EVALUATION OF THE SIGNAL-TO-NOISE RATIO REQUIRED TO ACHIEVE THE SAME PERFORMANCE IN ENGLISH AND MANDARIN CHINESE

by

Yi Ye

MD, Luzhou Medical College, 1998

Master in Audiology, West China Medical School, 2006

Submitted to the Graduate Faculty of

School of Health and Rehabilitation Science in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

SCHOOL OF HEALTH AND REHABILITATION

This dissertation was presented

by

Yi Ye

It was defended on

July 22, 2014

and approved by

John Durrant, PhD, Professor Emeritus, Department of Communication Science and

Disorders

Sussan Shaiman, PhD, Associate Professor, Department of Communication Science and

Disorders

Alan Juffs, PhD, Associate Professor, Department of Linguistics

Christopher Brown, PhD, Assistant Professor, Department of Communication Science and

Disorders

Dissertation Advisor: Catherine Palmer, PhD, Associate Professor, Department of

Communication Science and Disorders

EVALUATION OF THE SIGNAL-TO-NOISE RATIO REQUIRED TO ACHIEVE THE SAME PERFORMANCE IN ENGLISH AND MANDARIN CHINESE

Yi Ye, PhD

University of Pittsburgh, 2014

Difficulty communicating in noise is a common complaint for people with hearing loss. When communicating in noise, speakers increase the intensity level of their voice and alter the stress patterns of their speech not only to monitor their own voice but also to be heard by others. Speech that increases in intensity for the purpose of self-monitoring and being understood in noise is called Lombard speech. Few studies have assessed communication performance with Lombard speech in noise which closely reflects the real-life communication situation. In addition, the characteristics of Lombard speech may be different(among) languages with different characteristics and identifying features so the few results available for English listeners may not apply to listeners of other languages.

This study evaluated the performance of English speaking and Mandarin Chinese speaking individuals listening to English and Mandarin Chinese speech in corresponding babble noise. Speech materials were the IEEE sentences in English and translated into Mandarin Chinese while controlling for phonological, grammatical, and contextual predictability. The sentences and 4-talker babble were recorded in a conversational manner and at a Lombard speech level produced while listening to 80 dB SPL of noise. The performance of 18 native English speakers and 18 native Mandarin Chinese speakers was evaluated. The SNR-50, the signal-to-noise level required to produce 50% performance, was the same for conversational and Lombard English indicating that there is not a particular benefit in producing Lombard speech to be understood. The reason to produce Lombard speech in English is to improve the signal-to-

noise ratio in order to facilitate improved communication. The results for the Mandarin Chinese listeners revealed a benefit when producing Lombard speech with the SNR-50 for Mandarin Chinese significantly different between conversational and Lombard speech. In noisy situations where increasing vocal intensity is expected, , Mandarin Chinese listeners appear to benefit from features preserved or enhanced through Lombard speech that English listeners do not access.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0   INTRODUCTION

Difficulty communicating in noise is a common problem for individuals with hearing loss including those who speak Mandarin Chinese. The majority of research examining the ability to hear in noise by individuals with hearing impairment has focused on developing signal processing techniques to improve the signal-to-noise ratio. It is unclear if Mandarin-Chinese speakers/listeners require comparable signal-to-noise ratios in order to perform similarly on speech in noise tasks as compared to English speakers/listeners. Current hearing aid signal processing strategies and communication recommendations are based on data from Germanic languages (e.g., English, Danish, German, etc). However, Mandarin Chinese is a typical tonal language in that the pitch contour over a syllable can distinguish word meanings. It is unclear if these same strategies apply to Mandarin-Chinese speakers/listeners because a paucity of data exist describing the impact of noise (other speakers talking) on the recognition of Mandarin Chinese.

Byrne et al (1994) compared the average long term speech spectrum (LTASS) of 12 different languages including tonal languages such as Mandarin Chinese, Cantonese and Vietnamese. The average long term speech spectrum is widely used in hearing aid fitting algorithms. They concluded that the LTASS was similar for all languages with the average level of normal speech at 71.8 dB SPL for males and 71.5 dB SPL for females. Therefore, they suggested forming an international long term average speech spectrum (ILTASS). Although the

1

goal of the Byrne et al study was to create a single LTASS (ILTASS), there were differences among the various languages with Mandarin Chinese producing the most intense LTASS with a level of 75.2 dB SPL.Speech produced in quiet conditions was recorded and LTASS was determined. . It is unclear how the LTASS of these languages would compare if spoken in a background of noise with the intention of being understood. Individuals trying to communicate in noise typically change their stress patterns, slow down their speech, and increase the intensity level produced depending on the acoustic properties of the communication environment. A number of differences between English and Mandarin Chinese including linguistic characteristics, vulnerability to noise, and differences in Lombard speech used to project in noisy environments would predict that Mandarin-Chinese listeners may need an enhanced signal-to-noise ratio compared to English listeners in order to preserve necessary speech cues to obtain similar performance.

If specific speech cues in Mandarin Chinese must be preserved by enhancing the signal-to-noise ratio when communicating in difficult listening situations, communication strategies and hearing aid fitting algorithms would need to be specific to Mandarin Chinese rather than relying on average data based on the Germanic languages or on an overall average of world languages. The aim of this study is to quantify the signal-to-noise ratio required for similar performance between Mandarin speakers and their English speaking counterparts.

In the sections that follow, the variables and considerations noted above are explored to make the case warranting an investigation of performance in noise for Mandarin-Chinese and English listeners.

1.1   LINGUISTIC CHARACTERISTICS OF MANDARIN CHINESE

1.1.1   Phonologic Characteristics

Traditionally when describing the phonology of Mandarin Chinese, the structure of the syllable is broken down into units called initials and finals. Here, the term "initial" is used to describe the beginning part of the syllable, usually a single consonant. The term "final" is used to describe the part following the initial, usually a vowel or combination of two or three vowels (diphthongs). As a tonal language, the tones used in Mandarin Chinese are embedded in vowels which belong to finals (Baxter, 1992; Cheng, 1966). Figure1.1 illustrates the phonological structure of a Mandarin Chinese syllable.
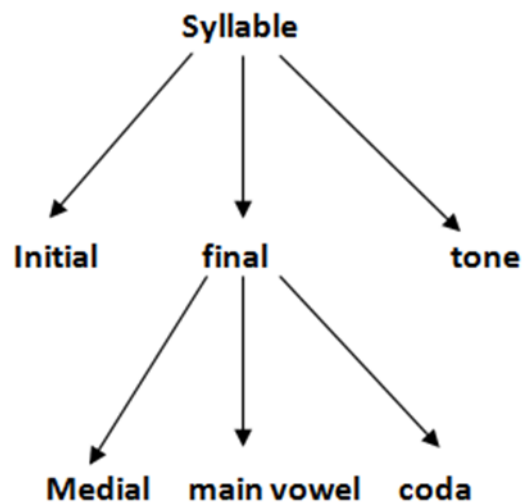
Figure 1.1 Phonological structure of a monosyllable of Mandarin Chinese

There are four possibilities of the phonological structure of such a monosyllable: (1) V (2) CV (3) VC (4) CVC. However, in Mandarin Chinese there are three additional points to note regarding these rudimentary syllabic constructions: (1) The initial consonant does not allow a

cluster, that is, CCV does not occur in the language (as in the pronunciation of the English word "stay"); (2) The final consonant is limited to either: the alveolar nasal [n] or the velar nasal [ŋ]; as distinguishable between the pronunciation of the words in Chinese "贫"[pin in tone 2](poor) and "平"[piŋ in tone 2](flat) (3) A tone (i.e., one of four tones) must be assigned to a syllable unless it is unstressed (Yip, 2000).

1.1.1.1 Initial consonant

In Mandarin Chinese, the initial is essentially the consonantal beginning of the syllable, or what is known as the syllable onset in standard phonology. There are twenty-three consonants in Mandarin, twenty-two of which can be initials. The velar nasal [ŋ] is the Mandarin consonant that cannot be an initial. There again are no consonant clusters in Mandarin; therefore the initial always comprises a single consonant. Some syllables do not have an initial consonant; in these cases the initial is described as zero initial. The following table (Table 1.1) describes the twenty-one initials of Mandarin using both Pinyin Romanization (right side of the column) and the phonetic symbols (left side of the column) used by the International Phonetic Association (IPA).

4

Table 1.1 Mandarin Chinese Initial Consonants

		Bilabial	Labiodental	Alveolar	Retroflex	Palatal	Velar
Stop	Aspirated	pʰ P		tʰ t			kʰ k
	Unaspirated	p b		t d			k g
Affricate	Aspirated			tsʰ c	Tʂʰ ch	tɕʰ q	
	Unaspirated			ts z	Tʂ zh	tɕ j	
Fricative	Voiceless		f f	s s	ʂ sh	ɕ x	x h
	Voiced						
Nasal	Voiced	m m		n n			ŋ ng cannot be initial
Approximant	Voiced			l l	ɻ r		W

Mandarin Chinese has eleven unvoiced affricates and fricatives, as opposed to English, which has only five unvoiced fricatives and affricates. Unlike English, in Mandarin Chinese it is aspiration and not voicing that is phonemically distinctive. Although it is also true that there are no inter-dental fricatives such as "θ" and "ð" in Mandarin Chinese, there are "c", "ch", "zh", "q" as well as "x" which do not exist in English. Therefore, there are relatively more high-frequency unvoiced consonants in Mandarin than in English. The audibility of high-frequency, unvoiced consonants would be compromised by challenging listening situations.

1.1.1.2 Finals

The final (known as the rhyme in standard phonology) represents the part of the syllable that occurs following the initial. There are thirty-seven finals in Mandarin, most of which contain only vowels. This final is sub-divided into medial (transition between the initial consonant and the nucleus), main vowel (nucleus) and coda (end of the final). Only two consonants can be included in the final: the alveolar nasal [n] and the velar nasal [ŋ]. These consonants can only

occur at the end of the final, in what is known as the coda position in standard phonology. The finals are presented in the following table (Table 1.2), listed as IPA (International Phonetic Association) symbols (Dunlop, 1997).

Table 1.2 Mandarin Chinese Finals

	A	ð	o		ai	ei	Au	ou	an	ðn	aŋ	ðŋ	
i	iA			Iɛ			Iau	iou	iɛn	In	iaŋ	iŋ	
u	uA		uo		uai	uei			uan	uan	uaŋ	uŋ	uðŋ
y				Y					yɛn	yn			

Compared to English, Mandarin Chinese has many more medials or diphthongs and the duration of the finals is longer which makes pronounced acoustic focusing on the final. This creates a situation where the final is less easily masked by noise as compared to the initial consonants in Mandarin Chinese.

1.1.1.3 Tones

Mandarin Chinese is a typical tonal language in that the pitch contour over a syllable can distinguish word meanings. On full syllables there are four tones. There is also a 5th neutral tone on a stress-less syllable. Being differentiators of meaning, their association with individual monosyllables is determined by convention. If the syllable is unstressed, there is no tone on this syllable. For example (see Figure 1.2): 1st tone ma 'mother', 2nd tone ma 'hemp', 3rd tone ma 'horse', 4thtone ma 'scold' and 5th unstressed ma, end-of-sentence particle for the formulation of a general question. Tonal information is conveyed by the final part of the syllable which is the vowel.

6
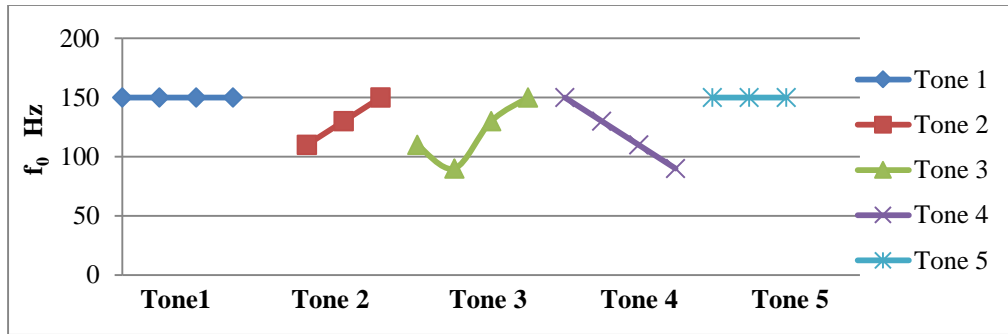
Figure 1.2 The Pitch-frequency contours as functions of time for five tones

Word stress in Chinese coincides with tones. An unstressed syllable naturally loses its tones. Generally speaking, all monosyllabic words which have lexical meanings are tonal and therefore stressed (e.g., dà: big, xiǎo: small, etc.). Only those monosyllabic words that are grammatical or functional in nature are toneless and therefore unstressed (e.g., de: a particle introducing an attributive, ne: an indicator for special questions, etc.). In any disyllabic word, if there is an unstressed syllable present, it is always the second and never the first. For example,

Xiūxi to rest   shétou tongue

Dǎsuan   to plan   lìhai terrible

In conclusion, there are several distinct phonological features of Mandarin Chinese syllables. Firstly, most syllables are formed by a single initial consonant and a final, the final is constructed by medial, main vowel and coda. This structure makes the acoustic energy temporally weighted on vowels. Secondly, since the tones are meaningful in Mandarin Chinese, pronouncing and hearing tones correctly is critical to communication. Therefore, there will be more low-frequency energy concentrated in   vowels than in other languages causing the vowels to be more resistant to the impact of noise. On the other hand, the consonant and vowel combination is so simple, ambiguity is a factor for the initial and the high-frequency consonant may be missed in a background of noise.

7

1.1.2 Monosyllable-based structure

Modern Standard Mandarin Chinese has about 7,000 commonly used characters (excluding 2,000 rare ones), most of which are monosyllabic morphemes. In linguistics, a morpheme is the smallest conceptual meaningful component of a word. Phonological habits of Mandarin Chinese make use of only 418 actual monosyllables, together with the total number of phonologically allowed monosyllables there are only about 1273 syllables (1st tone 330, 2nd tone 247, 3rd tone 312, 4th tone 353, and 5th and the unaccented tone 31) (Yip, 2000). The language is not codified with strict logic: not all the 418 monosyllables are used with all the four tones or found to be equally unstressed. This gives an average of 5.4 homophonous morphemes per syllable (Duanmu, 2007). Homophones are words that have exactly the same sound (pronunciation) but different meanings and spellings. Which means the limited number of monosyllables represents a much larger number of monosyllabic characters. For example (see Figure 3) : Yi in tone one could mean in character "一" (one), or "医" (medicine), or "伊" (her), or "依" (depend); Yi in tone two could mean in character "颐"(elegant), or "宜" (suitable), or "仪" (appearance), or "遗" (lost), or "姨" (aunt), or "怡" (joyful), or "夷"(smooth); Yi in tone 3 could mean in character "以" (according to), or "已" (already), or "蚁" (ant), or "漪"(ripples); Yi in tone 4 could mean in character "意" (mind), or "忆" (memory), or "易" (easy), or "毅" (decisive), or "亦" (also), or "义" (justice), or "益" (benefit), or "异" (different), or "艺" (art), or "翼" (wings), or "奕" (beauty), or "裔" (descendants), or "驿" (relay station), or "疫" (diseases), or "屹" (towering like a mountain peak). Sharing the same syllable causes a high degree of ambiguity. On the other hand, almost all Chinese characters have their own meanings (i.e., each character represents a morpheme or a smallest meaningful unit in the language). This is why many of the characters

8

can form mono-character words by themselves, and why the combination of several characters gives an almost unlimited number of poly-character words. For example, the combination of the two characters standing for "electricity" and "brain," respectively, becomes a new bi-character word standing for "computer". Therefore, monosyllable-based structure is usually taken as the first key to Mandarin speech recognition because accurate recognition of these 1273 Mandarin syllables already covers the whole language, including all possible characters and words.

1.1.3 Ambiguity Avoidance in Mandarin Chinese Words

Throughout the history of Chinese, there has been a dramatic decrease in the syllable inventory. For example, Middle Chinese (about AD 600) had over 3,000 syllables (including tonal contrasts), but modern Standard Mandarin Chinese has just 1,273. Thus, in the past one thousand years, the Chinese language has lost more than half of its syllables (Duanmu, 2007). This has created the large number of homophones described above. In addition, because the distribution is not even, some syllables represent more morphemes than others. For example, [yi4] ([i] with the fourth tone) represents 63 common morphemes, or over 90 morphemes if rare words are included. Given the large number of homophonous morphemes, many of which are independent words, it is not hard to imagine situations in which ambiguity arises and disyllabic expressions are used to avoid it. For example, in Standard Mandarin Chinese 'crow' and 'duck' are separate morphemes (written with different characters '鸦' and '鸭'), but they are both pronounced as [ya] in tone one. If one wants to say the former without ambiguity, one can use '乌鸦' [wu-ya] '(black) crow'. By adding the word '乌' [wu] 'black', additional color information is added to the key word 'crow'. For the ambiguity-avoidance approach, this is how

disyllabic words are created (Duanmu, 2007). There are several rules in forming the disyllable word of nouns, verbs, adjectives and adverbials, such as prefixes and suffixes, 敌人 [diren](enemy human being: enemy); 诗人 [shiren](poem human being: poet); or juxta-positional as 朋友[pengyou](friend friend: friend) , 艰难［jiannan］ (difficult, difficult: difficult); or a modification such as 重视[zhongshi] (heavily regards: attach importance to), 珍视[zhenshi] (treasure regards: cherish); or complemented such as 突出[tuchu] (stand out : protrude); 累坏 [leihuai](tired badly: be exhausted) (Duanmu, 2007).

According to all of these rules, most di-syllable words have a distinct contrast in their acoustical features in consonant, vowel and tones, together with their syntactic rules, which make di-syllable words much easier to recognize and to avoid ambiguity. At the same time, the di-syllable word makes time duration of the syllable longer and builds up a temporal cue which allows much easier segregation from background noise as compared to monosyllables. However, there is still some ambiguity in di-syllable words especially when there is a lack of contextual cues or with background Mandarin Chinese babble noise as compared to other languages and when they exist as song lyrics because their tonal information is compromised (Duanmu, 2007). For example, di-syllable words also have homophones. '香蕉' (banana) and '相交'(intersection) both pronounced as 'xiangjiao' in tone one. Contextual information is needed when the pronunciations of the disyllable words are the same.   Mandarin Chinese is a monosyllabic based language. Although di-syllable words are the major part of modern standard Mandarin Chinese vocabulary, hearing each mono-syllable correctly is still the key for efficient communication. The initial consonant in Mandarin Chinese will be the most fragile component when presented in a background of noise due to the higher frequency components. This will lead to additional ambiguity to the di-syllable words while communicating in noise.

1.1.4 Speech information rate of different languages

Language is a communicative system whose primary function is to transmit information. Regardless of the different linguistic strategies on which they rely, to efficiently convey messages to a group who share the same communication system is the global goal. Pellegrino, et al (2011) did a cross language comparison of 7 languages including British English and Mandarin Chinese on their information density (information density per syllable) and information rate. The other languages were French, German, Italian, Japanese and Spanish. In their study, they reported that Mandarin Chinese had the highest information density per syllable and the slowest syllable rate. They also reported that Mandarin Chinese is more complex in its syllables. Therefore, to reach a similar information rate, Mandarin-Chinese speakers use a strategy of speaking slower, making information denser and more complex in each syllable (Pellegrino et al., 2011). Results of their study indicated that correct recognition of each syllable is more crucial to reach efficient communication for people who speak Mandarin Chinese than for other languages.

In conclusion, Chinese has differences at the level of phonology, lexical and sentence structure as compared to English. The summary of those differences are shown in Table 1.3.

