SPEECH-IN-SPEECH RECOGNITION: EFFECT OF LANGUAGE UNCERTAINTY

Preeta Singh

13AUD019

This Dissertation is submitted as part fulfillment for the Degree of Master of Science in Audiology University of Mysore, Mysore

All India Institute of Speech and Hearing, Manasagangothri, Mysore- 570006

May, 2015

CERTIFICATE

This is to certify that this dissertation entitled **"Speech-in-Speech Recognition: Effect of Language Uncertainty"** is the bonafide work submitted in part fulfillment for the Degree of Master of Science (Audiology) of the student with Registration No: **13AUD019**. 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.

Dr. S. R. Savithri

Director
All India Institute of Speech and Hearing,
Manasagangothri, Mysore- 570 006.

Mysore,
May, 2015.

CERTIFICATE

This is to certify that this dissertation entitled **"Speech-in-Speech Recognition: Effect of Language Uncertainty"** has been prepared under my supervision and guidance. It is also certified that this has not been submitted earlier to any other University for the award of any other Diploma or Degree.

Ms. Geetha C

Guide
Lecturer in Audiology,
Department of Audiology,
All India Institute of Speech and Hearing,
Manasagangothri, Mysore- 570 006.

Mysore,
May, 2015.

DECLARATION

This is to certify that this Master's dissertation entitled **"Speech-in-Speech Recognition: Effect of Language Uncertainty"** is the result of my own study under the guidance of Ms. Geetha C, Lecturer in Audiology, Department of Audiology, All India Institute of Speech and Hearing, Mysore, and has not been submitted earlier in other University for the award of any Diploma or Degree.

Mysore,
May, 2015.

Register No: 13AUD019

ACKNOWLEDGMENTS

Firstly I dedicate this section to all the readers who are more interested in reading the dedication section rather than the content written on the rest of the pages. Thank you and don't worry, I totally understand….

I sincerely thank my guide Mrs Geetha C. mam . Mam if your reading this I am sorry I have been irresponsible, lazy, behind schedule, not serious, not dedicated to my work (time, space, and modesty compel me to stop here). In spite of this you have never scolded me and you always had that smile on your face whenever I encountered you in the college. I am grateful for your patience mam.

And mam, I haven't used 'BUT' and 'ALSO' in this section ☺
Next to my parents who have taught me about life's decision making and my friends who handle the consequence of the same. (P. S. I haven't learnt them right. Don't worry mom and dad…Hopes still there I am still 24 ….long way to go).

Raja rajan sir and kumaran sir, you guys have been a source of encouragement and motivation, by often asking "have you started with your dissertation, you are still here I thought u left AIISH, finished with your dissertation, remember your guide or have you forgotten, are you planning to do your dissertation?" thank you.

How can I forget you, if not for you this thing wouldn't exist..... thank you MS Word, it was a pleasure knowing you, you made my life easier (no wonder word is called Ms. Word). Also thank you... Youtube, who helped to narrow the gap between me and Ms Word, you were a continuous source of knowledge, entertainment and all that I am know. Thank you (tears).

It would be a sin if I don't thank, Mendeley, you showed tantrums while extracting the input hence, feeding input initially did shot circuit me a little, but your output is simply awesome. And your search engine is mind blowing.....sincere advice to juniors, plz use it u ll love it.

At last the library, which has been my source of shelter for hangout s with friends and social get-togethers and also knowledge since internship, there are no words to convey my heartfelt gratefulness for everything. Thank you (more tears, I ll miss you).

My friends would tear this page if I don't mention them here, so thank you guys abadmpsr, rekha, panki, pooja, shwetha, bhuvana, vino, vini, husna, padma, rohan, akshay, bibek, heman, kapali, cv, giri, guna, kiran, pg, shreyank, divya for adding memories during these years. In spite of having bad memory, I promise I ll take extra effort to remember these moments. Thank you (smiley face).

ABSTRACT

Speech-in-speech recognition is a common phenomenon, in a multilingual country like India. This phenomenon co-exists with unpredictable change in the background language, especially with English language. The present study aims to study the influence of native vs. non-native language babble, language uncertainty of the babble and role of signal-to-noise ratio on the Kannada sentence recognition. Forty Kannada-English Bilinguals were subjected to Kannada sentence recognition task in the presence of two-talker Kannada babble, two-talker English babble, two-talker mixed (Kannada + English) babble and two-talker interleaved (unpredictable change in background language) at +3 dB SNR, 0 dB SNR and -3 dB SNR. The level of presentation of target sentence was 75 dB SPL. Comparison of the recognition scores across four babble conditions revealed significantly better scores in presence of two-talker English babble when compared to two-talker Kannada babble, implying the release of masking in the presence of non-native language babble. Further, the presence of language uncertainty in the two-talker interleaved condition reduced the scores when compared to two-talker mixed condition. The influence of other variables like F_0 and LTASS differences influencing the speech-in-speech recognition was negligible and the release of masking can be attributed to the linguistic mismatch of the target and the babble. In addition the SNR was found to be a factor influencing speech-in-speech recognition.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Speech recognition in the presence of multiple language background and multiple talkers is a common phenomenon. The impact of such noise on the speech intelligibility depends on the acoustic characteristic of the noise, semantic characteristics, meaningfulness, linguistic and phonetic content, native or non native language, number of background talkers, gender of the speaker and language uncertainty (Dillon, 1983; Freyman, Helfer, & Balakrishnan, 2007; Kalikow, Stevens, & Elliott, 1977; Van Engen, 2010) among others.

Research studies have tried to examine the effect of these factors on the speech recognition scores (SRS) in presence of multi-talker babble (MTB), also known as speech-in-speech. The use of MTB which contains linguistic information would compete with the recognition of the target rather than cause spectral or temporal masking, in other words, speech recognition in the presence of MTB causes less energetic masking (EM) and more informational masking (IM). Studies have shown that the amount of IM can differ when the relationship between the target and masker stimuli is varied (Brouwer & Bradlow, 2014; Brouwer, Engen, Calandruccio, & Bradlow, 2012; Carhart, Tillman, & Greetis, 1969; Engen, 2012).

Several studies have been done to assess the effect of linguistic and phonetic similarity between the languages and listeners familiarity of the language in speech-in-speech recognition task. Authors have shown that unfamiliar language masker leads to more masking release compared to native language or a familiar language (Brouwer et al., 2012; Jain, Konadath, Vimal, & Suresh, 2014). Further, decreasing similarity between the target and masker has been found to decrease the SRS. For example, the English language sentence recognition in two-talker babble of Mandarin, resulted in better SRS as Mandarin is dissimilar to English, when compared to that of English babble (Calandruccio, Brouwer, Van Engen, Dhar, & Bradlow, 2013).

Unlike the above findings, studies on speech-in-speech recognition of Indian languages have shown varied results. Anitha (2003) studied the effect of ten-talker babble of different languages (Kannada, Hindi, Malayalam) on SRS of Kannada phonetically balanced words in 40 Kannada speakers. SRS in the presence of ten-talker babble was poorer when compared to SRS in speech noise, whereas no significant difference was obtained across different language babbles. The authors' opinion about the result was that the major factor for masking was the spectrum of the babble not the linguistic or semantic content in the babble.

However, in the above study, the number of talkers in the babble was too large to assess the effect of linguistic context. Nonetheless, results of the study done by Jain et al. (2014) showed that the Kannada speakers performed better in the presence of Kannada

MTB than non-native Malayalam MTB. The authors speculate the reason for the poorer performance by Kannada group in the presence of non-native MTB as attention based factors, where the participants are distracted by non-native language, leading to poorer scores in presence of non native MTB.

The above mentioned studies studied the effect of different Indian languages, whereas influence of a non-Indian language babble on recognition of Indian language is studied less in the literature. A study done by Vineetha, Suma, and Nair (2013) reported no difference in the SRS obtained by Kannada-English bilinguals in the perception of Kannada words in the presence of four-talker Kannada and English babble at 5, 0,-5,-10 and -20 dB SNR.

Various factors could have influenced the difference obtained among studies. The factors could be difference in type of test material (words, sentences), number of takers in the MTB, language familiarity and similarity between the target language and the background language.

Another factor which could affect the speech in speech masking is uncertainty of the language of the babble. Various studies using non-speech stimuli have studied the effect of uncertainty of the masker tone. They reported that the uncertainty adversely affects thresholds of tonal target (Durlach, Mason, Shinn-Cunningham, et al., 2003; Kidd, Mason, & Arbogast, 2002; Neff, Dethlefs, & Jesteadt, 1993; Watson, Kelly, & Wroton,

1976). Similarly, uncertainty can also be induced by changing the language of babble. The study done by Brouwer and Bradlow (2014) investigated the effect of variation in the target-background language relationship (contextual variation) on the speech-in-speech recognition on 48 native English speakers. Two experiments were carried out to test the SRS of English sentences. The first experiment had English sentence recognition in the presence of two-talker English babble, two-talker Dutch babble and interleaved language condition (interleaved English and Dutch babble). In the interleaved condition, the English and Dutch Babble were interleaved such that it switched languages 30% of the time. They reported poorer scores in the condition where babble was a mixture of both the languages and this was due to IM.

