Copyright
by
Jingjing Guan
2017

**The Dissertation Committee for Jingjing Guan Certifies that this is the approved
version of the following dissertation:**

**Speech Perception in Noise with Formant Enhancement for Older
Listeners**

Committee:

Chang Liu, Supervisor

Craig Champlin

Rajka Smiljanic

Julia Campbell

**Speech Perception in Noise with Formant Enhancement for Older
Listeners**

by

Jingjing Guan, B.S., M.ED.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May 2017

Acknowledgements

I would like to thank Dr. Chang Liu for being such an inspirational supervisor and friend. I truly appreciate his tireless encouragement and support throughout the doctoral program. In addition, I would like to acknowledge the Dissertation Committee, including Dr. Craig Champlin, Dr. Julia Campbell, and Dr. Rajka Smiljanic. Their comments and suggestions through all stages of this study have always been timely and indispensable.

I would also like to give special thanks to Dr. SuHyun Jin, Dr. Craig Champlin, and Dr. Doug Sladen for their generous support in my research rotations. I feel fortunate to have had the opportunity to work with researchers and clinicians in such a wonderful department. Additional thanks goes to Joan Balash, Dr. Angela Carey, and Dr. Amanda Zappler for their clinical supervision and introducing me to the exciting field of clinical audiology. I am also grateful to have practiced in off-campus clinics under my advisors, Drs. Zhang, Powell, May, Tian, Archer, and Pho.

I cannot thank enough the friends I met in Austin. Hoyoung, Lisa, Ying, and Laurel always stand beside me. Special thanks also to my husband, parents, and parents-in-law. None of this would have been possible without the long support of my family.

Lastly, I wish to thank the subjects who participated in the study. The research in this dissertation was conducted with the support of Jesse H. Jones Fellowship and Global Research Fellowship.

Speech Perception in Noise with Formant Enhancement for Older Listeners

Jingjing Guan, Ph.D.

The University of Texas at Austin, 2017

Supervisor: Chang Liu

Degraded speech intelligibility in background noise is a common complaint of listeners with hearing loss. The purpose of the current study is to investigate the effect of spectral enhancement of the second formant frequency (F2) on speech identification in noise for older listeners with hearing loss (HI) and with normal hearing (NH). This study also aims to explore whether F2 enhancement improves speech perception in noise across languages such as American English and Mandarin Chinese. Target words (e.g., color and digit) were selected and presented based on the paradigm of coordinate response measure (CRM) corpus of English and Chinese versions. Speech recognition thresholds with original and F2-enhanced speech in two-talker and six-talker babble were examined for English and Chinese listeners with NH and HI. As expected, listeners with NH had better performance on speech perception in noise than listeners with HI in almost all listening conditions. More importantly, thresholds of both NH and HI groups were improved for enhanced speech signals across languages, primarily in two-talker babble, but not in 6-talker babble. Compared with the NH group, listeners with HI showed significantly greater benefits in the most challenging condition (e.g., low signal-to-noise ratios). F2-enhancement benefits did not significantly correlate with memory abilities. Moreover, the speech intelligibility index (SII) model interpreted the F2-enhancement benefits for the two language groups in multi-talker babble. That is, the perceptual improvement was mainly

associated with the greater availability of acoustic cues of speech with F2 enhancement in the two-talker babble, but not in the six-talker babble. Overall, speech sounds with F2 enhancement may improve listeners' speech perception in noise across different languages, possibly due to a greater amount of speech information available.

Table of Contents

List of Tables	ix
List of Figures	x
List of Abbreviations	xiii
Chapter 1 Introduction	1
Chapter 2 Literature Review	3
2.1 Speech perception for hearing-impaired listeners.....	3
2.1.1 Reduced audibility	3
2.1.2 Reduced frequency resolution in speech perception.....	7
2.1.3 Impaired temporal resolution in speech perception	12
2.1.4 Aging-related cognitive decline associated with speech perception	18
2.2 Strategies to enhance speech perception in noise	19
2.2.1 Spectral enhancement of speech	20
2.2.2 Speech enhancement of this study	24
2.2.3 Evaluating the effectiveness of speech enhancement	26
2.3 Goals of this study	28
Chapter 3: Methods.....	29
3.1 Participants.....	29
3.2 Speech stimuli and noise.....	30
3.3 Stimulus presentation.....	32
3.4 Procedure	33
Chapter 4: Results	36
4.1 Demographic factor	36
4.2 Memory span scores	36
4.3 Speech recognition in noise with and without F2 enhancement.....	40
4.4 Benefits from speech enhancement	47
4.5 Correlation of age, hearing loss, memory span with SRTs and F2-enhancement benefits.....	50

4.6 Speech intelligibility index (SII) model.....	61
Chapter 5: Discussion	67
5.1 Speech recognition in noise for older listeners with and without hearing loss	67
5.1.1 Effects of noise masking on SRTs	67
5.1.2 Effect of hearing loss on SRTs	69
5.1.3 SII and SRTs	70
5.1.4 Age and SRTs	72
5.1.5 Cognitive memory and SRTs.....	72
5.2 Speech enhancement for older listeners	74
5.2.1 Masking release from temporal modulation in noise.....	74
5.2.2 Information masking interference and F2 enhancement.....	76
5.2.3 Language background and F2 enhancement.....	78
5.2.4 F2 enhancement and other enhancement technology	79
5.3 Summary of results	82
5.4 Future directions	82
5.5 Clinical application of F2 enhancement	83
Chapter 6: Conclusion.....	85
References.....	86

List of Tables

Table 1:	Correlation coefficients between age and absolute SRTs of original and enhanced speech for English HI listeners	50
Table 2:	Correlation coefficients between memory span and SRTs of original and enhanced speech for English HI listeners	51
Table 3:	Correlation coefficients between audibility thresholds and SRTs of original and enhanced speech for English NH and HI listeners	53
Table 4:	Correlation coefficients between PTA and SRTs of original and enhanced speech for English NH and HI listeners.....	54
Table 5:	Correlation coefficients between audibility thresholds and SRTs of original and enhanced speech for Chinese NH and HI listeners.....	57
Table 6:	Correlation coefficients between PTA and SRTs of original and enhanced speech for Chinese NH and HI listeners.....	58
Table 7:	Band importance function for one-third octave band SII procedure.	60

List of Figures

Figure 1:	Transfer functions showing the relationship between the SII and speech recognition score	7
Figure 2:	Schematic representation of a complex stimulus with two formants in a normal auditory system and an impaired auditory system.....	9
Figure 3.	Waveform showing the temporal envelope and temporal fine structure of speech sounds	13
Figure 4.	Audiograms for English and Chinese group with normal hearing and hearing loss	29
Figure 5.	Spectra of original speech and speech with 9-dB F2 enhancement..	30
Figure 6.	Psychometric function showing percent responses as a function of three different up-down tracking procedures.	33
Figure 7.	Memory spans of Forward Digit Span, Backward Digit Span, and Letter-Number Sequencing subtests for English NH and HI group	36
Figure 8.	Scatter plots showing the correlations between age and memory spans of three subtests for English NH listeners and English HI listeners	37
Figure 9.	Memory spans of Forward Digit Span, Backward Digit Span, and Letter-Number Sequencing subtests for Chinese NH and HI group	38
Figure 10.	Scatter plots showing the correlations between age and memory spans of three subtests for Chinese NH listeners and Chinese HI listeners	39

Figure 11.	SRTs for English NH and HI listeners in two-talker and six-talker babble at three challenges levels without and with F2 enhancement.	40
Figure 12.	SRTs for Chinese NH and HI listeners in two-talker and six-talker babble at three challenges levels without and with F2 enhancement.	43
Figure 13.	Benefits from F2 enhancement for English NH and HI listeners in two-talker and six-talker babble at three challenge levels	46
Figure 14.	Benefits from F2 enhancement for Chinese NH and HI listeners in two-talker and six-talker babble at three challenge levels	47
Figure 15.	Average benefits from F2 enhancement in two-talker babble for English and Chinese group across hearing status, and for English and Chinese group with normal hearing and hearing loss	48
Figure 16.	Correlations between memory spans and F2-enhancement benefits for English HI listeners in two-talker (left) and six-talker babble (right) at challenge level I.	52
Figure 17.	SII value as a function of correct percentages for English NH and HI listeners in two-talker and six-talker babble with/without F2 enhancement	61
Figure 18.	SII values as a function of correct percentages for Chinese NH and HI listeners in two-talker and six-talker babble with/without F2 enhancement	62
Figure 19.	SII values for original English and Chinese speech in two-talker and	

six-talker babble.....63

Figure 20. Waveforms for English two-talker and six-talker babble and
Chinese two-talker and six-talker babble.....67

List of Abbreviations

AI	Articulation Index
CEFS	Contrast Enhanced Frequency Shaping
CN	Chinese Group
CV	Consonant -Vowel
CVC	Consonant -Vowel- Consonant
dB	Decibels
dB HL.....	Decibels Hearing Level
dB SPL.....	Decibels Sound Pressure Level
EN	English group
F1	First Formant Frequency
F2	Second Formant Frequency
FFT.....	Fast Fourier Transform
HI	Hearing-Impaired listeners
HINT	Hearing in Noise Test
LTSSN	Long-term Speech Shaped Noise
MTB.....	Multi-talker Babble
NH.....	Normal-Hearing Listeners
PTA	Pure Tone Average
SD	Stanard Deviation
SE.....	Standard Error
SII.....	Speech Intelligibility Index
SNR.....	Signal-to-Noise Ratio

SRTSpeech Reception Threshold
T Threshold
TFS..... Temporal Fine Structure
WAIS-III Wechsler Adult Intelligence Scale-III

Chapter 1: Introduction

Hearing-impaired listeners may achieve relatively good performance in one-to-one conversation in a quiet listening condition when signal levels are high enough. However, their performance in noisy environments drops significantly even when they wear hearing aids that compensate for the loss of audibility (Moore, 2007). Previous research has shown that, besides decreased audibility, other factors also affect their performance of speech perception in noise, such as suprathreshold deficits (e.g., impaired frequency and temporal resolution; Moore, 1996; Moore, 2008) and aging-related cognitive decline (Kalluri & Humes, 2012; Humes & Young, 2016). The purpose of the present study was to investigate whether the technique of spectral enhancement (e.g., formant enhancement) could help older listeners with normal hearing and hearing loss improve speech perception in noise across different languages.

In general, speech perception in noise primarily depends on the signal-to-noise ratios (SNR) rather than the overall signal level (Liu & Kewley-Port, 2004; Phatak & Allen, 2007; Plomp & Mimpen, 1979). Listeners with hearing impairment (HI) require greater SNRs to achieve performance equivalent to listeners with normal hearing (NH). Speech reception thresholds (SRT) in noise can be 5 to 15 dB worse for HI listeners than for NH listeners, depending on the complexity of speech materials (e.g. spondaic words, sentences with high and low predictability), and the type of noise (speech-shaped noise, temporally-modulated noise, and multi-talker babble (Dubno et al., 1984; Humes, 2002; Phatak & Allen, 2007; Plomp & Mimpen, 1979). To some extent, hearing aids help to increase the SNR by different techniques, such as directional microphones and noise reduction circuits. Such hearing aid technology of reducing noise and improving overall SNRs may solve the audibility problem. However, such approaches may not restore suprathreshold deficits such

as reduced spectral and temporal processing, as well as cognitive decline. In this study, instead of increasing the overall SNRs, spectral enhancement was applied to compensate the reduced spectral resolution for listeners by selectively enhancing the second formant of speech (i.e., F2). Speech recognition tasks in noise were conducted to evaluate the effectiveness of F2 enhancement for older listeners with normal hearing and hearing loss across two different languages: English and Mandarin Chinese.

Chapter 2: Literature Review

This review covered two areas relevant to the proposed study. Firstly, research on speech perception for hearing-impaired (HI) listeners are reviewed, such as the factors accounting for reduced performance of speech perception in noise for HI listeners. Secondly, techniques to enhance speech intelligibility in noise and their effectiveness on speech perception are summarized.

2.1. Speech perception for hearing-impaired listeners

In general, HI listeners have more difficulty in understanding speech in noisy situations compared with normal-hearing (NH) listeners. For example, HI listeners demonstrate elevated speech reception thresholds (SRT) in noise compared to normal-hearing listeners (Plomp, 1986; Houtgast et al., 2008). The SRT estimates the lowest speech level (or speech-to-noise ratio, SNR) that is needed for listeners to correctly recognize 50% of the sentences (Plomp, 1986). Houtgast et al. (2008) reviewed previous studies and summarized several important factors that may account for poor scores of speech-in-noise tests for HI listeners: audibility; spectral and temporal processing; and cognitive factors involved in speech processing. The effect of these factors on speech recognition in noise is described individually below.

2.1.1 Reduced audibility

2.1.1.1 Effect of reduced audibility on speech perception

HI listeners show elevated hearing thresholds and thus require relatively high speech levels to ensure speech audibility. Audibility is usually represented by pure-tone thresholds, which can illustrate the amount of hearing loss due to deficits in the outer ear,

middle ear, the cochlea, and/or the central auditory system (sensory-neural system). The most common type of hearing loss for older listeners is known as sensorineural hearing loss, mostly resulting from the dysfunction of outer hair cells in the cochlea. As the active amplifier of the cochlea, outer hair cells' dysfunction can result in the decrease of hearing sensitivity of up to 55-60 dB (Patuzzi et al., 1989). With reduced audibility, HI listeners have substantial difficulty understanding speech in quiet because the spectral level of speech falls below their hearing thresholds, resulting in only partially intelligible speech. Previous studies have shown that reduced audibility, especially for hearing sensitivity at high frequency regions, can account for performance of speech recognition in a quiet environment for hearing-impaired listeners (Ching et al., 1998).