Table 1.3 Linguistic Difference of Mandarin Chinese and English that relate to listening in noise

	Mandarin Chinese Vs English
Consonant	More high frequency consonants
Vowel	More diphthongs and complicated combinations. Tones are super imposed on vowels and are important cues for lexical distinguishing
Lexical	Each mono-syllable is meaningful
Sentence	Higher information density

1.2   EFFECT OF NOISE ON SPEECH PERCEPTION

The definition of noise is an acoustic phenomenon that has random and aperiodic features with a continuous spectrum (Durrant & Feth, 2012). White noise, for example, contains all audible frequencies just like white light contains all visible color spectrum. The effects of white noise, pink noise and speech spectrum noise on speech perception have been broadly studied in hearing science, but communication embraces a variety of other interfering sounds. The much broader definition of noise, psychologically speaking, in any undesirable sound or signal (Durrant & Feth, 2012). The motocycle noise on the road, the alarm during the night, and the neighbor's garage music band could all be examples of noise. Other talkers' speech also may constitute noise for a particular listener. Noise is present everywhere except in the rare

anechoic/soundproof environment and appears likely to have increased greatly in modern industrial societies. Thus, speech interference is all too common today.

1.2.1   Noise and culture differences

When considering the noises that affect everyday communication, social and cultural factors must be taken into consideration. For example, population density, the industrialization of a society and different languages will affect the total background noise level in any given population.

Namba (1991) conducted a survey study on cross-cultural neighbors' noise problems.People of Japan, West Germany, the U.SA, China and Turkey were chosen for the study. The major comparison was on self-report level of audibility of neighbors' noises and the level of annoyance. The sounds chosen as most prominent among those which are audible and annoying in the respective countries are different. For example, in West Germany and in Turkey, sounds from the bathroom or toilet, those from a communal hall, stairways or lifts and those caused by handicraft activities were annoying; in Japan, sounds from mopeds or motorcycles and the loudspeakers of street vendors where annoying; in the U.S.A., sounds from neighbors' automobiles and pets were annoying; and in China, sounds from television, radio and stereos were annoying. The differences indicated in these responses may stem from differences in the structure of houses, in human relations among neighbors and in ways of life. For example, the percentage of individuals owning television, radio and stereo sets is similar in the five countries. The emphasis on these as a cause of annoyance among the Chinese respondents may owe to the fact that one apartment building is sometimes shared by several families. There is also a disproportional level of annoyance on neighbors' voices in Japan and China. It is interesting to

13

see the similar annoyance pattern in these two Asian countries. Taking the density of population into consideration, Japan has closer living arrangements than China but the Chinese participants reported louder conditions when considering neighbors' voices. This discrepancy might stem from differences in loudness needed for efficient communication of Mandarin Chinese as compared to Japanese (Namba et al., 1991).

The population of mainland China in 2011 was 1,345,670,000 with an area of 9,640,821km2 .Thus the average density is 140/km2. But most of the population is located in cities therefore the population density in cities is much greater than average. Together with the linguistic differences described previously, these data indicate that it is worth studying Chinese speakers/listeners separately from people who speak English when considering communication in noise.

1.2.2 Auditory masking by noise

In general, masking can be said to occur whenever the reception of a specified set of acoustic signals (e.g., speech) is degraded by the presence of other stimuli (e.g., noise)). The most basic definition of masking is the elevation in threshold for detecting the target caused by the presence of the masker. The impact of masking has been the focus of much research (for example, see Durlach, 2006). In recent years, there has been an increasing body of evidence to support the theory that auditory masking consists of two separate components that originate at different physiological levels and psychological levels that may roughly be divided into the categories of peripheral and central masking. Peripheral masking appears to be energetic-based. The simple impact that may be predicted from peripheral masking is complicated by substantial nonlinear properties of the cochlear partition. These effects result from overlapping patterns of excitation

14

along the basilar membrane and patterns of excitation of the fibers of the auditory nerve, thus degrading detectability of the target sound. Energetic masking depends on spectral features of both the target sound and the masker with effects beyond the arithmetic summation of the respective spectra resulting from the auditory system through which the signal passes.

In contrast, central masking is characterized as the inability to detect a target signal embedded in a context of other sounds at the level of the central auditory system even when the target signal is clearly audible. This is usually attributed to non-energetic masking, termed informational masking. When the masker is speech, processing of the information in the masker may interfere with processing of the target speech at any one of a number of perceptual (e.g., phonemic identification) or cognitive (e.g., semantic processing) levels. As a result, the listener may find it difficult to segregate the target from the masker because the information in the speech masker is interfering with the processing of information in the target talker's speech (informational masking). Notions related to uncertainty, similarity, attention, memory, etc., which are sometimes introduced in connection with such maskers, are now thought to be classified as informational masking (Durlach, 2006).

Scott, et al (2004) found that in different SNRs, there were different weightings in energetic and informational masking. They also reported that by comparing cortex neural processing differences, when using speech as the masker, the activation areas are the bilateral superior temporal gyrus while when using un-modulated noise as the masker . The recruitment of brain regions are remote from those classically associated with speech perception. The researchers believed that the activation in the bilateral superior temporal gyrus was associated with the informational masking condition. The activation of classical speech areas of the temporal lobes might delineate the neural basis of the informational masking and this activation

also might explain the interfering effects of unattended speech and sound on more explicit working memory tasks. This study was a novel demonstration of candidate neural systems involved in the perception of speech in noisy environments, and of the processing of multiple speakers in the dorsal-lateral temporal lobes (Scott et al., 2004). Results showed the neurological evidence of differences in people processing background noises when the noise is speech vs. non-speech. Based on this evidence it is worth seperately considering using speech as maskers and using noise as maskers depending on the goals of any individual study. Most people describing difficulty communicating in noise are referring to other people talking as the noise.

Which part of a syllable is most easily masked at the peripheral auditory system is a key question to understanding speech perception in noise. Cox et al (1987) carried out a study on intelligibility of average talkers in typical listening environments in English. Intelligibility of conversational speech for normal-hearing listeners was studied for three male and three female talkers. Four typical listening environments were used. They found that consonant place was the most poorly perceived feature, followed by continuance, voicing, and vowel intelligibility (Cox et al., 1987).

Mattys et al (2009) noticed that not all sources of information for word boundaries are equally affected by noise during their research on word segregation in noise. Similar to the conclusion of Cox et al (1987), they pointed out that juncture related prosodic cues, such as stress and F0 movements, are resistant to relatively high levels of noise (e.g., -5 to -10 dB signal-to-noise ratios) but coarticulatory cues, transitional probabilities and lexical-semantic knowledge show greater vulnerability. In different hearing environments, coarticulatory cues have different importance for speech segregation. When lexical-semantic information is available, reliance on coarticulatory cues is drastically reduced. In mild noise (5 dB SNR), where lexical-semantic

16

information is less readily available, coarticulatory cues exert their effect again. However, when the hearing environment gets worse (e.g., SNR ≤ 0 dB) word segregation does not depend on the cues of coarticulation. The effectiveness of lexical-semantic knowledge drops steadily as a function of noise level, probably reflecting the increasingly diffuse lexical activity resulting from inaccurately encoded sensory information (Mattys et al., 2009).

Parikh et al (2005) investigated the acoustic and perceptual effects of noise on vowel and stop-consonant spectra in English. They used multi-talker babble and speech-shaped noise and the acoustic analysis indicated that F1 was detected more reliably than F2 and the largest spectral envelope differences between the noisy and clean vowel spectra occurred in the mid-frequency band. They suggested that in extremely noisy conditions, listeners must be relying on relatively accurate F1 frequency information along with partial F2 information to identify vowels. Stop consonant recognition remained high even at -5 dB SNR despite the disruption of burst cues to additive noise, suggesting that listeners must be relying on other cues, perhaps formant transitions, to identify stops. The findings of the stop consonant recognition conflicted with Cox et al (1987), who suggested that other cues such as formant transitions are important in noise. Mandarin Chinese has more voiceless consonants. It is reasonable to hypothesize that the initial consonants of Mandarin Chinese are much more easily compromised in noise and may affect communication in noise for Mandarin-Chinese listeners.

Studebaker et al (1999) measured monosyllabic word recognition at higher-than-normal speech and noise levels. They found that speech intelligibility in noise decreased when speech levels exceeded 69 dB SPL with the signal-to-noise ratio remaining constant. The second important conclusion from this research was that the effects of speech and noise level were synergistic. That is, the negative effects of added noise level are greater when the speech level is

high, and vice versa. The Studebaker et al (1999) study tested only one ear which ignored possible two-ear benefits of listening in noise.

Byrne et al (1994) reported from their cross linguistic comparison on the average long term speech spectrum that the average level of normal speech was 71.8 dB SPL for male and 71.5 dB SPL for females. However, the most intense normal speech was Mandarin Chinese with the level at 75.2 dB SPL. Therefore, together with Studebaker's conclusion, for cultural and linguistic reasons, in environments where people mainly speak Mandarin Chinese, for example, Chinese restaurants or Chinese tea houses, people might find it much more difficult to communicate (Byrne D, 1994; Studebaker et al., 1999) when Mandarin Chinese is the signal of interest and the background noise.

1.2.3   Temporal masking

Temporal masking means masking effects between sounds that are not presented simultaneously. Temporal masking can occur whether the masker is presented before (forward masking) or after the probe stimulus (backward masking); the amount of masking depends on the relative intensities of the two sounds, their frequencies or spectral content, and the time interval between them (Durrant & Feth, 2012). For intervals beyond 200 ms, the amount of masking decreases dramatically in forward masking. In backward masking, the amount of masking is greatly reduced for intervals beyond 25 ms. Speech is a fluctuating auditory signal comprised of different phonemes. In real life communication people encounter fluctuating noise created by other talkers' voices most commonly. When taking fluctuating noise into consideration, the temporal issue requires further discussion. The shape of the temporal envelope of speech is typically dominated by plosive sounds, and the envelope of these sounds is characterized by

quick onsets (steep slope) and slow decays (shallow slope). Since the auditory system does better following the onsets instantaneously than the offsets of a signal, the onsets of other speech sounds will interact with the perception of target speech signals and result in temporal masking. Forward masking has a clear effect on speech intelligibility (Festen 1987). However, the effect of backward masking from fluctuating noise is still poorly understood. Its effect on speech intelligibility is still unclear and most likely not very large (Moore 2003).

As discussed above, different from English, Mandarin Chinese is a monosyllabic based language. Each syllable is formed by an initial consonant and a following vowel. These phonological differences contribute to the modulation pattern of the fluctuation of babble noise. If hearing-impaired people have more difficulty with fluctuating maskers, it would be important to investigate the difference in the signal-to-noise ratio required to obtain similar performance in English and Mandarin Chinese.

1.2.4 Glimpsing theory

Sounds with longer durations are easier to detect and recognize in loudness, pitch, and timbre. In fact, the auditory system also applies a time window analysis on sound. To continue sound processing over tens to hundreds of milliseconds, a single time window would have to be open for tens or even hundreds of milliseconds. However, there is no basis for a stimulus power integration mechanism, rather current thinking favors the notion of short power integration (10-millisecond) combined with multiple looks (Durrant & Feth, 2012). Glimpsing theory, as suggested by Cooke et al., 2006, provides a possible explanation for the lower signal-to-noise ratio needed than expected in listening in noise by suggesting integration of several temporal cues.

19

The effectiveness of energetic masking is highly dependent on its interaction with the acoustic characteristics of the target signal. Signal-to-noise ratio (SNR) represents the relationship between the dB level of the speech and the dB level of the noise. This ratio often is used as an indicator of how much difficulty a listener will have in noise (e.g., one person needs a +6 dB SNR to understand 50% of a signal and another person needs a +15 dB SNR to understand 50% of a signal). Cooke et al (2006) reported that SNR is not a perfect indicator for listening in noise. He suggested a two-way notation namely a spectral-temporal parameter to get a more precise analog of listening in noise at the peripheral level (Cooke, 2006). Speech is a highly modulated signal in time and frequency regions and energy with high informational value are concentrated in relatively few spectro-temporal regions. Even at quite adverse signal-to-noise ratios (SNRs) they are sparsely distributed, and significant masking in other regions has little impact on speech intelligibility. Speech is a highly-redundant signal, even for seemingly adverse energetic backgrounds. Some frequency regions survive the noise well and, if they can be detected and integrated, they often contain enough information to allow successful speech communication (Mattys et al., 2009). Based on the sparseness and redundancy, Cooke et al (2006) suggested the glimpses theory, where speech perception in noise is based on the use of glimpses of speech in spectral-temporal regions where it is least affected by the background noise. Even though the global SNR is identical in all masking conditions, the typical glimpse size differs. Rather than estimating local SNR directly, it is possible that listeners apply principles of auditory organization to group spectral-temporal regions based on properties such as the similarity of location or fundamental frequency. To group spectral-temporal regions also helps to track ongoing speech and this is more important when the back ground contains multi-talker's speech. Cooke et al (2006) exploited such glimpsing of speech and has demonstrated a close

match to listeners' performance in an English consonant identification task for both stationary and fluctuating noise. The glimpse model provided a precise prediction of intelligibility of English speech produced in noise (Lu et al., 2008).

Li et al (2007) studied important factors for glimpses on speech perception in noise in English. They found that the frequency location and total duration of the glimpses had a significant effect on speech information in the F1/F2 frequency region (0-3 kHz) for at least 60% of the utterance. The author pointed out that listeners can detect useful glimpses of speech at different times and in different regions especially in the F1/F2 of the spectrum and they can integrate those glimpses into the target speech (Li et al., 2007).

It is reasonable to suggest that the glimpse pattern of people who speak/listen to Mandarin Chinese might be different from individuals speaking/listening to English especially when the back ground noise is babble in Mandarin Chinese.

1.2.5   Informational masking by noise

Informational masking can include low-level sources of interference, such as masker-to-target misallocations of short spectro-temporal information (e.g., burst, frication), but it refers predominantly to the higher-level consequences of masking, independent of signal degradation, mainly: (1) Competing attention of the masker: The mere presence of a masker, whether it causes energetic masking or not, can lead to a depletion of attention resources needed for the main task. Simply put, this is the cost of the effort involved in ignoring the masker through stream segregation or selective attention. (2) Interference from a known language: The intelligibility of the masker itself can have an effect on performance in the main task. This type of informational masking is thought to be due to lexical semantic interference between masker and target. (3)

21

Higher cognitive load: This component of informational masking is usually based on the assumption that processing resources are limited. Consequently, any processing elicited by the masker will impair performance on the main task. This kind of informational masking especially applies to circumstances of divided attention, wherein participants are required to perform a task on both the target and the masker (Mattys et al., 2009).

The key issue of releasing the target from informational masking is to distinguish the target from back ground noise in order to follow the target. There are several factors such as visual cues, spatial cues, and fundamental frequency that can help release the target speech from the speech masker.

Helfer and Freyman (2005) explored the role of visual speech cues in reducing energetic and informational masking. The data they obtained indicated that visual cues provided more benefit for both recognition and detection of speech when the masker consisted of other voices (versus steady-state noise). Moreover, visual cues provided greater benefit when the target speech and masker were spatially coincident versus when they appeared to arise from different spatial location (Helfer et al., 2005).

Arbogast et al (2002) investigated the effect of spatial separation of sources on the masking of a speech signal. They used cochlear implant simulation software to divide the speech signal and the maskers into 15 logarithmically spaced envelope-modulated sine waves, and randomly assigned eight of these bands to the target speech and six other bands to the maskers. Again, this resulted in a stimulus that presumably produced little or no energetic masking but retained the potential target-masker confusions that would normally be present in multi-talker speech. For energetic masking both the target speech and maskers were assigned at the same bands. They found that in multi-talker speech, the advantage due to spatial separation of sources

is greater for informational masking than for energetic masking. While the signal and masker were sent to normal hearing subjects; the vibration of the basilar membrane is not as well separated with CI electrodes so the signal representation on the normal basilar membrane and the auditory nerve may have been different from what was expected (Arbogast et al., 2002).

When both the signal and masker are speech stimuli, the perception of spatial separation between the signal and masker can be sufficient for a significant speech recognition advantage to occur. Freyman et al. (1999) used localization dominance in the precedence effect to demonstrate this. When listeners were asked to identify nonsense sentences spoken by a female talker in the presence of speech-shaped noise, they showed an average 8-dB spatial release from masking with a 60° separation. In the presence of another female talker, the release was nearly 14 dB. This larger release was attributed to the presence of informational masking in the speech masker, which allowed the listener to use the perceived separation of sources as a cue to source segregation. This spatial release is larger than possible with head shadow and binaural interaction cues alone. Spatial release from the speech masker was 5 to 8 dB greater than that predicted by detection thresholds in the speech-shaped noise for the same location/perceived location configurations (Freyman et al., 1999).

The last but not the least important factor impacting signal and noise separation is fundamental frequency, especially when other cues by themselves are insufficient to isolate and attend to the target in the presence of the interference. Darwin et al (2003) investigated the effect of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers. They concluded that the greatest improvement in segregation came from different-sex talkers (Darwin et al., 2003).

What differences would be expected in the release from informational masking for different languages such as English and Mandarin Chinese? Fundamental frequency will be a major difference. As mentioned before, Mandarin Chinese has relatively longer vowel duration and must have tones on the vowels. These characters lead to low frequency acoustic energy which will be more important to the listeners as a cue. This leads to the listener trying to focus on the fundamental frequency but the competing speech produces increased variation of the fundamental frequency.

Wu et al (2005) found that release from informational masking, due to perceived target/masker spatial separation induced by the precedence effect, also occurs for tonal languages such as Mandarin Chinese. The degree of release was smaller for Mandarin Chinese compared to English. In noisy environments, especially multi-talker babble in Mandarin Chinese, people will have much more difficulty segregating the target speech (Wu et al., 2005). Therefore, if people want to have successful communication, they might try to rely on peripheral energy distribution. That is, to make their speech or voice spectrally-temporally prominent as compared to the back ground.

Yang et al (2007) studied the effect of voice cuing on releasing Mandarin Chinese speech from informational masking. A priori knowledge of the talker's voice and/or the content of the target speech improved speech recognition in a Mandarin Chinese cocktail-party environment. However, in this study there was no comparison to English, therefore, it is not clear if Mandarin Chinese benefit more on voice cuing or not (Yang et al., 2007). Mandarin Chinese is a tonal language where tonal variation is information bearing. To pay attention to the voice cue plays a much more prominent role in Mandarin Chinese communication than in English.

When energetic masking and informational masking are taken together, babble noise seems to bring more ambiguity to the Mandarin-Chinese listeners than to the English listeners. The modulation pattern of Chinese babble might have a different influence on speech perception due to temporal masking as well. In addition, the babble noise competes with the target on speech segregating. Therefore, in order to maintain some residual cues for the listener to segregate speech from babble noise and then continue to follow the target speech, people may need to speak louder in Mandarin Chinese.

1.2.6 Lombard speech

By modifying their vocal effort, speakers attempt to maintain a constant level of intelligibility in the face of degradation of the message by the environmental noise source. Some studies have reported intelligibility gains for Lombard speech (increased vocal effort) presented in noise when compared to normal speech in noise (Lu et al., 2008).

Lombard's seminal article (Lombard, 1911) describes the speech production of a patient with unilateral deafness when presented with an intense noise. The noise was presented first to the impaired ear, then to the normal ear, while the patient was being engaged in ordinary conversation. In the first case, the patient raised his voice slightly or not at all. However, with his good ear subjected to the noise, he immediately increased the vocal effort and fundamental frequency (F0), and reduced the vocal effort and F0 to the former level once the noise stopped.

Lombard speech can be thought to have two intentions: 1) to make the talker hear his or her own voice and 2) to make the intended listeners able to hear the talker.

