INFLUENCE OF AUDITORY CLOSURE AND WORKING MEMORY ON AUDIO-VISUAL PERCEPTION OF SPEECH

Deepashree S R

Registration No: 10AUD007

A Dissertation Submitted in Part Fulfillment for the Degree of

Master of Science (Audiology)

University of Mysore, Mysore

ALL INDIA INSTITUTE OF SPEECH AND HEARING

MANASAGANGOTHRI, MYSORE – 570006

May 2012

CERTIFICATE

This is to certify that this dissertation entitled **'Influence of Auditory Closure and Working Memory on Audio-Visual Perception of Speech'** is the bonafide work submitted in part fulfillment for the Degree of Master of Science (Audiology) of the student with Registration No. 10AUD007. This has been carried out under the guidance of a faculty of this institute and has not been submitted earlier to any other University for the award of any other 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 the dissertation entitled **'Influence of Auditory Closure and Working Memory on Audio-Visual Perception of Speech'** has been prepared under my supervision and guidance. It is also certified that this has not been submitted earlier in any other university for the award of any Diploma or Degree.

Mysore
May, 2012

Dr. Sandeep M.
Guide
Lecturer in Audiology
Department of Audiology
All India Institute of Speech and Hearing
Manasagangothri, Mysore - 570006

DECLARATION

This is to certify that this Master's dissertation entitled **'Influence of Auditory Closure and Working Memory on Audio-Visual Perception of Speech'** is the result of my own study under the guidance of Dr. Sandeep M., Lecturer in Audiology, Department of Audiology, All India Institute of Speech and Hearing, Mysore, and has not been submitted in any other University for the award of any Diploma or Degree.

Mysore Registration No. **10AUD007**

May, 2012

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Dedicated to

My parents

(Veena and Ramachandra)

ACKNOWLEDGEMENTS

I am grateful to my Lord, Swami Vivekananda for being with me always and in all ways. Without his blessings I can never ever think of achieving anything in my life.

I would like to express my sincere thanks to my guide Dr. Sandeep M., for his constant guidance and support. Sir, the way you handle multiple tasks and the way you understand your students' needs (before it is expressed) and guide them is amazing. You have let me to explore and learn. I have learnt a lot from you sir…

I would like to thank Dr. S. R. Savithri, Director, All India Institute of Speech and Hearing, for permitting me to carry out this research.

I am grateful to my Parents and my 'Jenugudu' family (Nagendra doddappa, Radha doddamma, Veda chikkappa, Rama chikkamma, Seethamani atte, Nagashree & Jayaram, Nataraj & Pallavi, Ramyashree & Dilip, Padmashre & Vasuki, Roopashree, Anushree, Sunayana, Raghunandan, Paartha, Pranathi and Pranav) for their support. I am very lucky to have such a wonderful family.

I am thankful to Megha ma'am and Sharath sir for their precious assistance during stimulus preparation in spite of their busy schedule.

Thanks to Priya ma'am and Sandeep sir for their support and for providing me laptop, which helped me, a lot during data collection and analysis.

Thanks to Nike and Santosh sir (statistician) for their timely help during dissertation.

A special thanks to Paddu mava, Savithri atte, Ajji and Sheshu for helping me during my hunt for subjects.

I would like to thank all my participants for their co-operation, especially Rukmini aunty and Asha aunty.

I would like to thank to all my teachers from my Kindergarten to Post-graduation for everything they have taught me.

I would like to thank my Music teacher Sridhar sir, because of whom I could join 'Purple fusion' group of AIISH.

I am grateful to Dr. S. R. Savithri ma'am and Ajiish sir for providing an opportunity to sing in front of Bharatha Ratna Dr. A. P. J Abdul Kalam, which was one of the most exciting moments of my life.

I would like to thank library faculty for their guidance.

I can never forget my mischievous 'Buddies' group (Sahana, Arpitha, Mythri, Deepthi, Priyanka & Raheela) for all their support and for making me smile in this hectic schedule. In these six years you people made me to have fun, each and every moment.

I have to remember my singing partner Apoorva, shuttle partner Shabnam, coffee partner Mahima and my special friend Suma Chatni. Thank you girls….

I am extremely thankful to my Incredible Classmates for all the memories they have given me.

I would like to thank my seniors and juniors for several memorable moments which I had at AIISH.

Chapter 1

INTRODUCTION

Speech perception is influenced by visual cues even though it is primarily an auditory process. In instances where auditory cues are compromised (such as in noisy environments or hearing impairment), visual input has been reported to significantly improve speech intelligibility by supplementing the missing auditory cues (Tye-Murray, Sommers & Spehar, 2007; Munhall, Kroos, Jozan & Vatikiotis-Bateson., 2004; MacLeod & Summerfield, 1987). The benefits provided by bimodal presentation of a speech signal are larger when auditory stimuli are degraded than when the speech signal is clear (Sumby & Pollack, 1954; O'Neil, 1954; Neely, 1956; Erber, 1969; Grant & Seitz, 2000; Rudmann, Mc Carley & Kramer, 2003; Bernstein, Auer & Takayanagi, 2004; Ross, Saint-Amour, Leavitt, Javitt & Foxe, 2007). But studies on relationship between the amount of redundancy and audio-visual (AV) speech perception show equivocal results (Sumby & Pollack, 1954; Erber, 1969; Summerfield, 1979; Middelweerd & Plomp, 1984; Shannon, Zeng & Wygonski, 1998; Anderson, 2006; Huffman, 2007).

Sumby and Pollack (1954) demonstrated that with increase in difficulty of auditory-only perception, the benefits obtained by combining the auditory and visual speech information also increased. Erber (1969) reported an improvement of 60% in word recognition scores from auditory-only condition to AV condition at -10 dB SNR for young adults. Similar results were reported using sentence materials (Middelweerd & Plomp, 1984; Summerfield, 1979). Whereas Anderson (2006) and Huffman (2007) reported that, the amount of AV integration did not vary across different auditory signal manipulations and hence, systematically removing information from the auditory stimulus does not necessarily affect the degree of integration benefit.

The studies on effects of age are equivocal. The study by Tye-Murray, Sommers and Spehar (2007) compared the dependency on visual cues between older adults with and without hearing loss and suggested that older adults with hearing loss may rely much more on the visual cues than those with normal hearing. The degree of hearing loss was also shown to affect the integration of auditory and visual syllables as reported by Grant, Walden and Seitz (1998). Further, the older adults are reported to be less successful in combining information across two or more sensory modalities (Shoop & Binnie,1979; Middelweerd & Plomp, 1984; Plude & Hoyer, 1985). But in contrast, Spehar, Tye-Murray and Sommers (2008) suggested that the inter-modal and intra-modal integration abilities are largely resistant to changes with age and hearing loss. The researchers have argued that age related changes in cognitive or central auditory processing abilities play a limited role in the poor recognition of speech (Sommers, 1997; Schneider, Danema, & Pichora-Fuller, 2002).

In the cognitive domain, Pichora-Fuller, Schreider and Daneman (1995) explained that speech understanding needs a constant encoding of information into and out of the working memory and loss of working memory may contribute to the age-related declines in the comprehension of speech which may be evident even in favorable listening conditions.

1.1 Justification for the Study

Earlier studies (Middelweerd & Plomp, 1984; Anderson, 2006) showed that dependency on the visual information increase with the decrease in the signal to noise ratio of the auditory signal. This means that the subjects depended on the information in the other modalities when the external redundancy is cut down through the auditory modality. Individuals with auditory processing disorders (APD) have deficits in auditory closure (ASHA, 2005) and show poor speech perception in noise. This is due to their reduced internal redundancy. Such individuals logically should be depending on visual information in everyday listening conditions lot more than persons without APD. However, the effect of reduced internal redundancy on AV perception of speech has not been studied experimentally.

Earlier studies (Cienkowski & Carney, 2004; Spehar & Tye-Murray, 2008) have also shown that advancing age does not degrade the audiovisual integration. Considering that auditory processing deficits are more prevalent in elderly individuals (Roberts & Lister, 2004; Martin & Jerger, 2005) these findings may partly support that APD does not degrade audiovisual integration. However, this will only be an inference and needs experimental investigation before concluding.

Hence, the present study was taken up to study effect of poor speech perception in noise on audio-visual integration. It was hypothesized that if selected individuals have normal or near-normal visual and auditory sensory systems, it is likely that their information extraction capabilities would be good. However, if the central processing mechanisms are affected, the integration performance may be suspected to be poorer than that otherwise. The present findings also throw light on the differential effects of internal and external redundancy, if any on audio-visual perception.

Because perception involves cognition, it was expected that reduced cognitive abilities, more specifically the working memory, could degrade the process of audio-visual integration. But the relation between working memory and audio-visual speech perception has not been studied earlier. Hence, the present study also aimed to study the relation between the working memory and audio-visual integration.

1.2 Objectives of the Study

The present study had the following two specific objectives:

1) To explore the relation between the speech perception in noise and audiovisual integration in speech perception.
2) To explore the relation between the working memory and audio-visual integration in speech perception.

Chapter 2

REVIEW OF LITERATURE

Speech perception is dependent on both auditory and visual processes, though the major contribution is from the auditory modality. Through the auditory modality, place of articulation, manner of articulation and voicing information about a speech sound will be obtained whereas visual modality provides information only about place of articulation (Anderson, 2006).

Speech perception through auditory mode alone can be impaired due to reduction of external redundancy such as in noisy environments or due to reduction of internal redundancy, as in case of hearing impairment or central auditory processing disorder (Walden, Busacco & Montgomery, 1993; Anderson, 2006**)**. In such instances where auditory cues are distorted, visual input has been reported to significantly improve speech intelligibility by supplementing the missing cues (Tye-Murray, Sommers & Spehar, 2007; Munhall, Kroos, Jozan & Vatikiotis-Bateson, 2004; MacLeod & Summerfield, 1987).

2.1 Effect of External Redundancy on Audio-visual Speech Perception

Evidence from previous research explains that speech intelligibility improves when listeners receive information from both auditory and visual modes in instances where the auditory cues are degraded by reducing external redundancy (Sumby & Pollack, 1954; Anderson, 2006; Huffman, 2007). But there are equivocal results about the effect of amount of redundancy on audio-visual (AV) speech perception.

It has been explained by earlier studies that the benefits provided by bimodal presentation of a speech signal are larger when auditory stimuli are degraded than otherwise (Sumby & Pollack, 1954; O'Neil, 1954; Neely, 1956; Erber, 1969; Grant & Seitz, 2000; Rudmann, Mc Carley & Kramer, 2003; Bernstein, Auer & Takayanagi, 2004; Ross, Saint-Amour, Leavitt, Javitt & Foxe, 2007).

In the study by Sumby and Pollack (1954) it was demonstrated that the addition of visual cues improved speech perception by an amount equivalent to a 5 to 18 dB increase in the signal-to-noise ratio (SNR) which accounts for improvement of up to 60% in word recognition. They also demonstrated that with increase in difficulty of auditory-only perception, the benefits obtained by combining the auditory and visual speech information also increased. Erber (1969) also reported an improvement of 60% in word recognition scores from auditory-only condition to AV condition at -10 dB SNR for young adults. Similar results for sentence materials were obtained by Middelweerd and Plomp (1984), and Summerfield (1979).

Anderson (2006) examined the amount of redundancy necessary for optimal AV integration in young adults using different degraded auditory cues. He reduced the speech signals to four spectral bands, effectively reducing the redundancy of the auditory signal. The performance of participants was explored under four conditions; 1) degraded auditory only, 2) visual only, 3) degraded auditory+visual, and 4) non-degraded auditory+visual. Results of this study indicated that degrading method effectively reduced redundancy in speech signals and also appeared to have an effect on listeners' reliance on different modality inputs. Listeners relied more on visual stimulus when auditory stimulus was degraded and they achieved higher performance in auditory+visual mode rather than degraded auditory alone or visual alone condition. The overall percentage of responses that reflect integration was very similar for normal and degraded auditory conditions, suggesting that the degree of integration is not influenced by the amount of redundancy in the auditory signal. He also explained that the waveform for speech is highly redundant as it contains far more information than is necessary for identification of a speech sound and hence, even a relatively small amount of temporal and spectral information is useful for the identification of speech sounds. This supports the study by Shannon, Zeng, and Wygonski (1998), which pointed out that even when spectral cues are absent in a stimulus, the phonemes can be perceived with high accuracy as consonant recognition was consistently less sensitive to the distortions than vowel recognition.

Huffman (2007) conducted a study on young adults to explore whether systematically removing information from the auditory stimulus results in greater usage of the visual input and thus greater integration. In this study, the CVC syllables were degraded by selectively removing fine-structure of spectrum but the temporal envelope characteristics of the waveform was maintained by obtaining output through 2, 4, 6, and 8-channel bandpass filters. Results of this study indicated that with increase in the number of channels, auditory-only performance of the listeners increased systematically as more number of channels provide more spectral information (less ambiguity is present). However, the amount of auditory-visual integration did not vary across different auditory signal manipulations even though there was auditory-visual integration for all conditions. This suggests that systematically removing information from the auditory stimulus does not necessarily affect the degree of integration benefit.

2.2 Effect of Internal Redundancy on Audio-visual Speech Perception

Speech perception through auditory mode can be affected because of changes in peripheral and/or central auditory system even when the external auditory stimulus has high fidelity. The changes in peripheral auditory system include hearing loss and changes in central auditory system include auditory processing deficits or cognitive deficits. There are equivocal results about the effect of hearing loss on AV speech perception.

Tye-Murray, Sommers and Spehar (2007) compared the dependency on visual cues between older adults with and without hearing loss. The results suggested that older adults with hearing loss may rely much more on the visual cues than those with normal hearing. They also imply that, when auditory stimulus is distorted, even subjects with normal hearing may find bimodal presentation useful. This is consistent with general finding that as the auditory SNR declines, the benefit of visual cue increases. The degree of hearing loss is also shown to reduce the integration of auditory and visual syllables in adults with hearing-impairment (Grant, Walden & Seitz., 1998).