Bacon et al. (1998) pointed out that reduced audibility could also explain why listeners with sensorineural hearing loss had poor performance in speech perception in temporally-modulated noise. They compared three groups (NH, sensorineural HI, and NH that simulated the elevated thresholds of HI) in several types of noisy backgrounds. The difference in speech perception in noise between the NH group and the simulation group was regarded as an estimate for the threshold-related component of the hearing loss (audibility), while the difference between the HI group and the simulation group was considered to be an estimate of the component due to suprathreshold deficits. Results revealed that compared to suprathreshold deficits, reduced audibility contributed more greatly to the variations of speech perception in temporally-modulated noise for HI listeners.

Hearing loss, as measured through pure tone thresholds, usually accounted for about 50-80% of variance in speech perception performance (Shrivastav et al., 2006). There were a strong association between hearing loss and speech performance in noise (for a review, Akeroyd, 2008). For example, PTA (pure tone average) above 1 kHz for HI

listeners was associated with speech reception in noise, while PTA at and below 1 kHz was related to speech reception in quiet (Smoorenburg, 1992). However, pure-tone audiometry to determine the degree of hearing loss is a simple measure of sensory processing. Some research reported that audibility, such as PTA, is not a reliable predictor for speech perception in noise (Vermiglio et al., 2012). The discrepancy on the effect of audibility loss (e.g., whether audibility loss is highly correlated with HI listeners' speech recognition in quiet and noise) may be due to the differences in speech materials (e.g., phonemes, syllables, words, and sentences), the type of noise (e.g., stationary noise, temporally-modulated noise, multi-talker babble), and the configuration and severity of listeners' hearing loss across these studies.

2.1.1.2 Quantifying decreased audibility and speech intelligibility

Compared to pure tone thresholds, the speech intelligibility index (SII) is a significantly stronger predictor for speech audibility (Stiles, 2010; Vermiglio et al., 2012). Unlike pure tone thresholds that are developed to describe the degree of hearing loss, SII incorporates the importance-weighted audibility of speech signal by assigning perceptual weights to individual frequency bands in different listening conditions, such as noise and reverberation.

SII is developed based on the Articulation Index (AI) and includes corrections to compensate negative effects of upward spread of masking and high presentation levels (ANSI, S3.5, 1997). The following equation describes the computation of the SII:

$$S = \sum_{i=1}^n I_i A_i \quad \text{Eq. 2.1}$$

where the value S refers to the SII (ranging from 0 to 1), and n is the index number of the frequency bands. I_i is the value of band importance function that characterized the weight of the i th frequency band for speech intelligibility. A_i is the value of the band audibility function that indicates the proportion of speech dynamic range in the i th frequency band above listener's hearing threshold.

In the computation of SII, the frequency range between 100 Hz and 9500 Hz is divided into several frequency bands in a variety of procedures: critical band procedure (21 bands), one-third octave band procedure (18 bands), equally-contributing critical band procedure (17 bands), and octave band procedure (6 bands). In each frequency band, the speech level and noise level are computed with listener's hearing thresholds included (e.g., A_i). The proportion of audible speech in a specific frequency region (A_i) is then multiplied by relative importance of frequency range (I_i) (see Table 7). Eventually, SII is summed over all frequency bands.

The SII score is an indicator of the availability (audibility) of speech information rather than a percentage score of speech recognition. An SII of 0 indicates that no speech information is available for listeners, while an SII of 1.0 implies that speech information is fully available. Basically, as the value of SII increases, speech perception score grows monotonically (see Figure 1). However, SII is not a direct measure of intelligibility. For example, an SII of 0.5, where 50% of speech cues is available, does not mean that listeners recognize speech sounds at 50% accuracy. As for some listeners, they may obtain almost 100% correctness of speech perception at an SII of 0.5. In order to obtain speech recognition scores from SII, a transfer function was examined. An example of the transfer function is illustrated in Figure 1. It should be noted that the transfer function of speech recognition score from the SII is dependent on the type of speech materials, such as phonemes, words, and sentences (Kryter, 1962; Hornsby et al., 2004).

In short, SII is a useful tool in predicting audibility and quantifying the noise masking on speech and will be used in this study to examine the effect of spectral enhancement on speech recognition in noise for older listeners with normal and impaired hearing.

Figure 1. Transfer functions showing the relationship between the SII and speech recognition scores.

2.1.2 Reduced frequency resolution in speech perception

In addition to the audibility issue, suprathreshold auditory deficits such as reduced frequency selectivity and temporal processing also account for the degradation of speech perception for hearing-impaired listeners (Houtgast et al. 2008). Perceptual deficits in the spectral and temporal domain for HI listeners mainly result from cochlear damage (Moore, 1996; Oxenham et al., 2003) that includes the loss of inner and outer hair cells.

2.1.2.1 Effects of reduced frequency selectivity on speech perception

The cochlea can be regarded as a bank of overlapping auditory filters (Tyler, 1986). Because the filter-sharpening effects of the outer hair cells in the cochlea are reduced due to hearing loss, auditory filters become remarkably broader in hearing-impaired individuals compared to normal-hearing individuals (Leek & Summers, 1993; Oxenham & Bacon, 2003). Generally, the worse hearing loss, the greater distortion of internal frequency representation. As illustrated in Figure 2A, in the spectral domain, formant frequencies are clearly displaced as the resonant peaks and are distinctive components of speech sounds like vowels, conveying critical acoustic information. The frequencies of the first two formants, F1 and F2, are primary acoustic cues for vowel categorization (Peterson & Barney, 1952). When speech signals pass through auditory filters in the cochlea, the spectral peaks of speech are well resolved in the normal cochlea (e.g., Figure 2A-2C), while they are smeared in the impaired cochlea, i.e., the spectral contrasts between peaks and valleys are significantly reduced (Figure 2D-2E). Such spectral smearing due to broader auditory filter bandwidths is one of the important factors accounting for hearing-impaired listeners' challenge in speech perception in quiet and noise (Leek et al., 1987; Liu & Kewley-Port, 2007; Woodall & Liu, 2013).

Figure 2. Schematic representation of a complex stimulus with two formants (F1 and F2) in a normal auditory system (top) and an impaired auditory system (bottom). Spectral smearing as shown in E.

In general, broader auditory filters in hearing loss degrade the performance of speech perception in quiet due to a reduction in formant peak amplitude (Liu & Kewley-Port, 2007; Woodall & Liu, 2013). Moreover, broader auditory filters allow more noise to come through to fill the spectral dips, resulting in dramatic reduction in spectral contrasts (Nitttrouer et al., 2015). Baer and Moore (1993) found that even a spectral smearing simulation of six-time broader auditory filters did not change speech perception in quiet for listeners with normal hearing. However, because of the spectral smearing, speech perception was adversely affected in the background noise with signal-to-noise ratio (SNR) of -3 dB, where the signal level is lower than noise level by 3 dB. These results indicated that spectral smearing resulting from broader auditory filters negatively affected speech perception, especially in noisy environments.

2.1.2.2 Effects of spectral contrasts on speech perception

Hearing aid devices can be used to restore audibility by increasing the intensity of speech at specific frequencies where individuals have hearing loss. However, simple amplifications on speech levels are not able to sharpen the traveling wave and replace damaged hair cells and do not fully resolve distorted components, such as spectral smearing. Thus, listeners with similar audiograms show great individual differences on benefits from hearing aids amplification. Hearing aids are helpful for some users in everyday communication, while other hearing aid wearers with similar hearing loss gain little benefit (Plomp, 1978,1986; Johnson et al., 2016).

Ter Keurs et al. (1992) investigated the relative importance of frequency resolution on speech intelligibility of phonemes and sentences in NH listeners. Impaired frequency resolution was stimulated by smearing the short-term spectral envelope of speech signals on a log-frequency scale with a Gaussian-shaped filter. Compared with original speech, understanding spectrally-smearred speech required significantly higher SNR values. Furthermore, they found that the identification of vowels appeared to be dependent more on spectral contrasts, as compared to the identification of consonants that had low intensity levels and short durations. Some back vowels with the first and second formants (F1 and F2) close to each other, were mostly confused with each other. Larger amplitude difference between peaks and valleys of formants (i.e. spectral contrast) certainly helped improve speech perception in noise (Leek et al., 1996). Liu and Eddins (2008) also investigated the effect of spectral smearing (e.g., reduced spectral contrasts of formant peaks) by applying a filter in the spectral modulation domain to vowel stimuli. They reported that vowel identification was degraded for normal-hearing listeners as the spectral contrasts was reduced. In particular, vowel identification scores were significantly correlated with the

spectral contrasts of F1 and F2, indicating the importance of the F1 and F2 resolution in vowel perception.

Previous studies showed that to reach 75% accuracy of vowel identification in quiet, NH listeners needed 1 to 2 dB peak-to-valley difference in speech formants, while listeners with moderate hearing loss required a much larger spectral contrast (e.g., 6-to-7 dB difference; Leek et al., 1987). Their further study (Leek et al., 1996) assessed the effect of reduced spectral resolution on vowel identification in noise. Results suggested that HI listeners with broadened auditory filters at 2000 Hz required higher spectral contrasts of formant peaks to recognize vowels in noise than did NH listeners. Thus, the reduction in the internal representation of speech spectral contrast for HI listeners may be one of the crucial factors underlying their poor performance of speech perception in noise.

As the studies above indicated the importance of formant peaks such as F1 and F2 on speech recognition, it has been demonstrated that the role of F1 and F2 may be different, in particular, F2 is quite important for speech perception in noise for hearing-impaired listeners. Firstly, the low-frequency spectral peak, such as F1, usually has relatively high amplitude and thus is well preserved in noise. Previous studies examined vowel formant discrimination in which the smallest change in formant frequency was measured. Results showed that F1 discrimination thresholds were similar between young and elderly listeners, and between listeners with normal and impaired hearing (Coughlin et al., 2008; Liu & Kewley-Port, 2007; Woodall & Liu, 2013). HI listeners had primary difficulty in discriminating differences of F2 frequency (Liu & Kewley-Port, 2007; Woodall & Liu, 2013). In noisy conditions, listeners relied on relatively accurate F1 information along with partial F2 to identify vowels (Parikh & Loizou, 2005). Second, the upward spread of masking, where a low-frequency noise masker or the F1 of speech signals may interfere high frequencies in speech like F2, also made speech perception harder for older listeners

(Martin & Pickett, 1970; Klein et al., 1990). Third, high-frequency hearing loss is a very common type of hearing impairment for old listeners. Poor audibility usually started decreasing at high frequency regions including F2 frequencies. Simply amplifying overall speech levels to a well-audible level (e.g. 90 or 95 dB SPL) may not significantly compensate for spectral smearing of formant peaks for HI listeners (Liu & Kewley-Port, 2007; Woodall & Liu, 2013). For example, thresholds of F2 frequency discrimination became slightly worse at 90 or 95 dB SPL than at 70 dB SPL for hearing-impaired listeners, likely due to their worsened frequency selectivity (e.g., broader auditory filter). In summary, preserved and audible F2 information is critical to construct the clear representation of spectral contrast in noise for HI listeners.

2.1.3 Impaired temporal resolution in speech perception

2.1.3.1 Importance of temporal cues for speech perception

Two types of acoustic temporal cues are important for discrete speech signals and speech perception: temporal envelope and temporal fine structure (TFS) (Rosen, 1992) (see Figure 3). Basically, temporal envelope (relatively slow variations in amplitude over time: 2 - 50 Hz) refers to information related to the amplitude envelope of speech (e.g. rise/fall time of consonants and vowels). TFS with rapid oscillations in amplitude with a rate close to the center frequency of the band (500 - 10,000 Hz), includes speech information such as vowel formants and format transitions between phonemes. Previous studies revealed that temporal envelope was important to speech perception in quiet (Shannon et al., 1995). In contrast, TFS contributed to understanding speech in noise (Nie et al., 2005; Lorenzi et al., 2006), especially for listening in the dips of the temporally modulated background noise (e.g. two-talker babble) (Strelcyk & Dau, 2009).

Figure 3. Waveform showing the temporal envelope and temporal fine structure of speech sounds.

Shannon et al. (1995) reported the significance of temporal envelope cues for speech perception. Speech signals were split into one to four frequency bands and the temporal envelope of each frequency band was extracted by rectification followed by low pass filters. Then, the envelope was modulated by broadband noise and summed across frequency bands. In this way, temporal envelope cues in each band were preserved but spectral energy was restricted with minimal TFS. Listeners could recognize words, phrases, and sentences using preserved amplitude envelopes of speech signals and their speech intelligibility remained as high as 80% with only three frequency bands. They also found that the temporal envelope with minimal TFS is sufficient to convey information on “voicing” and “manner” but not on “place” for consonant perception, consistent with previous findings (Rosen, 1992). Reduced perception in the “place” of articulation resulted from the limited spectral information of voiced speech where the harmonic structure was not present in noise-band simulations. Results demonstrated that the temporal envelope in a few frequency bands is sufficient for reliable intelligibility of speech in quiet.

Lorenzi et al. (2006) explored the usage of temporal cues (i.e., temporal envelope, and TFS) for young NH listeners, young HI listeners, and elderly HI listeners by filtering speech sounds so as to preserve either the temporal envelope or TFS. They found that young and elderly HI listeners could performed almost as well as young NH listeners with intact speech sounds and the processed speech sounds with only temporal envelope. However, both young HI and elderly HI listeners showed significant reduced abilities to use TFS. Furthermore, they evaluated young HI listeners' performance of identifying intact nonsense syllables in a temporally-varying noise with sinusoidal amplitude modulation at 8-Hz with a 100% depth. A significant positive correlation between the ability to use TFS and the masking release from temporal modulation of noise was observed for young HI listeners, indicating TFS may be critical for listening in the fluctuating background noise.