Over the past decades a large number of studies have focused on the impact of background noise on speech production by asking talkers with normal hearing to speak under

noisy conditions. The speech materials, the noises and the procedures of those studies vary and are summarized in Table 1.4These studies are considered in more detail in the sections to follow.

Table 1.4 Literature Review of Lombard Speech studies

Author	Title	Noises	Materials	Procedures
Pittman & Wiley, 2001	Recognition of Speech Produced in Noise	White noise 80 dB SPL Multi-talker noise in 80 dB SPL	English Sentence	To encourage each talker to speak in a manner that would maximize recognition, an assistant wearing headphones was seated outside the window of the sound booth and instructed to write the final word of each sentence. Each talker was told that the listener was unable to see the features of her face and was instructed to speak clearly, to read the sentences in order, and to wait for the listener to look up from the response sheet before proceeding.
(Lau, 1998)	The effect of type and level of noise on long-term average speech spectrum	babble-noise, traffic-noise, restaurant-noise 50 dB SPL, 65 dB SPL, and 85dB SPL	712 words Chinese newspaper article.	Participants read aloud articles in Cantonese naturally and ignored errors.

Table 1.4 (continued)

Author	Title	Noises	Materials	Procedures

Author	Title	Noise	Materials	Procedures
(Summers *et al.*, 1988)	Effects of noise on speech production: Acoustic and perceptual analyses.	white noise	15English words	Participants were told to read out the 15 words. They were told that the experimenter would be listening to their speech outside the booth while the recording was being made.
(Junqua, 1993)	The Lombard reflex and its role on human listeners and automatic speech recognizers	white-gaussian noise in 85 dB SPL	49 English words	Read out the words while recording.
(Tartter *et al.*, 1993)	Some acoustic effects of listening to noise on speech production	White noise in 35, 60, 80 dB SPL	English words	Participant read out those words.

Table 1.4 (continued)

Author	Title	Noise	Materials	Procedures

(Garnier *et al.*, 2006)	An acoustic and articulatory study of Lombard speech: global effects on the utterance	85 dB white noise 85dB cocktail party noise	French short sentence with a subject-verb-object(SVO) structure	Noise was presented through loudspeakers and participants were asked to read the sentences.
(Bond *et al.*, 1989)	Acoustic-phonetic characteristics of speech produced in noise and while wearing an oxygen mask	pink noise at a level of 95 dB SPL	English Words CID spondee words	Participants read the words.

Table 1.4 (continued)

Author	Title	Noises	Materials	Procedures
(Junqua *et al.*, 1998)	Influence of the speaking style and the noise spectral tilt on the lombard reflex and automatic speech recognition	Pink noise	English words	Participants read the words.
(Hansen *et al.*, 2009)	Analysis and Compensation of Lombard Speech Across Noise Type and Levels With Application to In-Set/Out-of-Set Speaker Recognition	Car noise 70 80 90 dB Large crowd noise 70 80 90 dB Pink noise 65 75 85 dB	English words	Participants read the words.
(Patel *et al.*, 2008)	The Influence of Linguistic Content on the Lombard Effect	multi-talker noise at 60 and 90 dB SPL	An interactive computer game based on short sentence.	Sixteen speaker–listener pairs engaged in an interactive cooperative game and seated separated at different rooms communicated only through head phones

Table 1.4 (continued)

Author	Title	Noises	Materials	Procedures
(Letowski *et al.*, 1993)	Acoustic properties of speech produced in noise presented through supra-aural earphones.	Wideband noise Traffic noise Multi-talker noiseat70 and 90 dB SPL	connected speech (131 word passage "My Grandfather")	Participants read the words.
(Korn, 1954)	Effect of psychological feedback on conversational noise reduction in rooms	White noise	Conversation in English	Participants participated in a conversation with the operator who was also wearing similar earphone fed by the same noise
(Webster *et al.*, 1962)	Effects of ambient noise and nearby talkers on a face-to-face communication.	Ambient thermal noise levels of 65, 75, and 85 dB	English Words	From 1 to 5 talker-listener pairs, talkers seated shoulder-to shoulder on one side of a table with listeners on the other, communicated word lists in conditions of quiet and ambient thermal noise levels of 65, 75, and 85 dB. Each talker read one word at a time to his listener-partner, who repeated back each word for verification by the talker.

Table 1.4 (continued)

Author	Title	Noise	Materials	Procedures
Castellanos *et al*, 1996)	An analysis of general acoustic-phonetic features for Spanish speech produced with the Lombard effect	white noise in 85 dB	Spanish Continuous speech corpus.	Read out materials as clearly as possible.
(Dreher *et al.*, 1957)	Effects of ambient noise on speaker intelligibility for words and phrases	Wide band noise at 70, 80, 90, 100 dB	25 spondee and 25 sentences in English	Participants read the words.

1.3   LOMBARD SPEECH

A speaker modifies his vocal output while speaking in the presence of background noise. This phenomenon is called the Lombard effect after the French oto-rhino-laryngologist Etienne Lombard, who first described the impact of noise on speech production (Lombard, 1911).

1.3.1   The main acoustic changes of Lombard speech

Compared to normal speech, the main acoustic changes of Lombard speech in English are:

- Increase in speech level

- Increase in fundamental frequency (F0)

- Shift of spectral energy to higher frequency (Bond et al, 1989a)

- Increase in vowel duration

- Shift in formant center frequencies for F1 (mainly) and F2 (Pittman & Wiley, 2001)
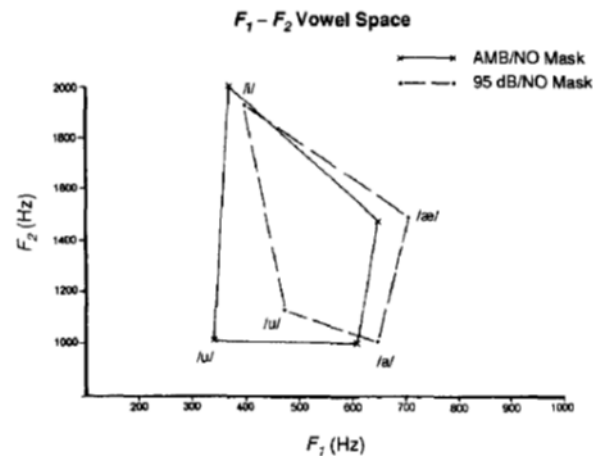
Figure 1.3 Shift in average center frequencies of F1 and F2 as function of a 95 dB pink noise for male speakers.

(Reprinted from "Acoustic--phonetic characteristics of speech produced in noise and while wearing an oxygen mask" by Z. S. Bond, Thomas J. Moore and Beverley, Gable, 1989, The Journal of the Acoustical Society of America, 85(2), 907-912. Copyright 2014 by Acoustic Society of America. Adapted with permission.)

Increases in F0 and F1 frequency in Lombard speech are physiologically related to the raised vocal effort. During Lombard speech production, the raising of sub-glottal pressure and the increase of tension in the laryngeal musculature needed to create a louder voice contribute to an increase in F0 (Schulman, 1989). Likewise, in order to increase speech level, the wider mouth opening, accompanied by lowering the jaw and the tongue, induces an increase in F1.

Lau (1998) investigated the effect of type and level of noise on long-term average speech spectrum by using Cantonese Chinese speech material. There were no significant differences in

32

Long Term Average Speech Spectrum (LTASS) from Cantonese to English which was consistent with Byrne et. al (1994). However, there was a significant difference in the Cantonese Chinese LTASS in quiet versus in noise at 50, 65, and 80 dB SPL. English was not tested. At least in Cantonese Chinese, the level of noise presented to the speaker impacted the final spectrum shape of the Lombard speech. This would support the need for creating Lombard speech materials for testing using the intended background noise level during recording in order to replicate the LTASS that would have been achieved in a real communication situation.

1.3.2   Factors impacting changes observed in Lombard speech

The degree of the acoustic changes observed in Lombard speech is influenced by many factors. These factors are described below.

   1. Noise level.

Noise level affects changes in word duration, vocal intensity and F0 (Dreher & O'Neill, 1957; Hansjörg &Mixdorff, 2006; McCullough et al., 1993; Mixdorff et al., 2006; Patel & Schell, 2008; Summers et al., 1988b; Webster &Klumpp, 1962). Dreher and O'Neill (1957) reported increases in word duration from 15 to 31% and a 6 to 9 dB increase in intensity in Lombard speech over speech produced in quiet. Pittman and Wiley (2001) reported that compared to quiet, the vocal levels produced in wide band noise and multi-talker noise increased an average of 14.5 dB; word duration also increased an average of 77 ms. Lau (1998) examined people who speak Cantonese. She reported that the overall level and the 1/3 octave band levels of speech produced in quiet are significantly different from those produced in noise conditions at 50 dB SPL, 65 dB SPL, and 85 dB SPL. The overall voice levels can increase by 3 dB or more for an increase of 10 dB in noise levels. Also, Lau (1998) compared her data of mean overall level of male and female

Cantonese speakers in quiet to Byrne et al (1994) who investigated 12 languages and reported no differences between the average long term speech spectrums reported in each study (Byrne D, 1994; Lau, 1998). Lau's data was a result of asking people to read a newspaper in Cantonese Chinese aloud which did not reflect conversational speech and there was no comparison of speech in noise between different languages. With the rise in noise level from 70 to 90 dB SPL, Letowski et al (1993) found an increase in F0 of between 10 to 20 Hz in English..

2. Types of noise also affect acoustic changes in Lombard speech.

Lombard speech has been studied in varying types of noise, including broadband noise, traffic noise, white noise, pink noise, and multi-talker noise (Bond et al., 1989b; Hansen &Varadarajan, 2009; Hansjörg & Mixdorff, 2006; Patel & Schell, 2008; Pittman & Wiley, 2001; Lau, 1998). Pittman and Wiley (2001) reported that the long term average speech spectrum in wide band noise and in multi-talker noise was not simply an elevated long term average speech spectrum in quiet. There is frequency tilt producing larger amplitudes at mid-high frequency (2.5 kHz) than at lower frequencies (0.2 kHz). The LTASS is different in multi-talker babble compared to wide band noise. Since there were only 6 talkers in this study, this difference could not be generalized as a widely existing phenomenon, but did suggest that the LTASS in Lombard speech produced in multi-talker babble could predict better performance in noise with Lombard speech than with conversational speech in babble presented at the same signal-to-noise ratio.

3. The role of the word in a sentence for Lombard speech.

Patel and Schell (2008) observed larger effects of F0 and duration for information-bearing word types (Patel & Schell, 2008). Rivers and Rastatter (1985) studied the effect of multi-talker noise on the production of stressed and non-stressed words in spontaneous speech. The noise

conditions included quiet, 90 dB of white noise, and 90 dB multi-talker noise. Across noise conditions, the average F0 for stressed words increased by 62 Hz from the quiet condition for both males and females while the average F0 for non-stressed words increased by only 33 Hz for males and 25 Hz for females (Rivers, 1985).

4. The language spoken.

Junqua (1996) reported that comparisons of French, Spanish, and Japanese to American English showed that the Lombard speech pattern of French is very similar to American English. However, the amount of variation in formant frequency F1 and F2 (Figure 1.3) between normal and Lombard speech was found to be more important in the case of American English compared to French, Spanish and Japanese. In the case of Spanish, while similar tendencies to that observed for American English and French were reported, some differences in the variation of the second formant were noted. The second formant generally increased. For Japanese, a shift to higher frequencies for the first and second formants when they are below 1500 Hz and a shift to lower frequencies for the higher formants when above 1500 Hz were reported (Jean-Claude, 1993; Junqua, 1996). Because there were too many degrees of freedom in these studies, it was difficult to identify the origin of the differences observed across languages. Furthermore, the dependence of the Lombard reflex on the speaker and the fact that these studies have been done with only a few speakers limit the comparison. Lau (1998) reported a Lombard speech study with people speaking Cantonese Chinese in different types of noises. She found similar intensity and spectrum changes on Lombard speech in Cantonese regardless of noise type. Her study did not compare Cantonese Chinese to other languages.

5. Speaker gender

Junqua (1993) reported that the influence of the Lombard effect on vocal effort and F0 was greater for male speakers than for females (Jean-Claude, 1993). But Patel and Schell (2008) failed to find this effect. Lau (1998) reported the Lombard effect on speech spectrum changes on males and females who speak Cantonese Chinese. She reported that the average male speech spectrum had larger SNR than the female speech spectrum at 500 Hz; the average female speech spectrum had larger SNRs than the male speech spectrum at 2 kHz and 4 kHz, especially for noise conditions at 80 dB SPL.

6. Type of task employed to study the Lombard effect.

The magnitude of the Lombard effect appears to be governed by the premium on intelligible communication (Lane et al., 1971). In most studies of Lombard speech, the majority of the tasks are reading words or sentences which have little, if any, intelligibility premium compared to communicative scenarios, in which a speaker and listener are engaged in a cooperative task. Thus, if speakers are not engaged in a communicative interaction, intelligibility may not be a concern, and the acoustic changes may not be comparable to those that would be made in natural conversations. In a study by Garnier (2007), individual talkers were asked to complete a non-interactive task alone or an interactive task with a speech partner. In both tasks the background was quiet or contained wideband noise. Noise-induced speech modifications such as increases in speech level, F0 and vowel duration as well as an increase in F1 and more spectral energy in higher frequencies were larger in the interactive task.

Other studies observed a larger increase in speech level due to the rise of noise level when speakers were communicating (Webster and Klumpp,1962; Gardner, 1964) compared to just reading texts (Dreher and O'Neill, 1957; Laneet al., 1970). The reaction to noise is not solely

a reflex, but rather consciously driven by other factors such as the speaker's effort to maintain effective communication in noise (Lu, 2010).

When ambient noise is present, speech production modifications might firstly be the consequences of overcoming the speakers' difficulty in monitoring their own productions due to both energetic and informational masking effects of the noise. Furthermore, if the speech is a communicative task, speakers also need to arrange spectral-temporal planning to reduce the amount of overlap with a modulated background noise. The effect will be stronger when the noise contains intelligible speech (Lu, 2010).

1.4   SUMMARY AND EMPIRICAL QUESTION

Table 1.5 provides a summary of the differences between Mandarin Chinese and English across linguistics, culture, noise masking, and Lombard speech which have been discussed in the previous sections. The differences identified are based on analysis of conversational speech and would predict that native Mandarin-Chinese speakers would require an enhanced signal-to-noise ratio as compared to English when communicating at a conversational level in noise. It is unclear if this prediction would apply to Lombard speech since there is a paucity of data in this area. Since it is Lombard speech that is most typically used in noisy situations, it is this condition that is of most interest.

Table 1.5 Summary of differences of Mandarin Chinese and English across linguistics, culture, noise masking and Lombard speech.

	Differences in Mandarin Chinese Vs English
Linguistic	1.More high frequency consonants.

Characteristics	2. More diphthongs and complicated combinations.
	3. Tones accompany the vowels and are important cues for lexical distinction.
	4. Each mono-syllable is meaningful, initial consonants play an important role in syllable discrimination.
	5. Higher information density for each sentence.
Culture	1.More population density
	2.The negative perception of added noise level is greater when the speech level is high and vice versa.
Masking	1.Energetic masking:
	Initial consonant compromised in noise which largely compromises communication.
	2. Information masking:
	Different segregation strategy focuses on information bearing monosyllabic based words.
	The degree of release from informational masking by using spatial cues was smaller for Mandarin Chinese.
	Attention to voice might be more important for English than tonal information.
Lombard Speech	Inadequate research so far on the comparison of Lombard speech of Mandarin Chinese and English

Although di-syllable vocabulary is the main unit of modern Mandarin, identification of every monosyllable correctly is the key to efficient communication. Language is actually a communicative system whose primary function is to transmit information. The unity of all languages is probably to be found in this function, regardless of the different linguistic strategies on which they rely. Efficient communication conveys the correct information to the target subject. Based on phonology, morphology and syntax differences, it is reasonable to hypothesize

that the performance in a given signal-to-noise ratio for people who speak English might differ from people who speak Mandarin Chinese. Signal processing aimed at reducing the impact of noise is developing rapidly to be applied to amplification systems for people with hearing loss. As these systems are refined and signal-to-noise ratio can be enhanced, it will be of interest to know what levels of performance to expect based on the languages spoken or what type of signal processing might be more appropriate based on a given language. Noise exists everywhere in our everyday life and communicators find themselves speaking loudly to convey information while at the same time their voice further contributes to the environmental noise around them. Based on the differences between English and Mandarin Chinese, the potential differences in the impact of noise on the languages, and the differences in Lombard speech used to communicate in noisy situations, the study question that arises is: Do native Mandarin-Chinese listeners need an enhanced signal-to-noise ratio to achieve the same performance when communicating in babble noise as American English listeners?

2.0   METHODS

The current research question asks: do native Mandarin-Chinese listeners need an enhanced signal-to-noise ratio to achieve the same performance when communicating in babble noise as American English listeners? To investigate the question, two linguistically distinct groups were examined. The following sections outline the characteristics of the speech stimuli, signal processing strategies, study participants, and procedures for administration of the protocol for the experiment.

2.1   STIMULUS

2.1.1   Speech materials

In order to better reflect everyday communication situations, sentence material was chosen as opposed to single word material. Table 2.1 lists several speech recognition tasks with sentences as the test material.

Table 2.1 Speech material chosen criteria

Speech materials	Sentences	predictability	Close to conversation	Phonetically balanced	Test efficiency
HINT test (Nilsson *et al.*, 1994)	American BKB (Bamford-Kowal-Bench) sentences(Bench *et al.*, 1979)	Short, highly redundant sentences rich with semantic and syntactic context; designed for a first-grade reading level	Short sentences for children	Yes	High
SPIN test	Low predictability sentences	Low	Yes	Yes	Low, only one key word for each sentence
QuickSIN test (Killion *et al*, 2004)	IEEE sentences	Provides mainly syntactic cues with only subtle semantic cues to aid in recognition.	Yes	Yes	High
CST (Connected Sentence Test)(Cox *et al.*, 1987)	Continues passages which contain 10 syntactically simple sentences, 7 to 10 words in length.	High	Yes	no	medium

Table 2.1 (continued)

Speech materials	Sentences	predictability	Close to conversation	Phonetically balanced	Test efficiency
BKB-SIN (Bamford-Kowal-Bench Speech-in-Noise Test)		Short, highly redundant sentences rich with semantic and syntactic context designed for a first-grade reading level	Short sentences for children	Yes	high

The IEEE (Institute of Electrical and Electronics Engineers) sentences in the QuickSIN test (Killion, Niquette et al. 2004) were chosen as the test material of the proposed study because they have low predictability, are phonetically balanced, and demonstrate test efficiency. The Harvard sentences or IEEE sentences (Rothauser et al., 1969) are phonetically-balanced sentences that use specific phonemes at the same vocabulary frequency in English. The IEEE sentences are more difficult than other sentence tests primarily because of the few contextual cues available to aid in key word identification. They are widely used in research in telecommunication and speech and acoustics research where standardized and repeatable sequences of speech are needed. These sentences were recorded in English for the English speaking participants and translated into Mandarin Chinese for the Mandarin-Chinese speaking participants. Recording methods will be described later in the document.

According to Bentler (2000), the 360 sentences in 9 lists of IEEE sentences did not provide evidence of psychological equivalence. Psychological equivalency means equivalent at

the difficulty level which was investigated through equivalent percent-correct and SNR scores at each presentation level (Bentler, 2000). The IEEE sentences were used to construct the QuickSIN test and SIN tests (Killion et al., 1993; Killion et al., 2004). Based on the criticisms from Bentler (2000), Killion (2004) did the final selection of 12 lists of sentences from the IEEE sentences for the QuickSIN based on selecting not only phonological but also psychological (difficulty) equivalence.