On the contrary, studies which have induced uncertainty in terms of spatial location of masker or number of talkers and gender of the talker of babble (Brungart & Simpson, 2004; Freyman et al., 2007; Jones & Litovsky, 2008) reported not much of an effect of uncertainty on the SRS.

The above experiments studying IM have been studied in various SNRs. They found that SNR does affect the amount of IM and EM. The EM increases with decreasing SNR (Calandruccio, Dhar, & Bradlow, 2010), whereas, IM was found to influence the most when linguistic content of both target and babble are audible, and to influence the least at easy SNR and difficult SNR where the target or the babble is least audible (Engen & Bradlow, 2007; Van Engen, 2010).

1.1. Need of the study

Due to the advent of globalization, the phenomenon of bilingualism/multilingualism has become common. In a multi-lingual country like India, English has coexisted with the indigenous languages since the British rule, and now it is one of the official languages of the country. English is used as a language of education in most schools, colleges and universities. In the domestic setting, along with English, other local language is also used and spoken. Hence, in the classrooms or in different social settings, the mixture and switching of codes have become inevitable (Harini & Shyamala, 2008).

The earlier studies on Indian languages have studied only the influence of an Indian language on another Indian language. These studies have used MTB either with 4, 5, 6, or 10-talker babble (Anitha, 2003; Jain, Kodanath, Vimal, & Suresh, 2014). In Anitha's study, the results revealed no significant difference in terms of language of the masker. This could be because of the number of speakers used in the babble. When the number of talkers is more, the background signal resembles the speech noise and hence, the semantics is lost. Therefore, there is more influence of EM than IM (Carhart, Johnson, & Goodman, 1975). In a study done by Van Egan and Bradlow (2007), significant difference was observed for two-talker babble than six-talker babble. Similarly, Freyman, Balakrishnan, and Helfer (2004) compared two-talker babble versus 3, 4, 6 and 10-talker babble and reported that maximal informational masking occurs in

two-talker babble background. Hence, it is important to study the effect of speech babble with lesser number of talkers in order to test the influence of language content of the masker.

In addition, there is evidence that the performance decreases if the language of the babble changes unpredictably. In Indian context, there is a lot of code mixing and code switching, especially with English words (Harini & Shyamala, 2008; Mathew, 2012) and hence, the unpredictable change in the background language is a common phenomenon. Hence, it is important to study the effect of English language background on speech recognition and the effect of unpredictable changes in the language. Further, the role of SNR cannot be neglected as it also plays an important role in influencing the amount of EM and IM in speech-in-speech recognition task (Calandruccio et al., 2010; Engen & Bradlow, 2007; Van Engen, 2010).

1.2. Aim of the study

The aim of the present study was to study the effect of native and non-native language babble on SRS of Kannada sentences, to study the effect of language uncertainty on the SRS of Kannada sentences, and to study the role of SNR on speech-in-speech recognition in Kannada-English bilinguals.

1.3. Objectives of the study

- To obtain SRS of Kannada sentences in the presence of two-talker Kannada babble, two-talker English babble, two-talker mixed (Kannada+ English) babble and two-talker interleaved (background language switching between Kannada and English unpredictably) babble at +3 dB SNR, 0 dB SNR and -3 dB SNR.
- To compare the SRS in presence of two-talker Kannada babble and two-talker English babble.
- To determine the effect of language uncertainty by comparing the SRS in the presence of two-talker interleaved babble with other two-talker babble conditions.
- To determine the effect of SNR on the SRS across the four two-talker babble conditions.

Chapter 2

Literature Review

Everyday listening situation demand us to extract the information from speech signals that are masked by competing maskers. Various researchers have studied the effect of various types of competing maskers like narrow band noise, broad band noise, pink noise, speech spectrum noise MTB on speech perception (Anitha, 2003; Carhart, Johnson, & Goodman, 1975; Carhart, Tillman, & Greetis, 1969; Cherry, 1953; Cullington & Zeng, 2008). Out of all, speech babble has been found to be more effective in masking as it enables us to simulate everyday listening situation and predict the outcomes (Carhart et al., 1969; Chen, Li, Li, Wu, & Moore, 2015; Cherry, 1953; Cullington & Zeng, 2008; Engen, 2012; Hall III, Grose, Buss, & Dev, 2002).

These competing noises may or may not have linguistic information. Based on the acoustic and semantic characteristics of the masker, Carhart (1969) and Pollack (1975) have made attempts to distinguish between energetic and informational masking in speech-in-speech recognition task, that is, speech recognition in the presence of speech as the background noise. Energetic masking (EM) refers to masking at the peripheral auditory system and is related to the audibility of the target signal. This causes partial or complete loss of information due to spectral and temporal overlap between the masker and the target (Brungart, Simpson, Ericson, & Scott, 2001). Informational masking (IM) refers to the masking beyond what contributes to energetic masking, also associated as

central masking. It is not about the presence of overlap of the signals, rather it is a competitive aspect interfering in the later processing of speech signal. In IM, some amount of EM also occurs (Brouwer et al., 2012; Brungart et al., 2001; Durlach, Mason, Kidd, et al., 2003). However, studies have shown that EM is less prominent in IM (Brungart, Chang, Simpson, & Wang, 2006; Chen et al., 2015). They found intelligible speech babble or synthesized speech signals give rise to greater masking than speech modulated noise or reverse speech.

The aim of the present study was to determine the effect of language, to determine the effect of language uncertainty, and to study the role of SNR using two-talker babble on speech-in-speech-recognition. Hence, the literature was reviewed and presented under the following headings:

2.1 Effect of language of the masker on speech recognition

- 2.1.1Native vs. non-native language
- 2.1.2 Familiar vs. unfamiliar language
- 2.1.3 Effect of accent

2.2 Role of uncertainty in speech-in-speech recognition task

2.3 Other factors influencing speech-in-speech recognition

- 2.3.1 Number of talkers
- 2.3.2 F_0 and gender of the talker
- 2.3.3 Attention and memory

2.1. Effect of language of the masker on speech recognition

Speech perception in various types of background noise has been reported to be affected differentially in the presence of background signal with linguistic content than without linguistic content (Chen et al., 2015). Studies have assessed speech perception of native language in the presence of native language, non native language, foreign language (unfamiliar) or accented speech as background noise (Brouwer et al., 2012; Calandruccio et al., 2013; Calandruccio & Zhou, 2014; Cooke, Lecumberri, & Barker, 2008; Jain, Kodanath, et al., 2014; Vineetha et al., 2013). Most of the studies have reported an improvement in speech recognition when the languages of masker and target are mismatched (Brouwer et al., 2012; Calandruccio et al., 2013; Engen & Bradlow, 2007), however, few studies have reported no difference (Mattys, Brooks, & Cooke, 2009; Vineetha et al., 2013).

The reason for influence of type of language masker is owed to two possible reasons. One is linguistic similarity (Brouwer et al., 2012), that is, more similar the target and masker speech harder it is to segregate into two separate streams. The second reason is the familiarity of the language of the babble, that is, there has been an improvement in the recognition due to inability of the listener to understand the masker speech, leading to lesser confusion or interference in the processing of the target speech (Brouwer et al., 2012; Van Engen, 2010).

2.1.1. Native vs. non-native language

Researchers have studied the linguistic contributions of native vs. non native language on speech perception. They have seen the effect of native language perception by bilinguals in the presence of non-native and native language babble (Calandruccio & Zhou, 2014; Engen & Bradlow, 2007; Jain, Kodanath, et al., 2014) and other set of studies reported recognition of non-native speech perception in presence of native and non-native babble (Brouwer et al., 2012; Mattys, Carroll, Li, & Chan, 2010; Van Engen, 2010; Vineetha et al., 2013).

Engen and Bradlow (2007) studied the recognition of English sentences by English-Mandarin bilinguals in the presence of two-talker English and Mandarin babble. Similarly, Calandruccio and Zhou (2014) studied English sentence recognition in English-Greek bilinguals in the presence of two-talker babble of English and Greek. Both the studies reported significant improvement in the speech recognition scores when linguistically mismatched target-masker condition was present even though the listener was able to obtain meaningful linguistic information from the masker speech.

Van Engen (2010) studied second language recognition in the presence of first language (L1) and second language (L2) two-talker babble. He reported that the non-native English speakers had more difficulty in recognition of English target sentences in the presence of English two-talker babble when compared to Mandarin two-talker babble. Further, the amount of release in the presence of two-talker Mandarin babble was less in

Mandarin listeners. Similarly, Brouwer et al. (2012) studied English sentence recognition in the presence of two-talker English and Dutch babble in Dutch-English bilingual listeners. The results revealed that there was a release from masking when the competing speech was different from target speech i.e., better scores of L2 recognition in the presence of L1 masker. The above finding indicates that both similarity between the target and masker, and the language experience of the listener contribute to the interference experienced during speech perception in noise.

On the contrary, Mattys et al. (2010) studied English phrase recognition in Cantonese-English bilinguals in the presence of competing speech in Cantonese and English language (one-talker). They reported no difference in the scores and postulated that the listeners in their study relied on the acoustic cues of the target to recognize and not on the lexical semantic competitor of the background language.