2.1.3.2 Temporal resolution and speech perception by listeners

Temporal resolution (or temporal processing ability) is defined as the auditory system's ability to respond to rapid changes in the envelope of the sound. Temporal resolution is important for listeners to follow acoustic signals in sequence, put information together into meaningful speech, and understand fast time-varying stimuli (e.g. speech). There are two main ways to evaluate the relationship between listeners' temporal resolution and speech perception: measuring listeners' temporal resolution and manipulating temporal features of speech signal or background noise.

First, several common psychoacoustic methods are utilized to examine temporal resolution for listeners: (1) Gap detection: to determine how much the gap length is needed for listeners to detect the gap between two sounds (Moore et al., 1992). (2) Temporal modulation detection (Viemeister, 1979): to reflect listeners' sensitivity to temporal changes in the amplitude envelope of sound. (3) Forward masking paradigm (Nelson &

Pavlov, 1989; Wojtczak & Viemeister, 2005): to understand the recovery ability of masking from the preceding peaks of sounds. A long recovery time may make speech information in dips less audible. (4) Duration discrimination: to measure minimal change in the duration that can be discriminated (Fitzgibbons & Gordon-Salant, 1995). In addition, some electrophysiological studies showed that amplitude modulation following response (i.e., Auditory Steady-State Response, ASSR) could also be a good indicator of listeners' envelope processing capabilities (Purcell et al., 2004; Grose et al., 2009).

Second, temporal characteristics of speech signals can be modified by time compression (Gordon-Salant & Fitzgibbons, 1999; 2001) or adding temporal jitter (Pichora-Fuller et al, 2007) and then the perception of speech with the temporal manipulation was measured to evaluate the role of speech temporal features on speech recognition. Fast-rate or time-compressed speech sounds impede different types of temporal cues to be used, while bringing disruptions in periodicity (i.e. increase in temporal jitter) can make signals less periodic by simulating the loss of temporal synchrony. Research demonstrated that even for listeners with normal hearing, weaker temporal resolution of speech adversely affected listeners' performance in speech perception (Gordon-Salant & Fitzgibbons, 1999; Pichora-Fuller et al, 2007).

On the other hand, the usage of temporal cues in noise for listeners in speech perception in noise can be evaluated by comparing with their performance in noisy conditions with different temporal features (e.g., stationary noise, sinusoidal amplitude-modulated noise, fluctuating noise, multi-talker babble, and the like). For example, in order to compare the amount of masking release listeners could obtain from different fluctuating maskers that differed only in temporal envelope, Bacon et al. (1998) employed four types of noise maskers: 1. the steady-state speech-shaped noise; 2. the speech-shaped noise that was modulated by the envelope of multi-talker babble (MTB); 3. the speech-shaped noise

that was modulated by the envelope of single-talker babble; 4. the speech-shaped noise that was modulated with a 10-Hz square wave with 100% modulation depth. The HINT (i.e., Hearing in Noise Test) sentences were then presented in the above four background conditions. They reported that normal listeners received largest benefits from temporal variations in the speech-shaped noise with the square wave, followed by the noise with the temporal envelope of single-talker babble. Masking release decreased for MTB condition with increasing number of talkers in the background noise.

2.1.3.3 Effect of impaired temporal resolution on speech perception

When older listeners process fast-rate or time-compressed speech sounds, speech processing has been demonstrated to be problematic (Gordon-Salant & Fitzgibbons, 1999; 2001). Both the effect of aging and effect of hearing loss were negative when older listeners or listeners with hearing loss perceived time-compressed speech sounds with 50 % - 60% time compression ratio (TCR) (Gordon-Salant & Fitzgibbons, 1999). Moreover, time compression resulted in a reduction mainly in consonant identification for older listeners with hearing loss (Gordon-Salant & Fitzgibbons, 2001).

The decreased temporal resolution of HI listeners may be one of the major factors accounting for their poor speech recognition in noise (Dubno et al., 2002; Jin et al., 2006, 2014). In Jin et al. (2006)'s study, the overall level of the speech and noise was adjusted based on each HI listener's auditory sensitivity. They found that HI and NH listeners had comparable performance in quiet and in steady noise. However, the HI listeners did significant better in gated noise compared with stationary noise, but obtained significantly smaller masking release from gated noise than NH listeners. In addition to the contribution of reduced audibility, reduced temporal resolution accounted for 19% of the variance in speech perception in noise performance (Jin & Nelson, 2010).

Moreover, previous studies have found that older listeners demonstrated poorer temporal resolution and had difficulty following time-varying envelope fluctuations (Jin et al., 2014). In addition to aging effect, hearing loss is another factor resulting in the reduction of temporal resolution and speech perception (Lorenzi et al., 2006). Listeners with hearing loss have decreased ability of using the temporal cues (i.e. temporal fine structure) and taking less advantage of temporal dips in background noise to enhance speech perception compared to younger or normal-hearing listeners.

Some researchers used the model of decreased internal representation of the temporal envelope because of cochlear compression dysfunction for HI listeners to explain how impaired temporal resolution affected speech perception in noise (Moore et al., 1996; Kale & Heinz, 2010). Deficits in phase locking and a loss of synchrony in neural firing might be the underlying mechanism of reduced temporal resolutions. Cochlear damage could bring in a reduction in compressive nonlinearity in the input-output function of basilar membrane, “loudness recruitment”, poorer frequency and temporal resolution (Moore, 1995, Oxenham et al., 2004). Cochlear hearing loss might magnify amplitude fluctuations and made them more prominent perceptually (Moore et al., 1996; Fullguille, 2003). That is, nonlinear behavior of the basilar membrane may bring in increased audibility during the temporal dips of the noise. However, HI listeners who have less or absent compressive nonlinearity, have reduced audibility during the silent periods of masking noise. Research showed that there was a potential strong link between cochlear compression dysfunction and psychoacoustic ability of using temporal fluctuation in recognizing speech (Oxenham et al., 2004).

Overall, these studies indicated that reduced temporal resolution was a critical factor to interpret HI listeners’ poor performance in speech perception in quiet and noise, particularly the time-varying noise.

2.1.4 Aging-related cognitive decline associated with speech perception

In addition to the loss of audibility and/or the deficits in suprathreshold spectral and temporal processing, aging-related cognitive decline plays another significant role in speech perception in noise for older listeners with normal hearing and hearing loss (CHABA, 1988; Humes & Young, 2016). Cognitive function, including working memory, attention, and speed of processing, shows age-related decline (Salthouse, 2001; Verhaeghen et al., 2003). Speech perception requires peripheral bottom-up sensory encoding, as well as central top-down cognitive processing, which is modulated by general capacity of an on-line working memory system (Ronnberg, 2003). For instance, some measures of speech perception tasks ask listeners to repeat sentences or words to the experimenter, which demand listeners to correctly encode acoustic sounds, identify and then retain in memory so as to repeat information. If older listeners have poor working memory and limited cognitive resources to store and allocate information during the listening task, even for amplified speech, they probably will not perform well in speech perception tasks, especially in noisy environments that impose higher sensory and cognitive loads.

Cognitive factors, such as working memory, could be another important predictor of individual differences in speech perception performance (Humes, 2002). For example, listeners with normal cognitive function performed better in speech perception in noise than listeners with reduced cognitive function (Akeroyd, 2008). The combination of aging-related cognitive decline, reduced audibility, broadening auditory filters, and decreased temporal processing account for the challenge of speech recognition in noise for old listeners, especially the older listeners with hearing loss.

2.2 Strategies to enhance speech perception in noise

Amplification systems, such as hearing aids, and assistive listening systems, have attempted to find effective ways to compensate for hearing loss and improve speech perception in noise. Historically, many manufacturers and scientists have made numerous attempts to develop signal processing strategies with the goals of improving perceived comfort, listening effort, speech clarity, and speech intelligibility for listeners with hearing loss. There are two main strategies to improve listeners' performance in perceiving speech in noise: noise reduction and speech enhancement. Basically, noise reduction approach aims to increase listening comfort and speech intelligibility by reducing noise interference, while speech enhancement approach is to enhance speech components.

In noise reduction approach, two common ways used to target and reduce the noise are: microphone technologies and noise reduction algorithms (see a review, Chung, 2004). Microphone technologies use an array of microphones to cover different frequency ranges so as to detect spatial differences between speech and noise. The technology assumes that the speech source is stationary and suppresses nonstationary interferences depending on the direction and frequencies of target sounds (Luts et al., 2004). For instance, if the noise comes from behind the hearing-aid user, noise would arrive at the rear microphone port first, followed by the front microphone port. Therefore, noise can be identified according to the time delay between microphone ports. Studies have shown this type of processing to be effective on speech perception in noise and thresholds of NH listeners were improved by approximately 3 dB (Leeuw & Dreschler, 1991); but this approach requires a large number of effective microphones. Second, noise reduction algorithms statistically detect the incoming signal and judge the presence or absence of the speech signal depending on distinct modulation features between speech and noise. Once the noise is identified, the

gain would be automatically reduced. However, it is hard to distinguish speech from noise when they are inextricably mixed (Chung, 2004).

In addition to noise reduction, speech enhancement is another approach to achieve a better speech-to-noise ratio. Because frequency resolution and spectral contrast have been proven to be critical in speech perception in noise, the general theory underlying speech enhancement algorithms targets at increasing spectral contrast to make formants more robust. Ideally, successful speech enhancement should produce statistically and practically significant improvement for listeners with sensorineural hearing loss. Several commercially-available hearing aids have included “speech enhancement” signal processing. However, Eddins and Hayes (2007) reported these techniques provide minimal benefit, as measured by subjective preference in paired-comparison procedures. Little significant improvement in speech intelligibility has been demonstrated to date. In this section, several studies of speech enhancement particularly spectral enhancement of speech are reviewed.

2.2.1 Spectral enhancement of speech

The approaches of spectral enhancement can be divided into two categories based on acoustic and physiological considerations. The algorithm of spectral enhancement is dependent on acoustic features of speech, and/or physiological findings on the internal representation of the speech spectrum for NH and HI listeners.

2.2.1.1 Acoustically-based approach

Two solutions from the acoustic approach attempt to restore smeared spectral resolution by reducing formant bandwidth (i.e. spectral sharpening) (Summerfield, 1985) or increasing the spectral contrast by manipulating weighted spectral filters (i.e. spectral

expansion) (Bunnell, 1990; Stone & Moore, 1992; Baer et al., 1993; Franck et al., 1999; Lyzenga et al., 2002; Yang et al., 2003; Chen et al., 2011).

Using the spectral sharpening approach, Summerfield (1985) narrowed formant bandwidths in the synthesis of CVC (consonant-vowel-consonant) syllables to sharpen the signal spectrum and increase amplitude of formant peaks in order to reach greater peak-to-valley contrasts. However, consonant identification scores for listeners with normal hearing or hearing loss did not improve with the peak-to-valley contrast enhancement. This technique was easy to be implemented on synthesized stimuli. But the results indicated that the spectral sharpening approach may not produce selective improvements for listeners to correctly identify phonemes by narrowing formant bandwidths.

Using a spectral expansion approach, Bunnell (1990) applied an adaptive filter to enhance spectral contrast. The spectral envelope was estimated over consecutive and overlapping (50%) 25.6-ms speech segments, followed by a weighted inverse filter. Listeners with sensorineural hearing loss achieved approximately 5% increase in stop-consonant identification in quiet. Using a slightly different approach, in order to produce a normal excitation pattern in an impaired ear, Simpson et al. (1990) implemented an adaptive filter bank designed to model the filtering properties and internal excitation patterns at the human auditory periphery. The digital signal processing of Simpson et al. (1990) involved windowing short-term segments of the stimulus, Fourier transformation, smoothing with a function analogous to an excitation pattern, enhancement by convolution with a difference of Gaussian (DoG) function, and reconstruction via an inverse Fast Fourier transform (FFT) procedure. DoG enhancement resulted in amplification of the peaks and attenuation of adjacent valleys of the excitation pattern. As a result, sentence intelligibility in long-term speech shaped noise (LTSSN) for listeners with moderate to severe hearing loss showed small but significant improvements.

An analog version of the Simpson and colleagues (1990) method was implemented by Stone and Moore (1992) to speed up signal processing. They used a 16-channel analog filter bank on a logarithmic audio-frequency scale with bandwidths approximating those of auditory filter estimates and integration windows from 9.5 to 30 ms such that channels with local maxima were weighted positively and neighboring channels were weighted negatively. This procedure effectively enhanced the peak-to-valley contrast; however, there was no significant improvement in speech recognition compared to the baseline condition.

Following a systemic exploration of parameters in the Simpson and colleagues (1990) algorithm, Baer et al. (1993) reported a small (1.8%) but significant improvement in sentence recognition with enhancement in listeners with hearing loss. Based on Baer and colleagues method, Yang et al. (2002) developed an algorithm by reconsidering the calculation of enhancement magnitude spectrum and excitation pattern to make it computationally efficient. But, there was no formal testing of speech recognition. In order to improve overall speech quality, Rout (2006) employed a similar approach and used a broader range of frequencies for spectral enhancement up to 8000 Hz instead of 5000 Hz used in previous studies (Bunnell, 1990; Simpson et al, 1990). However, listeners preferred the sound quality without speech enhancement modifications, and sentence recognition scores for hearing-impaired listeners did not significantly improve as a result of these modifications. Extending the same approach with single and multi-channel amplitude compression, Frank et al. (1999) reported significant improvement in vowel identification but not consonant identification in LTSSN. Similarly, Lyzenga et al. (2002) found no effect of spectral expansion alone, but when combined with spectral shaping (high-frequency amplification to prevent upward spread of masking), sentence reception thresholds in LTSSN improved by about 1.0 dB in NH listeners.