In the QuickSIN test, five key words are labeled for each sentence. For example, the sentence would be presented as "A white silk jacket goes with any shoes.(一件白色絲綢外套可以和任何鞋子搭配.)" [yijian baise sichou waitao keyi he renhe xiezi dapei] These sentences were translated into Mandarin Chinese (see Appendix A). Making equivalent sentences in Mandarin Chinese and English is very difficult. Every aspect of equivalence was controlled as much as possible with the goal of eliminating any possible differences in test materials as a potential explanation for any differences that might be found in performance. In order to verify the comparability of the QuickSIN test sentences and their translated Mandarin Chinese version, several aspects of the test were verified in the English and Mandarin Chinese version. The parameters that were evaluated are listed in Table 2.2 and the procedures are detailed below.

Table 2.2 Insuring equivalence in difficulty between the English IEEE sentences and the Mandarin Chinese translation.

	English	Mandarin Chinese
Word token frequency	At the same vocabulary frequency in English	Within range of word token frequency of conversational vocabulary
Phonetic equivalence	Phonetically-balanced (Rothauser et al., 1969)	Compared to Tang et al's (1995) data Vowel R=.827 Consonants R=.809
VP initial structure (Head-final structure)	Subject→ Predicate →Object	
Contextual predictability	Low: 90% Medium: 6.7% High: 3.3%	Low: 89.7% Medium: 6.0% High: 4.3%

2.1.1.1 Word token frequencies

The Mandarin Chinese version of IEEE sentences' word token frequencies were maintained at a similar lexical frequency range as compared to the original English version. A comparison was accomplished by examining word token frequency as calculated from SUBTLEX which is a set of files providing word and character frequency measures based on a corpus of film subtitles (33.5 million words). The reason to choose this corpus instead of a more common corpus which comes from written or printed vocabulary data was the vocabulary in SUBTLEX is more closely

related to conversational vocabulary (Cai et al., 2010). The measurement parameter is W-CD% (word-contextual diversity percent). It means word token frequency measured by contextual diversity in percentage of observed films. According to Adelman et al (2006), it was CD (contextual diversity) instead of word frequencies that determined word-naming and lexical decision times. The larger the number in CD indicates that the word could have more contextual diversity and would be harder to identify (Adelman et al., 2006). Figure 2.1 shows the distribution of W-CD% for the Mandarin Chinese translated IEEE sentences. The data reveal two major peaks. One is located at less than 20% (majority of the words) which is mainly comprised of the di- or tri-syllable words and the other is located at 90% which is mainly formed by the monosyllable words. The less contextual diversity implies that the listener will have an easier time identifying the correct word. Therefore, most di- and tri-syllable words are not difficult words to identify in translated IEEE sentences. It was not surprising to see the high contextual diversity in monosyllabic words based on the previous literature review which indicated that these words are often ambiguous. The majority of the words included in the sentences fall at 20% or below indicating that they should be familiar to the listeners, similar to the English version of the test.
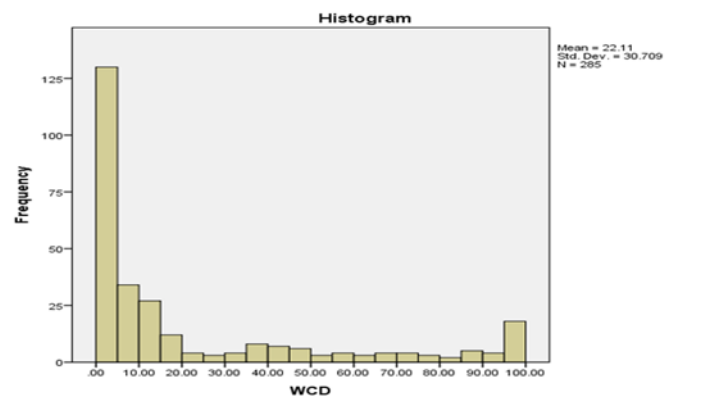
Figure 2.1 W-CD% of Translated Mandarin Chinese IEEE sentence key words

2.1.1.2 Phonological balancing

In order to test the phonological balance of the translated speech materials, the key words are labeled in Pinyin (a method to label the pronunciation of these words) and divided in initial consonant and following vowels and are shown in Table 2.3 and Table 2.5 . The percentage of each consonant and vowel was compared to Tang et al's (1995) data from a statistical analysis of phonological distribution of Mandarin Chinese. Pearson correlation was accomplished with SPSS 20, the correlation co-efficient R of vowels was .827 (p<0.001) and the R of consonants was .809 (p<0.001). Both results indicate a high correlation which means the translated key words in Mandarin Chinese are adequately phonologically balanced. The initial consonants and nucleus of the key words of English IEEE sentences are listed in Table 2.4 and Table 2.6 in American English pronunciation. The high frequency affricate and fricatives "c", "ch", "zh", "q" as well as "x" account for 20% of all initial consonants in Mandarin Chinese which are not pronounced in English. Therefore, Mandarin-Chinese listeners typically experience more high frequency consonants which may be fragile in noise.

Table 2.3 Consonants distribution in Mandarin Chinese IEEE sentence key words in percentage

	B	P	T	D	K	g
Number	42	20	25	51	17	22
Percentage	7.38	3.51	4.39	8.96	2.99	3.87
Reference range	5.15	0.98	2.45	D	1.83	5.50
	Sh	Ch	X	Z	Zh	J
Number	26	17	40	25	27	32
Percentage	4.57	2.99	6.91	4.39	4.75	5.62
Reference range	7.66	2.75	4.86	3.01	7.18	6.98
	F	S	C	Q	N	M
Number	14	12	12	19	7	22
Percentage	2.46	2.1	2.11	3.34	1.23	3.87
Reference range	2.45	1.08	1.15	3.11	2.53	3.74
	H	R	L	w+y+0		
Number	26	8	36	47+10+2		
Percentage	4.57	1.4	6.33	10.3		
Reference range	4.42	1.94	5.69	12.45		

Table 2.4 Consonants distribution (initial onset) in English IEEE sentence key words in percentage

	b	P	T	d	K	g	tʃ	dʒ	
Percentage	6.27	7.0	5.9	6.3	4.1	0	1.5	3.0	
	f	V	Θ	ð	S	z	ʃ	ʒ	
Percentage	7.4	0.3	1.5	1.1	7.0	0	3.0	0	
	M	N	h	r	J	w	L		
Percentage	3.3	2.2	4.1	3.3	1.1	5.2	5.5		
	Pl	Bl	kl	gl	Pr	br	Tr	Dr	kr
Percentage	0.37	1.1	1.5	1.1	0.7	0.37	0.37	2.6	1.5
	Fl	Sl	fr	θr	Sw	sp	st	Sk	sm
Percentage	1.1	1.5	1.8	0.73	0.37	1.5	3.7	0.74	0.74
	Str	Skr	skw						
Percentage	1.5	0.37	0.37						

Table 2.5 Vowel distribution of translated Mandarin Chinese IEEE sentence keywords in percentage

	I	e	a	o	U	ai	Ei
Number	77	31	19	3	52	26	7
Percentage	13.80	5.56	3.41	0.54	9.32	4.66	1.25
Reference range	15.21	12.38	3.89	0.54	7.11	2.83	1.28
	Ie	ue	er	in	Ia	ui	iu
Number	9	6	1	9	15	13	10
Percentage	1.61	1.08	0.18	1.61	2.69	2.33	1.79
Reference range	2.42	1.01	0.28	1.95	1.09	2.75	2.60
	Uo	ou	un	ao	An	ua	en
Number	17	17	4	31	37	5	14
Percentage	3.05	3.05	0.68	5.56	6.63	0.90	2.51
Reference range	4.40	1.88	0.89	3.10	3.41	0.44	3.62
	Ing	iao	ian	iang	ong	ang	uang
Number	29	22	33	7	17	12	8
percentage	5.20	3.94	5.91	1.25	3.05	2.15	1.43
Reference range	3.05	2.06	4.10	1.80	4.18	2.87	0.65

Table 2.5 (continued)

	uan	uai	Eng				
Number	10	1	16				
percentage	1.79	0.18	2.87				
Reference range	0.85	0.32	3.09				

Table 2.6 Nucleus distribution of English IEEE sentence key words in percentage

	ɪ	ɛ	æ	ʌ	ʊ	i	E	
Percentage	9.0	8.67	6.0	3.67	1.33	9.0	6.0	
	U	O	ɔ	ɑ	ə	ɚ		
Percentage	3.0	4.0	2.33	1.0	1.67	1.33		
	ɔɪ	aɪ	aʊ	ɪɹ	ɛɹ	ɔɹ	ɑɹ	
Percentage	0.33	4.67	3.0	0.33	1.0	2.0	2.0	
	æn	ɪn	ɛn	On	ʌn	ɚn	ɔn	aʊn
Percentage	1.67	2.0	2.67	0.67	1.67	0.33	0.33	0.33
	æŋ	ɪŋ						
Percentage	0.33	1.0						
	ɪl	ɛl	il	El	ɔl	ɔɪl	aɪl	ɚl
Percentage	0.67	1.33	0.67	2.0	2.0		0.67	0.33
	ɛm	ʌm	aɪm					
Percentage	0.33	0.67	0.33					

2.1.1.3 VP intial structure

When labeling the key words for the Mandarin Chinese sentences, there was not 100% word to word correlation due to syntactical differences especially in the cases of attributives, adverbials, and prepositions. For this reason, some of the key words labeled under the English sentences could not be replicated in the Mandarin Chinese version. For example, "The sink is the thing in which we pile dishes." (水槽是我们用来堆叠盘子的地方.) In this translation, the word "which" did not translate directly into Mandarin Chinese, therefore, this word was not labeled as a key word in this study. The grammar of IEEE sentences is clear and there is little ambiguity. The

sentences have a VP initial structure (head-final structure). When translating the sentences into Mandarin Chinese, the VP initial structure (head-final structure), meaning the subject should be followed by the predicate and then the object, was strictly maintained. However, the location of attribute and adverbial might change according the grammar of Mandarin Chinese. Two native Chinese speakers checked the sentences and if either judge indicated that a sentence did not sound natural, the sentence was deleted in the prepared materials. The sentences came from IEEE sentence lists 1, 2, and 9 (Bentler, 2000).These sentences were translated into Mandarin Chinese because it was found that only those lists were equivalent for group comparison (Bentler, 2000).

2.1.1.4 Context predictability test

In order to test context predictability of the key words in IEEE sentences from the QuickSIN and the context predictability of the translated Mandarin Chinese version, 50 native English speakers and 50 native Mandarin-Chinese speakers participated in a test of the material. These subjects were not familiar with these words lists (audiologists and audiology students were excluded) and did not participate in the main part of the experiment. One key word was randomly omitted from each sentence and the participants were asked to fill in the blank. Each participant received only one example of the sentence with a particular word missing. Another individual would receive the same sentence with a different word missing, etc. Individuals received the lists in a paper and pencil format with instructions to fill in the blanks as best as possible in one sitting. The activities took approximately 10 minutes. Examples:

"A white silk jacket goes with any _____.

A white _____ jacket goes with any shoes.

A ＿＿ silk jacket goes with any shoes.

A white silk ＿＿ goes with any shoes.

A white silk jacket goes with＿ shoes."

"一件白色丝绸＿＿可以和任何鞋子搭配.

一件 ＿＿ 丝绸外套可以和任何鞋子搭配.

一件白色＿＿外套可以和任何鞋子搭配.

一件白色丝绸外套 可以和任何＿＿搭配.

一件白色丝绸外套可以和＿＿ 鞋子搭配."

Percentage correct prediction was calculated and the frequencies of correctly predicted words are shown in figure 2.2. Alexandre et al. (2011) established categories of predictability for the purpose of comparing sets of materials and populations. They indicated that correct identification of <50% could be interpreted as "low predictability", 50-70% correct was "medium predictability" and >70% was "high predictability". Table 2.7 provides the percent of 300 key words scored in each of these ranges and illustrates that predictability based on context was almost identical between the English and Mandarin Chinese versions of this test.

ENG

Mean = 18.27%
Std. Dev. = 23.384%
N = 300

Frequency

125
100
75
50
25
0

.00%   30.00%   60.00%   90.00%   120.00%

ENG

CHN

Mean = 18.07%
Std. Dev. = 24.377%
N = 300

Frequency

125
100
75
50
25
0

.00%   20.00%   40.00%   60.00%   80.00%   100.00%   120.00%

CHN
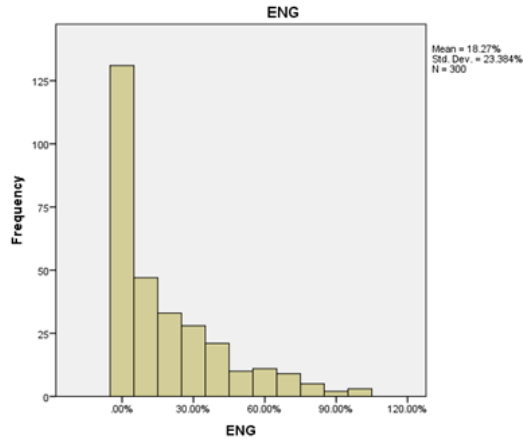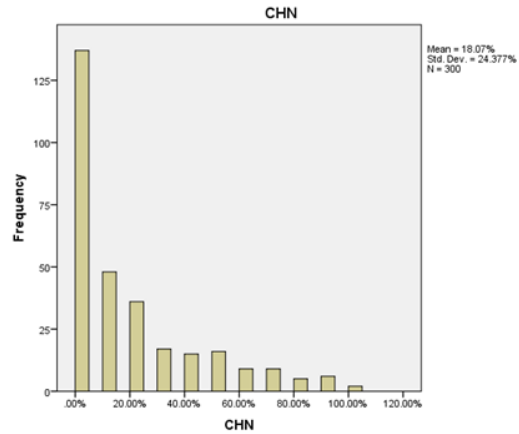
Percentage correct   Percentage correct

Figure 2.2 Distribution frequency histograms of the context prediction score in English (ENG)

and Mandarin Chinese (CHN)

Table 2.7 Predictability of key words and classification of predictability level

Predictability level		English	Mandarin Chinese
Low	<50%	90%	89.7%
Medium	50-70%	6.7%	6.0%
High	>70%	3.3%	4.3%

Paired T-tests did not reveal a significant difference in context predictability between English and Mandarin Chinese sentences in the study speech material (t=.129, 299, p=.898). Almost 90% of the key words in both languages had low predictability. As shown in Table 2.8, for those key words having medium and high predictability, there were 48 key words in English where the contextual predictability was larger or the same as the correspondent Mandarin Chinese. Only 11 out of 300 key words had higher contextual predictability in Mandarin Chinese

51

than in English sentences. Therefore, by translating the English sentence material into Mandarin

Chinese, there was no increase in contextual predictability for the key words in the sentences.

Table 2.8 Keywords with different context predictability

ENG>CHN N=41				Eng<CHN N=11				ENG=CHN N=7			
Path	路径	80%	10%	sheet	纸	33%	70%	tear	撕下	60%	60%
Near	靠近	60%	0%	glasses	杯	20%	70%	post	粘贴	60%	50%
Air	空气	80%	40%	drifts	逐流	40%	60%	sense	觉	80%	70%
waters	水域	60%	0%	along	随波	0%	80%	both	两	90%	90%
before	之前	60%	40%	cut	削	0%	70%	sun	太阳	90%	90%
miss	错过	60%	0%	sharp	尖	80%	90%	foot	脚	100%	80%
straw	稻草	83%	10%	method	办法	17%	90%	store	商店	60%	70%
sink	水槽	60%	20%	rain	大雨	50%	100%				
which	用来	70%	0%	space	空间	40%	60%				
which	用来	70%	0%	roused	惊醒	70%	80%				
dishes	盘子	100%	0%	sleep	睡	70%	100%				
smell	嗅	67%	50%								
Up	起	90%	60%								
dice	骰子	60%	0%								
better	更好	80%	0%								
than	比	70%	0%								
helped	帮助	80%	20%								
told	讲	100%	30%								
ends	头	100%	60%								
If	如果	90%	30%								
swayed	摇摆着	60%	50%								
stayed	还在空中	67%	40%								
hot	炎热	80%	40%								
Sun	阳光	67%	30%								
secret	保密	80%	0%								
many	很多	60%	30%								
pole	杆	60%	50%								
floor	地板	70%	20%								
Tell	讲的	100%	50%								
with	用	67%	20%								
Fell	落下	60%	20%								
worse	糟糕	60%	0%								
made	使得	100%	0%								
Peg	挂钩	80%	0%								
hole	洞	70%	0%								

Table 2.8 (continued)

start	开始	100%	20%		
deeply	深深得	80%	30%		
toad	蟾蜍	80%	50%		
frog	青蛙	100%	90%		
Tell	区别	100%	60%		
apart	开	100%	0%		
deep	熟	90%	40%		

2.1.2 Stimuli recording

One female native English speaker and one female native Mandarin Chinese speaker were recruited for the sentence recording. Both speakers spoke standard American English and standard Mandarin Chinese, respectively. This quality of speech was judged by three native speakers of each language. Acceptance of the speaker required all three judges to agree that the speaker represented standard American English or standard Mandarin Chinese.

For recording purposes, the speakers were seated in a double-walled sound treated booth with their mouth 5 inches (13 centimeters) away from a mounted CAD U1 dynamic microphone (20 Hz-20 kHz). The microphone was routed to a PC digital recorder with settings for a mono recording at a 44.1 kHz sampling rate with Adobe Audition CS5.5. A 1 kHz pure tone was

recorded at the beginning as a calibration tone for later playback calibration. The sensitivity of the microphone was adjusted to prevent any peak clipping of the speaker's voice. A speaker and a researcher (listener) were seated face to face in the sound treated room. A small table separated these two people by 80 cm and the CAD U1 microphone was placed on the small table and faced the speaker. Sentences were recorded under two conditions: speaking in quiet and speaking in babble noise. While recording speaking in babble noise, the speaker wore insert earphones in both ears and the listener also wore insert earphones. Babble noise was sent through the insert earphones to the speaker. When speaking with the babble noise, the following instructions were given. "Please memorize each sentence before you say it. Once you have memorized the sentence, please look at the listener seated across from you and repeat the sentence three times without looking back at the paper. While you are speaking the sentences, you will hear noise through the insert earphones. The listener is hearing noise as well so please speak so your sentence can be heard correctly. The listener will be writing down what you say. Please continue to do this until you have repeated all of the sentences." When speaking in a quiet environment, the instruction was "Please memorize each sentence before you say it. Once you have memorized the sentence, please look at the listener seated across from you and repeat the sentence three times without looking back at the paper. The listener will be writing down what you say, so please speak so your sentence can be heard correctly. Please continue to do this until you have repeated all of the sentences."

Recording noise that might be picked up by the microphone was eliminated by Adobe Audition CS5.5. Before the speech recording started, 5 seconds of silence was recorded as a background noise floor. The Adobe Audition analyzed this noise and removed this noise throughout the recording without compromising the speech recording.