Spehar, Tye-Murray and Sommers (2007) studied inter and intra-modal integration in three groups of participants: young adults with normal hearing, older adults with normal hearing, and older adults with mild to moderate hearing loss. Comparisons between the two groups with normal hearing assessed the effects of age, independent of hearing loss while the comparisons between the two elderly groups assessed the effects of hearing loss. Intra-modal integration ability was assessed using both monotic and dichotic presentations of bandpass filtered sentences (presentation level- 50dBSPL for subjects with normal hearing & 70dBSPL for subjects with hearing impairment). Inter-modal integration ability was assessed using auditory only, visual only and audio-visual presentations of Iowa sentences (presentation level- 60dBSPL for subjects with normal hearing and 80dBSPL for subjects with hearing impairment). Results suggested that both types of integration ability are largely resistant to changes with age and hearing loss. In addition, intra and inter-modal integration were not correlated. They reported that this lack of correlation could be due to absence of common mechanism for both intra and inter-modal integration. That is, there are distinct or separate processes for the integration of audiovisual information and the fusion of spectral information within the hearing system.

In addition to factors like hearing loss, speech reading ability may also influence the AV perception. Brault, Gilbert, Lansing, McCarley and Kramer (2010) tested whether visual cues and expanded auditory bandwidth improves speech intelligibility in older adults with normal hearing and those with mild – moderate hearing loss. Results showed that in quiet listening conditions, bimodal presentation improved the intelligibility of spoken words only for participants with hearing loss and good lip reading proficiency even though bimodal presentation improved speech intelligibility for all participants in noise. The effect of stimulus bandwidth was observed only in the presence of noise. This suggests that bimodal presentation may improve speech intelligibility only when auditory information is degraded either by hearing loss or presence of noise.

2.3 Effect of Age on Audio-visual Speech Perception

In the literature it is shown that, the effect of aging on the benefit from AV speech perception could be attributed to the age-related differences in auditory sensitivity, speech reading, integrating auditory and visual speech information, or some combination of these factors. The efforts to study the effect of aging on the ability to integrate auditory and visual speech information is complicated by declines in both audio and visual speech perception as a function of age (Shoop & Binnie, 1979;

Middelweerd & Plomp, 1987; CHABA, 1988; Walden, Busacco, and Montgomery, 1993; Cienkowski, 1999; Cienkowski & Carney, 2002, 2004;).

In study by Shoop and Binnie (1979), the visual recognition abilities of middle-aged and elderly adults with normal hearing was compared using the CID Everyday Sentences. They found that as age increased, percent correct identification of key words in the visual only condition decreased. From this finding it can be indirectly inferred that the benefit of AV perception shall reduce with aging. Middelweerd and Plomp (1984) compared the SNR at which 50% recognition of Dutch sentences was achieved in auditory-only and in AV conditions for young adults and elderly. The addition of visual information resulted in a 4.6 dB better threshold for a group of young subjects and 4.0 dB better threshold for a group of elderly subjects when compared to auditory presentation alone. Hence, older adults seem to have a slight disadvantage for using the visual information.

Plude and Hoyer (1985) suggested that the older adults are less successful at integrating information than younger adults. Based on their findings, Cienkowski and Carney (2002) suggested a conflicting auditory and visual speech stimulus to be a better option for testing integration performance of older adults. They also explained that if older adults with normal or near-normal visual and auditory sensory systems were selected, their information extraction capabilities could be similar to those of younger adults. In spite of it, if the integration performance of the older adults were poorer than that of young adults, then changes in central processing mechanisms may be suspected. Plude and Doussard-Roosevelt (1989) also reported that older adults are less successful in combining information across two or more sensory modalities.

Walden, Busacco, and Montgomery (1993) examined the benefit derived from visual cues in AV speech recognition in middle-aged and elderly men with moderate to severe hearing loss, using consonant-vowels (CVs) and sentences. The testing was conducted in the presence of speech shaped noise only for sentences and the noise levels were adjusted to obtain approximately 40 to 50% correct scores in the auditory condition. The results showed that visual enhancement (the difference between the AV and A conditions) did not differ between the two groups for sentences. But the elderly subjects performed poorer than middle-aged in the visual-only condition which prevented comparing the groups under conditions of equivalent unimodal performance. Also the scores in the AV condition for sentences were near ceiling for both the groups resulting in similar AV benefit and hence null effects of age was not interpreted properly. The patterns of consonant confusions were similar for the two groups.

Grant, Walden and Seitz (1998) reported that individuals who are less successful in integrating unimodal (auditory or visual) information, such as older adults, had poor bimodal perception of speech.

Cienkowski (1999) reported that older adults were not consistently successful at integrating information across sensory modalities. Results from this investigation of AV integration of speech stimuli have been inconclusive as the hearing loss was not well controlled. Cienkowski and Carney (2002, 2004) tested whether older adults are as successful as younger adults at integrating auditory and visual information for speech perception and whether integration is related to speech reading performance. There were three groups; young adults with normal hearing and vision, older adults with normal and near normal hearing and vision, and young controls, whose hearing thresholds were shifted with noise to match the older adults. Auditory and AV identification of syllables with conflicting auditory and visual cues and a lip reading test was completed by each participant. The results of this study showed that there were no significant differences between the groups for integration at the syllable level, but there were differences in the response alternatives chosen. Young adults with normal hearing chose an auditory alternative whereas, older adults and controls chose visual alternatives. In addition, older adults demonstrated poorer lip reading performance than their younger counterparts and this was not related to successful integration at the syllable level. They concluded that when auditory and visual integration of speech fails to occur, the participants select an alternative response from the modality with the least ambiguous signal.

Sommers, Tye-Murray and Spehar (2007) examined the effects of aging on an individual's ability to benefit from combined auditory and visual speech information, when compared to listening or speech reading alone. They compared young and old adults with normal hearing for identification of vowel-consonant-vowels, words and sentences in auditory-only, visual-only, and AV conditions. The stimuli were presented in the presence of 20-talker babble and for each participant and stimulus type, the signal-to-babble ratios were set individually to produce approximately 50% correct in the A condition. Along with this, they also compared visual enhancement (VE) and auditory enhancement (AE) for all stimuli in both the groups. Results of the study showed that older and younger adults obtained similar scores for the auditory condition for all three types of stimuli, indicating that the procedure for individually adjusting signal-to-babble ratios was successful at equating auditory scores for the two age groups. However, older adults had significantly poorer performance in the AV and visual modalities than younger adults. But when age differences in the visual condition was controlled, there were no age differences in the ability to benefit from combining auditory and visual speech signals as indicated by AE and VE. The findings of the above study suggest that the poorer performance of older adults in AV condition was a consequence of reduced speech reading abilities than due to impaired integration capacities.

Overall, the literature shows that elderly individuals are less successful in integrating auditory and visual cues during AV perception. The poor performance has been attributed to either decline in the visual modality or central auditory processing mechanism. Although most of the studies showed that speech reading performance decreases with age, their results were affected either due to hearing loss or due to visual problem. Even when visual acuity was controlled in some studies, the relationship of speech reading performance and integration of auditory and visual cues for speech perception is not as well understood.

2.4 Effect of Cognitive Aging on Audio-visual Speech Perception

Previous studies (Grant et al., 1998; Tye-Murray, Sommers and Spehar, 2007; Brault, Gilbert, Lansing, McCarley and Kramer, 2010;) provide evidence for loss in peripheral hearing sensitivity to be the cause for many of the listening problems of elderly persons. Although degree of audibility strongly influences speech comprehension, some older adults have more difficulty than would be expected based solely upon their audiometric configurations. When observed closely, their common complaints suggests that at least some portion of speech-understanding difficulties of older adults can be attributed to age-related declines in cognitive abilities, changes in higher-order auditory processes, or a combination of the two.

It is shown in one of the earlier studies that older adults have decreased processing resources available to them and hence perform poorer than young adults on memory tasks (Craik & Byrd, 1982). It is suggested that there is an interdependence of aging sensory systems and cognitive functions (Li & Lindenberger, 2002). The construct of working memory is considered important in the study of cognitive aging even though there are other measures of processing resources such as perceptual speed and reasoning. Studies by Park and colleagues (1996, 2002) measured the working memory capacity in adults across life span and it was found that it was greatest in young adulthood and then decreases across life span. Martin and James (2005) proposed that age-related deficits in processing of inter-hemispheric information may bring about some of the listening problems in older adults. Hence, it is suspected that age-related changes in working memory may provide basis for decreased age-related performance on a range of cognitive tasks.

It has been explained that speech understanding needs a constant encoding of information into and out of the working memory and manipulation of the information stored in memory (Pichora-Fuller, Schreider & Daneman, 1995). Hence, loss of working memory hinders speech communication and may contribute to the age-related declines in the comprehension of speech which may be evident even in favorable listening conditions. In adverse listening conditions such as in the presence of noise, this may degrade further (Cohen, 1979, 1981; Light et al., 1982; Pichora-Fuller, Schreider & Daneman,1995).

Brault, Gilbert, Lansing, McCarley and Kramer (2010), tested whether bimodal (AV) speech presentation and expanded auditory bandwidth improves working memory performance for older adults by reducing the cognitive efforts required for speech perception. In the results it was seen that the expanded bandwidth and bimodal presentation may not always improve working memory performance even though it improves speech perception in difficult listening conditions.

Some of the earlier investigators have argued that age related changes in cognitive or central auditory processing abilities play a limited role in the poor recognition of speech (Sommers, 1997; Schneider, Danema, & Pichora-Fuller, 2002). Sommers (1997) explained that most of the attempts to correlate age-related changes in cognitive function with basic measures of speech perception had only limited success and offered several reasons for the lack of significant correlations. First, the speech-specific cognitive deficits contributing to speech-understanding difficulties may not be reflected adequately by the more commonly administered cognitive measures. Second, the listening conditions and stimuli commonly used in experiments may be less demanding than those typically encountered in natural communication environments.

Schneider et al. (2002) explained that age-related declines in cognitive functions are highly correlated with concomitant changes in peripheral sensitivity and hence the age related changes in speech understanding are a consequence of auditory declines.

Thus, from the above review, it is clear that the dependency on the visual information increase with degradation of the auditory signal even though there are mixed opinions about the effect of the amount of redundancy on integration. This means that subjects depend on the information in other modalities when external redundancy of the auditory modality is cut down. For participants with hearing loss and processing disorder, there are equivocal studies about the effect of internal redundancy on audio-visual speech perception. The studies on aging effects have been inconclusive due to poor control of the variables like hearing loss and visual problem. Further, there was no systematic study to understand the relationship between working memory and AV speech perception. Hence, the present study was taken up to scientifically study the relationship among working memory, speech perception in noise and AV speech perception.

Chapter 3

METHOD

In the present study, quasi experimental research design was used to test the null hypothesis that there is no effect of speech perception in noise and working memory on audio-visual speech perception. The following method was used to test the hypothesis.

3.1 Participants

Sixty six normal hearing adults in the age range of 18 to 70 years participated in the study. The selected participants had pure tone thresholds within 15 dBHL at octave frequencies between 250 Hz and 8 kHz (ANSI, 1996), and normal or corrected vision of 6/6. All the selected participants were native speakers of Kannada.

All the participants were assessed for their speech perception in noise and working memory (details available later in this chapter). Based on their performance in speech in noise and working memory tests, they were divided into 4 groups. The groups were divided by considering the 95% confidence intervals of speech in noise scores and the working memory scores of the 66 participants. The scores less than lower boundary were grouped as 'poor' and the scores higher than the upper boundary were grouped as 'good'. As a result, following were the groups formed.

Group I, named as Low speech in noise (LowSPIN) group had 20 participants with poor speech perception in noise i.e., speech identification scores of less than or equal to 72% (lower boundary score of confidence interval) at 0 dB SNR.

Group II, named as High speech in noise (HighSPIN) group had 36 participants with good speech perception in noise i.e., speech identification scores of more than 80% (upper boundary score of confidence interval) at 0 dB SNR.

Group III, named as Low working memory (LowWM) group had 14 participants with poor working memory i.e., working memory scores of less than or equal to 69% (lower boundary score of confidence interval).

Group IV, named as High working memory (HighWM) group had 21 participants with good working memory i.e., working memory scores of more than 76% (upper boundary score of confidence interval).

A written consent was obtained from each participant prior to their inclusion in the study.

3.2 Test Stimulus

Six lists of bi-syllabic Kannada (Dravidian language spoken in the state of Karnataka, India) words having ten words in each list were used as test stimuli. These word lists were developed specifically for the purpose of the present study.

3.2.1 Selection of the words

To begin with, 300 bi-syllabic Kannada words, frequently spoken by native speakers were collected from recorded speech samples, news papers and media interviews. From this list, 23 words were omitted as they had two or more clusters and were felt to be difficult to identify visually. This left another 277 words in the list.

The words were then given to 15 native speakers of Kannada to rate them according to familiarity on a 3-point scale (unfamiliar, familiar, & very familiar). The participants selected unfamiliar- if they were unaware of the word, familiar - if they were aware of the word but it did not occur frequently in conversation, very familiar - if the word occurred very frequently in conversation. Out of these, only the words which were rated 'very familiar' by more than 12 participants (80%) were considered for the next level. One hundred and sixty nine words were rated very familiar out of the 277 words.

3.2.2 Audio recording

The selected 169 words were audio recorded using a computer with adobe audition (version 1.5) software at a sampling frequency of 44,100 Hz and 16 bit digitization in an acoustically treated room. Audio recording was done using a unidirectional microphone connected to the computer. An adult female native speaker of Kannada, who was a professional voice user, uttered the words. Each word was uttered three times and out of the three samples, the one with best acoustic fidelity was selected. The stimuli were further edited for removal of noise and, hiss reduction. A gap of three seconds was introduced between consecutive words, using the same software. Root Mean Square (RMS) normalization was done for all words in order to minimize differences in the stimulus amplitude.