Instead of adding amplitude compression and high-frequency amplification, Chen, Baer and Moore (2011) added a spectral-change scheme in which spectral changes occur across frames over time, following previous studies that targeted on static spectral enhancement (Simpson et al., 1990; Baer et al., 1993; Baer & Moore, 1997). Findings indicated that this algorithm produced some benefits (8%) for listeners with sensorineural hearing loss, but only in the steady noise at SNR of -6 dB and no improvements or even degraded performance in steady noise with a higher SNR or in two-talker babble.

2.2.1.2 Physiologically-based measure

In addition to the enhancement algorithms based on acoustic spectrum as described above, a more biologically-based approach to enhance spectral contrast in the higher frequency range, called Contrast Enhanced Frequency Shaping (CEFS), was developed on the base of the representation of spectral contrasts in auditory neurons. Because of the cochlea damage, the formant peaks like F1, F2, and F3 are spectrally smeared. Moreover, the levels of F2 and F3 are usually suppressed due to the upward spread of masking from F1. To address this smeared representation, CEFS method aims to amplify the F2 and F3 region of the target vowels without amplifying the harmonics in the trough between F1 and F2. Miller et al. (1999) recorded neural responses to spectrally enhanced synthetic vowels in cats with noise-induced hearing loss by using the CEFS-modified stimuli, thereby restoring normal neural representation of F1 and F2, as well as high-pass filter resolution, ensuring that high-frequency components of stimulus were spectrally enhanced. They observed that application of the CEFS method resulted in much better representation of the F1 and F2 and suppressed activities of the neural fibers corresponding to the spectral valleys between F1 and F2, and between F2 and F3. The CEFS scheme, where spectral valleys are reduced and high frequencies are enhanced, might therefore improve speech

intelligibility. Extending the same CEFS method with multichannel compression, Bruce (2004) used auditory computational models of normal and impaired ears and reported no distortion of neural representation of formants in auditory neurons. However, there have not been further studies yet on the benefits of CEFS amplification on speech quality and intelligibility for human listeners with hearing impairment.

In summary, the perceptual effects of spectral enhancement range widely from small improvements to no improvement or even deterioration in speech tasks. For example, small (3-5%) or no improvement has been observed in vowel and consonant identification in noise (Summerfield, 1985; Bunnell, 1990; Franck et al., 1999). Some research has showed that there was no improvement in speech recognition in noise (Stone and Moore, 1992; Baer et al., 1993), while others reported that speech recognition thresholds (SRT) for sentence recognition in LTSSN were improved by 1 dB (Lyzenga et al., 2002). These results suggest that the effects of speech enhancement depend on many factors: audibility, spectral and temporal resolution, cognitive function, and speech enhancement algorithms (Summerfield, 1985; Bunnell, 1990; Franck et al., 1999; Stone & Moore, 1992; Baer et al., 1993; Lyzenga et al., 2002).

2.2.2 Speech enhancement of this study

Recently, a straightforward acoustic approach has been implemented in our laboratory to enhance the F2 formant peak in order to examine whether such F2 enhancement could improve vowel perception (Woodall & Liu, 2013). F2 enhancement was selected because it is a relatively simple signal processing method, directly enhancing the F2 region without disrupting other frequency ranges. This enhancement procedure includes several sequential steps as follows: FFT (Fast Fourier transform) analysis to render a 3-dimensional spectrogram (amplitude $\times$ time $\times$ frequency); F2 estimator; F2

enhancement; and inverse FFT to reconstruct speech signals. Using this method, Woodall and Liu (2013) enhanced the F2 peak of vowel sounds by 3, 6, and 9 dB. Thresholds of vowel formant discrimination, i.e., the smallest formant shift that can be detected, were measured for listeners with normal hearing and hearing loss to investigate the effect of spectral enhancement at F2. Vowel sounds were presented at two levels: 70 and 90 dB SPL with and without F2 enhancement. Results showed that listeners with normal hearing did not gain any benefits from either increasing speech level or F2 enhancement on vowel format discrimination. On the other hand, for HI listeners, simply increasing speech level was not an effective solution. Instead, F2 enhancement with an increased speech level provided the greatest benefit to HI listeners.

F2 enhancement was employed in this study because it has been demonstrated that F2 enhancement significantly compensates for spectral smearing of formant peaks and improves vowel perception for HI listeners (Woodall & Liu, 2013). Considering that vowels play an important role in sentence recognition for listeners with normal hearing and hearing loss (Fogerty, Humes, & Kewley-Port, 2012; Fogerty & Kewley-Port, 2009; Kewley-Port et al., 2007), it is reasonable to expect that F2 enhancement would also benefit older listeners to have better scores on sentence recognition in noise.

As explained above, many techniques of speech enhancements have not provided significant benefits in speech perception in noise for listeners with hearing loss (Stone and Moore 1992; Baer et al., 1993), possibly due to the fact that the speech enhancement technique was applied to both speech signals and noise simultaneously, simulating hearing-aid processing. When signal-to-noise ratios are relatively low, spectral enhancement may primarily strengthen the noise spectrum rather than the signal spectrum, making it difficult for listeners to benefit from the signal enhancement itself. In this study, only speech signals are enhanced, while background noise is unchanged, simulating a mobile-phone listening

scenario. In a typical mobile-phone listening condition, listeners need to focus on speech sounds from the phone in the background noise that is usually unchangeable. Although increasing the speech volume is a simple solution, it has several limitations: 1. in very challenging conditions (e.g., high-intensity noise), even the highest speech volume may not reach a sufficient SNR; 2. for older listeners, particularly with hearing impairment, increasing speech levels may cause loudness recruitment (Buus & Florentine, 2002). Thus, this study aims to investigate whether F2 enhancement may improve speech perception in noise for older listeners in the mobile-phone listening condition, i.e., enhanced speech presented in background noise.

2.2.3 Evaluating the effectiveness of speech enhancement

In order to evaluate the effectiveness of F2 enhancement on speech perception in noise in this study, speech recognition tasks were presented in two types of background noise: two-talker and six-talker babble. The babble noise was selected because it is common and challenging noise interference in daily life and hearing-impaired listeners suffer significantly from multi-talker babble noise (Dubno et al., 1984; Souza & Turner, 1994; Bacon et al., 1998). Previous studies have shown that different numbers of talkers that are presented in the background result in different masking effectiveness of speech perception (Freyman et al., 2004; Simpson & Cooke, 2005; Rosen et al., 2013). For example, listeners' consonant identification performance was worst for eight-talker babble. As the number of masking talkers decreased from eight to six to four to two, performance generally improved (Simpson & Cooke, 2005). However, such a decreasing trend is not monotonically when the number of masking talkers is large enough and cannot make further effect on listeners' performance (Simpson & Cooke, 2005; Rosen et al., 2013). Compared to two-talker babble, the temporal structure of six-talker babble is denser and

smoother, i.e., less temporal variations, while the two babbles also differed in informational masking. Thus, the interference of the two babbles is different due to their difference in acoustic and linguistic features. The effectiveness of F2 enhancement on listeners' performance is then expected to be different with the presence of the two babbles.

On the other hand, in order to investigate whether F2 enhancement could improve speech perception over different languages, speech recognition in two languages (i.e., American English and Mandarin Chinese) were examined for English-native speakers and Chinese-native speakers, respectively. Most research on speech enhancement algorithms have focused on English speech only (Bunnell, 1990; Stone & Moore, 1992; Baer et al., 1993; Franck et al., 1999; Lyzenga et al., 2002; Yang et al., 2003; Chen et al., 2011). In this study, Mandarin Chinese was selected, because Mandarin Chinese differs significantly from English at a variety of levels such as acoustic, phonetic, phonological, and linguistic level. In addition, Chinese is a tone language, while English is not. Previous research indicated that English speech has more temporal modulations than Chinese speech (Calandruccio et al., 2010). As studies have focused on the importance of Mandarin tones in Mandarin speech recognition (Chen et al., 2014), this study was to examine whether the enhancement in the spectral domain could improve Mandarin speech perception in noise. Given that Mandarin tones play a limited role in Mandarin sentence recognition (Chen et al., 2014) implying the importance of phonetic cues, it is reasonable to predict that spectral enhancement on Mandarin speech could benefit Mandarin-native speakers' speech perception in noise. If F2 enhancement is beneficial across different languages, the development and implementation of such an algorithm has the potential to positively impact millions of assistive hearing device users.

2.3 Goals of this study

This study aims to investigate whether F2 enhancement improves speech recognition for older listeners with normal hearing or hearing loss. In addition, this study explored whether such benefits of F2 enhancement could be extended to other languages such as Mandarin Chinese. Three specific aims were tested in this dissertation. (1) The effect of F2 enhancement on sentence recognition in two-talker and six-talker babbles were measured for older listeners with normal hearing and hearing loss. It was hypothesized that by selectively amplifying F2, speech recognition performance would improve across listening conditions and also across listeners, regardless of their hearing status (normal hearing or hearing impaired), and regardless of their language background (English or Mandarin Chinese). (2) Speech recognition performance and the improvement of speech recognition by F2 enhancement were expected to associate with listeners' hearing thresholds and cognitive memory abilities. (3) The speech intelligibility index (SII) model was examined and expected to interpret the perceptual benefits from F2 enhancement across listeners.

Chapter 3: Methods

3.1 Participants

30 English-native speakers and 30 Chinese-native speakers were recruited from Austin, Texas, USA, and Shanghai, China, respectively. Listeners with normal hearing had pure-tone thresholds ≤ 25 dB HL at octave intervals between 250 and 4000 Hz and pure-tone threshold ≤ 35 dB HL at 8000Hz. Listeners with hearing loss had hearing levels ranging from mild to moderately severe hearing loss with no more than 65 dB HL in the range of 250 - 4000 Hz. Specifically, English group included 12 listeners with normal hearing (2 male; 10 female) and 18 listeners with sensorineural hearing loss (11 male; 7 female) with the average age of 64.8 years old ($SD = 8.4$). In contrast, in the Chinese group, there were 9 listeners with normal hearing (1 male; 8 female) and 21 listeners with sensorineural hearing loss (8 male, 13 female) with the average age of 62.6 years old ($SD = 3.9$). Behavioral audiometry (air and bone conduction) using a GSI-61 audiometer were completed on all participants. Air conduction hearing thresholds were obtained by using Telephonics TDH-50P supra-aural headphones. Right ears were usually selected as the target ear of the speech perception experiments for most listeners (EN: 25/30; CN: 24/30), while left ears were used (EN: 5/30; CN: 6/30) only when there was a conductive hearing loss or history of a surgery on the right ear. Mean audiograms for the test ears were plotted in Figure 4 for English and Chinese groups with NH and HI.

Figure 4. Audiograms for English and Chinese group with normal hearing and hearing loss.

Questionnaires about the history of hearing loss were administered by all participants. Background information revealed that main causes of hearing loss were aging and/or noise exposure. Only two participants in English group had around 3-year hearing aid experience, while all other participants in the two groups had no hearing aid experience. Consent forms were provided to each participant. All procedures were approved by the Institutional Review Board of the University of Texas at Austin.

3.2 Speech stimuli and noise

Sentences from the coordinate response measure (CRM) corpus with the form “ready [call sign] go to [color] [number] now.” served as the target speech signals. For example, “ready **Charlie** go to **blue three** now”. CRM was a measure of speech intelligibility relevant to military environments that was developed by Moore (1981). The

CRM sentences consisted of three key words: call sign, color, and number. In this study, there were English and Chinese versions of CRM corpus recorded by an English-native male speaker and a Chinese-native male speaker, respectively. Four call signs were Baron, Charlie, Eagle, Hopper for the English version, and were Zhang Jun, Li Ting, Wang Li, Zhao Hua for the Chinese version. Four colors, blue, red, green, and white, were assigned for both English and Chinese versions. The numbers range from 1 to 8 except the number 7 that has a bisyllabic structure, different from the other numbers with a monosyllabic structure in English. Thus, there were a total of 112 combinations of four call signs, four colors, and seven numbers for each language. Correct target word identification was based on the identification of the color-number combination in the present study.

The original sentences from CRM corpus served as the base stimuli for spectral enhancement. There was another set of speech stimuli with F2 enhancement by 9 dB. As shown in Figure 5, the spectral enhancement of F2 was manipulated for the CRM speech materials.

Figure 5. Spectra of original speech (solid line) and the speech with 9-dB F2 enhancement (dashed line).

First, a three-dimension spectrogram (amplitude $\times$ time $\times$ frequency) was created in a MATLAB program for each sentence with a sampling rate of 44,100 Hz, 10-ms windows, and a 50% overlap between the short windows. Second, at each time frame of the spectrogram (e.g., a spectrum of the 10-ms window), the peaks of F1, F2, and F3, and valleys of V1, V2 and V3 were identified according to the range of F1, F2, and F3 of English and Chinese vowels (e.g., 800-2500 Hz for F2). Third, the level of F2 peak was amplified by 9 dB with the surrounding components proportionally enhanced at each time frame. If no F2 peak was located (e.g., consonant segments), no amplification was applied in that spectrum. Finally, a new sentence with F2 enhancement by 9 dB was generated by using inverse FFT.

Two types of noise were selected as the masker for each language group: two- and six-talker English babble; and two- and six-talker Chinese babble, respectively. Multi-talker babble was generated by recoding two (one male and one female) or six (three males and three females) adults native talkers reading a section in a child's book of English and Chinese versions, and then being mixed. The contents were chose from the “Technology” section of the English and Chinese versions of the book “The New Children's Encyclopedia” (Lock, 2009). No enhancement processing was conducted to the multi-talker babble.

3.3 Stimulus presentation

For each trail, the target CRM sentence was played temporally in the middle of four-second masker that was randomly selected from a 30-second multi-talker babble. The overall level of babble was fixed at 70 dB SPL.