In order to record high quality speech, insert earphones were used to deliver the noise to the speaker. However, this recording situation did not reflect the real life condition of people communicating in noise. The insert earphone brought two effects to the communicators. Firstly, it created an occlusion effect which made it easier for the speaker to monitor her own voice. Secondly, the insert earphone produced an attenuation effect for the communicator by blocking the air conduction of her own voice (Tufts et al., 2003). Killion et al. (1985) reported that one's own voice could get to 100 dB SPL in the ear canal especially for the closed vowel "EE" and "OO" and the intensity of "EE" was larger than "OO" when placing an ear mold or ear plug in the ear canal. The amplification by the vibration of cartilage and bone occurs at 250 Hz, 500 Hz and 1000 Hz (Killion et al., 1985).   In order to minimize the occlusion effect, a test of occlusion effect was conducted for each of the speakers. According to Zwislocki (1953) and Killion (1988), the location of the insert earphone produces different levels of occlusion effect where shallower insertion results in larger occlusion effect (more enhancement of sound) and deeper insertion results in a smaller effect (Killion et al., 1988; Zwislocki, 1953).   A probe microphone was placed in the ear canal of the speakers and the RMS of the sound from the individual's ear canal while pronouncing a vowel was displayed on the Verifit Probe Microphone system monitor. This was done by comparing the RMS level of the probe microphone and the reference microphone which was located at the entrance of the ear canal.   The speaker was required to pronounce a sustained "ee" in a normal speaking level and loud speaking level. The level was displayed on the Verifit monitor in order to allow the speaker to monitor her speech production. The insert earphone was adjusted (inserted more deeply) and the measurement was repeated until the occlusion effect shown on the Verifit system did not reduce. The goal was to produce an occlusion effect of less than 3 dB. This constituted the final placement of the insert earphone for

all of the sentence recordings. The attenuation effect of the earphones was not compensated for but existed for all subjects in all conditions therefore creating a relative impact across recording conditions.

2.1.3   Analysis of speech recording

All sound files were edited in Adobe Audition CS5.5. A quiet noise floor was recorded as a reference and finally any background noise was removed by Adobe Audition. Only one sentence of several repeats was chosen for the final sentence list. To determine which of the sentences was included in the test sentence list, the sentence was judged to be clear and fluent and word speed fell within the range of 200 to 350 syllables per minute for English and 150 to 300 syllables per minute for Mandarin Chinese. François et al. (2011) conducted a cross language comparison of speech rate of seven languages including English and Mandarin Chinese and found the syllabic rate of English is 6.19 syl/sec and Mandarin Chinese is 5.18 syl/sec. After the appropriate syl/sec rate was verified, the judgment of fluency and neutrality of the sentences that met the syl/sec requirement was made by three native English speakers and three native Mandarin Chinese speakers, respectively. Fluency was rated on a scale from 1 to 5 with 5 being completely fluent. Every judge scored each sentence and sentences having two judgments higher than 3 were saved. The sentence receiving the highest score was used in the experiment. The rating scale was described by the following: "1. The speech is too slow/fast and doesn't sound like a normal dialog. 2. The speech rate is Okay but there are some unexpected breaks in the sentence. 3. The speech rate is Okay and there are no unexpected breaks but some of the words are not pronounced clearly.4. The speech rate is Okay, there are no unexpected breaks and every word is

pronounced clearly. 5. The speech rate is perfect, there are no unexpected breaks in the sentence and the words are pronounced clearly and it is very easy to follow the sentence."

2.1.4   Babble noise

Four native English speakers (two males and two females) and four native Mandarin-Chinese speakers (two males and two females) were asked to read alternate lists of IEEE sentences in English and Chinese, respectively. Recording took place in the same sound treated booth used for stimulus material recording and the same procedure and instructions were used to produce the sentences. English and Mandarin Chinese four-talker babble was created from these recordings as follows. For each talker, two IEEE sentences were repeated out loud with 85 dB SPL of white noise delivered via EA-3A insert earphones as well as repeated out loud without any noise. The reason to record with the noise was to induce Lombard speech and the reason to record the speech without noise was to replicate a quieter conversational situation. While talking in a noisy environment, everyone uses a talking style that is different from talking in quiet. Sentences were different for each talker. The recordings took place in a double walled sound treated room through a USB dynamic microphone to Adobe Audition CS5.5. For each talker, two sentences (a different pair of sentences for each talker) were concatenated to ensure the duration of the noise tracks would exceed the duration of all target sentences. 500-ms of silence was added to each talker's file in order to stagger the talkers once they were mixed together. All four talkers' recordings were mixed, and the initial 500-ms of the mixed file was removed to eliminate segments that did not contain all four talkers (Van Engen et al., 2007). Figure 2.3 displays the frequency analysis of the four babble noise. The frequency distribution of the babble noise of

Lombard English, Lombard Mandarin Chinese, conversational English, and conversational Mandarin Chinese.
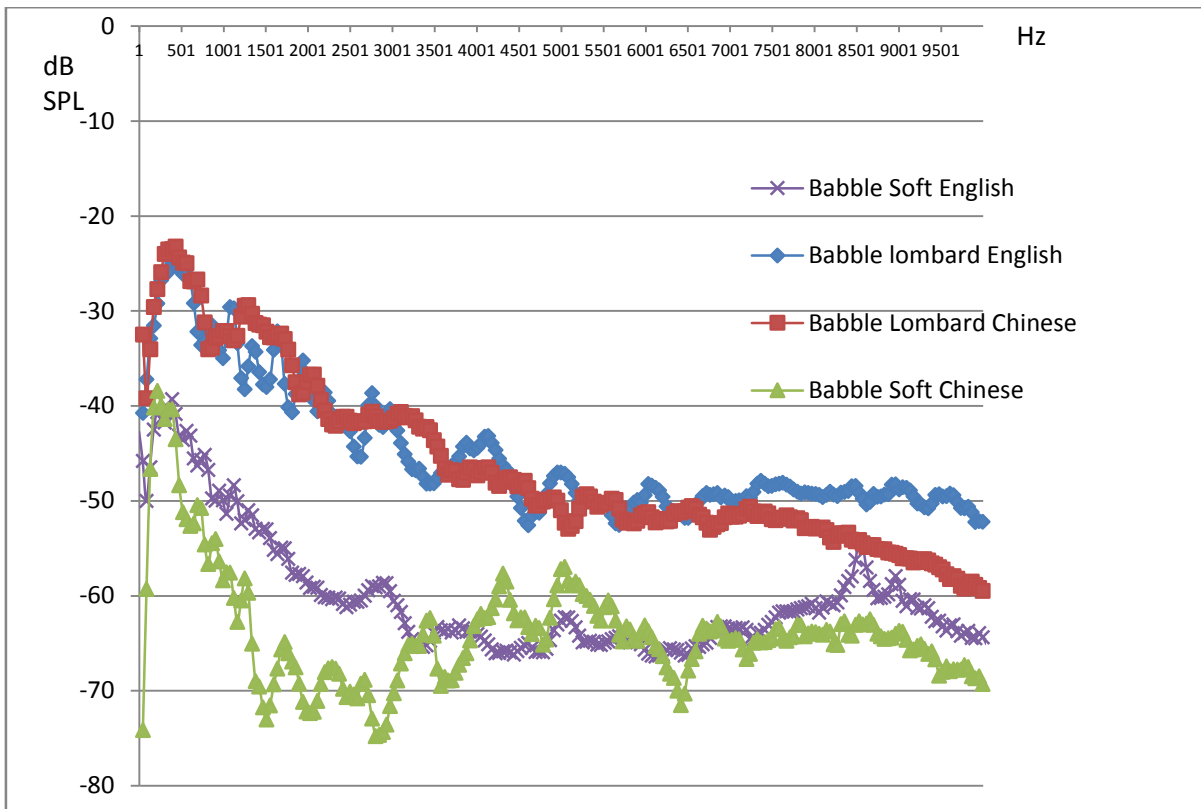
Figure 2.3 Frequency analysis of made babble noise

2.2   SUBJECTS

2.2.1   Study design

A mixed design with two major effects was used for this study. The main effects were languages (English vs. Mandarin Chinese) and speech pattern (conversational speech recorded in quiet vs. Lombard speech recorded in loud noise). In order to compare the effect of languages, the comparison was made for English at conversational level vs Mandarin Chinese at conversational

level as well as English Lombard Speech vs Mandarin Chinese Lombard Speech. In order to compare the effect of speech pattern, comparison was made between conversational speech and Lombard speech in English as well as conversational speech and Lombard speech in Mandarin Chinese. The presentation level was at 65 dB SPL and 85 dB SPL, respectively. ANSI S3.5-1997 reports 65 dB SPL for normal conversation level and 85 dB SPL for higher presentation levels that represent Lombard Speech level (ANSI, 1997). Signal-to-noise ratio started at 0 dB, and then the noise level was decreased in 3 dB steps and ended at 15 dB SNR. Percent correct was recorded. In total, there were 5 different signal-to-noise ratios for each presentation level (65 dB SPL and 85 dB SPL) and for each language. Compared to Killion et al., (1993), who used a 5 dB increment in the signal-to-noise ratio, this study used 3 dB steps because this showed more detail of the perception growth function (KillionVillchur, 1993). Using this growth function, SNR50 (the signal-to- noise ratio required to obtain 50% correct) was mathematically determined.

2.2.2   Power analysis and sample size

Eighteen normally hearing native English speakers and 18 normally hearing native Mandarin-Chinese speakers were recruited to test the null hypothesis that the population means of the English group and Mandarin Chinese group are equal with the probability of (power) .8. The type I error probability associated with the test of this null hypothesis was .05.

These participants were recruited from University of Pittsburgh students. All participants were between the ages of 18-25 years old. All participants were not fluent in a second language. Participants were given a hearing screening at 500, 1k, 2k and 4k Hz and hearing thresholds were at 20 dB HL or better.   A screen of word recognition in noise was accomplished by presenting

spondees in speech shaped noise at 65 dB HL in English and Mandarin Chinese, respectively. Subjects passed at 60% in -3 dB SNR. The English version of spondees came from the Homogeneous Spondees produced by Hearing and Speech Sciences Laboratory at Brigham Young University. The Mandarin Chinese version of spondees came from the KXC-1 (The Phrases Clearness Testing List-1) and the Chinese Phrases Testing List-1.   The word recognition inclusion criteria insured that individuals entered into the study had word recognition ability that would allow them to complete the study task. All participants consented in accordance with the guidelines of the Institutional Review Board (IRB#: PRO13020570) of the University of Pittsburgh.

2.2.3   Inclusion and exclusion criteria

Age: The age range was between 18 to 42 years old. The reason to limit the upper age at 42 years was that Mandarin Chinese Education requiring standard Mandarin Chinese at primary school began in 1978 which would include individuals age 42 years and below.

Hearing criteria: All participants were required to pass a hearing screening of 20 dB HL at 500, 1k, 2k, 4k and 8k Hz.. There was no self-reported history of ear diseases or brain tumors for any participant.

Gender: Both genders participated in the study and there was no requirement for gender balance.

Linguistic background: All participants were not fluent in a second language. The definition of fluency in a second language was being capable of speaking, listening and understanding the language without any difficulty compared to his/her first language. Some bilingual individuals hear poorly in noise (von Hapsburg & Bahng, 2009) so no bilingual

individuals were included in this study. The participants were recruited from the Pittsburgh area. In order to choose mono-lingual Mandarin Chinese participants, participants had arrived in the United States within a half year. They had less time in the English environment and all reported difficulties in listening to and understanding English.

2.3   PROCEDURE

Once the insert earphone calibration was completed and the participant had met the screening criteria, the listener was asked to complete the experiment. The sentences were presented through a Beltone 2000 audiometer by connecting a two channel CD player to the audiometer's input and the output was routed thought an ER-1 insert earphone. An ER-1 earphone was used because the frequency response simulates the open ear canal frequency response (Killion, 1984). This mimics the condition of hearing these sentences in the sound field where the signal would pass through the open ear canal. Each subject heard one sentence at each signal-to-noise ratio in order to familiarize themselves with the task. During the test there were 5 sentences presented at each signal-to-noise ratio (thereby creating 25 scored items). The sentence presentation level was 65 dB SPL for conversational speech (sentences recorded without babble noise presented to the speaker) and 85 dB SPL for Lombard speech (sentences recorded with babble noise presented to the speaker). In order to create the various signal-to-noise ratios, the recorded babble noise (English listeners always heard English babble and Mandarin-Chinese listeners always heard Mandarin-Chinese babble) was added to the signal. Signal-to-noise ratios were presented at 0, 3, 6, 9, 12, 15 with the babble noise level varying and the test signal staying constant. Babble noise was presented through the same bilateral ER-1A insert earphones.

Participants were required to repeat each sentence word by word exactly during the break between each sentence. There was no time limit for the participants to repeat the sentence. The repeated sentences were recorded through the CAD 1U microphone and saved on a computer. During the break, there was no stimulus presented through the insert earphones. Three judges of native English speakers and three judges of native Mandarin-Chinese speakers were asked to listen to the recording of responses and score the key words they heard from the participants. Scoring was reached by consensus if there were any discrepancies. One point was given to each key word repeated correctly and there was no half credit to make the judgment more objective. Number of correct key words at each signal-to-noise ratio was documented. The order of section of sentences (see Appendix A, 12 sections in total and 5 sentences in each section with each section including only one signal-to-noise ratio, therefore, six sections were needed to have the whole test in either conversational or Lombard speech) for presentation was randomized and two different recorded sentences including Lombard speech (sentence recorded with babble noise) and conversational speech (sentences recorded without noise) were counter balanced between participants (See Appendix B).

The instruction to the participants was: "While you are completing the test, you will hear a woman say several sentences. Together with these sentences, there will be background noise. Please ignore the background noise and focus on the talker's sentence. At the end of each sentence there is a break, please repeat the sentence you heard".

The maximum level of babble noise was 85 dB SPL. Lebo et al. (1994) recorded the noise levels from 27 restaurants in the San Francisco Bay Area, California and they concluded that the noise level ranged from 60 to 80 dBA depending on what restaurants they investigated. California cuisine had louder noise levels ranging from 74 to 80 dBA but in elegant restaurants

the noise level was only 60 to 66 dBA (Lebo et al., 1994). Yu, et al., (2002) did a study in Hong Kong on occupational noise exposure and hearing impairment among employees in Chinese restaurants and showed that the average level of sound in the service areas of Chinese restaurants was 75.9 dBA with a standard deviation of 5.6 (Yu TSI, 2002). Taking those studies together, the worst listening situation in the USA and Chinese restaurants had a noise level of 80 dBA. The experiment required a 0 dB SNR condition, therefore, the maximum noise level was 85 dB SPL to match the maximum level of the speech to get a 0 dB signal-to-noise ratio.

In order to control the exact presentation level at the ear drum, the real-ear-to-dial difference was measured for each subject. Speech spectrum noise generated from the audiometer was presented through the ER-1 insert earphone. A probe microphone was inserted into the ear canal close to the ear drum. A real-ear measurement mode was used on the Verifit (Audio Scan RM 500 Series) by choosing the real time speech mode. The root-mean-square level in dB SPL of the speech spectrum noise at the ear drum was measured by the probe microphone and displayed on the screen. The target presenting level at the ear drum was 85 dB SPL and adjustment of the dial of the audiometer was made until this was achieved. Figure 2.3 shows the entire procedure for the experiment.

Figure 2.4 Study procedure replicated for 18 English listeners and 18 Mandarin Chinese listeners.

3.0   RESULTS

This study investigated whether Mandarin-Chinese listeners need an enhanced signal-to-noise ratio to achieve the same performance when communicating in speech babble noise as compared to American English listeners. The potential needs for an enhanced SNR in conversational listening level was established through a review of the differences in phonology, morphology, syntax, cultural considerations. It was less clear how Lombard speech differences between Mandarin Chinese and American English might impact signal-to-noise ratio requirements for similar performance because of the paucity of data in this area. Data collected to answer this question included percent correct for low context sentences in various signal-to-noise ratios using conversational (produced in a quiet condition) and Lombard speech (produced while listening to loud noise) with Mandarin Chinese and American English listeners. Speech perception scores of Mandarin-Chinese and English listeners are listed in Table 3.1 and Table 3.2. All analyses were conducted using these data. Analysis focused on the comparison of performance-intensity functions for each group in conversational and Lombard speech conditions. Additionally, the SNR-50 (the signal-to-noise ratio required to achieve 50% correct) was determined for English and Mandarin-Chinese listeners in the conversational and Lombard speech conditions. This is a common method for direct comparison of different groups of listeners or different test materials.

Table 3.1 Speech perception score in Mandarin Chinese

		0	3	6	9	12	15
S1	Conversational	0%	80%	96%	96%	92%	88%
	Lomard speech	76%	100%	96%	96%	96%	96%
S2	Conversational	0%	80%	96%	100%	100%	100%
	Lomard speech	72%	100%	100%	100%	96%	96%
S3	Conversational	32%	80%	96%	84%	96%	100%
	Lomard speech	80%	92%	100%	100%	96%	96%
S4	Conversational	60%	72%	72%	96%	72%	100%
	Lomard speech	92%	96%	92%	100%	100%	100%
S5	Conversational	16%	52%	100%	80%	100%	92%
	Lomard speech	88%	92%	100%	96%	96%	100%
S6	Conversational	76%	72%	88%	96%	96%	92%
	Lomard speech	76%	100%	100%	100%	100%	100%
S7	Conversational	60%	76%	100%	100%	100%	100%
	Lomard speech	36%	100%	96%	100%	100%	100%
S8	Conversational	4%	52%	96%	96%	100%	100%
	Lomard speech	88%	92%	100%	100%	100%	100%
S9	Conversational	96%	80%	88%	100%	100%	88%
	Lomard speech	80%	100%	96%	96%	100%	100%
S10	Conversational	60%	60%	96%	96%	100%	100%
	Lomard speech	72%	80%	96%	96%	100%	100%
S11	Conversational	20%	24%	92%	92%	96%	100%
	Lomard speech	64%	96%	100%	100%	100%	100%
S12	Conversational	0%	48%	76%	96%	100%	100%
	Lomard speech	36%	96%	100%	100%	100%	100%

Table 3.1 (continued)

S13	Conversational	36%	72%	84%	96%	80%	100%
	Lomard speech	8%	52%	88%	96%	100%	100%
S14	Conversational	8%	56%	60%	96%	100%	100%
	Lomard speech	40%	76%	100%	100%	100%	100%
S15	Conversational	4%	68%	72%	96%	100%	100%
	Lomard speech	44%	76%	100%	100%	96%	100%
S16	Conversational	12%	40%	44%	72%	84%	80%
	Lomard speech	24%	80%	100%	96%	100%	100%
S17	Conversational	24%	88%	88%	92%	96%	100%
	Lomard speech	24%	64%	88%	100%	100%	100%
S18	Conversational	36%	56%	92%	100%	100%	100%
	Lomard speech	40%	88%	96%	100%	100%	100%

Table 3.2 Speech perception score in English

		0	3	6	9	12	15
S1	Conversational	40%	56%	68%	92%	96%	100%
	Lomard speech	32%	100%	88%	96%	100%	100%
S2	Conversational	64%	56%	60%	92%	100%	100%
	Lomard speech	40%	88%	100%	100%	100%	100%
S3	Conversational	44%	96%	80%	88%	88%	88%
	Lomard speech	72%	64%	84%	76%	100%	96%
S4	Conversational	56%	64%	84%	100%	100%	88%
	Lomard speech	72%	88%	88%	96%	100%	100%
S5	Conversational	60%	76%	100%	100%	100%	96%
	Lomard speech	28%	88%	100%	100%	100%	100%
S6	Conversational	80%	84%	92%	92%	100%	100%
	Lomard speech	88%	80%	100%	96%	100%	100%

Table 3.2 (continued)