3.2.3 Increasing homogeneity with respect to audibility

The variability with respect to audibility was reduced and inturn the homogeneity across spondaic words was increased using the standard procedure (Hirsh, Reynolds, and Joseph, 1954). Ten speech and hearing professionals who had Kannada as their native language and with a minimum of three years training were selected for this procedure.

To begin with, speech recognition threshold (SRT) was found out for all ten subjects using the procedure by Tillman and Olsen (1973), a descending method for SRT measurement. Based on this procedure, threshold was found out for paired-words in 10dB steps, and 10 dB above this was considered as 'starting level'. From the starting level, the intensity was decreased in 2 dB steps until five of the last 6 responses were incorrect. The SRT was obtained by subtracting the number of correct responses after the threshold testing was begun at the starting level from the starting level of the threshold testing and adding one.

SRT= starting level - no. of correct responses after starting level + 1

After obtaining SRT, the paired-words were presented at +4, +2, 0, -2, -4 and -6 dBSL (ref: SRT), and the participants were asked to orally repeat the words. The whole list of 169 words was presented at each sensation level (SL) and it was randomized during each presentation. The responses obtained from the ten listeners were analyzed to identify 'Easy' and 'Hard' words. The words which were missed once or less by all listeners when the list was presented at +4, +2, 0, -2, -4 and -6 dBSL were considered as 'Easy words'. Whereas the words which were missed five or more times by all listeners when presented at +4, +2, 0, -2, -4 and -6 dBSL were considered as 'Hard words' (based on procedure by Hirsh et al., 1952). All the 'Easy' and 'Hard' words were eliminated to increase the homogeneity with respect to audibility. This finally resulted in sixty bisyllabic Kannada words.

3.2.4 Calculation of frequency of occurrence speech sounds

The frequency of occurrence of all the speech sounds in the list of sixty bisyllabic Kannada words were calculated in order to find out whether the list of bisyllabic words was phonetically balanced. Although this was not mandatory for testing the objectives of the current study, the investigator felt that a phonetically balanced list would have been an advantage while generalizing the results of the present study. Hence, the frequency of occurrence of sounds in the list was tested using the method used by Ramakrishna, Nair, Chiplunkar, Ramachandran and Subramanian (1962).

Based on the above method, all the speech sounds of the word list were tabulated with vowels written horizontally in top row and consonants written vertically in the first column as shown in the Table 3.1. The different speech sounds in the list were tabulated as follows. The second row and second column immediately following the vowels and consonants were left without any designation. The speech sounds were represented in the table by means of dots placed in the cell of respective speech sound appropriately. Each word was analyzed syllable by syllable. Each syllable having vowel and consonant was located by a dot in the cell corresponding to the intersection of the row headed by consonant and column headed by vowel. For example, the word /ka:ge/ was analyzed as follows: The first syllable 'ka' was denoted by a dot in the cell corresponding to the intersection of the row headed by consonant 'k' and the column headed by vowel 'a'. The second syllable 'ge' was denoted by a dot in the cell corresponding to the intersection of the row headed by consonant 'g' and the column headed by vowel 'e' (Table 3.1).

If vowels and consonants occurred by themselves in a word, they were represented by dots in undesignated (second) row and column respectively. For example, in the word /akka/, the vowel 'a' was represented by a dot in the cell immediately below 'a' without the consonant heading and the consonant 'k' was represented by a dot in the cell immediately to the right of 'k' without the vowel heading. The final syllable 'ka' was denoted by a dot in the cell corresponding to the intersection of the row headed by consonant ‘k’ and the column headed by vowel ‘a’ (Table 3.1).

Table 3.1: *Representative scheme for determining the relative frequencies of occurrence of different speech sounds in the sixty word list. The example shown is for the words /kage/ and /akka/*

V / **C**		**A**	**a:**	**i**	**i:**	**u**	**u:**	**E**	**e:**	**ai**	**o**	**o:**	**Ou**	**am**	**aha**
		●													
k	●	●	●												
k^h															
g								●							
g^h															

All the words in list were analyzed similarly and denoted in the form of dots. So, each dot in a cell represented a syllable (two speech sounds), except the dots in second row and column which represented only one sound. The number of occurrence of any particular vowel was the total number of dots in the column of cells headed by that vowel. Likewise, number of occurrence of any particular consonant was the total number of dots in the row of cells headed by that consonant. The total number of speech sounds in the list was equal to sum of number of dots in the second row and column and twice the number of dots in the remaining cells.

The probability of any speech sound was obtained by taking ratio of number of occurrence of speech sound in the list to the total number of speech sounds in it. For example, the number of occurrence of speech sound ‘a’ was thirty two and the total number of speech sounds was two hundred and ninety seven. So the probability of ‘a’ was 0.1879 and the relative frequency of vowel ‘a’ was 18.79%. The relative frequencies for all speech sounds in the bisyllabic word list were calculated in the same way. The relative frequencies of speech sounds obtained in the present study were then compared with relative frequencies of speech sounds obtained by Ramakrishna et al., 1962 (Table 3.2). Both the relative frequencies were similar and hence the present list was phonetically balanced.

Table 3.2: *Relative frequencies of speech sounds in the present study (column B & D) and the relative frequencies of speech sounds obtained by Ramakrishna et al., 1962 (column A & C) in percentages*

Speech sounds	**Col A Frequency (%)**	**Col B Frequency (%)**	**Speech sounds**	**Col C Frequency (%)**	**Col D Frequency (%)**
a	18.79	10.7	ṭ	0.95	1.68
a:	4.74	4.7	ṭʰ	0.01	0.00
i	7.30	6.06	ḍ	1.26	1.34
i	0.70	1.01	ḍʰ	0.01	0.00
u	6.28	6.39	ṇ	0.65	0.67
u:	0.84	1.01	t̪	4.50	2.69
e	4.00	5.72	t̪ʰ	0.21	0.00
e:	1.27	1.00	ð	5.20	2.69
ai	0.15	0.00	ðʰ	0.43	0.00
o	1.06	1.34	n	4.90	2.69
o:	0.62	2.02	p	1.29	2.35
ou	0.04	0.00	pʰ	0.05	0.00
am	2.55	1.68	b	1.20	3.36
ah	0.02	0.00	bʰ	0.62	0.00
k	2.66	3.76	m	2.00	2.02
kʰ	0.08	0.00	j	2.47	1.68
g	3.22	2.69	r	5.47	4.04
gʰ	0.01	0.00	l	3.12	3.03
ŋ	0.00	0.00	v	3.85	1.68
ʧ	0.66	0.67	ʃ	0.66	0.67
ʧʰ	0.00	0.00	ʃʰ	0.52	0.00
ʤ	0.25	1.01	s	2.21	3.03
ʤʰ	0.00	0.00	h	1.31	2.02
ɳ	0.00	0.00	ḷ	1.70	1.01

Finally, from the list of sixty words, six lists having ten words in each list were obtained. The total number of vowels and consonants in the list which were obtained from the Ramakrishna et al.'s method were distributed equally across the six lists. Thus, it was attempted to keep similar frequency of occurrence of vowels and consonants in all the six lists of bisyllabic words were made similar. The six lists of 10 bisyllabic words each used in the present study are given in Appendix I.

3.2.5 Audio and video recording

Five adult, female, native speakers of Kannada, who were speech and hearing professionals with clear articulation, were selected and the video of final six lists were recorded by a professional videographer using a digital video camera. A white screen was used as a background and the participants were instructed to avoid bright clothing to avoid distractions in the visual stimuli. The participants were also instructed to produce the words clearly without exaggerating the articulators, reduce eye blinks and avoid head movements, while recording. The words which were articulated unclearly were recorded twice. Few frames of a representative video depicting the articulation of word /jeeva/, clarity of video recording and the background is shown in Figure 3.1.

Simultaneously, along with the video recording, the audio of the speech stimuli was recorded digitally for all the five participants. The audio recording was done using a collar microphone, at a sampling frequency of 44,100 Hz and 16 bit digitization, using Praat Software (version 5.1.31). Out of the 5 samples (recordings of 5 individuals) of stimuli, best sample was selected based on clarity of stimuli visually as well as auditorily. The audio and video recordings were done in a sound treated room.

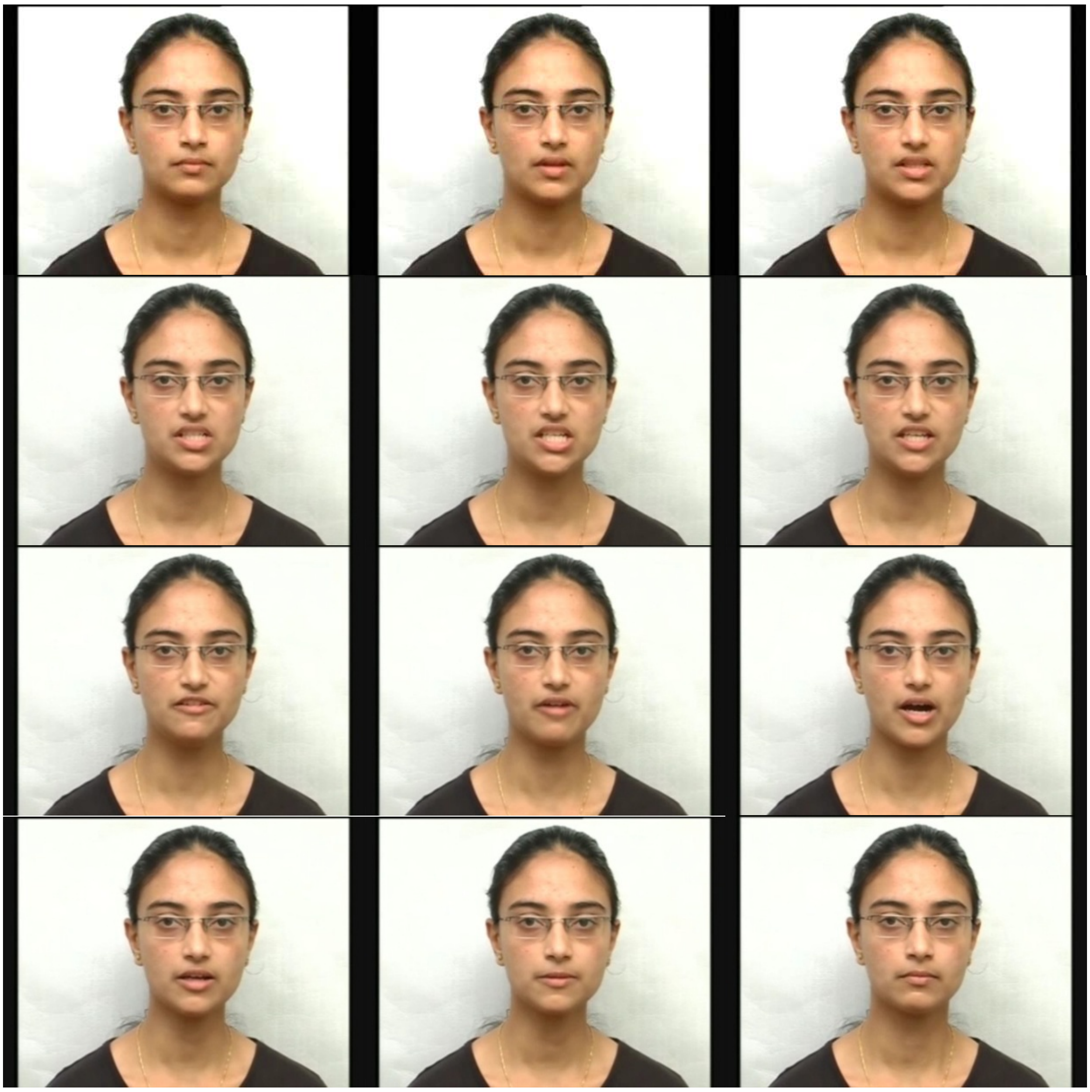

Figure 3.1: Few frames of a representative video depicting the articulation of word /jeeva/, clarity of video recording, and the background.

The auditory stimuli were edited using adobe audition (version 1.5) software for noise and hiss reduction and a gap of five seconds was introduced between the words. All the words were normalized to a constant scaling factor. To test in the degraded conditions, all the six lists were further superimposed with speech noise at +5 dB, 0 dB, -5 dB and -10 dB SNRs. The waveforms of a sample word /bennu/ in quiet, at +5 dB, 0 dB, -5 dB and -10 dB SNRs are shown in the Figure 3.2.