The difference in the finding in the literature discussed above could be due to difference in the type of test materials. All the above studies except the study done by Mattys et al. (2010) used sentences as target stimuli and have two-talker babble, and used open recognition task. Whereas, Mattys et al. (2010) used a competing talker as a masker and two word phrases as target stimuli and the response mode was closed set. Mattys et al. (2010) stated that they could have got the difference in the performance if there was higher cognitive load on the task, thereby slowing the cognitive process and reducing the speech perception ability in the non- native language.

Few studies have been done on speech-in-speech recognition in Indian languages and they reported no such benefit from linguistic mismatch (Vineetha et al., 2013). Vineetha et al. (2013) reported no difference in the scores of Kannada-English bilinguals in the perception of Kannada words in the presence of four-talker Kannada and English babble at 5, 0,-5,-10 and -20 dB SNR. The reason for this could be that the language proficiency of the non-native language could be equal to the native language. The other reason for no significant difference in the study done by Vineetha et al. (2013) could be scores reaching ceiling effect at all SNRs except at -20 dB SNR, and the number of talkers being high. Hence, the present study aimed at using a MTB with lesser number of talkers i.e. two-talker babble to investigate the effect of English on speech recognition of Kannada.

2.1.2. Familiar vs. unfamiliar language

Engen and Bradlow (2007) studied English sentence recognition in native English monolingual listeners in the presence of two-talker babble in English and Mandarin. They reported that native English listeners received a release of masking in English recognition in presence of two-talker Mandarin versus English background babble. Similarly, such release of masking has been reported in other languages like Dutch (Calandruccio et al., 2013), Croatian (Calandruccio et al., 2010) and Spanish (Lecumberri & Cooke, 2006). The above studies have looked into the influence of native versus unfamiliar language.

Further, studies have also investigated the masking release for foreign speech maskers that vary in the degree of linguistic similarity to the target. Calandruccio et al. (2013) investigated target-masker linguistic similarity in three conditions: identical target-masker (English in English recognition); linguistically close target-masker (English in Dutch recognition); and linguistically distant target masker (English in Mandarin recognition). English and Dutch belongs to the same linguistic family of Indo-European, whereas, Mandarin belongs to the Sino-Tibetan family. They reported that the performance of monolingual English speakers to be most affected in the presence of English masker followed by Dutch and least by Mandarin. Hence, smaller masker release is observed when a linguistically similar language is used to mask the target.

However, in the above study, the authors reported that this cannot be wholly accounted to the IM as the difference in the spectral properties does exist between languages which could contribute to less or more EM along with IM. In order to find experimental evidence that the release from masking is only due to IM not only to EM, the LTASS of all the three two-talker babble was extracted and two types of noise were generated. One was steady state spectrally matched noise and the other one was temporally modulated white noise matched to the three two-talker babble. The scores obtained in the spectrally matched and temporally modulated noise was significantly better than two-talker English babble and two-talker Dutch babble, whereas no difference was found between the matched noises and two-talker Mandarin babble. They concluded that Mandarin babble masking was solely due to EM rather than linguistic dissimilarity,

whereas, contribution of EM in the presence of English and Dutch babble was less and overall masking can also be attributed to the IM not alone EM.

Studies on influence of unfamiliar Indian languages have shown varied results. Anitha (2003) studied the effect of ten-talker babble of different languages (Kannada, Hindi, Malayalam) on SRS of Kannada phonetically balanced words on 40 Kannada speakers. SRS in the presence of ten-talker babble was poorer when compared to SRS in speech noise, whereas no significant difference was obtained across different language babbles. The authors' opinion about the result was that the major factor for masking was the spectrum of the masker not the linguistic or semantic content in the babble. Hence, using a less number of talker babble could have produced IM.

However, the results of the study done by Jain et al. (2014) showed that the Kannada speakers performed better in the presence of Kannada six and ten-talker babble than non-native Malayalam six and ten-talker talker babble. The authors speculate the reason for the better performance by Kannada group in the presence of native MTB as cognitive factors, where the participants get distracted by non-native language, leading to poorer scores in presence of non native MTB.

The above mentioned studies (Anitha, 2003; Jain, Kodanath, et al., 2014) report no release of masking in the presence of non-native language. Various factors could have led to these results. The numbers of talkers used were six-talker and ten-talker babble.

This could have led to more of EM than IM, leading to no significant release of masking across different language of babble. It could also be that the languages chosen for the experiment are from the same family causing no difference in the masking. If the LTASS of the different language babble or F_0 of the talkers were compared in the above studies, a proportional conclusion could be drawn about the amount of IM. Hence, the present study has compared the F_0 and LTASS across the two-talker babble conditions and the target sentences.

2.1.3. Effect of accent

The accent influences the non-native language acoustically and phonetically (Sirsa & Redford, 2013). Calandruccio et al. (2010) reported that the English sentence recognition scores in native English monolinguals were significantly better in the presence of Mandarin accented English babble than native English two-talker babble. This was due to influence of Mandarin accent on English language causing a release in masking. They also reported a larger LTASS difference between accented babble and the English babble and reported lesser difference between accented babble and Mandarin babble.

2.2. Role of uncertainty in speech-in-speech recognition task

One of the factors with respect to IM using speech stimuli is 'uncertainty' which is explored less in the past. The word 'uncertainty' here relates to the less predictable nature of the masker. By inducing the uncertainty in the task, it becomes difficult as it alters the attention of the listener to the unpredictable change of the babble. The uncertainty could be in terms of spatial location, language, content, gender of the MTB.

Brouwer and Bradlow (2014) studied varying levels of masker uncertainty of language. Forty eight native American-English listeners were subjected to three experimental tasks of open set English sentences recognition in the presence of two-talker babble of English and Dutch. In the first condition, they were subjected to two-talker English babble and Dutch babble. In the second condition, they were subjected to recognition in the presence of interleaved (English and Dutch) two-talker babble. In the third condition, the interleaving was done with and without English two-talker babble. In both second and third condition, the sentence recognition scores were calculated separately in the presence in mixed two-talker English babble and mixed two-talker Dutch babble. In each condition, 68 sentences were presented. They reported no change in the scores of English in two-talker English babble interleaved condition when compared to two-talker English babble. However, they reported an increase in the English scores in two-talker Dutch interleaved condition when compared to two-talker Dutch babble condition. The interleaved condition that was with and without two-talker Dutch babble enhanced the scores of the English in Dutch babble. Brungart and Simpson

(2004), Freyman, Helfer, and Balakrishnan (2007) and Jones and Litovsky (2008) have measured the effect of predictability of the number and locations of speech maskers speech recognition scores and the spatial release of masking in a cocktail party scenario. The study by Brungart and Simpson (2004) used a dichotic task and the variability was introduced in terms of freezing the masker in terms of talker, content or both. The above mentioned studies report a small reduction in performance when uncertainty was introduced; however, the reduction was insignificant.

2.3. Other factors influencing speech-in-speech recognition

2.3.1. Number of talkers

The number of talkers is one of the major variables that affect the influence of IM. The masking effect of MTB is reported to be the strongest when there are two to four talkers (Broersma, 2012; Cullington & Zeng, 2008; Engen & Bradlow, 2007; Hall III et al., 2002; Simpson & Cooke, 2005). This effect becomes less prominent as the number of talkers further increases and saturates when there are twelve or more talkers as it reduces the semantic interference on perception of target speech (Cullington & Zeng, 2008; Simpson & Cooke, 2005).

Miller's classic study of masking (Miller, 1947) was the first to investigate the effect of number of talkers in the MTB. Miller measured the intelligibility of words with 1, 2, 4, 6, and 8-talker speech babble. He found that the difference in masking effect for a single talker over two talkers was equivalent to an SRT difference of about 8 dB.

Babble with 4, 6, 8 talkers produced an additional 3-4 dB of masking over the two-talker condition. He found a monotic decrease in performance as the number of talkers increased. Similarly, Carhart et al. (1969) found that a two-talker masker was more effective than continuous white noise in masking the recognition of spondee words.

The effectiveness of masker reduces even when the number of talkers are too less. Carhart, Johnson, and Goodman (1975) found a single-talker masker resulted in less masking than speech noise and suggested that the single-talker masker was relatively ineffective because it contained abundant temporal 'windows' during which the target word could be processed. Similar results were also found by Brungart et al. (2001).

Carhart et al. (1975) measured the intelligibility of spondees in the presence of speech babble with 1, 2, 3, 16, 32, 64, 128 talkers, and modulated noise. The difference between the speech babble and modulated noise was found to be 6.2 dB with the one talker and 7.2 dB with the two-talker and the maximum (9.8 dB) was reached with three-talkers. Thereafter, the difference decreased, stabilizing at about 3 dB with 64 talkers. Hence, they concluded that the two, three and four-talker babbles to be effective for IM.

Hoen et al. (2007) also found that there was non-monotic increase in speech recognition scores as number of talker increased, i.e., significantly poorer scores in four-talker babble when compared to six or eight talker condition, and no difference between

six and eight talker babble was found. They concluded that lexical masking occurs for low number of talkers and diminishes with more talkers.

Similarly, Boulenger, Hoen, Ferragne, Pellegrino, and Meunier (2010) measured the recognition of target words in terms of reaction time using real-time word recognition paradigm in presence of MTB with 2, 4, 6, and 8 talkers. They reported a decrease in performance as the number of talkers increase. They found significantly faster reaction times in two-talker condition and the reaction time reduced systematically as number of talker increased from four to eight, but, it was not significant. Hence, they stated that with increase in number of talkers there was increased spectral and temporal saturation.

