a task which is more demanding such as testing in the presence of noise.

Keywords: *McGurk effect, Audio-visual integration, Voicing, Congruent, Incongruent.*

Introduction

Speech perception is an audio-visual phenomenon, in both 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/ dubbed onto the articulatory movement for /ga/, results in the perception of /da/. This suggests that auditory-visual speech is perceived as a 'whole' perceptual unit rather than as a separate unimodal feature (Green & Kuhl, 1991). There are different factors that could affect the process of audiovisual integration. One of the major factors is age. McGurk and MacDonald (1976), in their earlier 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) and increasing audiovisual integration over a period of time (McGurk & Macdonald, 1976; Rosenblum, Schmuckler & Johnson, 1997; Wightman, Kistler, & Brungart, 2006).

This pattern of result has been replicated in studies done by different researchers (Massaro, 1984; Massaro, Thompson, Barron & Laren, 1986; Sekiyama & Burnham, 2004; Desjardins, Rogers & Werker, 1997; Dupont, Aubin & Menard, 2005; Hockley & Polka, 1994; rightman et al., 2006). The reason for this may be a result of ongoing peripheral vision development (Massaro, 1984). Massaro et al. (1986) also found that children are poorer lip readers than adults, which results in lesser visual influence rather than attention playing a role.

The age by which visual influence completes has also been researched upon. 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, Iyer (2005), found that younger children (7 year olds female & 11 years old male listeners) were more likely to use auditory cues for speech perception, while older children (11 years old female & 15 years old male listeners) used both auditory and visual information in poor listening conditions. There is another study which has compared adults with children till the age of 11 years, assuming that there will not be much changes in the auditory integration process above 11 years of age (Sekiyama & Burnham, 2004).

However, there could be changes in audiovisual integration even after 11 years of age which also may vary depending on the gender. Result obtained by Iyer's study is in support of this view. From the above, it can be observed that, there are equivocal results seen with respect to age by which children achieve adult-like performance. Hence, studying audiovisual integration across age group, that is, in children even above 11 years of

[1]Email: preetisahu50@gmail.com,
[2]Lecturer in Audiology, Email: geethamysore.cs@gmail.com

age will help in better understanding of developmental changes of audiovisual integration process in normal hearing participants. This can also be useful to evaluate and rehabilitate different pathological conditions.

The auditory visual integration as demonstrated by McGurk effect was also found to be dependent 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 & Burnham, 2004). The reason for this is that Japanese language has simple phonological structure and hence, promotes better auditory only speech perception early in development. Thus, while Japanese children may experience audiovisual interaction to some extent, they may not integrate visual information because the phonological environment does not demand it. Hence, there clearly, is an influence of the language on the integration ability.

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

However, the studies evaluating the audio-visual integration have mainly investigated Western and Japanese population. The phonological system of Indian languages and the articulatory experience of Indian population are very different from English and Japanese languages. The conclusions drawn from those studies may not hold good to the Indian population, as the phonological structure of Indian languages are very different.

In the Indian context, a research study perceptually evaluated the auditory integration of ten Hindi speaking normal individuals in the age range of 18 to 25 years in CVC context (Sarika, 1995). The results of this study revealed that the auditory and visual mode influence each other. Further, when the stimulus was in the back of the oral cavity, the subject were likely to perceive the in between stimulus, i.e., neither auditory nor visual stimulus. As it can be noted, in the above study, the integration was assessed only in adult listeners and on a very small sample size.

Khan, Salian, Rajashekhar and Dhamani (2008) studied normal hearing and cochlear implant children to see the impact of temporal envelop and fine structure on audiovisual integration. In addition, audiovisual integration was also studied by Wasim, Chavan, Deema and Dhamani (2010) in dyslexic, poor academic performers and normal reading children. Apart from these, there are no other studies, to our knowledge, which have actually evaluated the developmental changes of audiovisual integration in the Indian population.