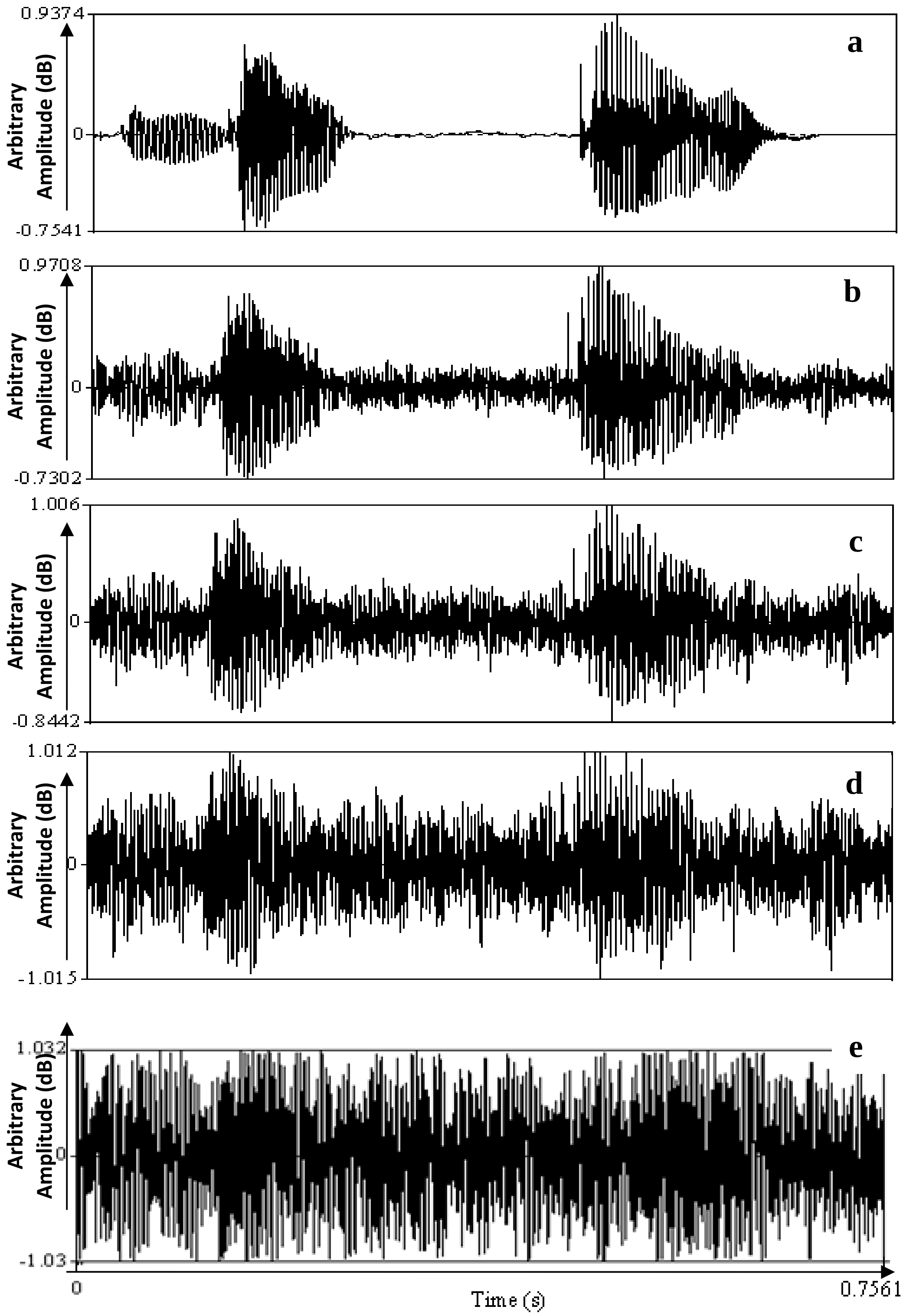

Figure 3.2: The waveforms of a sample word /bennu/ in quiet (a), at +5 dB (b), 0 dB (c), -5 dB (d) and -10 dB (e) signal to noise ratios.

The visual stimuli of the selected subject were edited using Windows movie maker software, to introduce a gap of five seconds between the words. The original audio of the video sample was muted and the auditory stimuli which was recorded using Pratt software was overlapped with the visual stimuli. This was necessary as the original audio was distorted and had high background noise.

To test for homogeneity in auditory and visual stimuli among the lists, all the lists were presented to five Kannada native speakers in three modalities; only Auditory (A), only Visual (V) and Auditory-visual (AV). The pilot comparison showed that the scores obtained for all the lists were similar.

3.3 Instrumentation

In the present study, a calibrated Madsen Orbiter-922 type I diagnostic audiometer with TDH-39 headphones was used to estimate the air-conduction thresholds and administer speech audiometry. A laptop computer with windows movie maker was used for video editing. The Pratt and Adobe Audition softwares were used for recording, editing and presenting the test stimuli. A digital video camera was used to record visual stimuli and a Sony MX78 omni-directional collar microphone was used to record auditory stimuli. Headphones were used to present auditory stimuli in 'A' mode and 'AV' mode.

3.4 Test Environment

All tests were administered in an acoustically treated room with noise levels at permissible limits (ANSI S3.1, 1991).

3.5 Test Procedure

The procedure started with preliminary evaluations which included case history, puretone audiometry and speech audiometry. For all the participants, the puretone air conduction thresholds (0.25, 0.5, 1, 2, 4, and 8 kHz) and speech recognition thresholds were obtained monaurally for both the ears. Only the individuals who fulfilled all the subject selection criteria (mentioned in section 3.1) were chosen for the study. After preliminary evaluation, the procedure included assessment of speech perception in noise, working memory and audio-visual integration.

3.5.1 Assessment of speech perception in noise

Speech perception in noise (SPIN) was binaurally tested at 0 dB signal to noise ratio (SNR). The phonemically balanced (PB) word list developed by Yathiraj and Vijayalakshmi (2005) was used as signal and was presented along with the speech noise. The participants were asked to repeat the words and total number of words repeated correctly was noted down. The percentage of correct responses was calculated by dividing number of correctly repeated words by the total number of words and multiplying this factor by 100.

3.5.2 Assessment of working memory

The procedure used to measure working memory capacity was adapted from versions of the operation span task used by Kane, Hambrick, Tuholski, Wilhelm, Payne and Engle (2004). Guidelines recommended by Conway, Kane, Bunting, Wilhelm and Engle (2005) were followed during administration and scoring. To be consistent with the Conway et al.'s terminology, the task consisted of 'items' that varied in difficulty, which was manipulated by varying the number of 'elements' per item.

The operation span task consisted of 5 items and 20 elements. The number of elements per item varied from 2 to 6. Each element consisted of a mathematical operation which included addition and division, and a Kannada word (E.g. (6/3) + 7 = 8? 'Ka:ge') The participant's task was to read the mathematical problem aloud, then say ‘yes’ or ‘no’ to indicate whether the given answer was correct or wrong, and then say the word. After all the elements in an item were presented, the participants were required to repeat all the words in that item. The difficulty of the items was randomized such that the number of elements was unpredictable at the outset of an item.

The scoring for mathematical problem and words were done separately. The participants who scored less than seventeen out of twenty (80%) in mathematical problem were not considered for analysis as results of those subjects are not valid. For each correct item, one mark was given, only if all the elements were repeated correctly. If some of the elements were incorrect, the number of correct elements was divided by total number of elements. For example, if three elements out of five elements of an item were correct, then a score of 0.6 was given. Finally, the scores of all five items were added and divided by five to obtain the final score for working memory. If the scores obtained in five items were 1, 0.8, 0.5, 0.16, 1 and the total obtained was 3.46, then the final score for working memory will be 0.693 (3.46 divided by 5). The scores for working memory ranged between 0 and 1.

3.5.3 Assessment of audio-visual integration

Prior to beginning of the test, the participants were instructed about the type of the stimuli, presentation mode and the response task. The testing was conducted in quiet as well as in the presence of noise. In the presence of noise, there were four different signal to noise ratios used; +5dB, 0dB, -5dB and -10 dB SNR. While in quiet the stimuli were presented in 3 modalities (auditory only, visual only & auditory-visual), in the presence of noise, the stimuli were presented only in 2 modalities (auditory only, & auditory-visual), at each SNR (Figure 3.3). The ‘V’ mode performance was assessed in the quiet condition only.

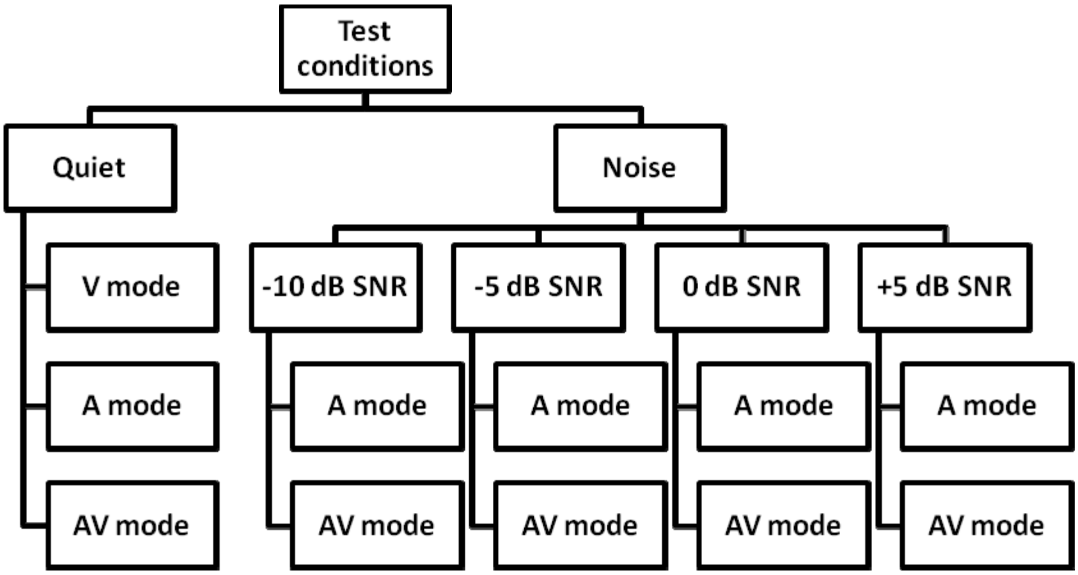

Figure 3.3: Block diagram of the experimental paradigm.

Finally there were three conditions in quiet (A, V, AV) and eight conditions in the presence of noise (A, AV in +5dB, 0dB, -5dB & -10 dB SNR). In each condition, only one list of 10 bi-syllabic words out of six lists was presented. The lists were randomized when they were needed to be presented for the second time. The audio stimuli were presented binaurally through headphones at most comfortable level (MCL). The visual stimuli were presented from wide screen laptop with fifteen inches and the distance between the participant and laptop was maintained at fifteen inches. The audio of the video stimuli was muted while testing in V mode. The mode of presentation was randomly chosen. The subjects responded by repeating the words and the words repeated were noted down by the clinician.

3.6 Response Analysis

In each condition (quiet, -10 dB, -5 dB, 0 dB, & +5 dB SNR) the total number of words repeated correctly out of ten words was noted separately for A, V, and AV mode. The AV speech perception was quantified by calculating visual enhancement

(VE, the benefit obtained from adding a visual signal to an auditory stimulus) and auditory enhancement (AE, the benefit obtained from adding an auditory signal to a visual-only stimulus) scores. The VE and AE were calculated according to the equations mentioned below. The scores for VE and AE ranged between -1 to +1.

$$\mathbf{VE = (AV - A)/\ 1 - A}$$

$$\mathbf{AE = (AV - V)/\ 1 - V}$$

3.7 Data Analysis

The data obtained from the participants were subjected to the following analysis to test the objectives of the study.

1. To correlate the SPIN scores with the VE and AE scores across different stimulus conditions.
2. To correlate the working memory scores with the VE and AE scores across different stimulus conditions.
3. To correlate SPIN and working memory scores in different stimulus conditions.

Chapter 4

RESULTS

The primary objective of the study was to analyze whether the speech perception in noise (SPIN) and working memory (WM) influence audio-visual (AV) perception of speech. The independent variables were SPIN and WM while the dependent variable was AV speech perception. SPSS (version 16) was used for the statistical analysis of data. Descriptive statistics, Repeated measures ANOVA, correlation, independent t-test and one-way ANOVA were the statistical tests used for the purpose. The results thus obtained are ordered under the following headings

4.1. Effect of modality on speech perception scores

4.2. Effect of stimulus condition on AV speech perception

4.3. Relation between speech perception in noise and AV speech perception

4.4. Relation between working memory and AV speech perception

4.5. Relation between speech perception in noise and working memory

4.1 Effect of Modality on Speech Perception Scores

To begin with, the speech perception scores of 66 participants in the three modalities were analyzed to obtain the mean and standard deviation scores. This was done separately for each stimulus condition (-10 dB, -5 dB, 0 dB, +5 dB SNR, & quiet). The mean (raw score out of 10) and standard deviation of speech perception scores in the 3 modalities at different signal-to-noise ratios are given in Table 4.1.

Table 4.1: *Mean and standard deviation (S.D) speech perception scores in the visual, auditory, and auditory-visual modes in quiet, -10 dB, -5 dB, 0dB, 5 dB SNR*

Stimulus Condition	Visual mode Mean (S.D)	Auditory mode Mean (S.D)	Audio-visual mode Mean (S.D)
-10 dB SNR	CNT	1.12 (0.87)	3.65 (1.96)
-5 dB SNR	CNT	3.58 (1.61)	6.82 (1.66)
0 dB SNR	CNT	7.74 (1.25)	9.21 (0.94)
5 dB SNR	CNT	9.71 (0.70)	9.85 (0.40)
Quiet	1.09 (1.17)	9.95 (0.21)	10.0 (0.0)

Note: CNT- could not be tested as presentation of auditory stimulus was inevitable in these conditions.

In general, the mean speech perception scores were higher in AV-mode compared to A and V-mode. The scores were least in the V-mode. This was true in quiet as well as in all the SNR conditions. The mean scores also decreased with decrease in SNR in the A and AV modality. In the visual modality, only the scores obtained in quiet was used for modality-wise comparison in all the stimulus conditions. The difference in the mean scores was tested for the statistical significance on Repeated measures ANOVA. The results of Repeated measures ANOVA are given in Table 4.2

Table 4.2: *Results of repeated measures ANOVA comparing speech perception scores across the 3 modalities in the 5 stimulus conditions*

Stimulus condition	F	df (error)
-10 dB SNR	108.881 *	2 (130)
-5 dB SNR	363.733*	2 (130)
0 dB SNR	1718.676*	2 (130)
5 dB SNR	3216.879*	2 (130)
Quiet	3762.289*	2 (130)

*Note: * - $p < 0.01$*

The results of repeated measures ANOVA showed a significant ($p<0.01$) main effect of modality on speech perception scores in all the 5 stimulus conditions. As there was a significant main effect, the pair-wise comparison was tested on Bonniferroni post-hoc test. The results of post-hoc test showed significant difference across all three modalities (A-V, A-AV & V-AV) at 0, -5 and -10 dB SNR. However, in quiet and +5 dB SNR the significant difference was found only between A-V and V-AV modalities.

4.2 Effect of Stimulus Condition on AV Speech Perception

In order to test the effects of independent variables on AV speech perception, the scores were first converted as ***visual enhancement*** and ***auditory enhancement*** scores. The visual enhancement (VE) score was a quantity of the benefit obtained from adding a visual signal to an auditory stimulus score, whereas the auditory enhancement (AE) score was a quantity of the benefit obtained from adding an auditory signal to a visual-only stimulus. The raw scores obtained in the 3 modalities were used to compute VE and AE scores, separately for each individual. This was done for all the 5 stimulus conditions. The mean and standard deviation of VE and AE scores in the 5 stimulus conditions are shown in Figure 4.1.