To conclude, majority of the studies on IM suggested to use two, three or four-talker babble (Brouwer & Bradlow, 2014; Brouwer et al., 2012; Calandruccio, Brouwer, Van Engen, Dhar, & Bradlow, 2013; Calandruccio & Zhou, 2014; Van Engen, 2010; Wu et al., 2015 among others). Higher number of talkers leads to more EM and less IM as the lexical and semantic content of the masker would be less intelligible to compete with the target.

2.3.2. F_0 and gender of the talker

Studies have revealed that difference between the F_0 of masker and target leads to lesser overall masking. Brungart et al. (2001) studied the speech in speech masking using two, three and four-talker babbles varying in terms of talker and gender. They revealed

that the best scores were obtained when the talker of target and masker were different and the poorest were obtained when the same talker spoke both masker and target. The authors speculate that the overall masking in the above conditions could be due to more EM as both target and masker are spoken by the same talker. It could also be due to more IM occurring due to qualitative similarities between the target and masker leading to more interference at the lexical-semantic recognition of target (Brungart et al., 2001; Cooke et al., 2008). Similar finding have been reported by Chen et al. (2015) and Wu et al. (2015).

With respect to gender, they reported of poorer scores when the target and masker were spoken by the same gender, and results showed deterioration of score when one of the talker in the masker was replaced by opposite gender talker. He reported that this could be because of the distraction of the opposite gender voice which induced divided attention towards to masker and target and hence reduced scores (Mattys et al., 2009; Mattys & Wiget, 2011).

The study done by Cullington and Zeng (2008) reported that the maskers recorded by female talkers are less susceptible to masking when compared to male and child talkers. Further, Bradlow and Bent (2002) also reported that the target recorded by female talker is less susceptible to noise than male recorded target speech.

On the whole, the vocal characteristic of the speaker of the target and the masker affects IM differentially by influencing both peripheral processing (peripheral masking) and higher level processing (attention factors, lexical and semantic interference). Hence, in the present study, the gender of the talkers of the target and babble was kept constant throughout the conditions.

2.3.3. Attention and memory

The attention could vary due to various reasons in normal listeners. For instance, the listeners' attention could be divided by the presence of a distracter and these distractions could affect IM (Cooke et al., 2008; Garcia Lecumberri & Cooke, 2006). The distractor could be the change in the semantic content or could be cross modality distracters, for e.g., multi tasking conditions which reduce attention capacity due to divided attention. Kahneman (1973) (as cited in Mattys, Davis, Bradlow, & Scott, 2012) reported that attention resources get depleted, if the task has to be concurrently executed with a speech task. Along with attention resources, memory demands can also influence the performance. For instance, listening to various talkers sequentially, is shown to engage more working memory resources than listening to a single target (Nusbaum & Morin, 1992 as cited by (Mattys et al., 2012). Authors reported that the representation of speech maintained in working memory is likely to be phonological (Baddeley, 1986 as cited in Mattys et al., 2012) which means that reduced memory capacity affects sub-lexical processes as well. Hence, reduced attention or divided attention can affect sub-lexical and lexical processes thereby affecting speech perception.

After reviewing some of the factors influencing speech-in-speech recognition, it can be concluded that IM is a higher level processing which is affected by peripheral processes and also parallel higher level processing. Some of the main factors are language of the babble and the uncertainty in the language of the babble. Studies on speech-in-speech recognition of Indian languages have shown varied results and as per the author's knowledge, influence of English on Indian language using two-talker babble has not been explored in the past. Hence, the present aims to determine the effect of native and non-native language babble on SRS of Kannada sentences, to study the effect of language uncertainty on the SRS of Kannada sentences and the role of SNR in Kannada-English bilinguals.

Chapter 3

Method

3.1. Participants

A group of 40 participants in the age range of 18-30 years (mean = 24.9; SD = 3.153; 20 males and 20 females) were selected based on the following criteria.

3.1.1. Selection criteria:

- All the listeners were native speakers of Kannada with English being the second language exposed at least from 5^{th} grade.
- All the participants had a minimum of X^{th} grade education in English medium School and Kannada as the second language.
- The participants were considered as Kannada-English bilinguals if they obtained a score of two or above in English in the International second language proficiency rating scales (ISLPR) developed by Wylie (2006).
- They had hearing sensitivity less than or equal to a four frequency puretone average (at 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz) of 15 dB HL (WHO, 2008).
- They had either 'A' or 'As' type of tympanogram with ipsi and contralateral reflexes present within 100 dB HL (Jerger, 1970).
- They had SPIN scores of 60% or above at 0 dB SNR.
- They had presence of oto-acoustic emissions in both ears.

- They reported of no other otological or neurological symptoms, and no other speech or language problems.

3.2. Instrumentation

- Routine audiological evaluation was carried out using a calibrated two channel diagnostic audiometer (GSI-61) for puretone threshold estimation and speech audiometry. TDH-39 headphones coupled with MX/41AR ear cushions and Radio ear B-71 bone vibrator were the transducers used.
- Calibrated GSI-tympstar (Grason-Stadler Incorporation, USA) clinical immittance meter was used for tympanometry and acoustic reflex measurement.
- For the experimental task, a HP Notebook with software Adobe Audition v3 was used for recording, signal generation, processing and mixing. MOTU Microbook II, an audio interface, was connected to the notebook and was used to record the Kannada and English passage for the construction of the babble.

3.3. Material used

- The speech recognition thresholds were obtained using Kannada paired words developed at the Department of Audiology.
- The Kannada Phonemically Balanced word list developed by (Yathiraj & Vijayalakshmi, 2005) was used to find SRS for routine audiological evaluation and SPIN testing.

- Kannada sentence list developed by Geetha, Kumar, Manjula, and Pavan (2014) was used to find SRS for sentences in quiet and in presence of two-talker babble. This test consists of twenty five homogeneous lists with ten sentences under each list.
- Kannada sentences from standardized passage of 300 words in Kannada (developed by Savithri, S and Jayaram (2005)) and English sentences from the standardized English rainbow passage (Fairbanks, 1960).

3.4. Test environment

The test was carried out in an air conditioned sound treated double room suite with ambient noise levels within permissible limits (ANSI, 1991).

3.5. Procedure

The test procedure was carried out in two different phases:

Phase 1: Development of two-talker English babble, Kannada babble, interleaved babble and mixed babble

Phase 2: Measurement of SRS in the presence of four two-talker babble conditions

3.5.1. Phase 1: Development of two-talker Kannada babble, English babble, interleaved babble and Mixed babble in 3 different SNR condition.

For the construction of the four two-talker babble, Kannada sentences from standardized passage of 300 words in Kannada developed by Savithri and Jayaram (2005) were recorded by two native speakers (one male and one female). The recording microphone was placed 20 cm in front of the mouth of the speaker and the speaker was asked to articulate the words clearly. The passage was recorded digitally in a sound proof booth using MOTU Microbook II and mixed using Adobe Audition version 3 at a sampling rate of 44.1 kHz with 24-bit resolution. The same procedure was used to record English sentences using the standardized English rainbow passage (Fairbanks, 1960).

The sentences recorded in Kannada by each talker were concatenated, with no silent intervals between the sentences to create a string of sentence. The order of concatenation was varied such that the sentences spoken by male and female talker were not the same at a point of time. This was also done for English sentences spoken by one male and one female. For each of the four two-talker babble conditions and three SNRs (4*3), twelve strings of sentences of each talker (one male and one female native Kannada speakers, and one male and one female non-native English speaker) were created. The sentences were normalized before and after concatenation. The above procedure was done using Adobe audition version 3.

Kannada sentences spoken by two speakers' one male and one female were mixed to create two-talker Kannada babble. Similarly, two-talker English babble was constructed. For two-talker Mixed babble, English sentences spoken by one speaker and Kannada sentences spoken by another speaker of the opposite gender was mixed. Whereas the two-talker interleaved babble was constructed by interleaving (changing) the two-talker Kannada babble and two-talker English babble.

3.5.1.1. Construction of two-talker Kannada babble, English babble and mixed babble

The two-talker Kannada babble was constructed by mixing two strings of Kannada sentences spoken by one male and one female talker using Adobe audition version 3. Out of the twelve strings of Kannada sentences spoken by two talkers, three strings of sentences of each talker were used to create three sets of two-talker Kannada babble for the three SNR condition. Similarly three strings of English sentences of each talker (one male and one female) were mixed to create three sets of two-talker English babble.

The two-talker mixed babble was constructed by mixing one string of Kannada sentences and one string of English sentences such that both are spoken by opposite gender and not the same gender. For example, one string of Kannada sentence spoken by

male talker was mixed with one string of English sentence spoken by female talker. For the three SNR condition, three sets of two-talker mixed babble were created.

The two-talker Kannada babble, English babble and mixed babble were mixed with the target standardized Kannada sentences spoken by a female speaker. Each set of two-talker babble was mixed with ten target sentences such that the onset of babble precedes the onset of word by 1second and continues till 1 second even after the end of the last sentence. The inter target interval of 6 second was given in order to give sufficient time for the oral response. The level of the target sentences was fixed at 75 dB SPL and level of the babble tracks were varied to produce a target to babble ratio of +3 dB SNR, 0 dB SNR and -3 dB SNR.