Speech and noise were presented at a sampling rate of 24,414 Hz to the right ears of listeners via SONY MDR-7506 headphones for both language groups. Speech signals

and multi-talker babble were resampled from 44,100 to 24,414 Hz to be compatible with the hardware. The English participants were seated in a sound-treated booth in the Speech Psychophysics Laboratory at the University of Texas at Austin (UT). Stimulus presentation was controlled by TDT (Tucker-Davis Technologies, Alachua, Florida) modules, including a two-channel 24-bit real-time processor (RP2.1), programmable attenuator (PA5), signal mixer (SM5) and a headphone buffer (TDT HB7). The data for the Chinese listeners were collected in a quiet test room in the Psycholinguistic Laboratory at Shanghai Jiao Tong University (SJTU). The stimulus presentation was controlled by the TDT mobile processor (RM1).

Sound pressure levels of speech and noise were measured and calibrated in an AEC201-A IEC 60318-1 ear simulator by a Larson-Davis sound-level meter (Model No. 2800) with a linear weighting band.

3.4 Procedure

Listeners were seated in front of an LCD monitor, which displayed 28 response alternatives ($4 \text{ colors} \times 7 \text{ numbers}$) as text boxes corresponding with each color-number combination. The 28 responses were shown by a layout of four colors (rows) $\times$ seven numbers (columns). For each trial, listeners were asked to respond by clicking the button corresponding with their response choice as soon as possible after each stimulus presentation.

By using the adaptive up-down tracking procedure, the thresholds of speech recognition were obtained based on three tracking algorithms: a one-down, two-up; a one-down, one-up; and a two-down, one-up adaptive rules, estimating points on the psychometric function corresponding to 29.3%, 50%, and 70.7 % correct, respectively, as shown in Figure 6 (Levitt, 1971). The three percentage points represented three different

challenge levels (T-29.3%: challenge level I, most difficult level; T-50%: challenge level II; T-70.7%: challenge level III, easiest level) to cover the dynamic range of the psychometric function. For each block, the threshold was computed as the average over the last even number of reversals with the first three reversals excluded. The threshold for each condition was obtained by averaging thresholds for two blocks of 56 trials (i.e., each target stimulus is presented twice). The third block was completed if the thresholds for the first two blocks differed by more than 3 dB. Overall, the order of the twelve experimental conditions (3 up-down procedures $\times$ 2 masker types $\times$ 2 enhancement conditions) were randomized for each listener. All procedures were implemented by the software Sykofizx®.

Figure 6. Psychometric function showing percent responses as a function of three different up-down tracking procedures (i.e., challenge level I; II; and III).

Cognitive tests of working memory, i.e., the digit span subtests of Wechsler Adult Intelligence Scale-III (WAIS-III, Wechsler, 1997) were conducted on each participant. It consisted of three sections: Forward Digit Span, Backward Digit Span, and Letter-Number Sequencing subtests. English and Chinese version of these three memory subtests were recorded by a female native-English, and a female native-Chinese speaker, respectively. During the testing, recording materials were presented at a comfortable level through a laptop. For each part, a series of items were played at the rate of one per second. Participants were asked to repeat the numbers in the order by which the items were presented (i.e. Forward Digit Span subtest) or in the reverse order (i.e. Backward Digit Span subtest) right after each audio presentation. As for the Letter-Number Sequencing subtest, participants needed to internally manipulate a series of random letters and numbers that were mixed up, and orally reported the numbers in the ascending order first, followed by the letters in the alphabetical order. For instance, when audio presentation was “8-D-6-G-1”, participants were expected to report correctly as “1-6-8-D-G”. There were two trials at each series length for Forward and Backward Digit Span subtests, while three trials were included for each series length of the Letter-Number Sequencing subtest. The test continued from the short to long series length until all trials at the series length were failed. According to the regression technique (Jenkins et al., 2000; Emery et al., 2007), the memory span was estimated by the series length at which the probability of answering correctly was 0.5. Firstly, a linear regression function was created based on the probability of the longest series length where the participant could recall all trials correctly ($p = 1.0$), and the probability of the shortest series length at which all trials were answered incorrectly by the participant ($p = 0.0$). Then, individual’s memory span was determined from the regression equation when $p = 0.5$. All three memory subtests lasted around 20 minutes.

Chapter 4: Results

4.1 Demographic factor

In the English group, there were 12 NH listeners with the average age of 60.4 years old ($SD = 5.4$), and 18 listeners with hearing loss at the average age of 67.7 years old ($SD = 8.9$). An independent-sample t-test showed that there was a significant group difference in age ($t = -2.804$, $df = 27.859$, $p < 0.05$) for the English group. Compared to English NH group, English HI listeners were significantly older than English NH listeners.

Nine NH listeners with a mean age of 69.9 years old ($SD = 2.4$) and twenty-two HI listeners with a mean age of 63.3 years old ($SD = 4.2$) were in the Chinese group. The t-test demonstrated that age did not significantly differ between the Chinese NH and HI groups ($t = -1.977$, $df = 24.928$, $p > 0.05$).

4.2 Memory span scores

4.2.1 English group

Scores for three subtests of memory span were illustrated in Figure 7 for the English NH and HI group. The average of Forward Digit Span, Backward Digit Span, and Letter-Number Sequencing were 6.8 ($SD = 1.3$), 5.2 ($SD = 1.0$), and 4.6 ($SD = 0.8$) for English NH listeners, and 6.6 ($SD = 1.2$), 5.0 ($SD = 1.1$), and 4.5 ($SD = 0.8$) for English HI listeners, respectively. A two-factor (between-subjects factor: listener group; within-subjects factor: type of memory subtests) analysis of variance (ANOVA) was conducted with memory span as the dependent variable for English listeners. Only the main effect of type of memory subtests was significant ($F_{2, 54} = 123.756$, $p < 0.05$), while the main effect of listener group and two-factor interaction effect were not significant (both $p > 0.05$). It suggested that English NH and HI groups had comparable memory spans in all three subtests (all $p >$

0.05). Further comparisons indicated that the average span of Forward Digit was significantly longer than either the span of Backward Digit or the span of Letter-Number Sequencing, while Backward Digit span was significantly longer than the span of Letter-Number Sequencing for both English NH and HI listeners (all $p < 0.05$).

Figure 7. Memory spans of Forward Digit Span, Backward Digit Span, and Letter-Number Sequencing subtests for English NH and HI group

For English listeners, there were significant correlations among the scores of these three memory subtests: r (forward, backward) = 0.800, $p < 0.05$; r (forward, sequencing) = 0.826, $p < 0.05$; and r (backward, sequencing) = 0.748, $p < 0.05$, indicating that listeners with longer forward spans usually had longer backward spans and longer spans of letter-number sequencing, and vice versa. In addition, there were no significant correlations between age and scores of any of the three memory span tests (all $p > 0.05$, see Figure 8).

Figure 8. Scatter plots showing the correlations between age and memory spans of three subtests (Forward Digit Span, Backward Digit Span, and Letter-Number Sequencing) for English NH listeners (left) and English HI listeners (right).

4.2.2 Chinese group

The average spans of Forward Digit (NH: 6.4, $SD = 0.5$; HI: 6.9, $SD = 1.5$), Backward Digit (NH: 4.6, $SD = 0.9$; HI: 4.6, $SD = 0.7$), and Letter-Number Sequencing (NH: 3.2, $SD = 0.8$; HI: 3.3, $SD = 0.7$) are shown in Figure 9. A two-factor (between-subjects factor: listener group; within-subjects factor: type of memory subtests) analysis of variance (ANOVA) was conducted with the scores of memory span as the dependent variable for Chinese listeners. Results suggested that memory span was only significantly affected by the main effect of type of memory subtests ($F_{2, 42} = 74.634, p < 0.05$). The main effect of listener group and interaction effect were not significant (both $p > 0.05$). For both Chinese NH and HI groups, the span of Forward Digit was found to be longest, followed by the span of Backward Digit, and the span of Letter-Number Sequencing was shortest among the three subtests (all $p < 0.05$).

Figure 9. Memory spans of Forward Digit Span, Backward Digit Span, and Letter-Number Sequencing subtests for Chinese NH and HI group

The correlation for the Chinese listeners between Forward Digit Span and Backward Digit Span was significant ($r = 0.516, p < 0.05$). However, the scores of Letter-Number Sequencing subtest did not significantly correlate with either Forward Digit Span or Backward Digit Span (both $p > 0.05$). In addition, the age of the Chinese NH and HI listeners did not show a significant correlation with their scores of the three memory span tests (all $p > 0.05$, see Figure 10).

Figure 10. Scatter plots showing the correlations between age and memory spans of three subtests (Forward Digit Span, Backward Digit Span, and Letter-Number Sequencing) for Chinese NH listeners (left) and Chinese HI listeners (right).

4.3 Speech recognition in noise with and without F2 enhancement

4.3.1 English group

The speech recognition thresholds (SRT) for the English NH and HI groups were plotted as a function of listening challenge levels (see Figure 11). A four-factor (between-subjects factor: listener group; within-subjects factor: type of noise $\times$ challenge level $\times$ F2 enhancement) analysis of variance (ANOVA) was conducted with the SRT as the dependent variable. There were significant main effects of challenge level ($F_{2, 56} = 134.117$, $p < 0.05$), and F2 enhancement ($F_{1, 28} = 11.799$, $p < 0.05$); a significant interaction between challenge level and type of noise ($F_{2, 56} = 39.944$, $p < 0.05$), and a significant interaction between F2 enhancement and type of noise ($F_{1, 28} = 12.079$, $p < 0.05$). The main effect of listener group was marginally significant ($F_{1, 28} = 4.092$, $p = 0.053$), while all other main effect, and the multi-factor interaction effects were not significant (all $p > 0.05$).

Figure 11. SRTs for English NH and HI listeners in two-talker and six-talker babble at three challenges levels without and with F2 enhancement

Because of the significant interaction effects of type of noise with challenge level; and F2 enhancement, three-factor (between-subjects factor: listener group; within-subjects factor: challenge level $\times$ F2 enhancement) ANOVAs with SRT as the dependent variable were conducted to examine the simple main effect of F2 enhancement under two-talker, and six-talker babble, respectively. In the two-talker babble, the main effects of challenge level ($F_{2,56} = 101.551, p < 0.05$), and F2 enhancement ($F_{1,28} = 16.224, p < 0.05$) were significant, while the main effect of listener group was marginally significant ($F_{1,28} = 4.092, p = 0.053$). All other two-, and three-factor interaction effects were not significant (all $p > 0.05$). Tukey post hoc tests suggested challenge level III (T-70.7%: SRT of - 2.0 dB) had the significant higher SRT than challenge level II (T-50%: SRT of -6.5 dB), and challenge

level I (T-29.3% : SRT of -11.3 dB), while challenge level II had a significant higher SRT than challenge level I. Additionally, Tukey post hoc tests also indicated that F2 enhancement significantly improved the SRT by 1.0 dB (e.g., SRT at -6.1dB without F2 enhancement vs. -7.1 dB with F2 enhancement). Moreover, English HI group showed a marginally significantly higher SRT by 2.5 dB in comparison with English NH group.

In the six-talker babble, three-factor ANOVAs showed that only the main effect of challenge levels ($F_{2,56} = 101.551, p < 0.05$) was significant, whereas the other main effects, and multi-factor interaction effects were not significant (all $p > 0.05$). The following Tukey post hoc tests indicated that the challenge level III (SRT of -3.8 dB) had the highest SRT, followed by challenge level II (SRT of -6.1 dB), while challenge level I (SRT of -8.4 dB) had the lowest SRT.

Due to the significant interaction effect of the challenge level and type of noise, to examine the simple main effect of the type of noise (two-talker and six-talker babble), another three-factor (between-subjects factor: listener group; within-subjects factor: type of noise $\times$ F2 enhancement) ANOVAs with SRT as the dependent variable were conducted at each challenge level. At the challenge level I, results suggested only the main effect of type of noise ($F_{1,28} = 26.019, p < 0.05$), and F2 enhancement ($F_{1,28} = 10.341, p < 0.05$) were significant. The two-talker babble noise (SRT of -11.3 dB) had significantly lower SRTs compared with six-talker babble noise (SRT of -8.4 dB) ($p < 0.05$). Furthermore, the SRT was significantly improved from -9.4 dB to -10.3 dB after F2 enhancement ($p < 0.05$).

At the challenge level II, the three-factor ANOVA showed that only the main effect of F2 enhancement ($F_{1,28} = 7.640, p < 0.05$), and the interaction effect between type of noise and F2 enhancement ($F_{1,28} = 7.188, p < 0.05$) were significant. Post-hoc pair wise comparisons by using a Bonferroni adjusted critical alpha level suggested that F2

enhancement significantly lowered SRTs only in two-talker babble ($p < 0.05$), but not in six-talker babble ($p > 0.05$).

At the challenge level III, the three-factor ANOVA showed that the main effect of types of noise ($F_{1,28} = 15.624, p < 0.05$), and the interaction effect between type of noise and F2 enhancement ($F_{1,28} = 7.323, p < 0.05$) were significant. Simple effects analysis with Bonferroni adjustment suggested that the thresholds in two-talker babble were significantly higher than those in six-talker babble. Moreover, there was a significant difference between F2 enhancement and original speech in two-talker babble ($p < 0.05$), but not in six-talker babble ($p > 0.05$).

4.3.2 Chinese group

As shown in Figure 12, SRTs across Chinese NH (top) and HI (bottom) listeners were presented as a function of challenge levels for two enhancement conditions and two types of noise. A four factor (between-subjects factor: listener group; within-subjects factor: challenge level $\times$ type of noise $\times$ F2 enhancement) ANOVA of the thresholds was conducted. Thresholds were significantly affected by challenge level ($F_{2,56} = 334.68, p < 0.05$); type of noise ($F_{1,28} = 140.70, p < 0.05$); and F2 enhancement ($F_{1,28} = 8.558, p < 0.05$). Of the six two-factor interactions, only one was significant: challenge level $\times$ type of noise ($F_{2,56} = 12.51, p < 0.05$). All the four three-factor interactions were not significant (all $p > 0.05$). But, the four-factor interaction (challenge level $\times$ type of noise $\times$ F2 enhancement $\times$ listener group) was significant ($F_{2,56} = 5.05, p < 0.05$).