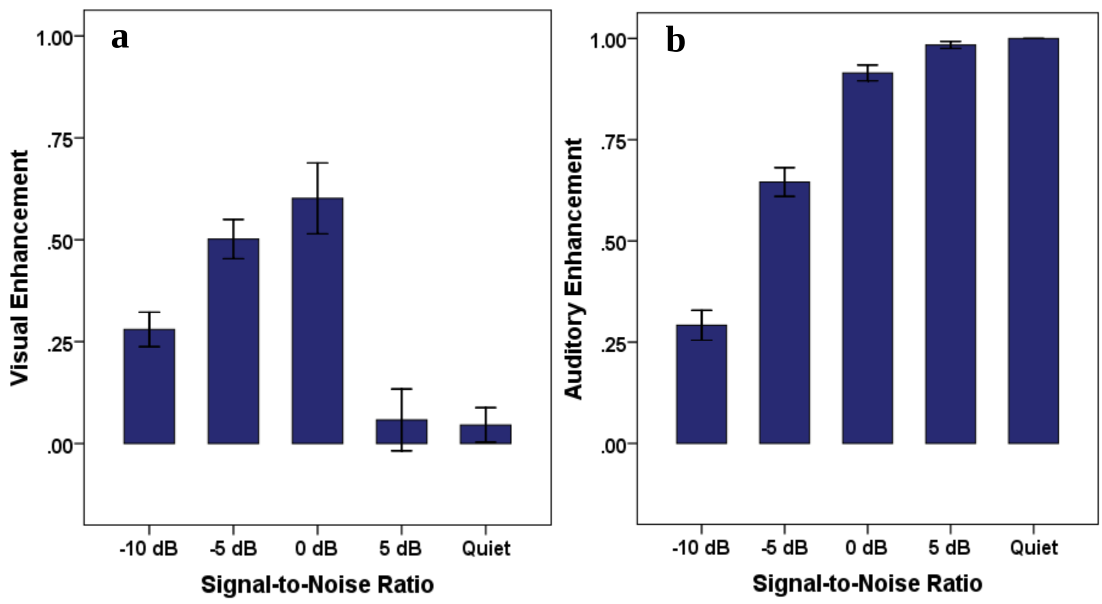

Figure 4.1: Mean and standard deviation of visual enhancement (a) and auditory enhancement (b) scores in quiet, -10 dB, -5 dB, 0 dB, and +5 dB SNR.

Note: Visual and auditory enhancement scores would vary between -1 and 1.

The inspection of the Figure 4.1 shows that there are mean differences among the 5 conditions in both AE and VE scores. The mean VE scores were maximum at 0 dB SNR and decreased with either increase or decrease in SNR. The AE scores on the other hand increased with increase in SNR. The mean differences observed were statistically tested for the effect of condition on repeated measures ANOVA. The results showed a significant main effect of condition on VE [$F(4,260) = 46.907$, $p<0.001$] and AE scores [$F(4,260) = 491.085$, $p<0.001$]. Consequently, pair-wise comparison was tested using Bonferroni post-hoc test and the results are represented in Table 4.3.

Table 4.3: *Pair-wise comparison of mean visual enhancement and auditory enhancement across quiet, -10 dB, -5 dB, 0 dB, and 5 dB SNR*

Parameter	Condition	-5 dB SNR	0 dB SNR	5 dB SNR	quiet
VE	-10 dB SNR	S	S	S	S
	-5 dB SNR		S	S	S
	0 dB SNR	S		NS	S
	+5 dB SNR	S	NS		NS
	quiet	S	S	NS	
AE	-10 dB SNR	S	S	S	S
	-5 dB SNR		S	S	S
	0 dB SNR	S		S	S
	+5 dB SNR	S	S		S
	quiet	S	S	S	

Note: S - p < 0.05, NS – p > 0.05

The results of post-hoc test showed a significant difference in the mean scores across all conditions in AE. However in VE, there was no significant difference between quiet and +5 dB SNR, and, 0 and +5 dB SNR.

4.3. Relation between Speech Perception in Noise and AV Speech Perception

4.3.1 Correlation between SPIN and AV Speech Perception

To understand the relation between SPIN and AV speech perception, the correlation was obtained for both VE and AE with SPIN scores in different stimulus conditions. Figure 4.2 (a-e) shows the relation between VE and SPIN scores while Figure 4.3 (a-e) shows relation between AE and SPIN scores. The data in Figure 4.2 (a-e) shows a positive correlation between the 2 variables in (a) and (b) but no evident relationship in (c). In (d) and (e) in majority of the individuals, VE did not change with SPIN scores.

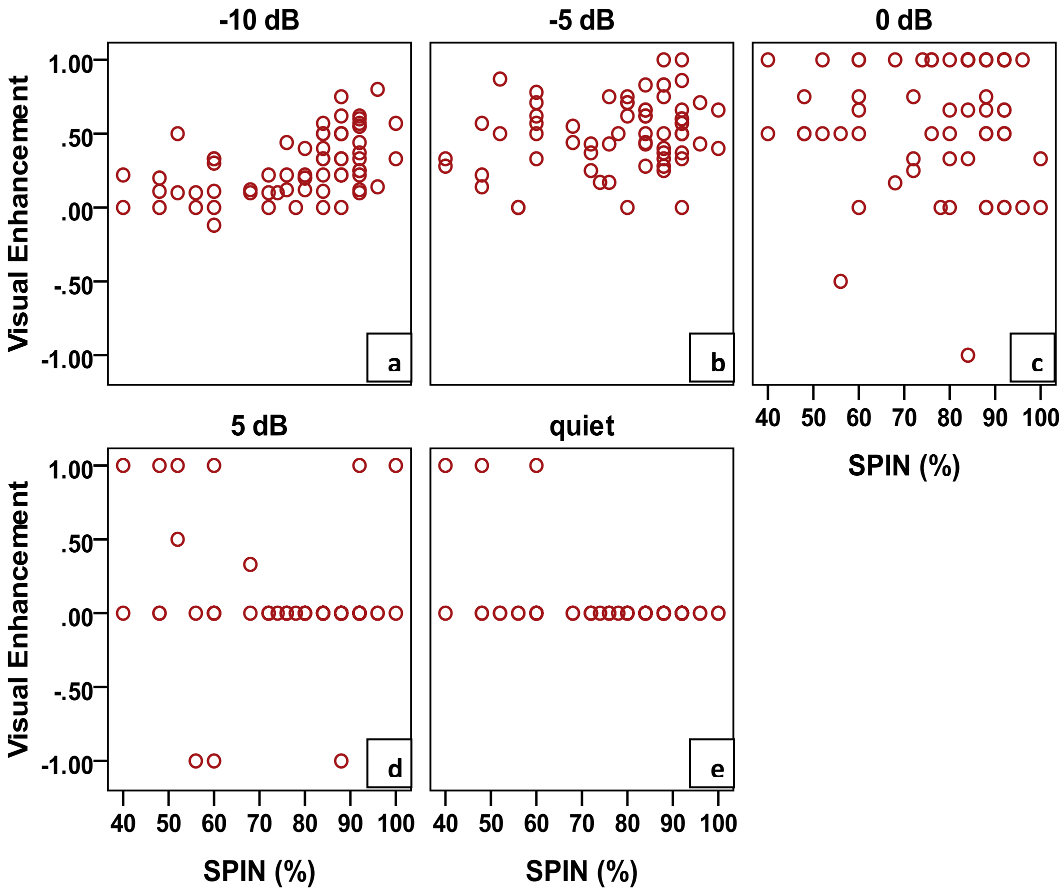

Figure 4.2 (a-e): Scatter plots representing correlation between visual enhancement and SPIN scores in quiet (a), -10 dB (b), -5 dB (c), 0 dB (d), and +5 dB SNRs (e).

Note: Visual and auditory enhancement scores would vary between -1 and 1.

The data in Figure 4.3 (a-e) also shows a trend similar to Figure 4.2 (a-e). That is, a positive relationship was observed between the 2 variables in (a) and (b) but no such relationship could be observed in (c), (d) and (e) (AE did not change much with SPIN scores).

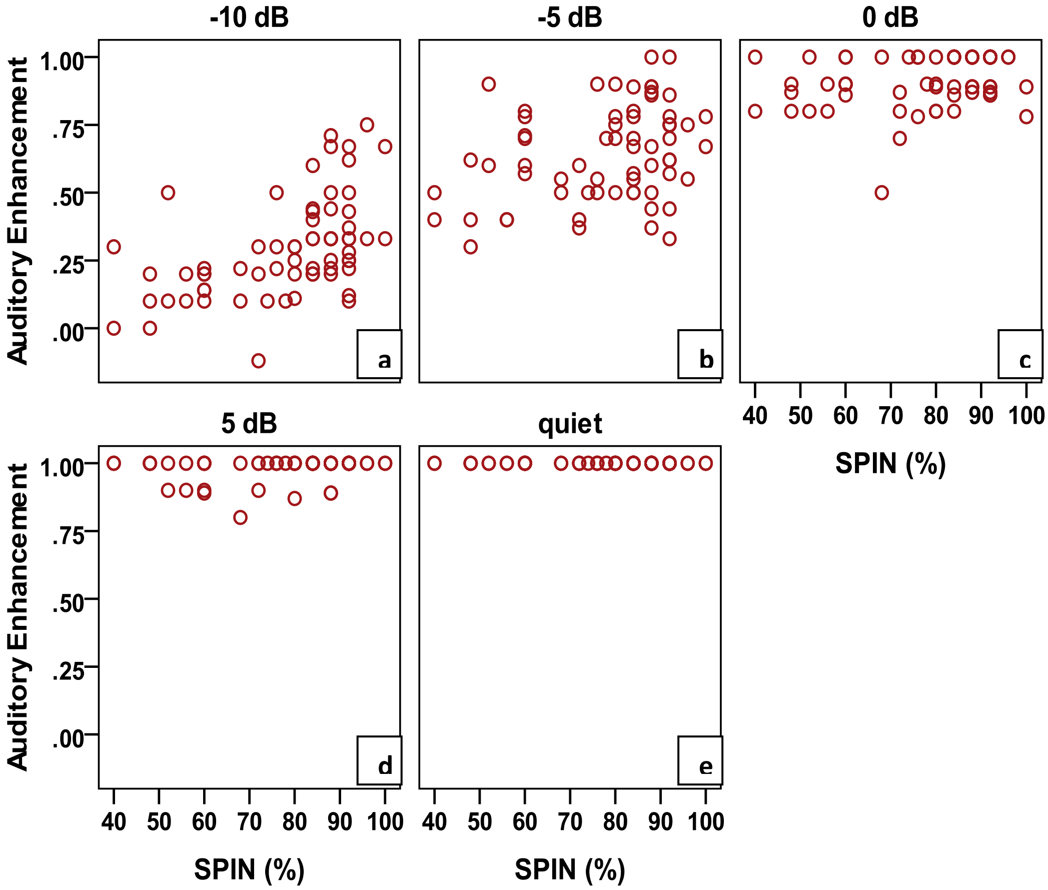

Figure 4.3 (a-e): Scatter plots representing correlation between auditory enhancement and SPIN scores in quiet (a), -10 dB (b), -5 dB (c), 0 dB (d), and +5 dB SNR (e).

Note: Visual and auditory enhancement scores would vary between -1 and 1.

The data were statistically tested for correlation between two variables on Pearson product moment correlation and the results are given in Table 4.4. The results of the correlation showed a significant low positive correlation at -10 dB and -5 dB SNR but a significant low negative correlation in quiet condition between mean scores of SPIN and VE. On the contrary, correlation of mean differences between SPIN and AE showed that there was a significant low positive correlation at -10 dB SNR but a significant low negative correlation in quiet and 5 dB SNR conditions.

There was no significant correlation in other conditions for both VE and AE with SPIN scores.

Table 4.4: *The results of correlation test correlating auditory enhancement (AE) and visual enhancement (VE) with SPIN in quiet, -10 dB, -5 dB, 0dB, and +5 dB SNR*

Stimulus condition	**SPIN and VE**		**SPIN and AE**	
	Pearson correlation	Significance	Pearson correlation	Significance
-10 dB SNR	0.496**	0.000	0.512**	0.000
-5 dB SNR	0.253*	0.041	0.335**	0.006
0 dB SNR	0.003	0.979	0.243*	0.049
+5 dB SNR	-0.166	0.184	0.207	0. 095
Quiet	- 0.407**	0.001	CNT	CNT

*Note: *- $p < 0.05$, **- $p<0.001$, CNT- could not be tested as all the participants had the same AE score of 1.*

4.3.2 Comparison between Poor and Good SPIN Groups

The influence of SPIN on AV speech perception was further tested using a second method wherein the participants were divided into two groups based on their SPIN scores (good SPIN & poor SPIN groups) and compared with each other for their AV speech perception. The grouping was done based on the confidence interval (at 95%) determined for the SPIN scores of 66 participants of the study. The lower-boundary and the upper-boundary of the interval for the SPIN scores thus obtained were found to be 72% and 82% respectively. The participants with scores lesser than the lower boundary (72%) were grouped as 'poor SPIN' and those with scores higher than the upper boundary (82%) were grouped as 'good SPIN'. The participants with scores within the interval (72-82%) were not considered for any comparisons.

The VE and AE scores of the resultant two groups were then compared to investigate whether there was any difference in the AV speech perception between the two groups. It was hypothesized that the presence of significant difference in VE scores would indicate that SPIN influences AV speech perception. The mean and the standard deviation of the VE and AE scores of the two SPIN groups (good & poor), in each stimulus condition (quiet, -10 dB, -5 dB, 0dB & 5 dB SNR) are shown in Figure 4.4.