Hence three sets of each two-talker Kannada babble, English babble and mixed babble mixed with the different set of target sentences were created to test SRS of Kannada sentences in these SNR.

3.5.1.2. Construction of two-talker interleaved babble

In order to construct two-talker interleaved babble, three sets of the two-talker English and two-talker Kannada babble constructed in section 3.5.1.1. were used. One set of two-talker Kannada and one set of two-talker English babble were interleaved such that it changes from Kannada to English or vice versa 30% of the time. The duration of interleaving condition was altered with respect to duration of the target sentences. For

example, for the first two target sentences, the background language was two-talker Kannada babble and for the next three target sentences, the background changed to two-talker English babble. The next three sentences had two-talker Kannada babble and the last two sentences had two-talker English babble. Hence, for one set of interleaved condition, the language of the babble changed thrice for ten target sentences and the pattern of language change or uncertainty was different for each of the SNR condition.

The three sets of interleaved babble condition were constructed and group normalized along with the stimuli constructed in Section 3.5.1.1., and the two-talker babble was mixed with the target sentences in the same procedure as mentioned in section 3.5.1.1.

The sentences were not repeated for any of the conditions to avoid practice effect, and the order of presentation of each condition and the SNR were randomized to eliminate the order effect.

Finally, the three sets of each two-talker Kannada babble, English babble, mixed babble and interleaved babble with target sentences were constructed in three SNR and were presented to the listeners as illustrated in Figure 3.1.

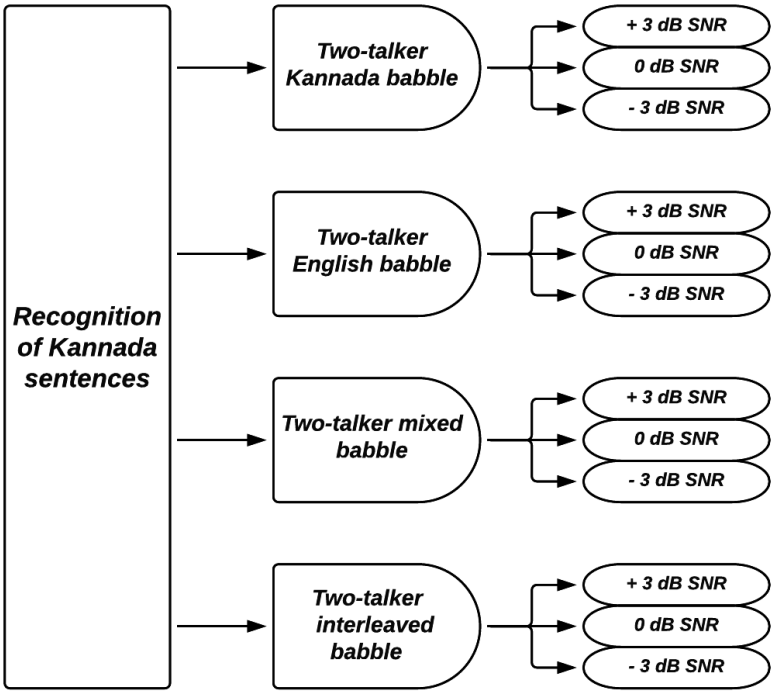

Figure 3.1: Illustration of the test conditions.

3.5.2. Phase 2: Measurement of SRS in the presence of four two-talker babble conditions

The participants were seated in a comfortable chair and the sentences were presented through HP notebook and calibrated i-Ball circum aural headphones diotically. All the participants were instructed to repeat the target sentences orally. The examiner scored the words identified correctly in the score sheet containing the target sentences. Every sentence in the sentences list had four key words and scoring was based on the correct identification of the key words in the sentence. The maximum number of keywords for each SNR condition was 40.

3.6. Statistical analysis

The scores obtained (out of 40) were entered in the SPSS version 17 for each SNR in each condition and non-parametric tests were used to analyze the data. The Friedman test was used to find the significant main effect of different babble and these SNRs on the SRS, if any. Further Wilcoxon signed rank test was done to find the significant difference among different pair of babbles, SNR's, if any.

Chapter 4

Results

The objectives of the present study were to develop two-talker Kannada babble, two-talker English babble, two-talker mixed babble and two-talker interleaved babble, and to compare the SRS of Kannada sentences in Kannada-English bilinguals in the presence of the above babbles in three SNRs.

4.1. Development of four different speech babbles

Four different speech babbles were developed, that is, Kannada babble, English babble, mixed babble and interleaved babble. All the babble had one male and one female talker. As the LTASS and the F_0 of the babble in comparison to that of the target stimuli has been reported to be factors affecting the speech recognition, the descriptive analysis of LTASS and F_0 of the babble used in the present study are given below. The LTASS of the target Kannada sentences and four two-talker babbles are shown in Figure 4.1.

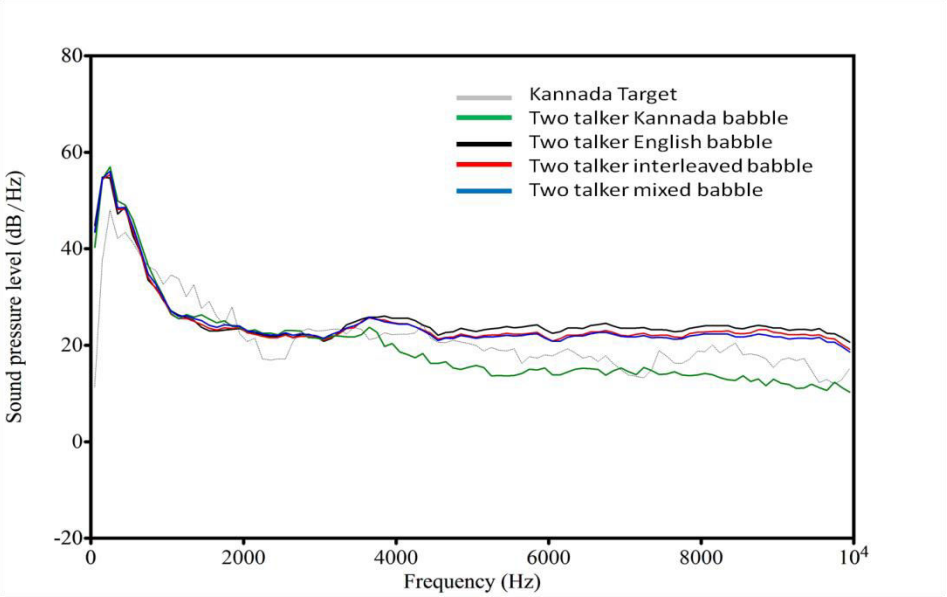

Figure 4.1: LTASS of different babbles and target sentence.

As the figure depicts, the LTASS of two-talker English, mixed and interleaved babble were similar and when the LTASS of these were compared to the target sentences, the target had lesser energy at high frequencies and equal energy at low frequencies. However, the LTASS of target stimuli was comparable to the Kannada babble at all frequencies. .

The F_0 of the target stimuli (spoken by a female) and F_0 of two-talker babble in Kannada (male and female) and English (one male and one female) were 225 Hz, 145 Hz, 210 Hz, 130 Hz and 205 Hz, respectively. The difference between the F_0 of the target Kannada sentences and the male speaker of the Kannada babble was 80 Hz and female speaker of the Kannada babble was 15 Hz, and the difference with male speaker of the English babble was 95 Hz and female speaker of the English babble was 20 Hz.

4.2. Measurement of SRS in the presence of different babbles

The scores of sentence recognition obtained for each of the three SNRs under four two-talker babble conditions were statistically analyzed using SPSS version 17.0 software. The mean, SD, median and the range of the SRS for each condition at different SNRs are displayed in Table 4.1.

Table 4.1: *Mean, median, SD and range of the SRS in the presence of two-talker Kannada, English, mixed, interleaved babble in three SNRs.*

Condition	SNR (dB)	Mean	SD	Median	Min-Max
Two-talker Kannada babble	+3	38.00	1.414	38.00	35-40
	0	32.90	2.540	33.00	27-37
	-3	30.23	2.537	30.00	25-35
Two-talker English babble	+3	39.43	.675	40.00	38-40
	0	37.80	0.992	38.00	36-40
	-3	34.93	2.188	35.00	28-37
Two-talker mixed babble	+3	38.00	.883	39.00	37-40
	0	34.60	2.479	35.00	27-39
	-3	31.68	2.777	32.00	26-35
Two-talker Interleaved babble	+3	37.65	1.955	38.00	32-40
	0	36.60	2.098	37.00	30-40
	-3	33.25	2.771	33.50	24-37

Note: Maximum number of key words = 40.

The data were subjected to normality tests and the 'p' value was less than 0.05 for all conditions. Hence, non-parametric tests were used to statistically analyze the data.

4.3. Effect of native vs. non-native language

Friedman test was done to determine the effect of the four two-talker conditions on the SRS across these SNRs. The results of the test revealed a significant main effect of different babble condition at each SNR [+3 dB SNR (χ^2 (3) =45.2, $p<0.05$), 0 dB SNR (χ^2 (3) =83.9, $p<0.05$) and -3 dB SNR (χ^2 (3) =71.676, $p<0.05$)]. At all the SNRs, except at +3 dB SNR, the rank order of sentence recognition in the presence of two-talker babble in decreasing order are two-talker English babble, interleaved babble condition, mixed condition and two-talker Kannada babble condition. At +3 dB SNR, the rank order of sentence recognition in the presence of two-talker babble in decreasing order are two-talker English Babble, two-talker mixed babble, two-talker Kannada babble and two-talker Interleaved babble.