With respect to the stimuli used for evaluating audiovisual integration, most of the times, stop consonants were used. Further, it was found that when the bilabial was paired with velar consonants, McGurk effect was found to be more intense (McGurk & MacDonald, 1976, Massaro, 1984; Massaro et al., 1986; Grant & Seitz 1998; Hockley & Polka, 1994; Sekiyama & Burnham, 2004; Dupont et al., 2005), which indicates that place of articulation plays a role in the integration process.

The voicing cue, one of the important features in speech perception, could also be perceived through visual modality via observation of lip movement. However, only few number of studies were found in this aspect. Schwartz, Berthommier and Savariaux (2004) compared identification scores for voicing in audio-only and audiovisual conditions. They used pre-voicing continuum from /ba/ to /pa/. They found improvement in perception of voicing when visual cue was added with audition. In the present study, we were interested to see, if these results could be obtained for participants across age groups using a non-continuum stimulus.

Hence, the aim of the present study was to explore the effect of age on audiovisual integration across six age groups (from age range 6 to 20 years), for male and female listeners, using the following conditions: Audio only; Visual only; Audiovisual congruent; and Audiovisual incongruent condition to test the audiovisual integration through McGurk phenomenon and audiovisual integration of voicing cue.

Method

The method consisted of 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 and duration of testing.

Phase I

Selection of Participants: In the first phase, selection of participants was done using a checklist and various tests. They are 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 years. There were 170 children and 30 adults. All of them had normal speech and language skills and had no history of complaint of any neurological deficits.

Further selection was based on a series of screening and diagnostic tests to rule out any auditory processing difficulties and hearing difficulties. Depending on the age group, the tests administered for selection varied. Out of the 200 participants selected, only 150 met the inclusion criteria based on the test results. Thus, these 150 participants were enrolled for the second phase.

Screening checklist for Central Auditory Processing (SCAP): This check list was developed by Yathiraj and Mascarenhas (2002, 2004). This checklist was administered on 170 participants from different schools. The participants with scores less than 6 were considered to have passed the checklist, as scores less than 6 indicates no processing problems. Out of 170 subjects, 140 participants scored less than 6 and were taken up for visual acuity test. The rest (30 in number) of the participants who failed were eliminated from the study at this stage itself.

Vision acuity test: These 140 children, who passed SCAP, and 30 adult participants, were tested for visual acuity by using tumbling 'E' chart (Taylor, 1976). 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. All 170 participants (140 children and 30 adults) passed the visual acuity test. Hence, all of them underwent a routine audiological evaluation.

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 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 below 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, SPIN and GDT were administered except for 15 children. These 15 children were below seven years of age. SPIN could not be administered on these children, as they had difficulty understanding instructions. GDT also could not be administered on them, as it has been standardized only for children above seven years of age. Thus, for these 15 children, between 6 to 7 years of age, the results of SCAP and routine audiological evaluation only were considered for inclusion. The remaining 148 (163 minus 15) subjects were considered for SPIN and GDT testing.

Speech in noise test: Speech in noise test (SPIN) was carried out for 148 participants. The phonemically balanced (PB) Kannada word list developed by Yathiraj and Vijayalakshmi (2005) was used for this. Speech identification scores in quiet and in the presence of ipsilateral noise were found out. The level of test stimulus and noise were 40 dB SL (ref. SRT).

Scoring was done by calculating percent correct responses. The difference between percent correct response in quiet and that 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.

Gap detection test: Gap detection test (GDT), developed by Shivprakash (2003) was administered on these 148 participants, to further confirm normal auditory processing abilities. This test consists of 56 trials, including 8 catch trials of broad band noise which contained gaps. 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 normative 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 years of age were included in the second phase of the study. Hence, there were a total of 150 participants for the next phase. These 150 participants were then divided into 6 groups based on the age (in years), with 25 participants in each group. The details of different groups and age are given in the Table 1.