S7	Conversational	56%	88%	100%	100%	100%	100%
	Lomard speech	56%	80%	80%	92%	100%	96%
S8	Conversational	48%	96%	100%	100%	100%	100%
	Lomard speech	64%	96%	76%	100%	100%	100%
S9	Conversational	84%	84%	84%	100%	92%	96%
	Lomard speech	76%	80%	88%	96%	100%	100%
S10	Conversational	72%	72%	88%	88%	96%	92%
	Lomard speech	40%	72%	96%	88%	96%	88%
S11	Conversational	52%	80%	84%	84%	96%	96%
	Lomard speech	24%	88%	92%	100%	96%	96%
S12	Conversational	80%	80%	96%	100%	100%	100%
	Lomard speech	32%	64%	96%	100%	100%	100%
S13	Conversational	60%	96%	92%	96%	100%	100%
	Lomard speech	44%	80%	100%	100%	100%	92%
S14	Conversational	40%	88%	92%	72%	96%	92%
	Lomard speech	88%	64%	96%	92%	92%	100%
S15	Conversational	76%	80%	80%	96%	96%	100%
	Lomard speech	36%	76%	80%	92%	96%	96%
S16	Conversational	60%	96%	96%	92%	96%	100%
	Lomard speech	60%	88%	76%	92%	92%	88%
S17	Conversational	80%	92%	100%	92%	96%	100%
	Lomard speech	36%	68%	84%	84%	100%	92%
S18	Conversational	60%	64%	92%	100%	88%	100%
	Lomard speech	56%	76%	96%	100%	100%	100%

3.1   PERFORMANCE-INTENSITY FUNCTION
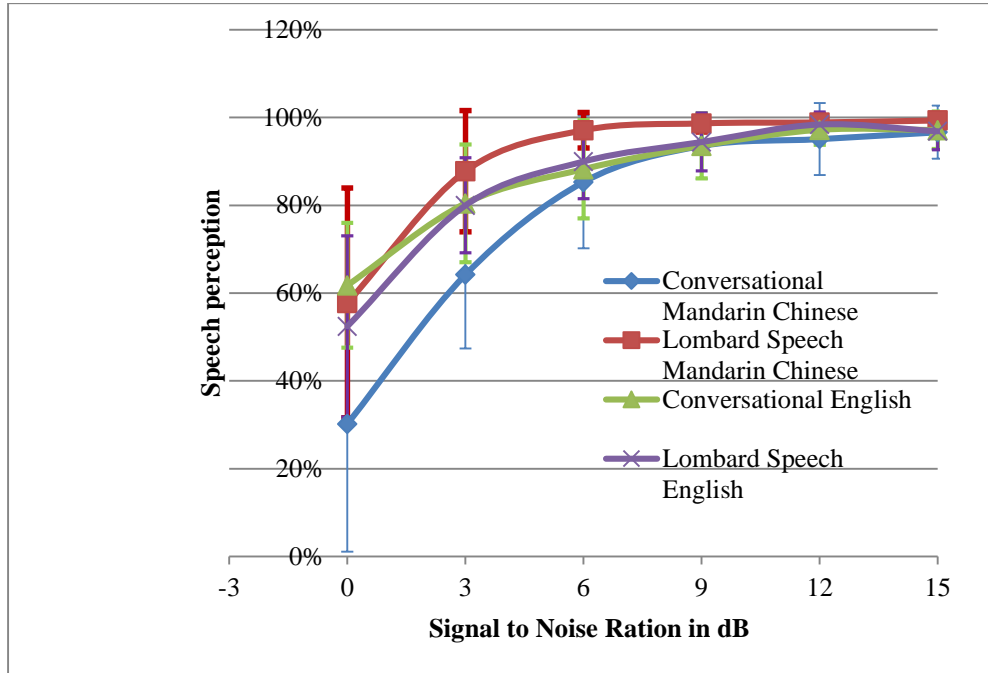
Figure 3.1 Average speech perception for four conditions as a function of signal-to-noise ratio. Bars represent 2 standard deviations.

The performance-intensity functions are displayed in Figure 3.1. These data represent the average percent correct score at each SNR for American English listeners and Mandarin-Chinese listeners in the conversational and Lombard speech conditions. Multivariable analysis of variance (MANOVA) was used to evaluate the main effect of language on speech perception. There was a significant main effect between Mandarin Chinese and English on speech perception (F=2.506, p=.028, partial $ŋ^2$ =.567). The interaction was also significant. Comparisons of interest (conversational English vs conversational Mandarin Chinese; conversational English vs Lombard English; conversational Mandarin Chinese vs Lombard

69

Mandarin Chinese; Lombard English vs Lombard Mandarin Chinese) are described in the following sections. Independent paired t-tests were used to compare these specific conditions of interest. A Bonferroni correction was applied to the analysis with p set to <.008 (.05/6) in order to account for the multiple comparisons across signal-to-noise ratios in this experiment.

3.1.1   Comparison of conversational English to Conversational Mandarin Chinese

Figure 3.2 Performance-intensity function of the Conversational English and Conversational Mandarin Chinese., Bars represent 2 standard deviations.

*-- significant difference.

Figure 3.2 reveals that Mandarin-Chinese listeners require an enhanced signal-to-noise ratio to perform similarly to English listeners when listening to conversational level speech in babble noise at the two most difficult SNRs (0 and +3) included in the investigation. An independent paired T-test was completed to compare these specific conditions (Table 3.1).

Table 3.3 Independent T-test comparing English and Mandarin Chinese conversational speech

SNR	Mean Difference	df	t	Sig (2-tailed) <.008
0	.31556	34	4.129	.000*
3	.16222	34	3.196	.003*
6	.02889	34	.651	.520
9	.00000	34	.000	1.000
12	.01556	34	.727	.472
15	.004444	34	.255	.800

Significance at the p<.008 based on the Bonferroni correction.

3.1.2 Comparison between conversational English and Lombard English
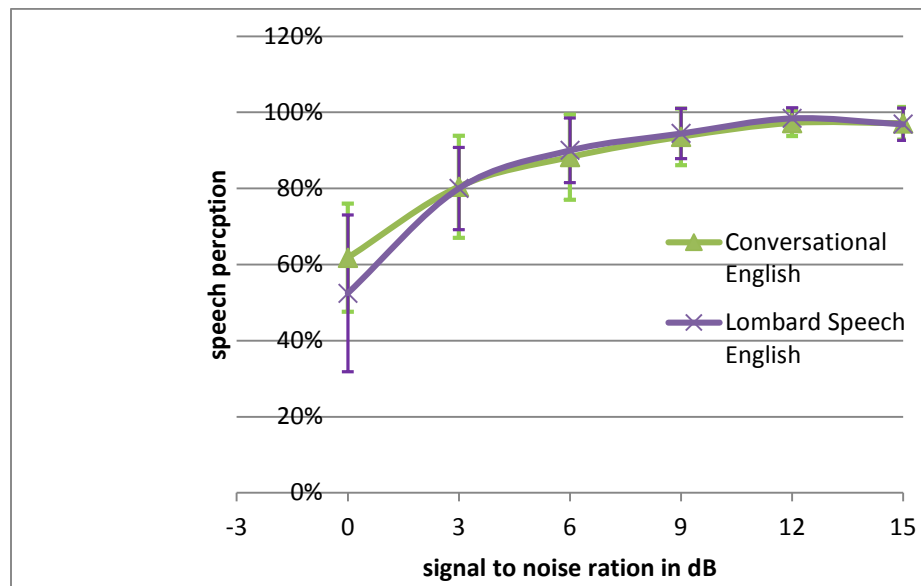
Figure 3.3 Performance-intensity function of the Conversational English and Lombard English. Bars represent 2 standard deviations.

The speech perception results from conversational English and Lombard English (Figure 3.3) were subjected to an independent paired T-test in order to assess differences in performance at each SNR (Table 3.2). No differences at any SNR were found for conversational versus Lombard English.

Table 3.4 Paired T-test comparing English conversational speech and English Lombard speech

SNR	Paired Mean differences	df	t	Sig (2-tailed) <.008
0	-.09333	17	-1.507	.150
3	-.00444	17	-.095	.926
6	.01778	17	.497	.625
9	.00889	17	.475	.641
12	.01111	17	1.230	.236
15	-.00222	17	-.156	.878

Significance at the p<.008 based on the Bonferroni correction.

3.1.3 Comparison of Conversational Mandarin Chinese and Lombard Mandarin Chinese
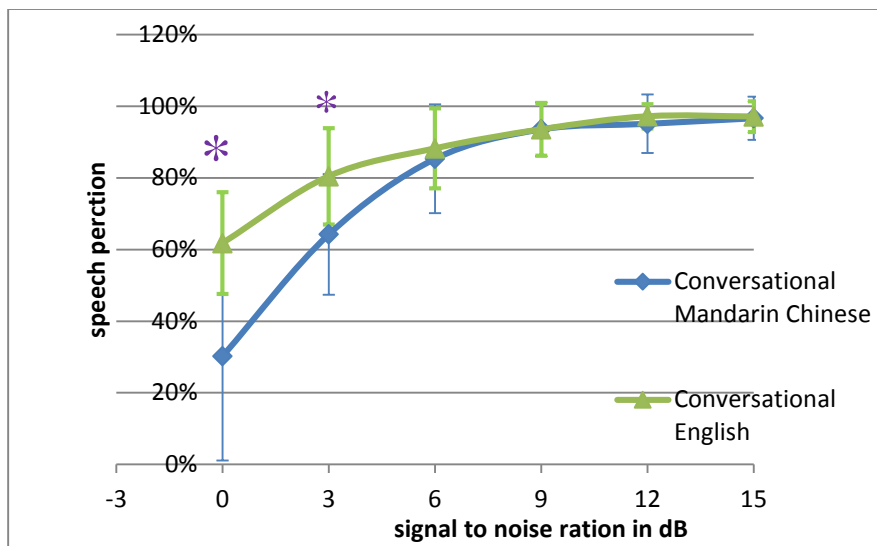

Figure 3.4 Performance-intensity function of the Conversational Mandarin Chinese and Lombard Mandarin Chinese. Bars represent 2 standard deviations. *-- significant difference.

An independent paired T-test (Table 3.3) revealed that conversational Mandarin Chinese required an enhanced SNR to match the performance obtained in Lombard Mandarin Chinese (Figure 3.4). In all but the most favorable SNRs (12 and 15 dB SNR) included in this study, the speech perception results for conversational Mandarin Chinese were significantly poorer than Lombard Mandarin Chinese.

73

Table 3.5 Paired T-test comparing Mandarin Chinese Lombard speech and Mandarin Chinese conversational speech

SNR	Paired Mean differences	df	t	Sig (2-tailed) <.008
0	.27556	17	3.355	.004*
3	.23556	17	4.484	.000*
6	.11778	17	3.118	.006*
9	.05111	17	3.053	.007*
12	.03778	17	1.836	.084
15	.02667	17	1.844	.083

Significance at the p<.008 based on the Bonferroni correction.

3.1.4 Comparison between Lombard Mandarin Chinese and Lombard English

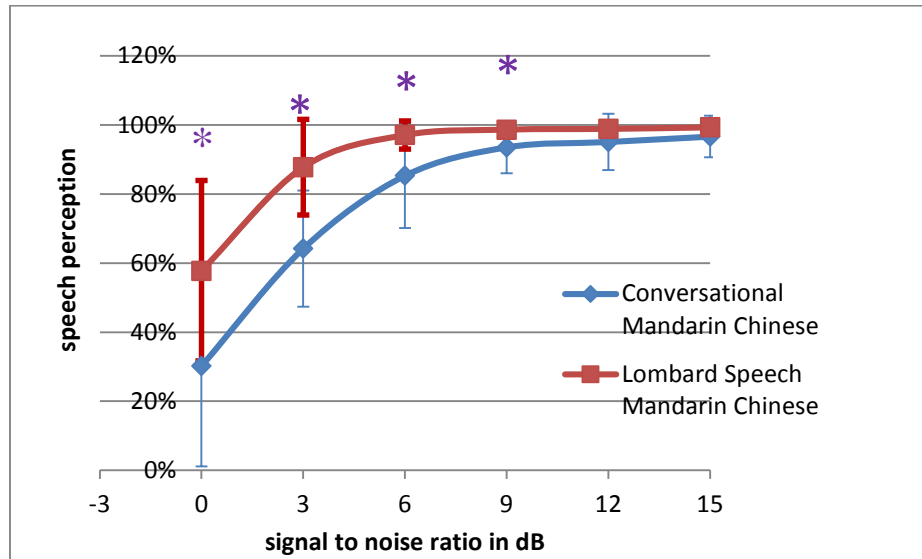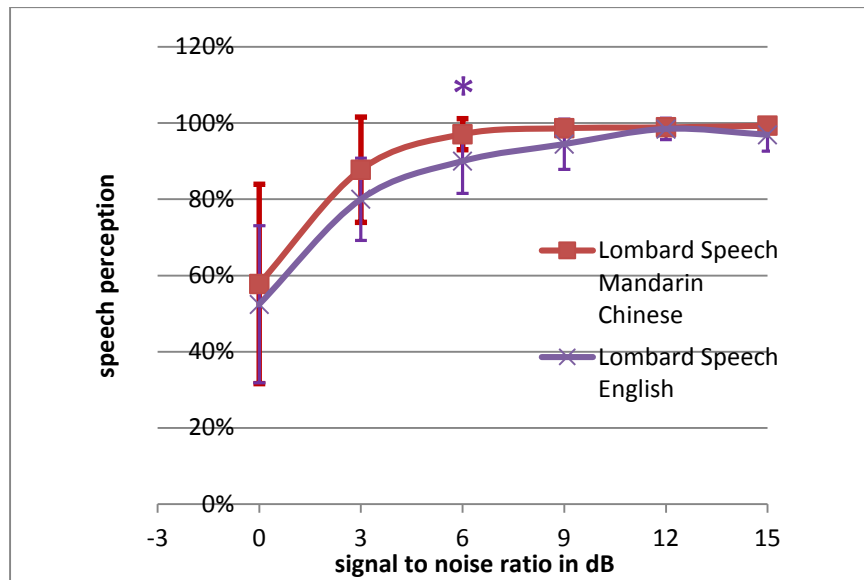**Figure 3.5 Performance-intensity function of the Lombard Mandarin Chinese and Lombard English. Bars represent 2 standard deviations. * - significance**

74

An independent paired T-test (Table 3.4) revealed significant differences for speech perception at +6 for Lombard Mandarin Chinese and Lombard English listeners (Figure 3.5) with the Lombard Mandarin-Chinese listeners performing better. In other words, the Lombard English listeners needed an enhanced signal-to-noise ratio in the moderate noise condition included in this study to perform similarly to the Mandarin Chinese listeners.

Table 3.6 Independent T-test comparing English and Mandarin Chinese Lombard Speech

SNR	Mean Difference	df	t	Sig (2-tailed) <.008
0	-.05333	34	-.679	.501
3	-.07778	34	-1.881	.069
6	-.07111	34	-3.206	.003*
9	-.04222	34	-2.604	.014
12	-.0044	34	-.564	.577
15	-.02444	34	-2.3	.028

Significance at the p<.008 based on the Bonferroni correction.

3.2 SNR-50

The dB SNR needed to achieve 50% (SNR-50) correct was calculated by fitting the speech perception data into a Probit regression analysis. SNR-50 is widely accepted as a parameter to measure the performance of listening in noise (Gelfand et al., 1988; Killion, 1997; Plomp et al., 1979; Nilsson et al., 1994b). Probit Analysis is designed to model the probability of response to

a stimulus. It is a type of generalized linear model that extends the linear regression model by linking the range of real numbers to the 0-1 range. The algorithm for the Probit regression model used by SPSS 21 is:

$\pi i = c + (1-c)F(zi)$

where

πi is the probability the ith case experiences the event of interest

zi is the value of the unobserved continuous variable for the ith case

F is a link function.

c is the natural response rate.

The SNR-50 was calculated by estimating the median. Since the speech perception score will increase as the signal-to-noise ratio increases until it reaches 100% , the data are not normally distributed and there is no central tendency. The raw data of this study is not a perfect fit for the Probit curve because scores do not reach 0%, but estimation with this method is possible based on the existing raw data. Fitting the raw data into a Probit model in SPSS 21 provides an estimation of SNR-50 at each listening condition (see Table 3.5).

Table 3.7 Estimated SNR-50 by Probit Analysis for four speech conditions.

	SNR-50
Conversational English	-1.50
Conversational Mandarin Chinese	1.12
Lombard English	-1.28
Lombard Mandarin Chinese	-3.30

Significant differences were determined using the confidence interval of the estimated SNR-50 (see Table 3.6). In order to compare these SNR-50s, the estimated median from Probit Analysis by SPSS 21 at each condition was paired. Then one SNR-50 was subtracted from its correspondent SNR-50 in the pair. The 95% confidence limits were calculated by SPSS 21in order to reject the H0 that there was no significant differences between the paired SNR-50s. Table 3.7 shows each pair of comparison and the corresponding confidence limits. Significant differences between SNR-50's were found between conversational English and conversational Mandarin Chinese, conversational Mandarin Chinese and Lombard Mandarin Chinese, and Lombard Mandarin Chinese and Lombard English. There was no difference between conversational and Lombard English.

Table 3.8 SNR-50 statistical comparisons using the Probit Analysis are shown for the condition comparisons of interest in this study.

	(I) Listening Condition	(J) Listening Condition	95% Confidence Limits		
			Estimate	Lower Bound	Upper Bound
PROBIT	Mandarin Conversation	Mandarin Lombard	4.423*	3.120	5.849
		English Conversation	2.622*	1.470	3.844
	Mandarin Lombard	English Lombard	-2.022*	-3.354	-.748
	English Conversation	English Lombard	-.222	-1.415	.968

* Significantly different

4.0 DISCUSSION

The general findings of this study can be summarized with the Table 4.1.

Table 4.1 Main comparisons of interest in this study, prediction based on the literature review and the results from this study.

Comparison	Prediction based on literature review(> meaning more difficulties at matched SNRs)	Results of the Experiment (> meaning more difficulties at matched SNRs)
Conversational English vs. Conversational Mandarin Chinese	Conversational Mandarin Chinese > Conversational English (based on literature review of difference in the languages)	Conversational Mandarin Chinese > Conversational English at 0 dB SNR (30% vs 60%) and +3 dB SNR (65 vs 80%) SNR-50 was significantly different.
Conversational English vs. Lombard English	Conversational English >Lombard English (Pittman & Wiley, 2001)	No significant differences
Conversational Mandarin Chinese vs Lombard Mandarin Chinese	Conversational Mandarin Chinese >Lombard Mandarin Chinese (Based on prediction of English, no other data available)	Conversational Mandarin Chinese > Lombard Mandarin Chinese at 0 dB SNR (30 vs 60%), +3 dB SNR (65 vs 85%), +6 dB SNR (85 vs 100%), +9 dB SNR (95 vs 100%) SNR-50 was significantly different.
Lombard English vs Lombard Mandarin Chinese	Lombard Mandarin Chinese > Lombard English (based on the expected pattern for conversational speech since there was a paucity of data for Lombard speech)	Lombard English > Lombard Mandarin Chinese at +6 dB SNR (90 vs 100%) SNR-50 was significantly different.

4.1 SPEECH PATTERN AND LINGUISTIC CHARACTERISTICS THAT MIGHT AFFECT COMMUNICATION DIFFICULTIES IN BABBLE NOISE.

The literature review provided support for predicting that conversational Mandarin Chinese would require an enhanced signal-to-noise ratio to achieve similar performance to English listeners based on linguistic, cultural, and noise masking differences. This prediction was supported by the data at the most difficult signal-to-noise ratios where fragile speech cues would be expected to be most impacted. This provides support for the speech materials and procedures used in the study. The comparison is artificial in that a typical Mandarin-Chinese listener cannot switch to English in this situation to improve communication. Therefore the data related to Lombard speech, the speech that actually would be used in difficult communication situations is of more interest.