Figure 12. SRTs for Chinese NH and HI listeners in two-talker and six-talker babble at three challenges levels without and with F2 enhancement

Because the two types of noise differed significantly in overall performance, and the interaction effect of type of noise, challenge level, F2 enhancement, and listener group was significant, separate analyses were conducted for the two-talker and six-talker babble. In the two-talker babble, a three-factor (between-subjects factor: listener group; within-subjects factor: challenge level $\times$ F2 enhancement) repeated measures ANOVA with SRT

as the dependent variable showed that thresholds of two-talker babble were significantly affected by challenge level ($F_{2, 56} = 171.763, p < 0.05$) ; F2 enhancement($F_{1, 28} = 7.702, p < 0.05$), and marginally significantly affected by listener group ($F_{1, 28} = 4.122, p = 0.052$). None of the two-way, and three-way interactions for two-talker babble were significant (all $p > 0.05$). Thresholds for F2 enhancement were significantly better than thresholds for original speech in two-talker babble by 0.5 dB ($p < 0.05$). Additionally, the SRT of Chinese NH listeners was lower than the SRT of English HI listeners by 2.3 dB in the two-talker babble with a marginal significance.

In the six-talker babble, the three-factor ANOVA results suggested that the only main effect of challenge levels was significant ($F_{2, 56} = 195.589, p < 0.05$), while other main effects, and multi-factor interaction effects were not significant (all $p > 0.05$). Tukey post hoc tests indicated the SRT for challenge level III (SRT of -4.4 dB) were significantly higher than the SRT for both challenge level II (SRT of -7.1 dB), and challenge level I (SRT of -9.9 dB), while thresholds for challenge level II were significantly higher than thresholds for challenge level I (all $p < 0.05$).

Similarly, in order to measure the simple main effect of type of noise, three-factor (between-subjects factor: listener group; within-subjects factor: type of noise $\times$ F2 enhancement) repeated measures ANOVAs for each of the three challenge levels were conducted, respectively. At the challenge level I, results showed that only the main effect of type of noise ($F_{2, 56} = 107.152, p < 0.05$), and interaction effect of type of noise, F2 enhancement, and listener group ($F_{1, 28} = 6.728, p < 0.05$) were significant. Post hoc analysis for multiple comparisons where p values were adjusted using a Bonferroni correction, suggested that in the two-talker babble, SRTs significantly improved for F2 enhancement in relative to SRTs of original speech only for Chinese HI listeners, not for Chinese NH listeners. Overall, Chinese HI listeners had significantly higher SRT by 3.2

dB than Chinese NH listeners in SRTs of original speech ($p < 0.05$), but there was no group difference in SRTs with F2 enhancement ($p > 0.05$). In addition, no significant effect of listener group and F2 enhancement was found in the six-talker babble (both $p > 0.05$). Overall, two-talker babble had significantly lower SRT than six-talker babble ($p < 0.05$).

At the challenge level II, the three-factor ANOVA showed significant main effects of listener group ($F_{1,28} = 4.511, p < 0.05$); and type of noise ($F_{1,28} = 99.978, p < 0.05$), but no significant main effect of F2 enhancement ($p > 0.05$). Of the two-factor and three-factor interaction effects, only the interaction effect of type of noise and F2 enhancement was significant ($F_{1,28} = 5.367, p < 0.05$). Results indicated the Chinese HI listeners had significantly higher SRTs by 1.9 dB than Chinese NH listeners ($p < 0.05$). In addition, a significant decrease of thresholds with F2 enhancement occurred only in two-talker babble at $p < 0.05$ adjusted by Bonferroni, but not in six-talker babble ($p > 0.05$). It should also be noted that thresholds for two-talker babble were significantly lower than six-talker babble across original and F2 enhancement conditions ($p < 0.05$).

At the challenge level III, the three-factor ANOVA suggested that SRTs were significantly affected by the main effect of type of noise ($F_{1,28} = 31.464, p < 0.05$); and F2 enhancement ($F_{1,28} = 6.619, p < 0.05$). The average SRTs significantly improved by 0.6 dB for F2 enhancement compared with the SRTs of original speech ($p < 0.05$). Moreover, two-talker babble had significant lower SRT by 2.3 dB than six-talker babble ($p < 0.05$).

In summary, English and Chinese listeners showed similar data patterns in SRTs across listening conditions. For example, generally, both language groups had significant lower SRTs in two-talker babble than in six-talker babble in most conditions. Particularly, in two-talker babble, SRTs significantly improved with F2 enhancement for both English and Chinese group. However, no significant changes in SRTs were observed before and after F2 enhancement in six-talker babble for both English and Chinese listeners. In

addition, in two-talker babble, either English or Chinese NH listeners had marginally significantly lower SRTs than their counterparts with hearing loss.

4.4 Benefits from speech enhancement

Benefits from F2 enhancements were calculated by subtracting SRTs for original speech (e.g., no F2 enhancement) from SRTs with F2 enhancement. The three-factor (between-subjects factor: listener group; within-subjects factor: challenge level $\times$ type of noise) ANOVAs were conducted with the benefits from F2 enhancement as the dependent variable for English and Chinese group, respectively.

4.4.1 English group

The results revealed that the amount of benefits from F2 enhancement was only significantly affected by type of noise ($F_{1,28} = 12.079, p < 0.05$), but not other main effects or two-way, and three-way interaction effects. Furthermore, English listeners gained more benefits from F2 enhancement in two-talker babble than in six-talker babble across different challenge levels, as shown in Figure 13.

Figure 13. Benefits from F2 enhancement for English NH and HI listeners in two-talker and six-talker babble at three challenge levels.

4.4.2 Chinese group

Similarly, the three-factor repeated measures ANOVA revealed that there was only a significant interaction effect of challenge level, type of noise, and listener group ($F_{2, 56} = 5.046, p < 0.05$), while there was no significant main effect of any of the three factors and two-factor interactions (all $p > 0.05$). In order to measure the simple main effect of listener group, firstly, two-factor (between-subjects factor: listener group; within-subjects factor: types of noise) ANOVAs were conducted at each challenge level using a Bonferroni adjusted critical value. Further analysis suggested that at challenge level I, in the two-talker babble, Chinese HI listeners had significant larger benefits from F2 enhancement by 1.8 dB than Chinese NH listeners (see Figure 14). In the six-talker babble, no significant listener group difference or interaction effects were found. At the challenge level II, results revealed that the amount of benefits from F2 enhancement was significantly higher in the

two-talker babble compared to the six-talker babble. Moreover, there were no significant effect by listener group or type of noise observed at the challenge level III (all $p > 0.05$).

Figure 14. Benefits from F2 enhancement for Chinese NH and HI listeners in two-talker and six-talker babble at three challenge levels.

4.4.3 English vs. Chinese group

In order to compare F2-enhancement benefits between English and Chinese listeners, especially in two-talker babble, F2-enhancement benefits for English and Chinese listeners and English and Chinese HI group were plotted in Figure 15. Independent T-tests on F2-enhancement benefits in two-talker babble across three challenge levels between English and Chinese NH listeners, English and Chinese HI listeners, or overall English and Chinese listeners indicated that all above group differences were not significant (all $p > 0.05$).

Figure 15. Average benefits from F2 enhancement in two-talker babble for English and Chinese group across hearing status (“Total”), and for English and Chinese group with normal hearing (“NH”) and hearing loss (“HI”).

4.5 Correlation of age, hearing loss, memory span with SRTs and F2-enhancement benefits

In order to explore how listeners’ factors (e.g. age, hearing loss, and memory span) contribute to SRTs with/without F2 enhancement, and the amount of benefits from F2 enhancement that listeners received, analyses of correlation and regression analyses were conducted for English NH and HI subgroups, and Chinese NH and HI subgroups, separately.

4.5.1 English group

4.5.1.1 Age vs. SRTs, and Age vs. F2-enhancement benefits

Results revealed that there was no significant correlations between age and SRTs for original speech or speech with F2 enhancement in all 12 listening conditions (3 challenge levels $\times$ 2 types of noise $\times$ 2 kinds of speech with/without F2 enhancement) for English NH group (all $p > 0.05$). However, for English HI listeners, it was found that age was significantly correlated with SRTs in 7 of 12 listening conditions (see Table 1). Significant positive correlations suggested that English HI listeners with the increasing age had poorer performance with higher SRTs in above 7 listening conditions. After controlling audibility thresholds as the covariates, partial correlation analysis between age and SRTs showed that the significant relationships in Table 1 became non-significant in most listening conditions except r in the two-talker babble at the challenge level III for original speech ($r = 0.592, p < 0.05$).

Table 1. Correlation coefficients between age and SRTs of original and enhanced speech for English HI listeners (* $p < 0.05$)

	Two-talker babble					
	Original speech			Enhanced speech		
	I	II	III	I	II	III
Age	$r=0.559^*$	$r=0.577^*$	$r=0.549^*$	$r=0.527^*$	$r=0.520^*$	$p > 0.05$
	Six-talker babble					
	Original speech			Enhanced speech		
	I	II	III	I	II	III
Age	$p > 0.05$	$r=0.495^*$	$p > 0.05$	$r=0.562^*$	$p > 0.05$	$p > 0.05$

Furthermore, age did not show significant correlations with benefits from F2 enhancement in all 6 listening conditions (3 challenge levels $\times$ 2 types of noise) for both English NH and HI subgroups (all $p > 0.05$).

4.5.1.2 Memory span vs. SRTs, and Memory span vs. F2-enhancement benefits

As for English NH listeners, correlations between any of three memory subtests (Forward Digit, Backward Digit, and Letter-Number Sequencing) and either SRTs of original or enhanced speech, or benefits from F2 enhancement were not significant (all $p > 0.05$). However, as for English HI listeners, results showed memory span significantly associated with SRTs in some listening conditions. For instance, there were significant relationships between the memory span of Forward Digit and SRTs in 5 of 12 listening conditions (see Table 2). Additionally, significant correlations between memory span of Letter-Number Sequencing and SRTs were shown in 7 of 12 listening conditions (see Table 2). Negative correlations indicated that listeners with longer memory span, did better in speech recognition in noise with lower SRTs. However, partial correlation analysis between memory span and SRTs by controlling audibility thresholds as the covariates, demonstrated that almost all significant relationships in Table 2 were not significant any more except r in the six-talker babble at challenge II for original speech (forward vs. SRT: $r = -0.801, p < 0.05$; sequencing vs. SRT: $r = -0.818, p < 0.05$).

Table 2. Correlation coefficients between memory span and SRTs of original and enhanced speech for English HI listeners (* $p < 0.05$)

	Two-talker babble					
	Original speech			Enhanced speech		
	I	II	III	I	II	III
Forward	$p > 0.05$	$p > 0.05$	$p > 0.05$	$r = -0.609^*$	$r = -0.525^*$	$p > 0.05$
Sequencing	$r = -0.518^*$	$r = -0.526^*$	$p > 0.05$	$r = -0.638^*$	$r = -0.590^*$	$p > 0.05$
	Six-talker babble					
	Original speech			Enhanced speech		
	I	II	III	I	II	III
Forward	$p > 0.05$	$r = -0.593^*$	$p > 0.05$	$r = -0.530^*$	$r = -0.510^*$	$p > 0.05$
Sequencing	$p > 0.05$	$r = -0.654^*$	$p > 0.05$	$r = -0.605^*$	$r = -0.533^*$	$p > 0.05$

On the other hand, as for the significant correlations between memory span and F2-enhancement benefits, it should be noted that scores of Forward Digit Span significantly correlated with benefits from F2 enhancement in two-talker babble at the challenge level I ($r = 0.5, p < 0.05$) only for English HI listeners (see Figure 16, right panel). In addition, the correlations between scores of three memory subtests and F2-enhancement benefits in six-talker babble at the challenge level I were significant for English HI listeners: Forward Digit ($r = 0.5, p < 0.05$); Backward Digit ($r = 0.8, p < 0.05$); and Letter-Number Sequencing ($r = 0.6, p < 0.05$), indicating that English HI listeners who had longer memory span in either one of these memory subtests, gained significant more benefits from F2 enhancement in six-talker babble at the challenge level I (see Figure 16, left panel). Significant correlations in other listening conditions were not observed.

Figure 16. Correlations between memory spans and F2-enhancement benefits for English HI listeners in two-talker (left) and six-talker babble (right) at challenge level I.

4.5.1.3 Audibility vs. SRTs, and Audibility vs. F2-enhancement benefits

Correlation analyses between audibility threshold of each pure tone at octave frequencies (i.e., 250, 500, 1000, 2000, 4000, and 8000 Hz) and SRTs in each listening condition have been conducted for English NH and HI groups (see Table 3). Nevertheless, all significant correlations for both English NH and HI listeners in Table 3 were not significant after controlling memory span and age as the covariates.