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: Audio only (AO), Visual only (VO), Audiovisual congruent (AV+) and finally, Audiovisual incongruent (AV-).

AO and VO were unimodal conditions, and AV+ and AV- were bimodal conditions. In AV+ condition, the sounds presented through auditory and visual mode were the same, wherein, the AV- condition had conflicting auditory and visual stimuli. *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 participants were asked to practice uttering the syllables before recording. During the recording, the participants were 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.

Stimulus editing: These avi files 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 speaker, three best recordings out of the four recordings were selected for further editing. These stimuli were separated into 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 the syllables were normalized in order to avoid

Table 1: Details of different groups, Mean age and Standard deviation

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

the effect of intensity difference between different syllables.

Using these edited audio and video files, stimulus for AO, VO, and AV (+/-) conditions was 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 same, where as in the AV- condition, monosyllables in the two modalities (audio and video) were different. For studying the presence of McGurk effect, auditory /pa/ was dubbed onto the articulatory movements for /ka/, and auditory /ba/ was dubbed onto the articulatory movements for /ga/.

For assessment of audiovisual integration of voicing cue, the voiced and voiceless sounds were paired (for e.g., audio voiceless /pa/ was paired with video of the /ba/ and vice versa). This was done for all the six stop consonants. The details about the stimuli in each of these conditions are given in the Table 2.

Goodness test: The prepared stimuli were presented to ten adults in the age range of 16 to 30 years. The participants were asked to rate in a three point rating scale which included: 1) Clear and good quality; 2) Distorted but can be identified; and 3) Distorted and cannot be identified.

Out of three utterances, the utterance which was rated as clear was taken up for making final stimulus. In addition, the stimulus prepared for assessing the McGurk effect, that is, Aud /pa/ - Vis /ka/ and Aud /ba/ - Vis /ga/ were also presented to those participants. The 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.

Seating 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.

Stimulus presentation: Stimuli were presented at 40 dB SL (ref. SRT). In all the conditions, the task was to repeat the syllables presented in auditory and/or visual modality. The responses were open set responses. In each condition, the task was preceded by a practice phase and a short break was given, if necessary. This was most often necessary for the children in the Group I (age range 6 to 7.11 years). The children above this age group were not given any repetitions of the instruction or the task.

Table 2: The stimuli used in different conditions

AudioonlyAO	VisualonlyVO	Audiovisual(congruent)AV+	Audiovisual(incongruent)AV−ForMcGurkeffect	Audiovisual(incongruent)AV−Forvoicing
Auditory	Visual	Auditory−Visual	Auditory−Visual	AuditoryVisual
Aud/ba/	Vis/ba/	Aud/ba/+vis/ba/	Aud/pa/+vis/ka/	Aud/pa/+vis/ba/
Aud/da/	Vis/da/	Aud/da/+vis/da/		Aud/ta/+vis/da/
Aud/ga/	Vis/ga/	Aud/ga/+vis/ga/		Aud/ka/+vis/ga/
Aud/pa/	Vis/pa/	Aud/pa/+vis/pa/		Aud/ba/+vis/pa/
Aud/ta/	Vis/ta/	Aud/ta/+vis/ta/		Aud/da/+vis/ta/
Aud/ka/	Vis/ka/	Aud/ka/+vis/ka/		Aud/ga/+vis/ka/

The stimuli in the unimodal conditions were presented first, i. e, AO and VO conditions. After three hrs of break, testing was done in the bimodal conditions (AV+, AV-). This gap of three hrs was given 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 participants were instructed depending on the next task. The experiment (testing time) lasted for about 15-20 minutes all together.

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 two responses out of the three presentation were same, that was taken as final response, regardless of whether it was repeated correctly or not. The number of CV monosyllables responded correctly for each individual participant, in different conditions, was calculated.

Results 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: Auditory only (AO); Visual only (VO); Audiovisual congruent (AV+); and Audiovisual incongruent (AV-) condition.