It also was hypothesized that conversational English would require an enhanced SNR to achieve the same performance as Lombard English based on data from Pittman and Wiley (2001). Pittman and Wiley (2001) reported an average advantage of 15% in a speech recognition task when listening to Lombard speech as compared to conversational speech in a background of multi-talker babble at the same SNR. However, this prediction was not supported by the current data. Pittman and Wiley (2001) only included six subjects who were reading materials rather than using conversational speech with communication intent so methodological differences may explain differences in results between the two studies.

Since there has been a paucity of data related to Lombard Mandarin Chinese, the prediction for the comparison of conversational level and Lombard Mandarin Chinese was based on the prediction for English. The results support the pattern for Mandarin Chinese with individuals needing an enhanced SNR for conversational Mandarin Chinese to obtain similar

79

performance as compared to Lombard Mandarin Chinese across the majority of SNRs tested in this study. At the higher SNRs (+6 and +9) the average performances were between 85 and 100% so these may not be meaningful communication differences. For the SNR at +3, average conversational Mandarin Chinese performance was 30% while average Lombard Mandarin Chinese performance was 60% and at +6 dB SNR, the performance was 65 and 85%, respectively. These levels of performance and difference in performance would be meaningful in day-to-day communication.

The Mandarin-Chinese listeners receive a benefit from Lombard speech which is not realized by English listeners (i.e., conversational English is equivalent to Lombard speech performance at the same SNRs). In other words, the data from Lombard Mandarin Chinese indicate there is significant benefit to raising your voice if you are a Mandarin-Chinese speaker communicating in babble noise compared to conversational Mandarin Chinese.

A significant difference was found for the moderate SNR (+6) with Lombard Mandarin Chinese outperforming Lombard English, but not at the more difficult SNRs. The variability of the data at the more difficult SNRs (e.g., 0 and +3) is much greater than the variability at the moderate to positive SNRs (+6 to +15) for all conditions (conversational, Lombard, English, and Mandarin Chinese). It is not surprising that larger variability is associated with the most difficult listening situations where individuals may find that they are guessing at the words. The largest variability was for the Lombard conditions and may explain why significant results were not seen for these more difficult SNRs yet were recorded for the moderate SNR. At the moderate SNR, the performance in both groups was between 90 and 100% which revealed significant difference but would not be practically different in terms of communication.

The SNR-50 analysis allows estimates to be used based on the raw data to predict differences between materials or populations. In this manner, the Probit analysis ignores the actual variability in the data and predicts differences based on the median data. Differences in SNR-50 were found for the conversational English vs conversational Mandarin conditions; the conversational Mandarin vs Lombard Mandarin conditions; and the Lombard English vs Lombard Mandarin conditions. This last comparison may seem at odds with the data since significant results were not reported for the 0 and +3 dB SNR conditions where average scores were 55 to 80%, but this addresses the issue of the extreme variability associated with these data that is ignored in the SNR-50 prediction.

The expected difference found between conversational English and conversational Mandarin Chinese as well as the control of word token frequency, phonetic equivalence, head-final structure, and contextual predictability between the English and Mandarin Chinese sentences provides support for interpreting these data without worrying that some difference in the speech materials is impacting the pattern of results.

The data suggest that important cues for understanding Mandarin Chinese are lost if people are speaking quietly in a background of noise. The simple phonological structure with the initial consonant plus flowing vowel and high ambiguity at the monosyllabic level may account for the fragile nature of conversational Mandarin Chinese in noise. Since high frequency consonants are masked by noise and other redundant cues may not be distinct at conversational level, the target speech at conversational level may be hard to segregate from the babble noise. These data may provide an acoustic reason for the perception that in China people tend to speak loudly in public places. There appears to be communication benefit in producing louder speech while maintaining signal-to-noise ratio.

81

The next sections explore what properties of the Lombard speech in Mandarin Chinese may be accounting for the improved performance relative to conversational Mandarin Chinese and Lombard English. The purpose of this study was to describe the pattern of results for the four speech conditions (conversational English and Mandarin Chinese and Lombard English and Mandarin Chinese) rather than to provide adequate data to explain any differences found, but it is interesting to speculate about the differences that were revealed. Future research will need to be designed to explore the cause of the patterns found in this experiment.

4.2   SPEECH SEGREGATION IN BABBLE NOISE BY USING TEMPORAL CUES

This study controlled the level of the speech and babble by manipulating the presentation level of the babble to create five signal-to- noise ratios. The speech spectrums of the target speech and background babble noise were well matched by using similar speech babble (conversational speech of 4 talkers made of 4-talker conversational babble and Lombard speech of 4 talkers made of 4-talker Lombard speech babble). This use of multi-talker babble avoids spectrum mismatch that will occur with other background noises (Lu et al, 2008). In this manner, there are less frequency-amplitude cues to be accessed by the listener than if they were listening to a different background noise. Given that spectral cues largely were controlled in the speech and babble used in this study, it is worth investigating how temporal cues may have impacted the results. There are three types of temporal cues (see Table 4.2) that exist in the speech signal: envelope, periodicity and fine-structure (Rosen, 1992).

Table 4.2 Common temporal cues in linguistics (Rosen 1992).

	Linguistic Aspect
Envelope	Duration, rise time and fall time Prosodic cues
Periodicity	Tone, intonation, voicing identification
Fine-structure	Place of articulation Voicing and manner

Table 4.3 Differences in temporal cues that might predict better performance for Lombard Mandarin-Chinese listeners in babble noise as compared to Conversational Mandarin-Chinese listeners and Lombard English listeners.

	Envelope	Periodicity	Fine-structure
Lombard Mandarin Chinese Speech vs Lombard English Speech/Conversational Mandarin Speech	CV structure More mono-syllables in sentence More pauses in sentences	Enlarged variation in F0 could help with voice tracking	Weighted more energy at first formant transition helps to distinguish initial consonant that is critical for speech understanding in Mandarin Chinese

At a conversational level, initial consonants are easily masked and may induce ambiguity in Mandarin Chinese, therefore the speaker will tend to speak louder because this will preserve these cues in noise. Table 4.3 provides three areas of temporal processing that may be enhanced in Mandarin Chinese when Lombard speech is produced. These differences potentially could contribute to the pattern of results when comparing conversational Mandarin Chinese to Lombard Mandarin Chinese and when comparing Lombard Mandarin Chinese to Lombard English.

Cooke et al (2006) promoted the glimpse theory, where speech perception in noise is based on the use of glimpses of speech in spectral-temporal regions where it is least affected by the background noise. Most of the key words in Mandarin Chinese in the speech materials used in this study were di-syllablic words with each mono-syllable being shorter in duration than English. In addition, the 4-talker sentences that produced the background babble were specific to each language, therefore the Mandarin Chinese listeners had different glimpses as compared to the English listeners. Multiple shorter pauses with more monosyllabic words may have helped Mandarin Chinese listeners escape from the babble noise energy overlapping with the speech. A reduction in foreground-background overlap can be expected to lead to release from both energetic and informational masking for listeners. Examples are shown in Figure 4.1 by analyzing the same sentences in English and Mandarin Chinese with PRAAT5380 (www.pratt.org).
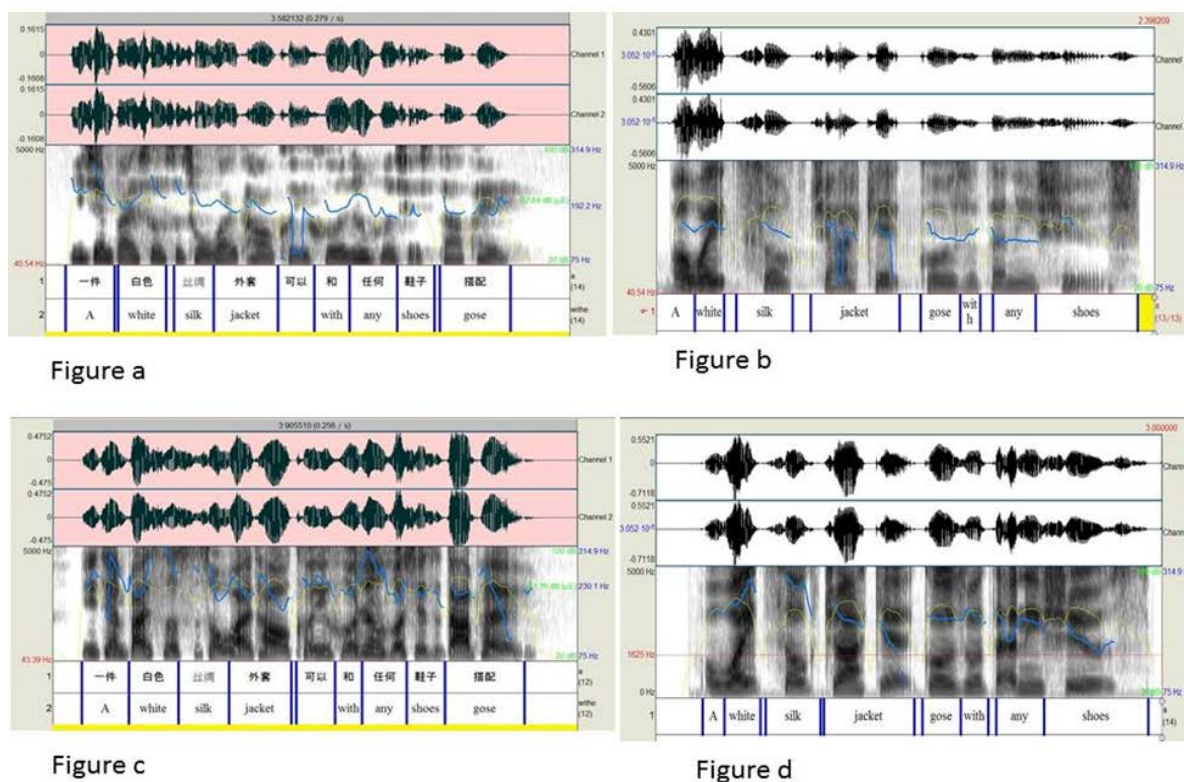
Figure 4.1 Phonological analysis of the sentence "A white silk jacket goes with any shoes". Figure a: conversational Mandarin Chinese; Figure b: conversational English; Figure c: Lombard Mandarin Chinese and Figure d: Lombard English.

As shown in Figure 4.1, the blue lines show the variation of fundamental frequency. In Lombard Mandarin Chinese the range of variation of fundamental frequency is larger than Lombard English. When translated from English to Mandarin Chinese, most key words turned into di-syllable words which are typical of modern Mandarin Chinese. This increased the total mono-syllable numbers as well as increased multiple pauses. Multiple shorter pauses with more monosyllables may help Mandarin Chinese listeners escape from babble noise energy overlapping with the speech target, thereby taking advantage of increased glimpses of speech in the noise compared to English listeners when speech is presented in a Lombard condition.

85

5.0   CONCLUSION

One major conclusion can be drawn from this study, Mandarin Chinese speakers may rely more heavily on Lombard speech than English speakers because the use of Lombard speech provides a direct benefit in communication in babble noise. The problem appears to be conversational Mandarin –the data in this study suggest that it is highly impacted by noise. This would make speakers want to raise their voices. In English there was no additional benefit to raising your voice in this study. It appears that the motivation to raise your voice in English would be to preserve SNR and thereby preserve the same level of function you had at a conversational level. There appear to be features related to Mandarin Chinese speech that make conversational speech fragile in noise and provide a significant benefit when the voice is raised even though SNR is not changed.

Future research will focus on identifying the cues that may account for the patterns of results identified in this study. This information will be valuable to individuals developing hearing assistance technology for individuals listening in Mandarin Chinese and English and for individuals interested in developing auditory training programs.

IEEE SENTENCES AND THE MANDARIN CHINESE TRANSLATION

Section 1

1.   A <u>white</u> <u>silk</u> <u>jacket</u> goes with <u>any</u> <u>shoes</u>.   (一件<u>白色絲綢外套</u>可以和<u>任何鞋子</u>搭配.)

2.   The <u>child</u> <u>crawled</u> <u>into</u> the <u>dense</u> <u>grass</u>.   (這<u>小孩爬行在茂密</u>的<u>草丛里</u>.)

3.   <u>Footprints</u> <u>showed</u> the <u>path</u> he <u>took</u> up the <u>beach</u>. (<u>脚印顯露</u>了他在<u>海滩</u>上<u>行走</u>的<u>路径</u>.)

4.   A <u>vent</u> <u>near</u> the <u>edge</u> brought in <u>fresh</u> <u>air</u>. (<u>靠近边缘</u>的<u>小孔</u>带来了<u>新鲜</u>的<u>空气</u>.)

5.   It is a <u>band</u> of <u>steel</u> <u>three</u> <u>inches</u> <u>wide</u>. (这是一个<u>三英寸宽</u>的<u>钢条</u>.)

Section 2

1.   <u>Tear</u> a <u>thin</u> <u>sheet</u> from the <u>yellow</u> <u>pad</u>. (从<u>黄色</u>的<u>本子</u>上<u>撕</u>下一张<u>薄</u>的<u>纸</u>.)

2.   A <u>cruise</u> in warm <u>waters</u> in a <u>sleek</u> <u>yacht</u> is <u>fun</u>. (乘坐<u>流线型</u>的<u>游艇</u>在暖<u>水域</u>中<u>巡游</u>是

<u>有趣</u>的.)

3.   A <u>streak</u> of <u>color</u> ran <u>down</u> the <u>left</u> <u>edge</u>. (一<u>条彩條</u>在<u>左</u>侧的<u>边缘垂落下</u>.)

4.   It was <u>done before</u> the <u>boy</u> could <u>see it</u>. (在<u>男孩</u>能够<u>看见它之前</u>就<u>做好</u>了.)

5.   <u>Crouch before you jump</u> or <u>miss</u> the <u>mark</u>. (在<u>跳跃之前要蹲下</u>不然会<u>错过标记</u>.)

Section 3

1. <u>Pitch</u> the <u>straw through</u> the <u>door</u> of the <u>stable</u>. (把<u>稻草</u>从<u>马棚</u>的<u>门扔</u>过去.)

2. The <u>sink</u> is the <u>thing</u> in <u>which</u> we <u>pile dishes</u>. (<u>水槽</u>是我们用来<u>堆盘子</u>的<u>地方</u>.)

3. <u>Post no bills</u> on this <u>office wall</u>. (这个<u>办公室</u>的墙上<u>不能张贴</u>任何<u>帐单</u>.)

4. <u>Dimes showered down</u> from <u>all sides</u>. (<u>硬币</u>如雨点般从四面八方<u>落下</u>.)

5. <u>Pick</u> a <u>card</u> and <u>slip</u> it <u>under</u> the <u>pack</u>. (<u>挑选</u>一张<u>卡片</u>并把它<u>平放</u>在这个<u>垫子下面</u>.)

Section 4

1. The <u>sense</u> of <u>smell</u> is <u>better than</u> that of <u>touch</u>. (<u>嗅觉</u>比<u>触觉</u>更灵敏.)

2. He <u>picked up</u> the <u>dice</u> for a <u>second roll</u>. (他<u>拾起骰子</u>为了扔<u>第二次</u>.)

3. <u>Drop</u> the <u>ashes</u> on the <u>worn old rug</u>. (<u>烟灰抖落</u>在<u>老的破旧</u>的<u>地毯</u>上.)

4. The <u>couch cover</u> and <u>hall drapes</u> were <u>blue</u>. (<u>沙发罩</u>和<u>大厅</u>的<u>窗帘</u>曾是<u>蓝色</u>的.)

5. The <u>stems</u> of the <u>tall glasses cracked</u> and <u>broke</u>. (这个<u>高脚杯</u>的<u>杯脚裂开</u>并<u>破碎</u>了.)

Section 5

1. To <u>have</u> is <u>better than</u> to <u>wait</u> and <u>hope</u>. (<u>拥有</u>比<u>等待</u>和<u>期望更好</u>.)

2. The <u>screen before</u> the <u>fire kept</u> in the <u>sparks</u>. (<u>火焰前面</u>的<u>屏幕隔离</u>着<u>火花</u>.)

3. <u>Thick</u> glasses <u>helped</u> him <u>read</u> the <u>pint</u>. (<u>厚厚的镜片帮助</u>他<u>阅读印刷品</u>.)

4. The <u>chair looked strong</u> but had <u>no bottom</u>. (这张<u>椅子看上去结实</u>但却<u>没有底座</u>.)

5. They <u>told wild tales</u> to <u>frighten</u> him. (他们给他<u>讲</u>些<u>胡编</u>的<u>故事</u>来<u>吓唬他</u>.)

Section 6

1. The <u>leaf drifts along</u> with a <u>slow spin</u>. (<u>叶子</u>慢慢<u>旋转</u>着<u>随波逐流</u>.)

2. The <u>pencil</u> was <u>cut</u> to be <u>sharp</u> at <u>both ends</u>. (这支<u>铅笔</u>被<u>削</u>得<u>两头尖</u>.)

3. <u>Down</u> that <u>road</u> is the <u>way</u> to the <u>grain farmer</u>.

(顺着那条路下去是去谷物农场主家的路.)

4. The best method is to fix it in place with clips. (最好的办法是用别针把它固定在原处.)

5. If you mumble your speech will be lost. (如果含糊地说话你的语言会听不清.)

Section 7

1. The kite dipped and swayed, but stayed aloft.　(这风筝摇摆着，下坠着，但是还在空

中飘着.)

2. The beetle droned in the hot June sun. (这只甲虫在炎热的六月阳光下发出嗡嗡声.)

3. The theft of the pearl pin was kept secret.(珍珠别针被盗這件事被保密着.)

4. His wide grin earned many friends. (他的大大的咧嘴笑赢得了很多朋友.)

5. Hurdle the pit with the aid of a long pole. (借助一根长杆撑跳过这个坑.)

Section 8

1. The sun came up to light the eastern sky.(太阳出来照亮了东方的天空.)

2. The stale smell of old beer lingers. (陈旧啤酒发出的腐败气味一直徘徊着.)

3. The desk was firm on the shaky floor.(这张桌子在摇晃的地板上很牢固.)

4. A list of names is carved around the base. (一组名字雕刻在基底部的周围。)

5. The news struck doubt into restless minds. （这个消息给那些不安的心灵带来了疑

慮。）

Section 9

1. Take shelter in this tent but keep still.（到帐篷里来躲避但是请保持安静.）

2. The little tales they tell are false. （他们讲的小故事是假的。）

3. Press the pedal with your left foot. （用你的左脚踩踏板。）

4. The black trunk fell from the landing.（黑色的树干从平台处落下。）

5. Cheap clothes are flashy but don't last.（便宜的衣服虽然华丽但却不能持久。）

Section 10

1. Dots of light betrayed the black cat. （光的亮点暴露了这只黑猫。）

2. Put the chart on the mantel and tack it down. （把图表放在壁炉那边并且钉下。）

3. The steady drip is worse than a drenching rain. （持续地下雨比倾盆大雨更糟糕。）

4. A flat pack takes less luggage space. （扁平的包装占有更小的行李空间。）

5. The gloss on top made it unfits to read. （表面的光泽使得不利于阅读。）

Section 11

1. The weight of the package was seen on the high scale. (上面的刻度显示了包裹的重量.)

2. The square peg will settle in the round hole.(这个方形的挂钩将要放置在圆形的洞里.)

3. The store was jammed before the sale could start. (这个商店在大减价开始以前就被挤

满了.)