Wilcoxon signed rank test was used for pair-wise comparison between the two-talker conditions at each SNR. The results revealed at 0 dB SNR and -3 dB SNR, there was a significant difference across all the conditions. At +3 dB SNR, there was a significant difference across all conditions except between two-talker English babble and two-talker mixed babble, and between two-talker Kannada babble and two-talker interleaved babble. The results of the pair-wise comparison are given in the Table 4.2, 4.3 and 4.4 for +3 dB SNR, 0 dB SNR and -3 dB SNR, respectively.

Table 4.2: *Results of Wilcoxon Signed test of SRS across different babble at +3 dB SNR.*

	Two-talker English babble	Two-talker Interleaved babble	Two-talker Mixed babble	Two-talker Kannada babble
Two-talker English babble		$\|Z\|=4.858$, $p<0.05$	$\|Z\|=2.659$, $p>0.05$	$\|Z\|=4.562$, $p<0.05$
Two-talker Interleaved babble			$\|Z\|=3.637$, $p<0.05$	$\|Z\|=.511$, $p>0.05$
Two-talker Mixed babble				$\|Z\|=2.932$, $p<0.05$
Two-talker Kannada babble				

Note: The unshaded box indicates no significant difference.

At +3 dB SNR, SRS in the presence of two-talker English babble was significantly higher than two-talker interleaved and Kannada babble conditions as shown in the Table 4.2. SRS in the presence of two-talker Kannada babble was significantly lower than two-talker mixed and English babble condition. However, SRS was significantly better in the presence of two-talker mixed condition than two-talker interleaved condition. There was no significant difference between SRS in the presence of two-talker English and mixed babble. In addition, no significant difference was found between SRS in the presence of two-talker Kannada and interleaved babble conditions.

Table 4.3: *Results of Wilcoxon Signed test of SRS across different babble at 0 dB SNR.*

	Two-talker English babble	Two-talker Interleaved babble	Two-talker Mixed babble	Two-talker Kannada babble
Two-talker English babble		$\|Z\|$=3.108, p<0.05	$\|Z\|$=5.271, p<0.05	$\|Z\|$=5.550, p<0.05
Two-talker Interleaved babble			$\|Z\|$=3.703, p<0.05	$\|Z\|$=4.701, p<0.05
Two-talker Mixed babble				$\|Z\|$=3.380, p<0.05
Two-talker Kannada babble				

Note: The unshaded box indicates no significant difference.

Table 4.4: *Results of Wilcoxon Signed test of SRS across different babbles at -3 dB SNR.*

	Two-talker English babble	Two-talker Interleaved babble	Two-talker Mixed babble	Two-talker Kannada babble
Two-talker English babble		$\|Z\|$=3.457, p<0.05	$\|Z\|$=5.180, p<0.05	$\|Z\|$=5.320, p<0.05
Two-talker Interleaved babble			$\|Z\|$=3.114, p<0.05	$\|Z\|$=5.181, p<0.05
Two-talker Mixed babble				$\|Z\|$=2.743, p<0.05
Two-talker Kannada babble				

Note: The unshaded box indicates no significant difference.

At 0 dB SNR and -3 dB SNR, all the conditions differed significantly from one another as given in the Table 4.3 and 4.4, respectively. The SRS in the presence of two-talker English babble was significantly better than all the other babbles, whereas the SRS in the presence of two-talker Kannada babble was significantly poorer than all the other

babbles. The SRS in the presence of two-talker interleaved babble was significantly poorer than two-talker mixed babble. The results of two-talker interleaved condition are elaborated in section 4.2.

4.4. Effect of Language uncertainty on SRS

The two-talker interleaved condition was used to study the effect of language uncertainty. The SRS in the presence of two-talker interleaved condition was compared to the other conditions. The Wilcoxon signed rank test results in Table 4.2, 4.3 and 4.5 shows that at all the SNRs, the SRS in the presence of two-talker interleaved condition was significantly poorer when compared to SRS in the presence of two-talker English and mixed condition. Further, the SRS in the presence of two-talker interleaved condition was significantly better when compared to SRS in the presence of two-talker Kannada condition at 0 dB SNR and -3 dB SNR. At +3 dB SNR, the SRS in the presence of two-talker interleaved condition was poorer than the SRS in the presence of two-talker Kannada babble; however, the difference was not significant.

4.5. Effect of SNR

Friedman test was done to determine the effect of SNR on SRS in each of the two-talker babble condition and the results of the test revealed a significant effect in all the four two-talker condition [two-talker English babble (χ^2 (2) =76.51, $p<0.05$), two-talker Kannada babble (χ^2 (2) =69.32, $p<0.05$), two-talker mixed babble (χ^2 (2) =68.82,

$p<0.05$) and two-talker interleaved babble (χ^2 (2) =57.28, $p<0.05$)]. Further, pair-wise comparison was done using Wilcoxon signed rank test to compare the SRS across SNRs for each condition. The results of this showed that the SRS at higher SNRs yielded better recognition scores, and the SRS decreased significantly as SNR decreased in all the four two-talker conditions. Table 4.1 depicts the decrease in median values of SRS as SNR decreases in all the four two-talker conditions. The results of Wilcoxon signed rank test are given in the Table 4.5.

Table 4. 5: *Results of Wilcoxon Signed rank test of SRS across three SNRs.*

Condition	**Comparison between SNR (dB)**		**|Z| values**
Two-talker Kannada babble	+3	0	5.521*
	+3	-3	5.519*
	0	-3	4.611*
Two-talker English babble	+3	0	5.316*
	+3	-3	5.544*
	0	-3	5.336*
Two-talker mixed babble	+3	0	5.457*
	+3	-3	5.522*
	0	-3	4.922*
Two-talker interleaved babble	+3	0	4.190*
	+3	-3	5.317*
	0	-3	5.128*

Note: *p <0.01

Chapter 5

Discussion

The present study evaluated the performance of Kannada-English bilinguals on recognition of Kannada sentences in the presence of two-talker Kannada babble, two-talker English babble, two-talker mixed and two-talker interleaved babble. In addition to this, the effect of language uncertainty was also studied by comparing the SRS in the presence of two-talker interleaved babble and SRS in the presence of other three non-interleaved babbles.

5.1. Effect of native vs. non-native language

Statistical analysis of the effect of native vs. non-native babble on the sentence recognition revealed a significant better Kannada sentence recognition scores in the presence of two-talker English Babble than two-talker Kannada babble in all the SNRs. Similar results have been found in other studies (Brouwer & Bradlow, 2014; Calandruccio & Zhou, 2014; Stibbard & Lee, 2006). These studies on bilinguals have reported masking release in the presence of two-talker non-native language babble. This is attributed to the linguistic mismatch contributing to less IM in the condition where non-native babble is presented. Whereas, Anitha (2003) and Vineetha, Suma, and Nair (2013) reported no release of masking in presence of non-native language MTB (Anitha, 2003; Vineetha et al., 2013). This could be due to the fact that these studies used four to

eight-talker babble, and hence, the spectral and temporal characteristic of the this would resemble a speech noise as the number of talkers used were more (Carhart et al., 1975; Cullington & Zeng, 2008; Simpson & Cooke, 2005) and hence, there was no IM.

In addition, the study done by Anitha (2003) used MTB of Indian languages of similar origin, thereby leading to lesser linguistic mismatch. The languages taken in the present study are from two different language families and are of different origin. Kannada language belongs to Dravidian family and English belongs to Indo-European family. Hence, the present study might have found masking release due to the linguistic mismatch.

However, the release of masking in the presence of non-native language babble (English babble) cannot be solely due to the difference in IM. There could be influence of energetic masking factors like F_0 of the speaker, gender of the speaker and LTASS of the language, and other linguistic differences like accent (Brouwer et al., 2012; Calandruccio et al., 2013, 2010; Calandruccio & Zhou, 2014). Before attributing the linguistic mismatch as the sole factor affecting the IM, the influence of the above mentioned factors need to be ruled out. The possible effects of these factors on IM in the present study are discussed below.

5.3.1. F_0 of the speakers

In the present study, the difference between the F_0 of speaker of the target sentences and F_0 of two speakers of Kannada babble was similar to the difference between F_0 of speaker of the target sentences and both the speakers of Kannada babble. Various authors have reported a difference in the masker and target in terms of F_0, and attributed the release of masking to the ability of the listener to segregate the masker and the target based on these differences rather than informational masking (Brouwer et al., 2012; Calandruccio et al., 2013, 2010; Calandruccio & Zhou, 2014).

However, in the present study, the difference in F_0 between target and masker in both conditions could have aided in the segregation of the target and babble, but the difference between the scores across conditions cannot be explained based on this, because the difference between the F_0 of target speaker and speakers of babble of both languages are almost same. Hence, F_0 difference does not influence the difference in scores between the two-talker English and Kannada babble, rather the linguistic mismatch has played a role in the release of masking (Brungart et al., 2001; Cooke et al., 2008) when Kannada sentences were presented in the presence of two-talker English babble.