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 six sounds. The number of CV syllables perceived correctly (that is, out of six syllables presented, how many were repeated corrected) by each individual was calculated for each group.

From the Table 3, it can be observed that the auditory-only performance is not very different across different age groups and gender, though the younger groups have performed a little poorer than the older groups.

In order to see, if there is a statistically significant difference across groups and the gender, 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 male and female participants [$F(1,138) = 0.462$, $p>0.05$] in the 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 results are consistent with the findings that the discrimination of the stop consonants have been observed in infants as young as 20 to 24 weeks of age (Moffitt, 1971). In the present study, all the participants were well above this age.

Visual only Condition

These were the responses of the participants for the unimodal visual only task. Here the participants were given only the visual stimuli (6 in number) 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 number of syllables correctly identified for all the groups and for both the gender are given in Table 4.

It can be observed from the Table 4 that the Group I (6 to 7.11 year olds) performed higher than all the other groups. Group VI (16 to 20 year olds) got the second highest scores (though the SD of this group is smaller than the Group I) followed by Group II. Groups III (10 to 11.11 year olds), IV (12 to 13.11 year olds) and V (14 to 15.11 year olds) performed poorer than the other three groups.

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

In summary, for VO condition, individuals between 6 to 7;11 years of age and 16 - 20 years of age had better performance. Individuals with 8 to 15;11 years of age performed poorer than other groups.

These results are not in agreement with other studies (McGurk & MacDonald, 1976; Desjardins et al., 1997;

Table 3: Mean and SD of number of correctly identified syllables across different age groups and gender for AO condition

Group	Age range (in yrs)	N	Mean for male participants	SD	Mean for female participants	SD	Total Mean	SD
I	6 to 7; 11	25	5.33	1.56	6.00	0.00	5.68	1.10
II	8 to 9;11	25	5.80	0.577	5.54	0.87	5.68	0.74
III	10 to 11:11	25	5.50	0.90	5.23	1.01	5.36	0.95
IV	12 to 13:11	25	6.00	0.00	6.00	0.00	6.00	0.00
V	14 to 15:11	25	5.58	1.44	6.00	0.00	5.80	1.00
VI	16 to 20	25	6.00	0.00	6.00	0.00	6.00	0.00

N=no.of participants.

Table 4: Mean and SD of number of correctly identified syllables across different age groups and gender for VO condition.

Group	Age range(in years)	N	Mean for male	SD	Mean for female	SD	Total mean	SD
I	6 to 7;11	25	4.92	0.96	3.84	0.61	4.36	1.65
II	8 to 9;11	25	3.08	1.44	3.08	1.48	3.44	1.47
III	10 to 11;11	25	3.08	1.78	2.92	1.25	3.00	1.50
IV	12 to 13;11	25	2.07	1.04	2.50	0.90	2.28	0.97
V	14 to 15;11	25	2.58	1.93	2.31	1.49	2.44	1.68
VI	16 to 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. Maximum no. of syllables- 6.

Table 5: Mean of number of correctly identified syllables across different age groups and gender in AV+ condition

Group	Age range (in years)	N	Mean for male	Mean for female	Total mean
I	6 - 7.11	25	6	6	6
II	8 - 9.11	25	6	6	6
III	10 - 11.11	25	6	6	6
IV	12 - 13.11	25	6	6	6
V	14 - 15.11	25	6	6	6
VI	16 - 20	25	6	6	6

Note: Here ?N? represents the no. of participants,. Maximum no. of syllables- 6.

Dupont, Aubin, & Menard, 2005; Hockley & Polka, 1994; Massaro, 1984; Massaro et al., 1986; Sekiyama & Burnham, 2004; Wightman et al., 2006), in which 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 to 7;11 years reflects good visual processing. Further, the children in this group were repeatedly instructed to pay attention which might have added to the better performance. In addition, the equal responses seen for children of 8 to 15;11 years may be because of 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, whereas it was done for the younger group. Adults do not require such monitoring and hence the performance was good.