4. The cleat sank deeply into the soft turf.(夹板深深地陷入在柔软的草皮内.)

5. A force equal to that would move the earth. (與那個力量等同的力量可以移動地球.)

Section 12

1. A toad and a frog are hard to tell apart. (蟾蜍和青蛙是难以区别开的.)

2. Peep under the tent and see the clown. (在帐篷下面偷窥看到了小丑.)

3. The sand drifts over the sill of the old house.（沙子堆积在这座老房子的窗台上.）

4. At night the alarm roused him from a deep sleep. （夜里警报声把他从熟睡中惊醒。）

5. Seven seals were stamped on great sheets. （七个印章被盖在大的表格上。）

Practice List A

1. The <u>lake sparkled</u> in the <u>red</u> <u>hot</u> <u>sun</u>. <u>湖泊</u>在<u>紅色烈日</u>下<u>閃光</u>。
2. <u>Tend</u> the <u>sheep</u> <u>while</u> the <u>dog wanders</u>. <u>當狗</u>在<u>散步</u>的時候<u>照看好羊</u>。
3. <u>Take two shares</u> as a <u>fair profit</u>. <u>拿兩份股份</u>作為一個<u>合理的利潤</u>。
4. <u>North</u> <u>winds</u> <u>bring</u> <u>colds</u> and <u>fevers</u>. <u>北風帶來了感冒</u>和<u>發燒</u>。
5. A <u>sash</u> of <u>gold</u> <u>silk</u> will <u>trim</u> her <u>dress</u>. <u>金色絲綢的腰帶</u>將用來<u>裝飾</u>她的<u>裙子</u>。
6. <u>Fake</u> <u>stones</u> <u>shine</u> but <u>cost little</u>. <u>假的石頭會閃光</u>卻<u>不值錢</u>。

Practice List B

1. <u>Wake</u> and <u>rise</u>, and <u>step</u> into the <u>green</u> <u>outdoors</u>. <u>醒過來</u>，<u>起床並且走</u>到<u>綠色</u>的<u>戶外</u>。
2. <u>Next</u> <u>Sunday</u> <u>is</u> the <u>twelfth</u> of the <u>month</u>. <u>下個星期天是</u>這個<u>月</u>的<u>十二號</u>。
3. <u>Every</u> <u>word</u> and <u>phrase</u> he <u>speaks</u> is <u>true</u>. 他<u>說</u>的<u>每個字句</u>都是<u>真的</u>。
4. <u>Help</u> the <u>weak</u> to <u>preserve</u> <u>their</u> <u>strength</u>. <u>幫助</u>這些<u>虛弱</u>的人<u>保持</u>他們的<u>體力</u>。
5. <u>Get</u> the <u>trust</u> <u>fund</u> to the <u>bank</u> <u>early</u>. <u>儘早讓信託基金投資銀行</u>。
6. A <u>six</u> <u>comes</u> up <u>more</u> <u>often</u> than a <u>ten</u>. <u>數字六</u>比<u>數字十出現得更頻繁</u>。

Practice List C

1. <u>One</u> <u>step</u> <u>more</u> and the <u>board</u> will <u>collapse</u>. <u>多一步</u>這個<u>板子</u>就要<u>塌了</u>。
2. <u>Take</u> the <u>match</u> and <u>strike</u> it <u>against</u> your <u>shoe</u>. <u>拿起火柴</u>在你的<u>鞋上劃一下</u>。
3. The <u>baby</u> <u>puts</u> his <u>right</u> <u>foot</u> in his <u>mouth</u>. 這個<u>嬰兒</u>把他的<u>右腳放進</u>的他的<u>嘴裡</u>。
4. The <u>pup</u> <u>jerked</u> the <u>leash</u> as he saw a <u>feline</u> <u>shape</u>. <u>當看到貓</u>的<u>模型</u>的時候這只<u>小狗跳了起來</u>並且<u>緊拉皮帶</u>。
5. <u>Leave</u> <u>now</u> and <u>you</u> will <u>arrive</u> on <u>time</u>. <u>現在就出發你會準時到達</u>。
6. <u>She</u> <u>saw</u> a <u>cat</u> in the <u>neighbor's</u> <u>house</u>.<u>她看見</u>一隻<u>貓</u>在<u>鄰居</u>的<u>房子裡</u>。

APPENDIX B

PRESENTATION RANDOMIZATION AND COUNTER BALANCE

Participant number	Section Order	Section Order
1	**7, 9, 12, 3, 11, 10**	4, 6, 1,2, 5, 8
2	4, 5, 10, 1, 9, 3	**7, 12, 6, 8, 2, 11**
3	**6, 7, 9, 12, 2, 1**	3, 4, 10, 8, 5, 11
4	6, 10, 5, 12, 8, 7	**11, 4, 2, 1, 3, 9**
5	**2, 3, 12, 8, 5, 4**	7, 11, 9, 1, 10, 6
6	4, 11, 6, 10, 2, 8	**5, 12, 7, 3, 1, 9**
7	**10, 9, 2, 11, 8, 5**	6, 1, 4, 3, 12, 7
8	7, 1, 11, 4, 12, 3	**8, 2, 10, 6, 9, 5**
9	**5, 12, 8, 7, 3, 1**	9, 10, 6, 4, 11, 2
10	1, 2, 6, 11, 12, 3	**9, 7, 10, 4, 5, 8**
11	**8, 2, 12, 11, 4, 1**	3, 7, 6, 10, 5, 9
12	4, 10, 1, 11, 9, 5	**7, 3, 2, 6, 8, 12**
13	**7, 10, 11, 3, 8, 4**	6, 12, 1, 5, 2, 9
14	2, 8, 11, 10, 9, 1	**6, 7, 4, 5, 3, 12**
15	**2, 6, 7, 9, 5, 8,**	1, 12, 10, 4, 3, 11
16	11, 4, 7, 10, 2, 8	**6, 1, 9, 3, 12, 5**
17	**7, 10, 12, 1, 4, 8**	5, 9, 3, 6, 2, 11
18	8, 7, 12, 9, 10, 3	**5, 6, 11, 2, 1, 4**

Notes: 1. Bold numbers representing Lombard Speech (counterbalanced between subjects).

2. There is no randomization on signal-to-noise ratio, the sequence of signal-to-noise

ratio is 15, 12, 9, 6, 3, 0 dB.

3. The sequence is repeated for 18 English speakers and 18 Mandarin Chinese speakers.

BIBLIOGRAPHY

Adelman, J.S., Brown, G.D.A., & Quesada, J.F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. Psychological Science, 17(9), 814-823.

Alexandre, Erika, Barreto, Simone dos Santos, & Ortiz, Karin Zazo. (2011). Predictability of sentences used in the assessment of speech intelligibility in dysarthria. Jornal da Sociedade Brasileira de Fonoaudiologia, 23(2), 119-123.

ANSI, A. (1997). S3. 5-1997, Methods for the calculation of the speech intelligibility index. New York: American National Standards Institute.

Arbogast, Tanya L., Mason, Christine R., & Kidd, Gerald, Jr. (2002). The effect of spatial separation on informational and energetic masking of speech. Journal of the Acoustical Society of America, 112(5 Pt 1), 2086-2098.

Baxter, William Hubbard. (1992). A handbook of Old Chinese phonology. Mouton de Gruyter.

Bench, J., Kowal, Å., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. British Journal of Audiology, 13(3), 108-112.

Bentler, R. A. (2000). List equivalency and test-retest reliability of the Speech in Noise test. American Journal of Audiology, 9(2), 84-100.

Bond, Z. S., Moore, Thomas J., & Gable, Beverley. (1989a). Acoustic--phonetic characteristics of speech produced in noise and while wearing an oxygen mask. The Journal of the Acoustical Society of America, 85(2), 907-912.

Bond, Z. S., Thomas, J. Moore, & Beverley, Gable. (1989b). Acoustic--phonetic characteristics of speech produced in noise and while wearing an oxygen mask (Vol. 85, pp. 907-912): ASA.

Brungart, D. S., Chang, P. S., Simpson, B. D., & Wang, D. (2009). Multitalker speech perception with ideal time-frequency segregation: effects of voice characteristics and number of talkers. Journal of the Acoustical Society of America, 125(6), 4006-4022.

Byrne D, Dillon H, and Tran K. (1994). An international comparison of long-term average speech spectra. Journal of the Acoustical Society of America, 96(4), 2108-2120.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. PLoS One, 5(6), e10729.

Cheng, Robert L. (1966). Mandarin Phonological Structure. Journal of Linguistics, 2(2), 135-158.

Cox, R.M., Alexander, G.C., & Gilmore, C. (1987). Development of the connected speech test (CST). Ear and Hearing, 8(5), 119s.

Darwin, J. Christopher, Brungart, S.Douglas , & Simpson, D. Brian. (2003). Effects of fundamental frequency and vocal-tract length changes on attention to one of two simultaneous talkers (Vol. 114, pp. 2913-2922): ASA.

Dreher, John J., & O'Neill, John. (1957). Effects of Ambient Noise on Speaker Intelligibility for Words and Phrases. The Journal of the Acoustical Society of America, 29(12), 1320-1323.

Duanmu, San. (2007). The phonology of standard Chinese. Oxford University Press.

Durlach, Nat. (2006). Auditory masking: Need for improved conceptual structure (Vol. 120): ASA.

Durrant, John D, & Feth, Lawrence L. (2012). Hearing Sciences: A Foundational Approach: Pearson Higher Ed.

François, Pellegrino, Christophe, Coupé, & Egidio, Marsico. (2011). Across-Language Perspective on Speech Information Rate. Language, 87(3), 539-558.

Freyman, Richard L , Helfer, Karen S , McCall, Daniel D, & Clifton, Rachel K (1999). The role of perceived spatial separation in the unmasking of speech (Vol. 106): ASA.

Garnier, Maeva;, Bailly, Lucie; , Dohen, Marion;, Welby, Pauline; , & Loevenbruck, Helene. (2006). An acoustic and articulatory study of Lombard speech: global effects on the utterance. Paper presented at the In INTERSPEECH-2006, Pittsubrgh, USA.

Gelfand, Stanley A, Ross, Leslie, & Miller, Sarah. (1988). Sentence reception in noise from one versus two sources: effects of aging and hearing loss. The Journal of the Acoustical Society of America, 83(1), 248-256.

Hansen, J., & Varadarajan, V. (2009). Analysis and Compensation of Lombard Speech Across Noise Type and Levels With Application to In-Set/Out-of-Set Speaker Recognition. Audio, Speech, and Language Processing, IEEE Transactions on, 17(2), 366-378.

Hansjörg Mixdorff, K. G., Martti Vainio. (2006). Time-domain Noise Subtraction Applied in the Analysis of Lombard Speech. Paper presented at the Speech Prosody 2006, Dresden, Germany.

Helfer, Karen S., & Freyman, Richard L. (2005). The role of visual speech cues in reducing energetic and informational masking (Vol. 117): ASA.

Jean-Claude, Junqua. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers (Vol. 93, pp. 510-524): ASA.

Junqua, J.-C. , Steven , C.F, & Field, Ken. (1998). Influence of the speaking style and the noise spectral tilt on the lombard reflex and automatic speech recognition. Paper presented at the ICSLP-1998, Sydney, Australia.

Junqua, Jean-Claude. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizers. The Journal of the Acoustical Society of America, 93(1), 510-524.

Junqua, Jean Claude. (1996). The influence of acoustics on speech production: A noise-induced stress phenomenon known as the Lombard reflex. Speech Communication, 20(1-2), 13-22. doi: Doi: 10.1016/s0167-6393(96)00041-6

Killion, M.C., & Villchur, E. (1993). Kessler was right-partly: But SIN test shows some aids improve hearing in noise. Hearing Journal, 46, 31-31.

Killion, M.C., Wilber, L.A., & Gudmundsen, G.I. (1988). Zwislocki was right. Hearing Instruments, 39(1), 14-18.

Killion, MC. (1984). New insert earphones for audiometry. Hearing Instruments, 35(7), 28.

Killion, MC, Wilber, LA, & Gudmundsen, GI. (1985). Insert earphones for more interaural attenuation. Hearing Instruments, 36(2), 34-36.

Killion, Mead C. (1997). Hearing aids: Past, present, future: Moving toward normal conversations in noise. British Journal of Audiology, 31(3), 141-148.

Killion, Mead C., Niquette, Patricia A., Gudmundsen, Gail I., Revit, Lawrence J., & Banerjee, Shilpi. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners.[Erratum appears in J Acoust Soc Am. 2006 Mar;119(3):1888]. Journal of the Acoustical Society of America, 116(4 Pt 1), 2395-2405.

Korn, T. S. (1954). Effect of Psychological Feedback on Conversational Noise Reduction in Rooms. The Journal of the Acoustical Society of America, 26(5), 793-794.

Lane, Harlan, & Tranel, Bernard. (1971). The Lombard Sign and the Role of Hearing in Speech. Journal of Speech & Hearing Research, 14(4), 677-709.

Lau, S.H. Polly. (1998). The Effect of Type and Level of Noise on Long-Term Average Speech Spectrum. A dissertation Submitted in partial fulfilment of the requirements for the degree of Master of Science in Audilogy at the University of Hong Kong, 1-33.

Lebo, C. P., Smith, M. F., Mosher, E. R., Jelonek, S. J., Schwind, D. R., Decker, K. E., . . . & Kurz, P. L. (1994). Restaurant noise, hearing loss, and hearing aids. Western Journal of Medicine, 161(1), 45-49.

Letowski, Tomasz, Frank, Tom, & Caravella, Jane. (1993). Acoustical Properties of Speech Produced in Noise Presented Through Supra-Aural Earphones. Ear and Hearing, 14(5), 332-338.

Lombard, E. (1911). Le Singe de l'Elevaton de la Voix (The sign of the rise in the voice). Ann. Maladiers Oreille, Larynx, Nez, Pharynx (Annals of diseases of the ear, larynx, nose and pharynx), 37, 101-119.

Lu, Youyi, & Cooke, Martin (2008). Speech production modifications produced by competing talkers, babble, and stationary noise (Vol. 124): ASA.

Mattys, S. L., Brooks, J., & Cooke, M. (2009). Recognizing speech under a processing load: dissociating energetic from informational factors. Cogn Psychol, 59(3), 203-243. doi: 10.1016/j.cogpsych.2009.04.001.

McCullough, J. A., Tu, C., Lew, H. L., McCullough, J. A., Tu, C., & Lew, H. L. (1993). Speech-spectrum analysis of Mandarin: implications for hearing-aid fittings in a multi-ethnic society. Journal of the American Academy of Audiology, 4(1), 50-52.

Mixdorff, Hansjörg, K. G., & Vainio, Martti. (2006). Time-domain Noise Subtraction Applied in the Analysis of Lombard Speech. Paper presented at the Speech Prosody 2006, Dresden, Germany.

Namba, Kuwano, Schick, Açlar, Florentine, & Zheng, Da Rui. (1991). A cross-cultural study on noise problems: Comparison of the results obtained in Japan, West Germany, the U.S.A., China and Turkey. Journal of Sound and Vibration, 151(3), 471-477.

Nilsson, M., Soli, S. D., Sullivan, J. A., Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994a). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. Journal of the Acoustical Society of America, 95(2), 1085-1099.

Nilsson, Michael, Soli, Sigfrid D, & Sullivan, Jean A. (1994b). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. The Journal of the Acoustical Society of America, 95(2), 1085-1099.

Patel, Rupal, & Schell, Kevin W. (2008). The Influence of Linguistic Content on the Lombard Effect. Journal of Speech, Language & Hearing Research, 51(1), 209-220.

Pellegrino, F., Coupe, C., & Marsico, E. (2011). A CROSS-LANGUAGE PERSPECTIVE ON SPEECH INFORMATION RATE. LANGUAGE, 87(3), 539-558.

Pittman, A. L.:, & Wiley, T. L. (2001). Recognition of speech produced in noise. Journal of Speech Language & Hearing Research, 44(3), 487-496.

Plomp, R, & Mimpen, AM. (1979). Improving the reliability of testing the speech reception threshold for sentences. International Journal of Audiology, 18(1), 43-52.

Rivers, C., Rastatter, M.P. . (1985). The Effects of multi-talker and masker noise on fundamental frequency variability during spontaneous speech for children and adults. The journal of Auditory Research, 25, 37-45.

Robyn, M. Cox, Genevieve, C. Alexander, & Christine, Gilmore. (1987). Intelligibility of average talkers in typical listening environments (Vol. 81, pp. 1598-1608): ASA.

Rosen, Stuart. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 336(1278), 367-373.

Rothauser, EH, Chapman, WD, Guttman, N., Nordby, KS, Silbiger, HR, Urbanek, GE, & Weinstock, M. (1969). IEEE recommended practice for speech quality measurements. IEEE Trans. Audio Electroacoust, 17(3), 225-246.

S.H.Lau, Polly. (1998). The Effect of Type and Level of Noise on Long-Term Average Speech Spectrum. A dissertation Submitted in partial fulfilment of the requirements for the degree of Master of Science in Audilogy at the University of Hong Kong, 1-33.

Schulman, Richard. (1989). Articulatory dynamics of loud and normal speech (Vol. 85, pp. 295-312): ASA.

Scott, Sophie K., Rosen, Stuart, Wickham, Lindsay, & Wise, Richard J. S. (2004). A positron emission tomography study of the neural basis of informational and energetic masking effects in speech perception. The Journal of the Acoustical Society of America, 115(2), 813-821.

Studebaker, G. A., Sherbecoe, R. L., McDaniel, D. M., Gwaltney, C. A., Studebaker, G. A., Sherbecoe, R. L., . . . Gwaltney, C. A. (1999). Monosyllabic word recognition at higher-than-normal speech and noise levels.[Erratum appears in J Acoust Soc Am 1999 Oct;106(4 Pt 1):2111]. Journal of the Acoustical Society of America, 105(4), 2431-2444.

Summers, W. Van , Pisoni, B.David, Bernacki, H. Robert, Pedlow I. Robert, & Stokes, A. Michael. (1988a). Effects of noise on speech production: Acoustic and perceptual analyses (Vol. 84, pp. 917-928): ASA.

Summers, W. Van, Pisoni, David B., Bernacki, Robert H., Pedlow, Robert I., & Stokes, Michael A. (1988b). Effects of noise on speech production: Acoustic and perceptual analyses. The Journal of the Acoustical Society of America, 84(3), 917-928.

Tartter, Vivien C., Gomes, Hilary, & Litwin, Elissa. (1993). Some acoustic effects of listening to noise on speech production. The Journal of the Acoustical Society of America, 94(4), 2437-2440.

Tufts, Jennifer B., & Frank, Tom. (2003). Speech production in noise with and without hearing protection. Journal of the Acoustical Society of America, 114(2), 1069-1080. doi: http://dx.doi.org/10.1121/1.1592165.

Webster, John C., & Klumpp, Roy G. (1962). Effects of Ambient Noise and Nearby Talkers on a Face-to-Face Communication Task. The Journal of the Acoustical Society of America, 34(7), 936-941.

Wu, Xihong, Wang, Chun, Chen, Jing, Qu, Hongwei, Li, Wenrui, Wu, Yanhong, . . . Li, Liang. (2005). The effect of perceived spatial separation on informational masking of Chinese speech. Hearing Research, 199(1-2), 1-10.

Yang, Zhigang, Chen, Jing, Huang, Qiang, Wu, Xihong, Wu, Yanhong, Schneider, Bruce A., & Li, Liang. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. Speech Communication, 49(12), 892-904.

Yip, Po-ching. (2000). The Chinese lexicon ; a comprehensive survey. London New York: Routledge.

Yu TSI, Wong TW. (2002). Occupational noise exposure and hearing impairment among employees in Chinese restaurants and entertainment sector in Hong Kong. Occupational Noise Exposure And Hearing Impairment. Health Services Research Committee.

Zwislocki, J. (1953). Acoustic attenuation between the ears. The Journal of the Acoustical Society of America, 25(4), 752-759.