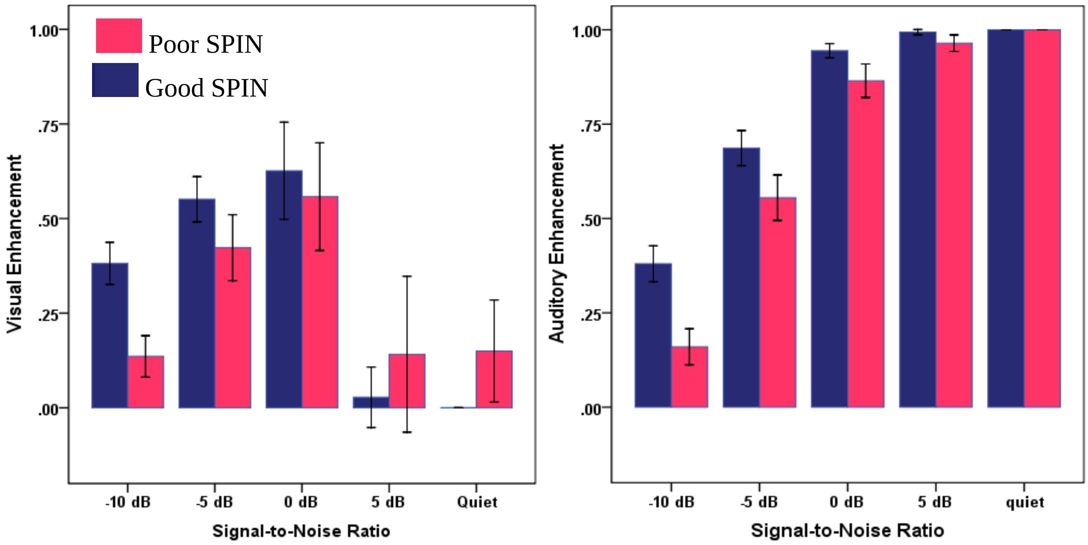

Figure 4.4: Graph representing mean and standard deviation of visual enhancement (Left panel) and auditory enhancement (Right panel) for good SPIN (Blue bars) and poor SPIN (Pink bars) groups across the 5 conditions.

Note: VE and AE scores would vary between -1 to +1.

The mean VE scores were higher in the good SPIN group compared to poor SPIN group except in +5 dB SNR and quiet, wherein the case was reverse. In contrast, the mean AE scores of good SPIN group was higher than that of poor SPIN group in all the stimulus conditions except in quiet, where the mean scores were equal. The mean scores of the two groups were statistically compared using independent samples t-test. The results of the t-test for VE and AE scores are given in Table 4.5.

The results showed that the scores for VE of good SPIN group were significantly higher in -10 dB and -5 dB SNR conditions and significantly lower in quiet, than the poor SPIN group. The mean differences in other stimulus conditions (0 dB & +5dB SNR) were not significant. On the contrary, the difference in the mean AE scores of the two groups was statistically significant in all stimulus conditions except in quiet.

Table 4.5: *The results of the t-test comparing the good SPIN and poor SPIN groups for their visual enhancement (VE) and auditory enhancement (AE) scores in quiet, -10 dB, -5 dB, 0dB and +5 dB SNR*

Stimulus Condition	**VE (df= 54) t**	**AE (df=54) t**
-10 dB SNR	4.755**	4.967**
-5 dB SNR	2.042*	2.814**
0 dB SNR	.557	3.119**
5 dB SNR	-1.003	2.585*
Quiet	-2.475*	CNT

*Note: *- p < 0.05, **- p< 0.01, CNT- could not be tested as both the groups had same mean and no standard deviation.*

4.4 Relation between Working Memory (WM) and AV Speech Perception

4.4.1 Correlation between WM and AV Speech Perception

The second independent variable of the study was working memory. Similar to effect of SPIN, the effect of working memory on AV speech perception was examined. The relation between WM and AV perception of speech across different stimulus conditions (quiet, -10 dB, -5 dB, 0dB, & +5 dB SNR) was examined by obtaining the correlation. The Figure 4.5 (a-e) shows the relation between VE and WM scores and the Figure 4.6 (a-e) shows relation between AE and WM scores.

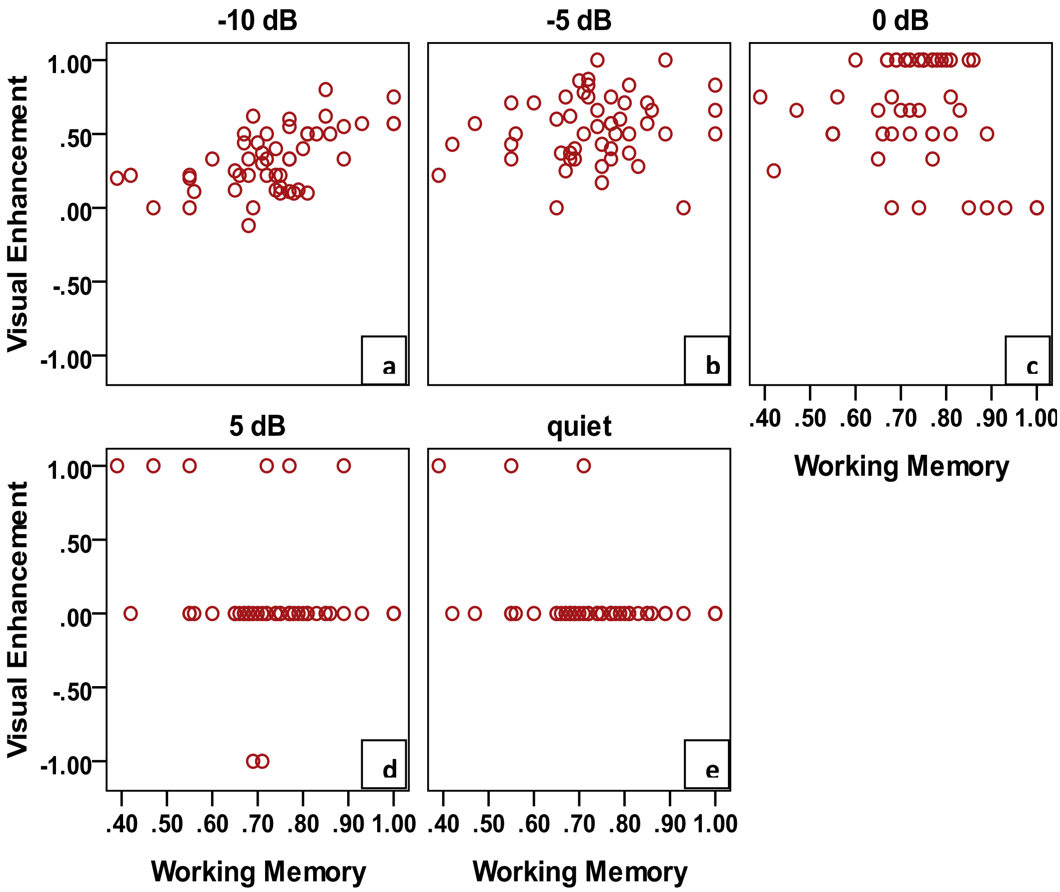

Figure 4.5 (a-e): Scatter plots representing correlation between visual enhancement and WM scores in quiet (a), -10 dB (b), -5 dB (c), 0 dB (d), and +5 dB (e) SNR.

Note: Visual and auditory enhancement scores would vary between -1 and 1 and WM scores ranged between 0 and 1.

As it is clear in the Figure 4.5 (a-e), there is a positive relationship between the VE and WM at -10 (a) and -5 dB SNR (b). Whereas in other conditions [quiet, 0 dB SNR and +5 dB SNR], such relationships were not observed. The Figure 4.6 (a-e) also shows a positive relationship between the 2 variables at -10 (a) and -5 dB SNR (b). But at 0 dB SNR (c), +5 dB SNR (d) and quiet (e), there was no relation between AE and WM. Also, in (d) and (e), the VE did not change with change in SPIN scores.

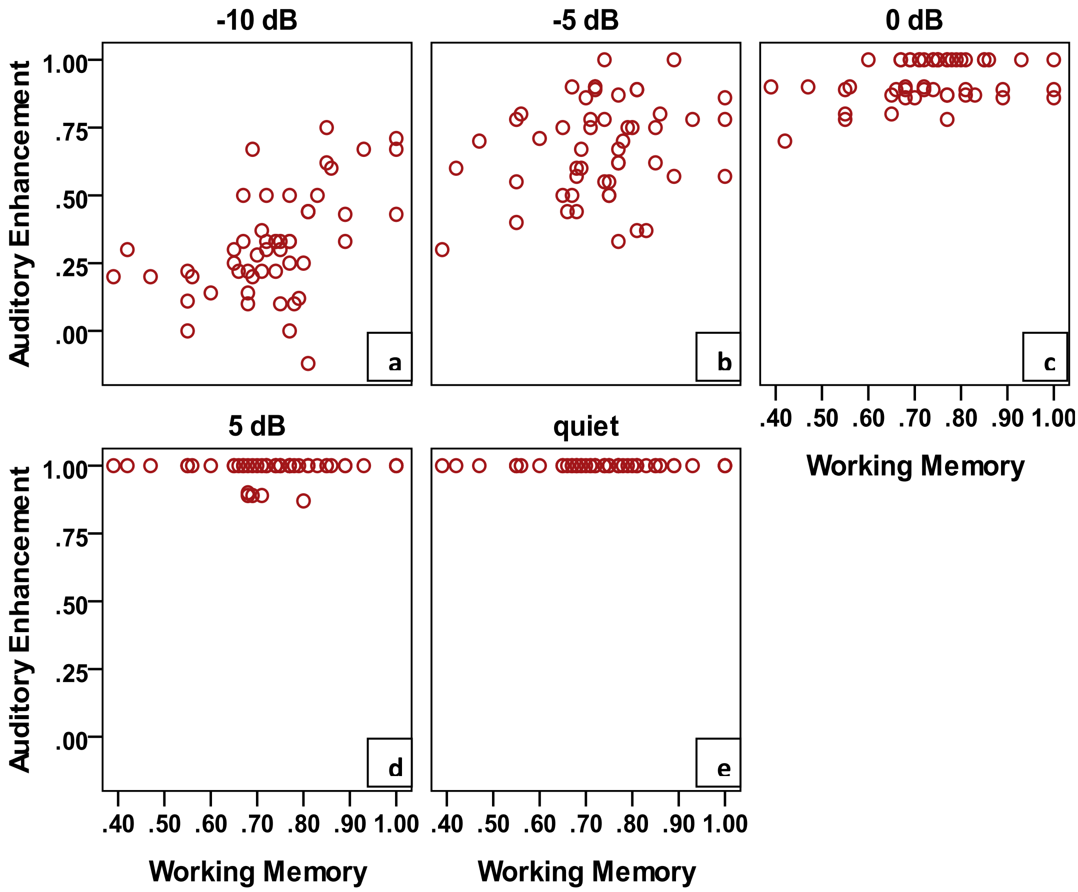

Figure 4.6 (a-e): Scatter plots representing correlation between auditory enhancement and WM scores in quiet (a), -10 dB (b), -5 dB (c), 0 dB (d), +5 dB (e) SNR.

Note: Visual and auditory enhancement scores would vary between -1 and 1 and WM scores ranged between 0 and 1.

The relationship between VE and AE with WM scores were statistically tested on Pearson product moment correlation test for different stimulus conditions (quiet, -10 dB, -5 dB, 0dB, and +5 dB SNR). The results of correlation test are given in Table 4.6.

Table 4.6: *The results of correlation test comparing visual enhancement (VE) and auditory enhancement (AE) with working memory in quiet, -10 dB, -5 dB, 0dB, and +5 dB SNR*

Stimulus condition	WM and VE		WM and AE	
	Pearson correlation	**Significance**	**Pearson correlation**	**Significance**
-10 dB SNR	0.572**	0.000	0.517**	0.000
-5 dB SNR	0.192	0.178	0.256	0.070
0 dB SNR	-0.307*	0.028	0.330*	0.018
+5 dB SNR	-0.208	0.143	0.040	0.781
Quiet	- 0.349*	0.012	CNT	CNT

*Note: *- p < 0.05, **- p< 0.01, CNT- could not be tested as both the groups had same mean and zero standard deviation.*

From the above Table 4.6, as we can witness clearly, there was a significant moderate positive correlation at -10 dB SNR and a significant low negative correlation in quiet and 0 dB SNR conditions between WM and VE. Whereas, correlation between WM and AE showed a significant moderate positive correlation at -10 dB SNR and a significant low positive correlation at 0 dB SNR conditions.

4.4.2 Comparison between Poor and Good WM Groups

In a second method of testing the relationship between WM and AV speech perception, the participants were divided into good WM and poor WM groups to test whether there was any difference between the two groups. The confidence intervals (95%) for WM scores were found out first. Then the participants with scores lesser than the lower boundary (70%) were grouped as 'poor WM' and those with scores higher than the upper boundary (76%) were grouped as 'good WM'. The participants with scores within the interval (70-76%) were excluded. The mean and standard deviation of VE and AE scores obtained for the 2 groups, across 5 conditions, are represented in Figure 4.7.