Comparison between the performance of the female and male listeners was made and the interaction between the group and gender interaction were analyzed. From the Table 4, it can be observed that between male and female participants there is not much of difference seen. Results of two way MANOVA revealed no significant differences for gender [F (1,138) = 0.175, p> 0.05].

Audiovisual Congruent Condition

In this condition response for the congruent stimuli pairs were recorded. The stimuli in auditory and the visual modality were same in terms of place of articulation and voicing in AV+ condition. 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. The results of this are given in the Table 5' should be added at this point. It is shown that all the groups performed equally and there was 100% responses. 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). Further, these were no gender differences seen in this condition.

Audiovisual Incongruent Condition

For studying the presence of McGurk effect, auditory /pa/ dubbed onto the articulatory movements for /ka/ was presented. McGurk effect was said to be present, if there was a fused response, that is, /ta/ in response to auditory /pa/ and visual /ka/ stimuli. 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 /ta/ and the total number of participants who perceived either /pa/ (auditory biased response) or /ka/ (visual biased response) were calculated. In the Table 6, the details of these responses across different groups are given.

From the Table 6, 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 from each other, except for the Group II, which has relatively poorer performance when compared to Groups III, IV, V and VI. 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 across groups. Only the pairs of groups between which there was significant difference were given in the Table 7.

Table 8: No. of male and female participants who gave fused responses across groups

Table 7 shows that there is a statistically significant difference in the performance of group II when compared to all the other groups. That is, 8 to 9;11 year olds, performed significantly poorer than that of children in the age range of 6 to7;11 years and that of 9 years and older children. Further, there was no significant difference between all the other groups (p>0.05), though the

Table 6: Fused and nonfused responses for the stimulus 'Aud /pa/ and Vis /ka/' (In each group, n=25)

Group	Total no. of participants perceived /ta/ Fused response	Total no. of participants perceived /pa/ Nonfused response	Total no. of participants perceived /ka/ Nonfused response
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 N= Total number of participants 'Aud' represents auditory mode and 'Vis' represents visual mode.

performance of the youngest group was a little poorer when compared to Group III, IV, V and VI. That is, the performance of 6 to 7.11 year olds and of children who were $\geq$ 9 years of age did not differ significantly.

This is contradicting to the results obtained in the studies (McGurk & MacDonald 1976; Rosenblum et al., 1997;, Desjardins et al., 1997; Dupont et al., 2005; Hockley & Polka, 1994; Massaro, 1984; Massaro et al., 1986; Sekiyama & Burnham, 2004; Wightman et al 2006) where, they have found poorer integration for children below 11years of age when compared to adults.

This may be because of the good VO scores by 6 to 7;11-year-olds in the visual only task. The innate differences of articulatory experience of the Indian children when compared to English speaking children may also be a reason. The support for this may be taken from the finding that the phonological structure of the language influences the audiovisual integration (Sekiyama & Burnham, 2008). It may also be that, the children taken in the present study had good phonological skills. However, this has to be tested for giving conclusive statement.

Analysis of the number of subjects who perceived /pa/ and who perceived /ka/, instead of a fused response, was also done. Table 6 provides the details of this. It is shown in the table that, only seven of 25 of the participants in Group II perceived /ta/, i.e., fused response. Most of the other participants in this group (14 in num-

Table 7: Pairwise comparison of fused responses

Stimulus Aud-Vis pa-ka Pair wise comparison					
Group	I-II	II-III	II-IV	II-V	II-VI
Z value	2.83	3.98	3.68	3.68	3.98.
P value	0.004**	0.000**	0.000**	0.000**	0.000**

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

Table 8: No. of male and female participants who gave fused responses across groups

Group	Fused response for Pa-ka stimulus: Total no. of participants given fused responses	Total no. of female who gave fused responses	Total no. of male who gave 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.

Table 9: Within group gender differences in the perception of fused response.