5.3.2. LTASS of the stimuli and the babble

In order to rule out EM in the present study, LTASS of different babble was analyzed. The LTASS has been reported to vary across languages (Byrne, Dillon, &

Tran, 1994). This difference in LTASS can also contribute to differential EM in the speech recognition task (Calandruccio et al., 2010; Calandruccio & Zhou, 2014).Studies have reported that babble with lesser energy leads to lesser masking (Calandruccio et al., 2013, 2010).

Conversely, in the present study, only very slight differences are present in the LTASS of the four two-talker babble when compared to that the target stimuli, and hence, its contribution can be considered less influential towards the difference in the scores across two-talker babble conditions. Hence, this reveals a strong support towards the role of IM in the masking of native language in the presence of native versus non-native language babble. However, measurement of LTASS using many speakers would have been better, and normalizing the LTASS before the experimental task could have been done in order to reduce the factors contributing towards EM while studying IM (Brouwer et al., 2012).

5.3.3. Effect of accent

Calandruccio et al. (2010) reported a significant difference between mandarin accented English and native English two-talker babble at difficult SNR. In the present study, English sentences in the babble were spoken by an Indian who has been residing in the state of Karnataka. The English produced by an Indian is recognized as General Indian English (Sirsa & Redford, 2013). The Indian English is a result of accent influence of the native language. Hence, effect of the General Indian English could have more

similarities acoustically and phonetically (Sirsa & Redford, 2013) with the native language i.e., Kannada language in our study.

Hence, the release of masking in the presence of two-talker English babble and more masking in presence of two-talkers Kannada babble in the present study could be attributed to difference in the lexical-semantic interference rather than acoustical and phonetic similarity contributed due to native accent influence.

5.4. Effect of language uncertainty on SRS

Another goal of the study was to find the effect of linguistic uncertainty of the babble on the recognition scores. The results revealed presence of both the languages in the babble (two-talker mixed and interleaved babble) condition has reduced scores than two-talker English babble and has better scores than two-talker Kannada babble. Among the mixed babble and interleaved condition, the addition of uncertainty in the interleaved condition has lead to significantly poorer scores than mixed babble condition at all SNRs.

Similar results were found in the study done by Brouwer and Bradlow (2014). They reported that the speech-in-speech recognition is sensitive to contextual variation in terms of the target-background language mismatch.

This could be due to the influence of cognitive factors like attention and working memory which affect the process of tuning in to target speech and tuning out of the

speech masker. The uncertainty induces distractions which affects the selective attention. Mattys et al. (2009) reported an influence of cognitive load in terms of dual attention and divided attention cause a slowing of their cognition process and hence, affecting the speech recognition ability. This explains the reduction in the scores of sentence recognition in the presence of two-talker interleaved condition.

5.5. Effect of SNR

In the present study, the recognition scores were significantly higher at higher SNRs than at lower SNRs in all the babble conditions. Studies have reported that with increase in SNR, there is less interference by the masker both in terms of EM and IM, thereby leading to better scores (Brouwer et al., 2012; Brungart et al., 2001; Cooke et al., 2008; Engen & Bradlow, 2007; Van Engen, 2010; Wu et al., 2015).

Another effect of SNR is that IM is seen only when both target and masker is audible to the listener. That is, at higher SNRs, the target is more audible than the masker, hence, the intensity cues facilitate in better segregation of masker and target (Engen & Bradlow, 2007) leading to less EM and IM. Whereas at difficult SNRs, the target is less audible when compared to the masker, inducing more EM and also there is less competition at the level of target speech recognition which eliminates the linguistic IM effects (Van Engen, 2010). Similar results have been found even in the present study.

Chapter 6

Summary and Conclusion

There has been a great influence of English on the Indian languages since the British rule and vice versa. In everyday listening situation, one is exposed to speech recognition in the presence of different language maskers and mixing of both the English and Indian languages commonly referred as code switching/code mixing. This scenario of speech-in-speech recognition gets influenced by many factors such as the language of the babble, language uncertainty of the babble and SNR.

The aim of the present study was to study the effect of native and non-native language babble on SRS of Kannada sentences, the effect of language uncertainty on the SRS of Kannada sentences, and to study the role of SNR on speech-in-speech recognition in Kannada-English bilinguals.

The SRS in 40 Kannada-English bilinguals in the presence of two-talker Kannada babble, two-talker English Babble, two-talker mixed babble and two-talker interleaved babble at +3 dB SNR, 0 dB SNR and -3 dB SNR was obtained. The influence of native language vs. non-native language was studied by comparing the SRS in the presence of two-talker Kannada babble and two-talker English Babble.

The results revealed that the linguistic mismatch did account for release in masking in the presence of two-talker English babble. In addition to this, the effect of language uncertainty was also studied by comparing the SRS in the presence of two-talker interleaved babble with the SRS in the presence of the other three non-interleaved babbles. The results revealed that addition of uncertainty deteriorated the SRS.

Further, with an increase in SNR, the scores across all four two-talker babble conditions increased systematically. However, there was no significant difference among the two-talker babble conditions seen at higher SNR (+3 dB SNR), as the target was less interfered by the two-talker babble and hence, there was lesser informational masking.

From the above results, it could be concluded that if the masker is a sentence from the native language, then the speech recognition could be poor when compared to that of non-native language. Further, unpredictable changes in the language of the masker affect the speech-in-speech recognition. This unpredictability is very common in the Indian set up and hence, this should be considered while carrying out hearing testing and hearing aid evaluation in the presence of speech babble. Further, the speech-in-speech recognition, in the present study, has been mainly due to informational masking and less of energetic masking. In addition, the SNR is found to be a significant factor affecting speech-in-speech recognition.

Clinical Implication:

- Measurement of speech recognition in the presence of babble would help to simulate the real-life situation. The results of the present study would help in selecting and improvising the masker that is used in the assessment of speech in the presence of background signal.
- It would also aid in counseling the persons with hearing impairment with respect to consequences of speech recognition in adverse listening conditions. The addition of language change in the background could have more deteriorating effects on recognition. Hence, use of one language is to be encouraged by the family members in multi-talker environment, especially for children with hearing impairment.

Future directions

- The influence of one Indian language on other Indian language using two-talker babble could be studied.
- A similar study could be carried out using the same speaker for the target and babbles to control the talker differences influencing the speech-in-speech recognition.
- Various Indian languages accented English speech can be used to study the influence of Indian accent on speech recognition.

References

Anitha, R. (2003). Effect of multi-talker babble of different languages on the speech recognition scores in Kannada. *Independent Project*, Done at the Department of Audiology,AIISH, Mysore.

ANSI. (1991). Maximum permissible ambient noise levels for audiometric test rooms. ANSI S3.1-1991, New York: American National Standards.

Boulenger, V., Hoen, M., Ferragne, E., Pellegrino, F., & Meunier, F. (2010). Real-time lexical competitions during speech-in-speech comprehension. *Speech Communication*, *52*(3), 246–253. doi:10.1016/j.specom.2009.11.002

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272–284. doi:10.1121/1.1487837

Broersma, M. (2012). Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes*, *27*(7-8), 1205–1224. doi:10.1080/01690965.2012.660170

Brouwer, S., & Bradlow, A. R. (2014). Contextual variability during speech-in-speech recognition. *The Journal of the Acoustical Society of America*, *136*(1), EL26–EL32. doi:10.1121/1.4881322

Brouwer, S., Engen, K. J. Van, Calandruccio, L., & Bradlow, A. R. (2012). Linguistic contributions to speech-on-speech masking for native and non-native listeners : Language familiarity and semantic content. *The Journal of the Acoustical Society of America*, *131*(2), 1449–1464. doi:10.1121/1.3675943

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2006). Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation. *The Journal of the Acoustical Society of America*, *120*(6), 4007. doi:10.1121/1.2363929

Brungart, D. S., & Simpson, B. D. (2004). Within-ear and across-ear interference in a dichotic cocktail party listening task : Effects of masker uncertainty. *The Journal of the Acoustical Society of America*, *115*(1), 301–310. doi:10.1121/1.1628683

Brungart, D. S., Simpson, B. D., Ericson, M. a., & Scott, K. R. (2001). Informational and energetic masking effects in the perception of multiple simultaneous talkers. *The Journal of the Acoustical Society of America*, *110*(5), 2527. doi:10.1121/1.1408946

Byrne, D., Dillon, H., & Tran, K. (1994). An international comparison of long-term average speech spectra. *The Journal of the Acoustical Society of America*, *96*(4), 2108.

Calandruccio, L., Brouwer, S., Van Engen, K. J., Dhar, S., & Bradlow, A. R. (2013). Masking Release Due to Linguistic and Phonetic Dissimilarity Between the Target and Masker Speech. *American Journal of Audiology*, *22*(1), 157–164. doi:10.1044/1059-0889(2013/12-0072)not

Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, *128*(2), 860–869. doi:10.1121/1.3458857

Calandruccio, L., & Zhou, H. (2014). Increase in Speech Recognition due to Linguistic Mismatch Between Target and Masker Speech: Monolingual and Simultaneous Bilingual Performance. *Journal of Speech, Language, and Hearing Research*, *57*(3), 1089–1097. doi:10.1044/2013

Carhart, R., Johnson, C., & Goodman, J. (1975). Perceptual masking of spondees by combinations of talkers. *The Journal of the Acoustical Society of America*, *58*(S1), S35. doi:10.1121/1.2002082

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual Masking in Multiple Sound Backgrounds. *The Journal of the Acoustical Society of America*, *45*(3), 694–703. doi:10.1121/1.1911445

Chen, J., Li, H., Li, L., Wu, X., & Moore, B. C. J. (2015). Informational masking of speech produced by speech-like sounds without linguistic content. *The Journal of the Acoustical Society of America*, *131*(4), 2914–2926.

Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975. doi:10.1121/1.1907229

Cooke, M., Lecumberri, M. L. G., & Barker, J. (2008). The foreign language cocktail party problem : Energetic and informational masking effects in non-native speech. *The Journal of the Acoustical Society of America*, *123*(1), 414–427. doi:10.1121/1.2804952

Cullington, H. E., & Zeng, F. (2008). Speech recognition with varying numbers and types of competing talkers by normal-hearing , cochlear-implant , and implant simulation subjects. *The Journal of the Acoustical Society of America*, *123*(1), 450–461. doi:10.1121/1.2805617

Dillon, H. (1983). The effect of test difficulty on the sensitivity of speech discrimination tests. *The Journal of the Acoustical Society of America*, *73*(1), 336. doi:10.1121/1.388815

Durlach, N. I., Mason, C. R., Kidd, G., Arbogast, T. L., Colburn, H. S., & Shinn-Cunningham, B. G. (2003). Note on informational masking. *The Journal of the Acoustical Society of America*, *113*(6), 2984. doi:10.1121/1.1570435

Durlach, N. I., Mason, C. R., Shinn-Cunningham, B. G., Arbogast, T. L., Colburn, H. S., & Kidd, G. (2003). Informational masking: Counteracting the effects of stimulus uncertainty by decreasing target-masker similarity. *The Journal of the Acoustical Society of America*, *114*(1), 368. doi:10.1121/1.1577562

Engen, K. J. Van. (2012). Speech-in-speech recognition : A training study Speech-in-speech recognition : A training study. *Language and Cognitive Processes*, *27*(7-8), 1089–1107. doi:10.1080/01690965.2012.654644

Engen, K. J. Van, & Bradlow, A. R. (2007). Sentence recognition in native- and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, *121*(1), 519–526. doi:10.1121/1.2400666

Fairbanks, G. (1960). The rainbow passage. In *Voice and articulation drillbook* (2nd ed., pp. 124–139). New York: Harper. Retrieved from https://books.google.com/books?id=qN1ZAAAAMAAJ&pgis=1

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, *115*(5), 2246–2256. Retrieved from http://scitation.aip.org/content/asa/journal/jasa/115/5/10.1121/1.1689343

Freyman, R. L., Helfer, K. S., & Balakrishnan, U. (2007). Variability and uncertainty in masking by competing speech. *The Journal of the Acoustical Society of America*, *121*(2), 1040–1046. doi:10.1121/1.2427117

Geetha, C., Kumar, K. S. S., Manjula, P., & Pavan, M. (2014). Development and standardisation of the sentence identification test in the Kannada language. *Journal of Hearing Science*, *4*(1), 18–26.

Hall III, J. W., Grose, J. H., Buss, E., & Dev, M. B. (2002). Spondee Recognition in a Two-Talker Masker and a Speech-Shaped Noise Masker in Adults and Children. *Ear & Hearing*, *23*(2), 159–165.

Harini, N. V, & Shyamala, K. C. (2008). Code mixing and Code Switching in simultaneous vs successive bilingual children. *Unpublished Master's Dissertation*, University of Mysore.

Hoen, M., Meunier, F., Grataloup, C., Pellegrino, F., Grimault, N., Perrin, F., … Collet, L. (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Communication*, *49*(12), 905–916. doi:10.1016/j.specom.2007.05.008

Jain, C., Kodanath, S., Vimal, B. M., & Suresh, V. (2014). Influence of native and non-native multitalker babble on speech recognition in noise. *Audiology Research*, *4*(1). doi:10.4081/audiores.2014.89

Jain, C., Konadath, S., Vimal, B. M., & Suresh, V. (2014). Influence of native and non-native multitalker babble on speech recognition in noise. *Audiology Research*, *89*(4), 9–13. doi:10.4081/audiores.2014.89

Jerger, J. (1970). Clinical Experience With Impedance Audiometry. *Archives of Otolaryngology - Head and Neck Surgery*, *92*(4), 311–324. doi:10.1001/archotol.1970.04310040005002

Jones, G. L., & Litovsky, R. Y. (2008). Role of masker predictability in the cocktail party problem. *The Journal of the Acoustical Society of America*, *124*(6), 3818–3830. doi:10.1121/1.2996336

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America*, *61*(5), 1337–1351. doi:10.1121/1.381436

Kidd, G., Mason, C. R., & Arbogast, T. L. (2002). Similarity, uncertainty, and masking in the identification of nonspeech auditory patterns. *The Journal of the Acoustical Society of America*, *111*(3), 1367. doi:10.1121/1.1448342

Lecumberri, M. L. G., & Cooke, M. (2006). Effect of masker type on native and non-native consonant percpetion in noise. *The Journal of the Acoustical Society of America*, *119*(4), 2445–2454. doi:10.1121/1.2180210

Mathew, M. (2012). Patterns of Code Switching in Children. *Language in India*, *12*, 512–550. Retrieved from https://scholar.google.co.in/scholar?q=Patterns+of+Code+switching+in+children&btnG=&hl=en&as_sdt=0%2C5#0

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load : Dissociating energetic from informational factors. *Cognitive Psychology*, *59*(3), 203–243. doi:10.1016/j.cogpsych.2009.04.001

Mattys, S. L., Carroll, L. M., Li, C. K. W., & Chan, S. L. Y. (2010). Effects of energetic and informational masking on speech segmentation by native and non-native speakers. *Speech Communication*, *52*(11), 887–899. doi:10.1016/j.specom.2010.01.005

Mattys, S. L., Davis, M. H., Bradlow, A. R., & Scott, S. K. (2012). Speech recognition in adverse conditions : A review. *Language and Cognitive Processes*, *27*(7), 953–978. doi:10.1080/01690965.2012.705006

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language*, *65*(2), 145–160. doi:10.1016/j.jml.2011.04.004

Miller, G. A. (1947). The masking of speech. *Psychological Bulletin*, *44*(2), 105.

Neff, D. L., Dethlefs, T. M., & Jesteadt, W. (1993). Informational masking for multicomponent maskers with spectral gaps. *The Journal of the Acoustical Society of America*, *94*(6), 3112–3126. doi:10.1121/1.407217

Pollack, I. (1975). Auditory informational masking. *The Journal of the Acoustical Society of America*, *57*(S1), S5. doi:10.1121/1.1995329

Savithri, S, R., & Jayaram, M. (2005). 300 words reading passages in Dravidian languages. *AIISH Research Fund Project*, Done at AIISH, Mysore.

Simpson, S. A., & Cooke, M. (2005). Consonant identification in N -talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, *118*(5), 2775–2778. doi:10.1121/1.2062650

Sirsa, H., & Redford, M. a. (2013). The effects of native language on Indian English sounds and timing patterns. *Journal of Phonetics*, *41*(6), 393–406. doi:10.1016/j.wocn.2013.07.004

Stibbard, R. M., & Lee, J.-I. (2006). Evidence against the mismatched interlanguage speech intelligibility benefit hypothesis. *The Journal of the Acoustical Society of America*, *120*(1), 433–442. doi:10.1121/1.2203595

Van Engen, K. J. (2010). Similarity and familiarity : Second language sentence recognition in first- and second-language multi-talker babble. *Speech Communication*, *52*(11), 943–953. doi:10.1016/j.specom.2010.05.002

Vineetha, C. V, Suma, R., & Nair, S. P. (2013). Effect of Bilingualism on Speech in Noise Perception in Young Adults. *Language in India*, *13*(6), 799–811.

Watson, C. S., Kelly, W. J., & Wroton, H. W. (1976). Factors in the discrimination of tonal patterns. II. Selective attention and learning under various levels of stimulus uncertainty. *The Journal of the Acoustical Society of America*, *60*(5), 1176–1186. doi:10.1121/1.381220

WHO. (2008). *World Health Organisation.Grades of hearing impairment. Prevention of Blindness and Deafness*. World Health Organization. Retrieved May 9, 2015, from http://www.who.int/pbd/deafness/hearing_impairment_grades/en/

Wu, X., Yang, Z., Huang, Y., Chen, J., Li, L., Daneman, M., & Schneider, B. A. (2015). Cross-Language Differences in Informational Masking of Speech by Speech : English versus Mandarin Chinese. *Journal of Speech, Language, and Hearing Research*, *54*(6), 1506–1524. doi:10.1044/1092-4388(2011/10-0282)target

Wylie, E. (2006). *International Second Language Proficiency Rating Scales*. Retrieved from www.balaibahasaperth.org/documents/ISLPR_scale_indonesian.pdf

Yathiraj, A., & Vijayalakshmi, C. S. (2005). Phonemically Balanced Word List in Kannada. *Departmental Project*, Developed in Department of Audiology, AIISH,Mysore.