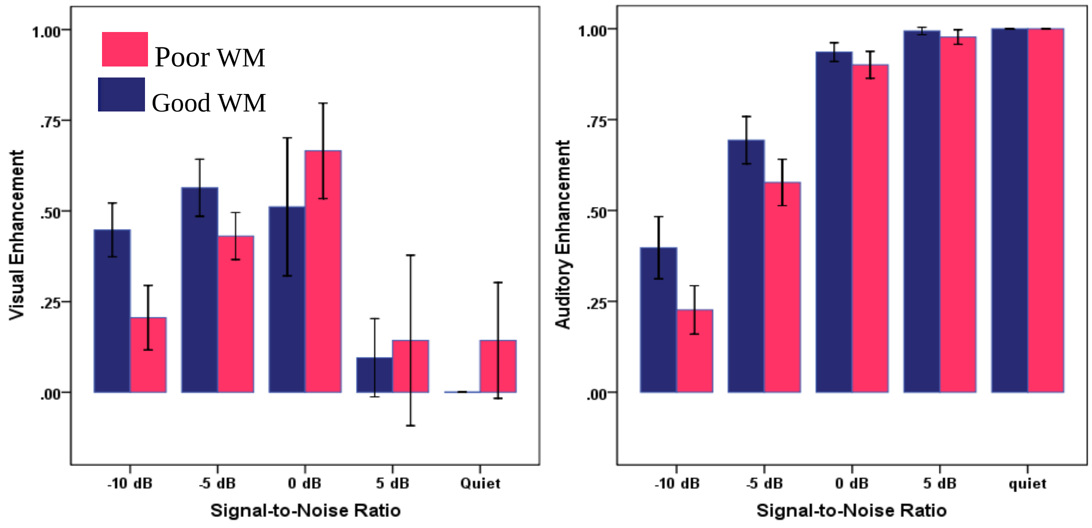

Figure 4.7: Mean, standard deviation of visual enhancement (Left panel) and auditory enhancement (Right panel) scores for good working memory (Blue bars) and poor working memory (Pink bars) groups in the 5 different stimulus conditions.

Note: Visual and auditory enhancement scores would vary between -1 and 1 and WM scores ranged between 0 and 1.

The above graph shows that the mean VE scores for the good WM group were better than those of poor WM at -10 dB, -5 dB and 0 dB SNR. But at +5 dB SNR and in quiet, poor WM group had better scores. Whereas, the mean AE scores of good WM group was higher than that of poor WM group in all the stimulus conditions except in quiet, where the mean scores were equal.

To test for significance in the mean differences between the two groups, independent samples t-test was administered. The results of the t-test for VE and AE scores across 5 conditions are given in Table 4.7.

The results of the independent samples t-test showed that the scores for VE and AE of good WM group were significantly different than that of poor WM group only in -10 dB SNR condition. Even though there were mean differences in other conditions, they were not statistically significant.

Table 4.7: *The results of the t-test comparing the visual enhancement and auditory enhancement scores for good WM and poor WM groups in quiet, -10 dB, -5 dB, 0dB and +5 dB SNR*

Stimulus Condition	AE (df: 33) t	VE (df: 33) t
-10 dB SNR	2.380*	3.421**
-5 dB SNR	2.007	1.986
0 dB SNR	1.312	-.986
+5 dB SNR	1.338	-.337
Quiet	CNT	-1.817

*Note: *- p < 0.05, **- p< 0.01, CNT- could not be tested as both the groups had same mean and zero standard deviation.*

4.4 Relation between Speech Perception in Noise and Working Memory

While analyzing the effect of WM on AV speech perception, across different SNR conditions, there is possibility of SPIN influencing the effect of WM. Hence, the relationship between the 2 independent variables was tested on correlation. Figure 4.8 shows the relation between SPIN and WM.

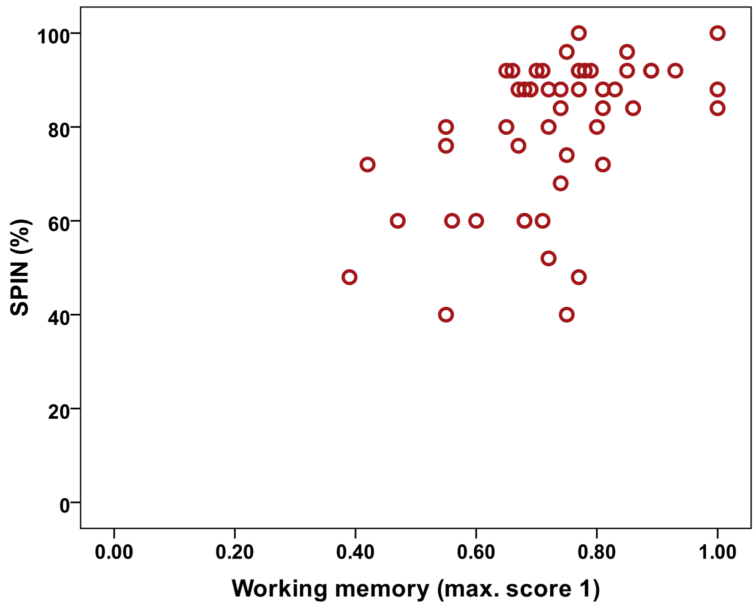

Figure 4.8: Scatter plot representing correlation between SPIN and WM groups.

The Pearson product moment correlation was found out to test for correlation between SPIN and WM groups. The results showed that there was a significant ($p<0.001$) moderate positive correlation ($r=0.508$) between the two independent variables.

Chapter 5

DISCUSSION

The perception of speech could be through either auditory (A-mode) or visual (V-mode) mode. Of the two modalities, auditory modality is the primary one as it provides information on place, manner and voicing of speech sounds. The dependency on the visual mode is only in instances where the information provided through auditory modality is insufficient. This happens due to reduction either in external redundancy or internal redundancy. The negative influences of reduction in both external and internal redundancy on speech perception have been well established in the literature. Although it is universally accepted that audio-visual mode (AV-mode) is beneficial over A-mode in adverse listening conditions, the underlying mechanisms of AV perception are not clearly understood.

The present study was an attempt to probe into the underlying mechanisms of AV perception in sensory and cognitive domain. In the sensory domain auditory closure was taken as an independent variable, while in cognitive domain working memory was the independent variable. The present experiment to analyze the effects of these two independent variables on AV perception, on 66 participants, showed some interesting findings. These findings for the clarity of presentation have been reported and discussed under the following headings;

5.1 Effect of modality on speech perception scores

5.2 Effect of stimulus condition on AV speech perception

5.3 Relation between speech perception in noise and AV speech perception

5.4 Relation between working memory and AV speech perception

5.5 Relation between speech perception in noise and working memory

5.1 Effect of Modality on Speech Perception

Evidence from previous research explains that speech intelligibility improves when listeners receive information from both auditory and visual modes in instances where the auditory cues are degraded by reducing external redundancy (Sumby & Pollack, 1954; Anderson, 2006; Huffman, 2007). The present findings are in consonance with these earlier reports. Overall performance of the participants was best in the AV-mode compared to auditory-alone or visual-alone modes. The performance was least in the V-mode. Auditory mode provides cues pertaining to the place, manner and voicing of a speech sound while in the visual mode one gets cues of only place of articulation, that too not completely (Anderson, 2006). Hence, it is logical to expect better performance in the A-mode compared to V-mode. Further, as the redundancy of the cues is more in the bimodal presentation, AV-mode showed better performance than that in A-mode or V-mode, which were unimodal presentations.

The results showed that the performance in the A-mode was comparable to that in AV-mode in quiet and at +5 dB SNR. This means that the necessary cues for an ideal performance are not cut in quiet and +5 dB SNR. However, the performance reduced with further reduction in the SNR both in A and AV-modes. As the reduction of redundancy was only in the auditory signal without distorting the visual signal, the reduction in the scores of A and AV-modes can be attributed solely to the reduction in the SNR of the auditory signal. Figure 5.1 represents the deterioration in the speech identification in A and AV-modes across different stimulus conditions.

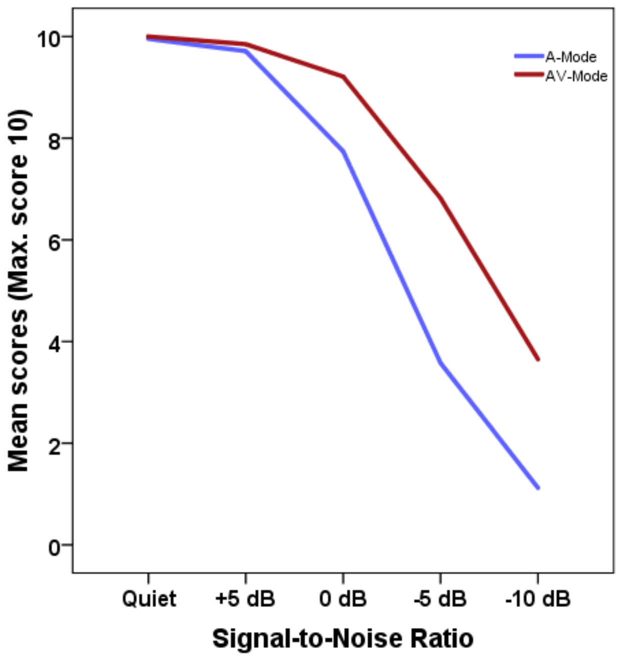

Figure 5.1: Mean identification scores in auditory mode (blue line) and audio-visual mode (red line) in quiet, -10 dB, -5 dB, 0 dB, and +5 dB SNR conditions.

As we can observe in the Figure 5.1, the performance of both A and AV-modes is same in quiet and decreases with decreasing SNR. But the decrease in performance for A-mode was more than that of AV-mode which is evident by the steepness of the curves. That is, the steepness of the A-mode (blue line) is greater than that of the AV-mode (red line). The difference between the A and AV-mode is due to complementary cues provided by the addition of visual cues, which however functions nonlinearly.

The increase in the difference between the performances in A and AV-modes with the decrease in SNR evidences, greater dependency on visual cues at lower SNRs. The finding is in agreement with the earlier studies (Sumby & Pollack, 1954; O'Neil, 1954; Neely, 1956; Erber, 1969; Grant & Seitz, 2000; Rudmann, Mc Carley & Kramer, 2003; Bernstein, Auer & Takayanagi, 2004; Ross, Saint-Amour, Leavitt, Javitt & Foxe, 2007) which showed larger benefits of bimodal presentations when the auditory signal was degraded. In the study by Sumby and Pollack (1954), it was demonstrated that the addition of visual cues improved speech perception by an amount equivalent to a 5 to 18 dB increase in the SNR which accounts for improvement of up to 60% in word recognition. They also demonstrated that with increase in difficulty of auditory-only perception, the benefits obtained by combining the auditory and visual speech information also increased. Although the present results showed a similar trend, the quantity of improvement at the worst SNR (-10 dB SNR) was only about 25%. The difference among the studies in the extent of improvement may be attributed to the different test materials used.

5.2 Effect of Stimulus Condition on AV Speech Perception

In the previous section (Section 5.1), it was learnt that AV-mode was better than A-mode only at lower SNRs. That is, although the redundancy of visual cues was not varied, it nonlinearly facilitated speech perception. Hence the absolute difference between the scores of AV and A-modes would not have given a clear picture of the visual enhancement (VE). For example, the importance of 10% facilitation by addition of V-mode will be different for 3 individuals who score 20%, 50% and 90% in the A-mode. An absolute difference between A and AV-mode would be 10% in all these cases, but logically the weightage of relative enhancement should be maximum for one who had 90% score in the A-mode followed by one with 50% and 20% score. This was achieved by calculating visual enhancement scores using the following formula

$$VE= AV-A/1-A$$

For the above example, the visual enhancement will be 0.12, 0.2, and 1 for the individuals with 10%, 50% and 90% score in the A-mode, respectively. With the similar logic the AE scores were derived from the performances in AV and V-mode using the formula, AE= AV-V/1-V. The AE score is a quantity of the benefit obtained from adding an auditory signal to a visual-only stimulus.

5.2.1 Effect of Condition on Visual Enhancement

The results showed that the VE was maximum at 0 dB SNR and decreased with either increase or decrease in SNR. The decrease in the visual enhancement at SNRs below 0 dB was an unlikely finding considering the trend observed from the absolute scores obtained in AV and A-modes. The absolute difference between scores of A and AV-mode indicated increasing facilitation with decreasing SNR. However, the results of VE are in contrary to this. The trend of VE scores represents the functional benefit obtained by an individual by adding the visual cues. The functional benefit progressively reduces with reduction in the performance of A-mode, which in turn is dependent on the SNR.

On the other hand, the same logic does not apply to the performances at +5 dB SNR and quiet. Above 0 dB SNR, performance again reduces with increasing SNR. This is attributed to the ceiling effect. Because the scores in the A-mode were already 100%, the resultant VE scores would become 0. The number of such individuals would increase with increase in SNR, resulting in progressively decreasing mean VE performance. Based on these findings, it may be inferred that any study intending to investigate visual enhancement shall take SNRs of 0 dB or lesser and not take the higher SNRs, in order to get a true picture of VE.

5.2.2 Effect of Condition on Auditory Enhancement

The AE was maximum in quiet (approximately 40 dB SNR) and decreased with decreasing SNR. This trend is logical as the enhancement shall increase with increasing redundancy of the auditory cues. To strengthen the notion, the correlation between A scores and AE scores was done in different SNRs. It was expected to get high positive correlation between the two variables in all the stimulus conditions if the stated logic was true. Table 5.1 gives the Pearson correlation co-efficients in the 5 stimulus conditions.

Table 5.1: *The Pearson correlation co-efficient correlating auditory enhancement scores and auditory mode scores in quiet, -10 dB, -5 dB, 0 dB, and +5 dB SNR conditions*

Stimulus conditions	Correlation co-efficient
-10 dB SNR	0.352**
-5 dB SNR	0.423**
0 dB SNR	0.536**
5 dB SNR	0.488**
Quiet	CNT

*Note: **- p<0.001, CNT- could not be tested as all the participants had the same AE score of one (1).*

The results showed that the correlation was maximum at 0 dB SNR and reduced at other conditions. Hence the above stated logic was defeated. The trend in Table 5.1 in turn shows the influence of VE as a primary variable. In the formula used for the calculation of AE, AV and V scores are the two parameters considered. As the V scores remained constant in all the stimulus conditions, the differences in the AE had to be because of AV which in turn was related to VE. As the trend of VE (Figure 4.1) across conditions is same as that of trend of correlation observed in Table 5.1, it can be inferred that it is this influence of VE over AV scores that varied AE across stimulus conditions. This was supported by the results of correlation of VE scores and AE scores at 0, -5 and -10 dB SNR. The data of +5 dB SNR and quiet were not considered as VE in these conditions was erroneous due to ceiling effect. The results of correlation showed a moderate positive co-efficient ($r=0.632$, $p<0.001$) between AE and VE scores.