Stimulus pa - ka	Within group gender 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*

ber), perceived /pa/ and four of them perceived /ka/. It is also shown in the table that, in all the other groups, almost equal number of participants perceived /pa/ and /ka/.

Hence, the 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 8 to 9;11 year olds, the auditory influence was more and they had significantly poorer audiovisual integration compared to all others.

This is in agreement with the results of the study done by Erderner and Burnham (2005). They found that there was a decline in the visual speech influence at around 7 years of age. Further, at the age of 8 years their integration jumped back to that of 6-years -old and their reading skill also was found to be increased substantially.

The reason attributed was that the reading skill starts as an automatic skill at 7 years of age. Hence, visual speech perception is not needed as a back support. Because of this they performed poorly.

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 gave fused response in this group. It was noticed that these seven children were of 9 years or older. All the others (who did not give fused responses) in the group were younger (around 8.2 years) than those seven children. Hence, children at around 8 years of age showed poorer integration in contrast to (at 7 years of age) Erdener and 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 years of age, in the present study, might have just developed reading as an automatic process. However, this is just a hypothesis, which needs to be tested.

Another objective of the study was to look for gender differences. The responses of male and female participants are given in Table 8. It was observed that more number of female participants have given fused responses when compared to male participants, except for the Group I. In the Group I, none of the female subjects gave fused responses and eight of male participants gave fused response. The test of equality of proportion was done in order to compare the performance between male and female participants across different age groups. It was found that, only for group I, V and VI, there was a significant difference between male and female participants, as given in Table 9. Females performed better when compared to male participants, except for group I where male did significantly better than

Table 10: Mean and SD of the numbers of syllables perceived across age groups in the audiovisual integration of voicing cue task

Group	Age range (in years)	N	MeanFor audio	SD	Meanfor video	SD
I	6 to 7;11	25	6.00	0.00	0.00	0.00
II	8 to 9;11	25	6.00	0.00	0.00	0.00
III	10 to 11;11	25	6.00	0.00	0.00	0.00
IV	12 to 13;11	25	6.00	0.00	0.00	0.00
V	14 to 15;11	25	6.00	0.00	0.00	0.00
VI	16 to 20	25	6.00	0.00	0.00	0.00

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

females. In the other groups, there were no significant gender differences found.

In most of the studies, female subjects were found to show a stronger McGurk effect than males (Iyer, 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; Iyer, 2005).

This is in agreement with the findings of the present study for children of 14 years of age and above. However, in three of six groups, there were no gender differences. This conflicting result may be because of the methodological differences with Iyer's study, in which integration was assessed in noisy condition.

Measurement of Audiovisual Integration for Voicing

In this condition, the voiced and voiceless sounds were paired, for example, audio voiceless /pa/ was paired with the video of /ba/ and vice versa. These were a total of six such pairs. The performance was checked to find out if there is an influence of visual stimulus in the perception of voicing when an incongruent auditory stimulus is given.

It can be observed from the Table 10 that when an auditory stimulus was voiced, participants perceived a voiced stimulus and when it was voiceless, participant perceived a voiceless sound. Because there was 100% response given by all the groups and the responses were same, no statistical analysis was done. All the groups repeated the auditory stimuli and there was no influence of vision. The findings of the present study are not in consonance with the previous studies (Schwartz et al., 2004).

Conclusions

To conclude, the developmental changes of audiovisual integration of the Indian children is different from the developmental changes reported in western studies. Further, it can be stated that the assessment of phonology, reading and writing skills of the children whilst looking into the audiovisual integration will give better insight about the process by itself. This study again confirms that the visual information facilities speech perception in the congruent condition. It can also be concluded that it is important to study the female and male participants separately and, also, to have a task which is more demanding such as testing in the presence of noise.

The developmental changes of audiovisual integration of the Indian children are different from the developmental changes reported in western studies. This demands more research to include cross linguistic studies in the Indian population. Secondly this study is one of the first of its kind in Indian population and the results of the study help in providing baseline for future research. Finally, it also adds onto the literature of audiovisual integration.