Table 3. Correlation coefficients between audibility thresholds and SRTs of original and enhanced speech for English NH and HI listeners (* $p < 0.05$)

Two-talker babble						
	Original speech			Enhanced speech		
	I	II	III	I	II	III
EN NH	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$
EN HI	$r = 0.541^*$ (1k Hz) $r = 0.506^*$ (2k Hz)	$p > 0.05$	$p > 0.05$	$r = 0.485^*$ (1k Hz)	$p > 0.05$	$p > 0.05$
Six-talker babble						
	Original speech			Enhanced speech		
	I	II	III	I	II	III
EN NH	$p > 0.05$	$p > 0.05$	$r = -0.580^*$ (500 Hz); $r = -0.582^*$ (1k Hz)	$p > 0.05$	$p > 0.05$	$r = -0.594^*$ (500 Hz);
EN HI	$p > 0.05$	$r = 0.473^*$ (1k Hz); $r = 0.552^*$ (2k Hz);	$r = 0.533^*$ (4k Hz);	$r = 0.543^*$ (500 Hz); $r = 0.569^*$ (1k Hz); $r = 0.528^*$ (2k Hz)	$r = 0.568^*$ (2k Hz)	$r = 0.514^*$ (250 Hz); $r = 0.586^*$ (500 Hz); $r = 0.547^*$ (1k Hz); $r = 0.531^*$ (2k Hz); $r = 0.492^*$ (4k Hz)

Four-frequency pure tone average (PTA) was calculated for English NH and HI listeners by averaging the thresholds of hearing at 500, 1000, 2000, and 4000 Hz. Significant correlations between PTA and SRTs were found in most listening conditions for English HI listeners, as shown in Table 4. No such significant correlations were observed for English NH listener. But, partial correlation analysis between PTA and SRTs by controlling memory span and age as the covariates, demonstrated that almost all

significant relationships in Table 4 were not significant any more except r in the six-talker babble at challenge II for original speech ($r = 0.639, p < 0.05$).

Table 4 Correlation coefficients between PTA and SRTs of original and enhanced speech for English NH and HI listeners (* $p < 0.05$)

	Two-talker babble					
	Original speech			Enhanced speech		
	I	II	III	I	II	III
EN NH	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$
EN HI	$r = 0.540^*$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$
	Six-talker babble					
	Original speech			Enhanced speech		
	I	II	III	I	II	III
EN NH	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$
EN HI	$p > 0.05$	$r = 0.589^*$	$r = 0.518^*$	$r = 0.553^*$	$r = 0.520^*$	$r = 0.657^*$

On the other hand, correlation analyses between audibility threshold at each octave frequency and F2-enhancement benefits in each listening condition have been conducted for English NH and HI groups. Results showed that in the English NH group, threshold of hearing at 500 Hz ($r = -0.694, p < 0.05$), and 1000 Hz ($r = -0.619, p < 0.05$) significantly related with F2-enhancement benefits in six-talker babble at the challenge level I, suggesting English NH listeners who had better hearing at 500, and 1000 Hz, significantly achieved larger amount of benefits from F2 enhancement in the six-talker babble. As for the English HI group, there was a significant positive correlation between threshold of hearing at 4000 Hz and F2-enhancement benefits in two-talker babble at the challenge level

I: $r = 0.469$, $p < 0.05$, which meant significantly more benefits from F2 enhancement were obtained for English HI listeners whose hearing was worse at 4000 Hz. Furthermore, PTA did not significantly related with F2-enhancement benefits in any listening condition for both English NH and HI listeners.

4.5.1.4 Regression analysis

The linear regression analysis was conducted for English HI listeners to investigate how two factors (Forward Digits Span, and hearing threshold at 4000 Hz) contributed to the benefits from F2 enhancement in two-talker babble at the challenge level I, which were found to be significantly correlated with both factors. The R squared indicated that approximately 22.0% of variations in benefits from F2 enhancement was predicted by hearing threshold at 4000 Hz, and 26.5% of variations was predicted by Forward Digits span, respectively. However, the two factors in the same linear regression could explain a significant proportion (i.e., 38.1%) of the total variations in benefits from F2 enhancement ($F_{2, 14} = 4.309$, $p < 0.05$).

In summary, age and PTA were not significant predictors for F2-enhancement benefits for English listeners, while audibility threshold at the specific frequency and memory span may be significantly related to F2-enhancement benefits in some listening conditions.

However, many significant correlations between contributing factors (i.e. age, memory span, PTA) and SRTs for original and enhanced speech were observed, particularly for English HI listeners. Age, and PTA did not significantly correlate with memory spans (see Figure 8), whereas age associated with PTA significantly ($r = 0.5$, $p < 0.05$). Considering that age, memory spans, and PTA may be confounded with the one

another when correlations were evaluated, partial correlation analysis were conducted to re-examine the relationship between each contributing factor (i.e. age, memory span, PTA) and SRTs for English HI listeners by controlling the other two factors as covariates. Partial correlation results suggested all above significant correlations for age, memory span, and PTA with SRTs for English HI listeners became non-significant, confirming that the effect of age were confounded with the effects of memory span, and PTA on SRTs for English HI listeners.

4.5.2 Chinese group

4.5.2.1 Age vs. SRTs, and Age vs. F2-enhancement benefits

Age was not found to be significantly correlated with either SRTs or enhancement benefits in any listening condition for both Chinese NH and HI groups (all $p > 0.05$).

4.5.2.2 Memory span vs. SRTs, and memory span vs. F2-enhancement benefits

For Chinese NH listeners or Chinese HI listeners, there were no significant correlations between any of three memory tests (Forward Digit, Backward Digit, and Letter-Number Sequencing) and either SRTs for original/enhanced speech or benefits from F2 enhancement in all listening conditions (all $p > 0.05$).

4.5.2.3 Audibility vs. SRTs, and audibility vs. F2-enhancement benefits

As shown in Table 5, correlations between audibility threshold of octave frequency and SRTs in some listening conditions were significant for Chinese NH and HI listeners. However, for Chinese NH listeners, partial correlation analysis by controlling memory span and age as the covariates, demonstrated that only two of the nine significant correlations were still significant (threshold at 1k Hz with SRTs in two-talker and six-talker

babble at the challenge level III for original speech: $r = 0.999$, $p < 0.05$; $r = 0.998$, $p < 0.05$). For Chinese HI listeners, only 2 of 13 significant correlations were still significant according to partial correlation analysis (threshold at 2 k Hz with SRT in two-talker babble at the challenge level I for enhanced speech: $r = 0.692$, $p < 0.05$; threshold at 2 k Hz with SRT in six-talker babble at the challenge level III for enhanced speech: $r = 0.626$, $p < 0.05$).

Table 5. Correlation coefficients between audibility thresholds and SRTs of original and enhanced speech for Chinese NH and HI listeners ($*p < 0.05$)

Two-talker babble						
	Original speech			Enhanced speech		
	I	II	III	I	II	III
CN NH	$p > 0.05$	$r=0.770^*$ (1k Hz)	$r=0.831^*$ (1k Hz)	$p > 0.05$	$r=0.767^*$ (1k Hz);	$r=0.848^*$ (1k Hz)
CN HI	$r=0.474^*$ (2k Hz);	$r=0.505^*$ (1k Hz) $r=0.715^*$ (2k Hz)	$p > 0.05$	$r=0.585^*$ (2k Hz)	$r=0.614^*$ (2k Hz)	$r=0.471^*$ (2k Hz)
Six-talker babble						
	Original speech			Enhanced speech		
	I	II	III	I	II	III
CN NH	$r=0.685^*$ (1k Hz)	$r=0.803^*$ (1k Hz)	$r=0.717^*$ (1k Hz)	$p > 0.05$	$r=0.856^*$ (1k Hz)	$r=0.682^*$ (2k Hz)
CN HI	$r=0.524^*$ (2k Hz)	$r=0.608^*$ (2k Hz)	$r=0.502^*$ (2k Hz)	$r=0.637^*$ (2k Hz)	$r=0.568^*$ (2k Hz)	$r=0.571^*$ (1kHz) $r=0.572^*$ (2k Hz)

Four-frequency (i.e., 500, 1000, 2000, and 4000 Hz) PTAs were calculated for Chinese NH and HI listeners. Correlations between PTA and SRTs of original and enhanced were illustrated in Table 6. Partial correlation analysis by controlling memory

span and age as the covariates, showed that four of the nine significant correlations was still significant for Chinese HI listeners, as marked as “*r*” in the parenthesis in Table 6.

Table 6 Correlation coefficients between PTA and SRTs of original and enhanced speech for Chinese NH and HI listeners (* $p < 0.05$)

	Two-talker babble					
	Original speech			Enhanced speech		
	I	II	III	I	II	III
CN NH	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$	$p > 0.05$
CN HI	$r=0.442^*$ ($r'=0.580^*$)	$r=0.653^*$ ($r'=0.676^*$)	$p > 0.05$	$r=0.527^*$ ($r'=0.761^*$)	$r=0.506^*$	$p > 0.05$
	Six-talker babble					
	Original speech			Enhanced speech		
	I	II	III	I	II	III
CEN NH	$p > 0.05$	$p > 0.05$	$r=0.886^*$	$p > 0.05$	$p > 0.05$	$r=0.671^*$
CN HI	$r=0.441^*$	$r=0.451^*$	$r=0.578^*$	$r=0.466^*$	$p > 0.05$	$r=0.544^*$ ($r'=0.669^*$)

On the other hand, results showed that for Chinese NH listeners, hearing threshold at 2000 Hz and F2-enhancement benefits in six-talker babble at the challenge level III were significantly correlated with each other: $r = -0.805$, $p < 0.05$. In six-talker babble at the challenge level III, hearing threshold at 250 Hz ($r = -0.463$, $p < 0.05$), and 4000 Hz ($r = 0.513$, $p < 0.05$) for Chinese HI listeners were significantly correlated with F2-enhancement benefits.

Nevertheless, there were no any significant relationships between PTA and F2-enhancement benefits in any listening condition for either Chinese NH or HI listeners,

suggesting that PTA was also not a successful predictor for F2-enhancement benefits for Chinese listeners.

4.6 Speech intelligibility index (SII) model

4.6.1. The computation of SII

In order to quantify the impact of individual listener's hearing loss on the audibility of speech, SII was calculated by incorporating listeners' hearing thresholds and physical features of speech signals and noise maskers (i.e., speech and noise spectra), and the weight of different frequency bands of speech materials. The procedure for SII calculation by using the frequency band importance function of one-third octave band (ANSI standard S3.5-1997) was described as follows:

First, according to the SRTs (i.e., signal-to-noise ratios) for original or enhanced speech in two-talker and six-talker babble that were obtained for each listening condition, the speech levels was calculated by the SRT plus the noise level of 70 dB SPL.

Second, hearing thresholds at each of the 18 one-third octave band frequencies were estimated for each listener using the linear interpolation or extrapolation methods with the audibility thresholds obtained at the octave frequencies from 250 to 8000 Hz.

Third, speech signals (e.g., the CRM materials) presented at the threshold levels (e.g., the three challenge levels) and the two- or six-talker babble at 70 dB SPL were analyzed with the sampling rate of 24,414 Hz and the time window of 50 ms to obtain the spectrogram of speech signals and maskers, respectively. The two-talker or six-talker was randomly selected from the noise materials with the same duration as the speech signal. FFT analyses were performed on each of the 50-ms frames for speech and noise, respectively.

Fourth, at each time frame of 50 ms, speech signal and noise masker were broken into 18 non-overlapping frequency bands. Then, the value of the band audibility function (i.e. A_i in the formula) in each frequency band was calculated by including the factor of the hearing threshold level and spectral levels of speech and noise spectrum. Regarding the band frequency importance function in Table 3, band audibility function multiplied with weighting function in each frequency band and then was summed up over 18 one-third octave frequencies to be an instantaneous SII value in each 50-ms window, following the Eq. 2.1. The SII value was then determined by averaging the SII values across all 50-ms windows.

Table 7. Band importance function for one-third octave band SII procedure (cited from ANSI standard S3.5-1997)

Band No.	Nominal midband frequency (Hz)	Band-importance
1	160	0.0083
2	200	0.0095
3	250	0.0150
4	315	0.0289
5	400	0.0440
6	500	0.0578
7	630	0.0653
8	800	0.0711
9	1000	0.0818
10	1250	0.0844
11	1600	0.0882
12	2000	0.0898
13	2500	0.0868
14	3150	0.0844
15	4000	0.0771
16	5000	0.0527
17	6300	0.0364
18	8000	0.0185

4.6.2 SII values for English and Chinese group

In order to reach the correct percentage as 29.3%, 50%, and 70.7%, the SII values, representing the proportion of available speech information extracted from the input signal for listeners, were calculated in different listening conditions for the English (see Figure 17) and Chinese (see Figure 18) groups.

Figure 17. SII values as a function of correct percentage (i.e., challenge levels) for English NH and HI listeners in two-talker and six-talker babble with/without F2 enhancement

Four-factor (between-subjects factor: listener group; within-subjects factor: challenge level $\times$ type of noise $\times$ F2 enhancement) repeated measures AVOVAs of SII values for each language group showed that SII values were only significantly affected by the main effect of challenge level (EN: $F_{2, 56} = 101.788$, $p < 0.05$; CN: $F_{2, 56} = 114.554$, $p <$

0.05) and the main effect of F2 enhancement (EN: $F_{1,28} = 18.084$, $p < 0.05$; CN: $F_{1,28} = 5.256$, $p < 0.05$) in two-talker babble for both language groups. Results suggested compared with SII values for original speech, SII values dropped significantly with F2 enhancement for English and Chinese listeners in two-talker babble, but not in six-talker babble.

Figure 18. SII values as a function of correct percentage (i.e., challenge levels) for Chinese NH and HI listeners in two-talker and six-talker babble with/without F2 enhancement.

4.6.3 SII values for original English and Chinese speech

SII values for original English and Chinese speech are shown in Figure 19. A three-factor (between-subjects factor: language; within-subjects factor: challenge level $\times$ type of noise) repeated measures ANOVA with SII values as the dependent variable was conducted for original speech across two languages. Results revealed that the main effects of language ($F_{1, 58} = 25.489, p < 0.05$), challenge level ($F_{2, 116} = 279.439, p < 0.05$), and type of noise ($F_{1, 58} = 110.020, p < 0.05$); all two-factor interaction effects between language and challenge level ($F_{2, 116} = 7.755, p < 0.05$), language and type of noise ($F_{2, 116} = 48.231, p < 0.05$), and challenge level and type of noise ($F_{2, 116} = 27.323, p < 0.05$); the three-factor interaction effect (language $\times$ challenge level $\times$ type of noise, $F_{2, 116} = 11.203, p < 0.05$) were all significant.