5.3 Effect of SPIN on AV Speech Perception

From the previous two sections, it is clear that AV speech perception is influenced by external redundancy of the auditory signal. Earlier studies have shown negative effects of advancing age on AV speech perception (Shoop & Binnie, 1979; Middelweerd & Plomp, 1987; CHABA, 1988; Walden, Busacco, & Montgomery, 1993; Cienkowski, 1999; Cienkowski & Carney, 2002; 2004). This age related decline could be either partially or completely due to reduced internal redundancy secondary to central auditory processing deficits. Hence it was of interest to study the relationship between internal redundancy and speech perception in noise. The relationship was tested using two methods; one, by correlating SPIN and AV speech perception. The other, by comparing the AV speech perception of good and poor SPIN groups.

The results of the correlation showed that below 0 dB SNR both VE and AE correlated with the SPIN scores. That is, VE and AE increases with increase in SPIN scores. Additionally, AE positively correlated with SPIN at 0 dB. A negative correlation between VE and SPIN is erroneous as most of the individuals had 0 VE at +5 dB SNR and quiet. Further, the comparison of AV speech perception among the good and poor SPIN groups also showed that the SPIN scores have a positive effect on speech perception through AV-mode.

From these findings it can be concluded that internal redundancy controlled by central auditory mechanisms, as evident through SPIN, is directly related to the AV speech perception. Hence, one should take SPIN into consideration while commenting on the benefit provided by AV speech perception.

5.4 Effect of Working Memory (WM) on AV Speech Perception

It has been explained that speech understanding needs a constant encoding of information into and out of the working memory and manipulation of the information stored in memory (Pichora-Fuller, Schreider & Daneman, 1995).

Similar to SPIN, the relationship between working memory and AV speech perception was analyzed by correlating the two parameters and by comparing good and poor working memory groups. This was done to decipher the underlying mechanisms of speech perception under the cognitive domain. Considering that the integration of information from two different modalities is a complex higher level task, cognitive factors such as working memory were expected to influence the AV speech perception.

The results of both correlation and group comparison showed evidence for working memory influencing AV speech perception in noise. The influence was particularly significant at lower SNRs and the relationship was direct. That is, AV speech perception was better in individuals who had better working memory.

The results of the present study are in consonance with the earlier report (Pichora-Fuller, Schreider & Daneman, 1995) which stated that the loss of working memory hinders speech communication and may contribute to the age-related declines in the comprehension of speech which may be evident even in favorable listening conditions.

From these findings it can be concluded that working memory, a cognitive factor, has a direct relationship with the AV speech perception and hence demands attention while interpreting on AV speech perception.

5.5 Relationship between SPIN and WM

As both working memory and speech perception in noise were influencing AV speech perception in a similar way, it was of interest to know how these two independent variables related to each other. The results of correlation showed that the speech perception in noise was better in individuals with good working memory and vice versa. However, it was only a moderate correlation which shows that they are not controlled by same physiological mechanism.

Chapter 6

SUMMARY AND CONCLUSION

The reduction in external redundancy and hearing loss is shown to have negative influences on speech perception in the auditory mode. Although it is universally accepted that AV-mode is beneficial over A-mode in adverse listening conditions, the underlying mechanisms of the benefit, in the sensory and cognitive domain are not clearly understood. In the sensory domain, along with the hearing sensitivity, the complex central auditory processing shall influence the AV speech perception. But the literature has only the studies that have taken hearing sensitivity as an influencing factor. Logically, central auditory processing shall be a predominant factor compared to hearing sensitivity, considering that AV speech perception requires challenging tasks like integration of cues from the two sensory modalities.

Further, as speech perception involves cognition in terms of constant encoding and manipulation of information stored in the memory, it is expected that reduced cognitive abilities could degrade the process of audio-visual integration. This notion is strengthened by the earlier findings of age-related decline of the benefit obtained from the AV speech perception. But the relation between working memory and audio-visual speech perception has not been studied earlier. Hence the objectives of the present study were to investigate the influence of the speech perception in noise and the working memory on audiovisual integration in speech perception.

Sixty six adult native speakers of Kannada in the age range of 18 to 70 years, having normal hearing and normal or corrected vision of 6/6, participated in the study. First, the participants were assessed for their speech perception in noise at 0 dB SNR using the PB word list developed by Yathiraj and Vijayalakshmi (2005) in order to obtain speech perception in noise scores. The working memory capacity was found out using an operation span task to assess the cognitive abilities. Further, audio-visual integration was assessed in quiet as well as in the presence of noise. There were three conditions in quiet (A, V, AV) and eight conditions in the presence of noise (A, AV in +5dB, 0dB, -5dB & -10 dB SNR). Six lists of bi-syllabic Kannada words (phonetically balanced) having ten words in each list, which were developed specifically for the purpose of the present study, were used as test stimuli (auditory & visual) and were presented randomly.

In the results, the mean speech perception scores were higher in AV-mode compared to A and V-mode. The performance of both A and AV- modes were same in quiet and decreased with decreasing SNR. But the decrease in performance for A-mode was more than that of AV-mode. The difference between the A and AV-mode is due to complementary cues provided by the addition of visual cues, which functions nonlinearly.

In order to test the effect of stimulus condition on AV speech perception, the visual enhancement and the auditory enhancement scores were determined using standard formulae. The mean VE scores were maximum at 0 dB SNR and decreased with decrease in SNR. This represents the functional benefit obtained by an individual by adding the visual cues, which progressively reduces with reduction in the performance of A-mode, which in turn is dependent on the SNR. Above 0 dB SNR, performance again reduces with increasing SNR due to the ceiling effect. The AE scores on the other hand increased significantly with increase in SNR.

The relation between speech perception in noise and AV speech perception was assessed using two methods, one by correlating SPIN and AV speech perception and the other by comparing the AV speech perception of good (>80%) and poor (<72%) SPIN groups. The results of the correlation showed that, below 0 dB SNR, both VE and AE had positive correlation with SPIN scores. Additionally AE positively correlated with SPIN at 0 dB. A negative correlation was obtained at +5 dB SNR and quiet as most of the individuals had 0 VE. Further the comparison of AV speech perception among the good and poor spin groups also showed that the SPIN scores have a positive effect on speech perception through AV-mode.

Similar to SPIN, the relationship between working memory and AV speech perception was analyzed by correlating the two parameters and by comparing good and poor working memory groups. The results of both correlation and group comparison showed evidence for working memory influencing AV speech perception in noise. The influence was particularly significant at lower SNRs and the relationship was direct. That is, AV speech perception was better in individuals who had better working memory.

As both working memory and SPIN influence AV speech perception, the relation between these two variables was tested. The results showed a moderate correlation which shows that they are not controlled by same physiological mechanism/s. Thus, it can be concluded that internal redundancy controlled by central auditory mechanism/s (as evident through SPIN) and the cognitive abilities (represented as working memory, a cognitive factor) has a direct relationship with the AV speech perception and hence demands attention while interpreting on AV speech perception.

Implications

The present study has important implications in rehabilitative audiology. In audiological clinics, AV-mode of speech perception is recommended for individuals with poor speech identification scores. The present findings showed that the WM and SPIN directly influence AV speech perception. Hence, one should consider WM and SPIN, assess for them, estimate the probable benefit with AV speech perception and accordingly recommend the same.

Future Directions

Future studies can establish relationship between SPIN, WM and AV speech perception in a group of individuals with diagnosed auditory processing disorder (auditory closure deficits). This may give independent effects of sensory domain and cognitive domain deficits on AV speech perception. Future studies can also study the independent effect of age by controlling for the effects of SPIN and WM.

REFERENCES

American National Standards Institute. (1991). *American National Standard Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms*. ANSI S3.1- (1991). New York: American National Standards Institute.

American Speech-Language-Hearing Association. (2005). *(Central) Auditory Processing Disorders* (technical report). United states: American Speech-Language-Hearing Association (ASHA).

Anderson, E. (2006). *Audiovisual Speech Perception with Degraded Auditory Cues*. Undergraduate honors thesis, The Ohio State University.

Bernstein, L.E., Auer, E.T., Jr., & Takayanagi, S. (2004). Auditory speech detection in noise enhanced by lipreading. *Speech Communication, 44(1),* 5-18.

Binnie, C.A., Jackson, P., & Montgomery, A. (1976). Visual intelligibility of consonants: A lipreading screening test with implications for aural rehabilitation. *Journal of Speech and Hearing Disorders, 41,* 530-539.

Brault, L.M., Gilbert, J.L., Lansing, C.R., McCarley, J.S., & Kramer, A.F. (2010). Bimodal stimulus presentation and expanded auditory bandwidth improve older adults' speech perception. *Human Factors, 52(4),* 479-491.

CHABA, Committee on Hearing and Bioacoustics, Working Group on Speech Understanding and Aging. (1988). Speech understanding and aging. *Journal of the Acoustical Society of America, 83,* 859–895.

Cienkowski, K. M. (1999). Auditory-visual speech perception across the lifespan [Doctoral Dissertation, University of Minnesota, 1999]. *Dissertation Abstracts International, 60(1),* 116.

Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear and Hearing, 23,* 439–449.

Cienkowski, K. M., & Carney, A. E. (2004). The Integration of Auditory-Visual Information for Speech in Older Adults. *Journal of Speech-Language Pathology and Audiology, 28(4),* 169-172.

Cohen, G. (1979). Language comprehension in old age. *Cognitive Psychology, 11,* 412-429.

Cohen, G. (1981). Internal reasoning in old age. *Cognition, 9,* 59-72.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide *Psychonomic Bulletin & Review, 12 (5), 769-786.*

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and cognitive processes* (pp. 191-211). New York: Plenum.

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

Grant, K.W., & Seitz, P.F. (1998). Measures of auditory-visual integration in nonsense syllables and sentences. *The Journal of the Acoustical Society of America, 4* (4), 2438-2449.

Grant, K. W., Walden, B. E., & Seitz, P. F. (1998). Auditory-visual speech recognition by hearing-impaired subjects: Consonant recognition, sentence recognition, and auditory-visual integration. *Journal of the Acoustical Society of America, 103,* 2677–2690.

Grant, K. W., and Seitz, P. F. (2000). The recognition of isolated words and words in sentences: Individual variability in the use of sentence context. *Journal of the Acoustical Society of America, 107,* 1000–1011.

Hirsh, I. J., Reynolds, E. G., and Joseph, M. (1954). Intelligibility of different speech materials. *Journal of the Acoustical Society of America, 26,* 530–537.

Huffman, C. (2007). The Role of Auditory Information in Audiovisual Speech Integration. Undergraduate honors thesis, The Ohio State University.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-217.

Li, K. Z. H., & Lindenberger, U. (2002). Relations between aging sensory/sensorimotor and cognitive functions. *Neuroscience and Biobehavioral Reviews, 26,* 777–783.

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

Martin, J. S. & Jerger, J. F. (2005). Some effects of aging on central auditory processing. *Journal of Rehabilitation Research & Development, 42*, 25–44.

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

Munhall, K.G., Kroos, C., Jozan, C., & Vatikiotis-Bateson, E. (2004). Spatial frequency requirements for audiovisual speech perception. *Perceptions & Psychophysics, 66 (4),* 574 – 583.

Neely, K. K. (1956). Effects of visual factors on intelligibility of speech. *Journal of the Acoustical Society of America, 28,* 1276-1277.

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

Park, D., Smith, A., Lautenschlager, G., & Earles, J. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging, 11*, 621-637.

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging, 17*, 299-320.

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

Plude, D. J., & Doussard-Roosevelt, J. A. (1989). Aging, selective attention, and feature integration. *Psychology and Aging*, *4*, 98- 115.

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.

Ramakrishna, B. S., Nair, K. K., Chiplunkar, V. N., Ramachandran, V., and Subramanian, R. (1962). *Some aspects of the relative efficiencies of Indian languages*. Ranchi, India: Catholic press.

Ross, L.A., Saint-Amour, D., Leavitt, V., Javitt, D.C. & Foxe J.J. (2007). Do you see what I'm saying? Optimal Visual Enhancement of Speech Comprehension in noisy environments. *Cerebral Cortex, 17(5),* 1147-53.

Rudmann, D.S., McCarley, J.S., & Kramer, A.F. (2003). Bimodal display augmentation for improved speech comprehension. *Human Factors, 45,* 329-336.

Schneider, B.A., Daneman, M., & Pichora-Fuller, M.K. (2002). Listening in aging adults: from discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology, 56(3),* 139–52.

Shannon, R.V., Zeng, F.G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270*, 303-304.

Shannon, R.V., Zeng, F.G., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *The Journal of the Acoustical Society of America*, 104 (4), 2467-2475.

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

Sommers, M.S. (1997). Speech perception in older adults: The importance of speech-specific cognitive abilities. *Geriatric Bioscience, 45*, 633-637.

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults, *Ear and Hearing, 26*, 263-275.

Spehar, B., Tye-Murray, N., & Sommers, M. S. (2008). Intra- versus intermodal integration in young and older adults. *Journal of the Acoustical Society of America*, *123 (5)*, 2858–2866

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

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

Summerfield, A. Q. (1981). "Some preliminaries to a comprehensive account of audio-visual speech perception". In B. Dodd & R. Campbell (Eds.), *Hearing by eye: The psychology of lip reading* (pp. 3-51). London: Erlbaum.

Tillman, T. W., & Olsen, W. O. (1973). Speech audiometry. In J. Jerger (Ed.), *Modern developments in audiology* (pp. 37–74). New York: Academic Press.

Tye-Murray, N., Sommers, M. S., & Spehar, B. (2007). Audiovisual integration and lip-reading abilities of older adults with normal and impaired hearing. *Ear and Hearing*, *28(5)*, 656-668.

Walden, B. E., Busacco, D. A., & Montgomery, A. A. (1993). Benefit from visual cues in auditory-visual speech recognition by middle-aged and elderly persons. *Journal of Speech and Hearing Research*, *36*, 431–436.