Acknowledgement

I would like to thank Prof. S.R. Savithri director of All India institute of Speech and hearing, Mysore and HOD of Audiology Dr. Animesh Barman who allowed carrying out the research and also to all the participants and their parents for their cooperation.

References

Carhart. R., & Jerger. J. (1959). Preferred Method of clinical Determination of Pure Tone Threshold. *Journal of speech hearing Disorder,24*, 330-345.

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

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

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.

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.

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

Erdener, V. D., & Burnham, D. K. (2005). Development of auditory-visual speech perception in English-speaking children: The role of language-specific factors. In E. Vatikiotis-Bateson, D. Burnham & S. Fels (Eds.), *Proceedings of Auditory-Visual Speech Processing International Conference* (pp.57-62). Vancouver Island, BC, Canada.

Grant, K. W., & Seitz, P. F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *Journal of the Acoustical Society of America, 104*(4), 2438-2450.

Green, K. P., & Kuhl, P. K. (1991). Integral processing of visual place and auditory voicing information during phonetic perception. *Journal of Experimental Psychology, 17*, 278-288.

Hockley, N., & Polka, L. (1994). A developmental study of audiovisual speech perception using the McGurk paradigm. *Journal of the Acoustical Society of America, 96*, 3309.

Iyer K. (2005). *Audiovisual Speech Perception Abilities of Children with Normal Hearing in Quiet and in Noise & Comparison to Participants with Cochlear Implants.* Unpublished Master of Speech language therapy research project. University of Auckland.

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

Khan, M. K. F. Salian. T. S., Rajashekhar. B. & Dhamani. I., (2008). *Effect of Temporal Envelope and Fine Structure cues in Auditory-Visual Integration in children using Cochlear Implants.* Unpublished oral paper presented in CIGICON, Bangalore.

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

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

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

Moffitt, A. (1971). Consonant cue perception by 20 to 24 weeks old infants. *Child development, 42*, 717-731.

Rosenblum, L. D., Schmuckler, M. A., & Johnson, J. A. (1997). The McGurk effect in infants. *Perception and Psychophysics, 59*, 347-357.

Sarika, M. (1995), *Cross- model perception of speech reading and auditory mode.* Unpublished Master's Dissertation, University of Mysore, Mysore.

Schwartz, J. L., Berthommier, F., & Savariaux, C., (2004). Seeing to hear better: Evidence for early audiovisual interactions in speech identification. *Cognition, 93*, 69-78.

Sekiyama, K., & Burnham, D. (2004). Issues in the development of auditory-visual speech perception: Adults, infants, and children. *Interspeech, 21*, 1137-1140.

Shivprakash, S. (2003). *Gap Detection Test - development of Norm.* Unpublished Master's dissertation. University of Mysore, Mysore.

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

Taylor, H. R. (1978). " Applying New Design Principles to the Construction of an Illiterate E Chart." *American Journal of Optometry and Physiological Optics, 55*, 348.

Wasim, A., Chavan. A. R., Deema, K. P. A., & Dhamani, I. (2010). *Auditory visual integration and dyslexia and poor academic performers.* An unpublished Poster paper presentation in ISHACON.

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.

Yathiraj, A., & Mascarenhas. K. (2002). Audiological profile of the children with suspected auditory processing disorder. Developed as a part of project titled *'Effect of auditory stimulation of central auditory processes in children with CAPD'* at the department of audiology. All India Institute of Speech and Hearing, Mysore.

Yathiraj, A., & Mascarenhas, K. (2004). Auditory profile of children with suspected auditory processing disorder. *Journal of Indian Speech and Hearing 18*, 6-14.

Yathiraj, A., & Vijayalakshmi, C. S. (2005). Phonemically Balanced word list in Kannada, Developed in Department of Audiology, AIISH, Mysore.