Figure 19. SII values for original English and Chinese speech in two-talker and six-talker babble

Further two-factor AVOVAs (between-subjects factor: language; within-subjects factor: challenge level) were conducted in two-talker and six-talker babble, separately. Post-hoc analyses by using a Bonferroni adjusted critical value indicated that English input signals had significant larger SII values than Chinese input signals in both two-talker and six-talker babble across the challenge levels (all $p < 0.05$).

Chapter 5: Discussion

This study investigated speech recognition in noise for older listeners with and without F2 enhancement and the contributions of age, cognitive function, and hearing thresholds to speech perception in noise with F2 enhancement for two language groups. In addition, speech intelligibility index (SII) were also computed to examine whether the SII model can interpret the benefits associated with F2 enhancement.

5.1 Speech recognition in noise for older listeners with and without hearing loss

It has been demonstrated that older listeners with normal hearing had comparable speech recognition abilities to younger listeners in a quiet listening condition (Dubno et al., 1984; Gordon-Salant et al., 2010). However, their performance dropped at a significantly greater rate and they needed higher SNRs to recognize speech under noisy or reverberant listening conditions compared to younger listeners with matched hearing thresholds (Dubno et al., 1984; Gelfand et al., 1988). It is well documented that aging may result in declines in cognition, hearing thresholds, spectral and temporal resolution, which greatly contribute to poor performance in speech perception (CHABA, 1988; Gordon-Salant et al., 1999, 2001; Pichora-Fuller et al, 2007, Houtgast et al., 2008). In this study, the effects of age, memory span, and audibility (e.g., hearing thresholds, SII) on speech recognition in noise for older listeners are also discussed.

5.1.1 Effects of noise masking on SRTs

In general, there are two types of masking in noise that interfere with listeners' speech perception: energetic masking and informational masking. Energetic masking usually produces interference at the periphery level in the auditory system when noise and speech are presented and overlap physically. In contrast, informational masking happens

at more central levels when meaningful background noise (e.g. competing speech) with linguistic features interferes with speech signal. Listeners might not easily separate, and identify speech signal from meaningful background noise like speech maskers.

Previous studies have revealed that masking effectiveness of noise is dependent on the acoustic and linguistic features of noise (Lewis et al., 1988). Compared with spectrally matched Gaussian noise, multi-talker babble made speech reception thresholds in noise worse by 2-3 dB (Keidser, 1991; Lewis et al., 1988). In this study, two-talker and six-talker babble were used as noise makers. Compared with six-talker babble, two-talker babble contained greater temporal variations (see Figure 20). Furthermore, linguistic content in two-talker babble was more audible than in six-talker babble, resulting in possible informational masking. Informational masking may bring in effective masking and produce a greater increase in SRTs, while higher temporal modulation (e.g., more and deeper temporal dips) can provide a greater release from noise masking, thus leading to lower SRTs. Thus, the final outcome of speech recognition in multi-talker babble is dependent upon the combination of the informational masking and temporal dip listening.

Van Engen and Bradlow (2007) found that English listeners' sentence recognition performance was poorer in six-talker babble than two-talker babble, regardless of the language of babble. Simpson and Cooke (2005) also reported lower consonant scores in six-talker babble than in two-talker babble. Results in this study revealed that both English and Chinese listeners had significantly lower SRTs in two-talker babble than in six-talker babble. Therefore, the lower speech recognition scores in six-talker babble than two-talker babble may mainly result from less temporal modulations in six-talker babble, yielding greater energetic masking and less masking release from the temporal modulation of babble on target speech signals.

Figure 20. Waveforms for English two-talker (right panels) and six-talker babble (left panels) and Chinese two-talker and six-talker babble.

5.1.2 Effect of hearing loss on SRTs

Compared to older listeners with normal hearing, older listeners with hearing loss needed somewhat higher SRTs in two-talker babble across all challenge levels. For instance, the SRT for English HI listeners was 2.5 dB higher than the SRT for English NH listeners in the two-talker babble, while the SRTs for Chinese HI listeners were higher by 2.3 dB than Chinese NH listeners. However, no significant difference between NH and HI listeners on SRTs was found in the six-talker babble for both language groups.

Furthermore, significant correlations between four-frequency PTA (i.e. 500, 1000, 2000, and 4000Hz) and SRTs in many listening conditions (see Table 4 for EN listeners

and Table 6 for CN listeners) indicated that the loss of audibility was a significant factor contributing to poor performance in speech recognition in noise for older listeners with HI.

Some researchers have argued that hearing loss at and below 1000 Hz affects speech perception in quiet, while hearing thresholds above 1000 Hz for HI listeners related with speech perception in noise (Smoorenburg, 1992). A more recent study confirmed that hearing thresholds were significantly associated with scores for speech perception in quiet, but not correlated with performance in noise conditions (Vermiglio et al., 2012). Those findings were somewhat different than the findings on PTA in the current study. In this study, correlation analyses did not show significant correlations between PTA and speech recognition performance in most listening conditions for English and Chinese listeners. The discrepancy might result from the differences of many factors across these studies such as age, degree and configuration of hearing loss, cognitive function of listeners, and speech materials. In fact, the partial correlation analysis for both language groups suggested that these factors had confounding effects with PTAs on SRTs. That is, after controlling the age and/or memory spans as covariates, partial correlations between PTAs and SRTs were no longer significant for English and Chinese listeners across almost all listening conditions. Therefore, PTAs were not successful indicators of HI listeners' speech identification in noise. In fact, in order to validate the accuracy of hearing thresholds in clinics, the traditional PTA with average hearing thresholds measured at 500, 1000, and 2000 Hz, is commonly used to interpret the individual variability of speech recognition threshold in quiet instead of noise.

5.1.3 SII and SRTs

In order to evaluate how much difficulty in speech recognition in noise was due to reduced audibility, the SII model allowed researchers quantify and predict the effect of

audibility on speech perception for hearing-impaired listeners (Vermiglio et al., 2012). That was because SII is calculated by using not only individual listeners' hearing thresholds but also the SNRs of incoming input signals in specific frequency bands across the spectrum to simulate outputs from normal and impaired periphery auditory systems for listeners with NH and HI. SII values provide information about the amount of speech information that is available for listeners to extract from speech signals presented in noise. According the third goal of this study, results confirmed that SII could interpret listeners' speech recognition performance and the perceptual benefits from F2 enhancement.

According to the monotonicity of the SII transfer function (see Figure 1), greater SII values usually correspond to higher correct percentage of speech recognition. In this study, findings confirmed that significantly higher SII values were needed for listeners to achieve higher correct percentage ($SII_{T-29.3\%} < SII_{T-50.0\%} < SII_{T-70.7\%}$). On the other hand, for both English and Chinese listeners, SII values in this study were sensitive acoustic differences between original and enhanced speech, particularly in two-talker babble. It has been found that SII values for enhanced speech were significantly lower than SII values for original speech. Moreover, two-talker babble had significantly higher SII values than six-talker babbles in most conditions because more acoustic cues were available in the two-talker babble because of temporal dips (see Figure 20). Results in this study indicated that with short-duration windows, the SII model can also be applied in fluctuating noise to interpret SRTs in multi-talker babble as well as the performance of speech perception in stationary quiet (Rhebergen & Versfeld, 2005; Smits & Festen, 2013).

Although the SII model provided the acoustic evidence for the improvement of speech recognition in the two- and six-talker babble, it did not account for suprathreshold deficits in auditory processing, such as reduced spectral and temporal acuity, and decreased

cognitive function due to aging. These factors needed to be included to explain speech recognition in different types of noise.

5.1.4 Age and SRTs

Age was not found to be significantly related to SRTs in all listening conditions for Chinese NH listeners, Chinese HI listeners, and English NH listeners. For English HI listeners, significant positive correlations between age and SRTs were reported in seven of twelve listening conditions. However, after controlling for hearing thresholds and memory spans, the correlations between age and SRTs became non-significant, indicating that age may covary with other factors to influence SRTs for elderly listeners, such as hearing thresholds, memory spans, and suprathreshold deficits. Another possibility could be that the age range of the elderly listeners was not broad enough, e.g., the oldest listener in this study was 85 years old for the English group and 74 years old for the Chinese group.

5.1.5 Cognitive memory and SRTs

Speech recognition in noise involves not only lower-level sensory and perceptual processes, but also higher-level cognitive processes, such as memory, attention, and language comprehension (Schneider et al., 2002). With reduced input of auditory sensory information due to cochlear dysfunction, the cognitive demand for HI listeners in adverse listening conditions became higher (Pichora-Fuller et al. 1995; Lunner, 2003). HI listeners may have to reallocate cognitive resources and make greater listening efforts to perceive speech in noise.

In this study, elderly listeners were observed to have the shortest spans of Letter-Number Sequencing and longest spans of Forward Digit for both the English and Chinese group, consistent with previous findings that aging leads to a much greater decline on

Letter-Number Sequencing than on either Forward or Backward Digit Span subtests in the third edition of the WAIS (WAIS–III) (Myerson et al., 2003; Emery et al., 2007). Forward Digit Span aimed to measure the short-time memory supported by long-term lexical knowledge. Backward digit span and letter-number sequencing cross-checked the working memory component by measuring listeners’ abilities to recall and manipulate the digits or letters. Letter-Number Sequencing subtest of digit and alphabet items was a complex working memory task. Compared to younger listeners, older listeners had significantly shorter spans of Letter-Number Sequencing (Emery et al., 2007), possibly because older adults have greater difficulty simultaneously manipulating and maintaining digit and letters in working memory.

In addition, it was found that shorter memory spans were significantly correlated with lower SRTs in limited listening conditions for English HI listeners, while no such significant relationships were observed in English NH listeners, Chinese NH listeners, or Chinese HI listeners. Many researchers argue that working memory capacity, which is responsible for processing task-relevant information (e.g. speech) and temporarily inhibiting non-relevant information (e.g. noise or competing talkers), is an important factor contributing to the performance of speech perception in noise (Pichora-Fuller et al. 1995; Schneider et al., 2002; 2007). In multi-talker babble, working memory helps listeners handle multiple streams of information and shift attention to the target speech signal from competing maskers. In this study, in multi-talker babble, better memory abilities might assist English HI listeners in some conditions to achieve better performance in speech perception in noise, when noise strongly masked the target speech in the peripheral and central auditory system.

In the current study, the lack of correlation between memory span and speech recognition performance in most listening conditions was contrary to the hypothesis. It

might be attributed to the following reasons. Firstly, the CRM tasks in this study only asked listeners to recall the last two words in a sentence, which may not require a high cognitive load. Secondly, the three memory subtests including digits and letters from WAIS-III might not be sensitive enough. Other working memory tasks closely related with speech understanding, such as reading span test (Lunner, 2003), could be included in the future studies.

5.2 Speech enhancement for older listeners

One primary goal of this study was to investigate whether acoustic enhancement such as spectral enhancement on speech signals could improve older listeners' speech recognition in noise. In this study, as expected, older listeners regardless of hearing status and native language background, performed significantly better in speech perception with F2 enhancement than without F2 enhancement in the two-talker babble, but not in the six-talker babble. Such F2-enhancement benefits for older listeners in both the English and Chinese groups demonstrated that the technology of F2 enhancement might help older listeners resist the negative effects of noise masking. Specific reasons underlying F2-enhancement benefits for older listeners might be related to the following aspects:

5.2.1 Masking release from temporal modulation in noise

The masking release from the temporal modulation in noise could be one important factor for the improvement in speech intelligibility for fluctuating (i.e., modulated) noise relative to stationary (i.e., unmodulated) maskers. The benefits from masking release could be interpreted in the terms of “dip listening”, where listeners can take advantage of spectral and/or temporal dips in the fluctuating noise to perceive target speech signals. The energy of the speech target exceeded that of the masker by a certain amount in either temporal or

spectral region because of these glimpses (Cooke, 2006). The benefits gained by the presence of those dips (e.g., masking release from the temporal variation in noise) may be affected by listeners' age, audibility, and suprathreshold deficits (frequency and temporal resolution) (Dubno et al., 2002; Takahashi & Bacon, 1992; Grose et al., 2009; Strelcyk & Dau, 2009; Bernstein et al, 2009).

There was a decline in various aspects of sensory temporal processing for elderly listeners (Gordon-Salant & Fitzgibbons, 1999; Pichora-Fuller et al., 1994; Dubno et al., 2002). Dubno et al. (2002) investigated the effect of aging on masking release for young and elderly listeners and presented threshold-matching noise to young listeners to rule out the effect of audibility. They observed that elderly listeners gained significantly smaller masking release and had poorer simultaneous and forward masking thresholds than young listeners when they recognized consonants in noise. Compared with young listeners, elderly listeners may be worse in "dip listening". On the other hand, additional hearing loss for elderly listeners may make it even more difficult to take advantage of temporal dips so as to achieve lower thresholds of speech recognition in temporally-fluctuating noise than their peers with normal hearing (Moore et al., 1999). HI listeners, even with mild hearing loss, showed significantly reduced masking release from temporal dips of noise than NH listeners (Takahashi & Bacon, 1992; Grose et al., 2009; Strelcyk & Dau, 2009). Compared with NH listeners, HI listeners showed 7-12 dB less masking release at the 50% correct performance level (Bernstein et al, 2009).

In this study, there was no significant difference between NH and HI listeners on F2-enhancement benefits observed in the English group. However, Chinese HI listeners achieved significantly larger benefits from F2 enhancement by 1.8 dB than Chinese NH listeners in the two-talker babble at challenge level I. Greater F2-enhancement benefits for Chinese HI listeners than their NH peers may result from F2-enhancement technology

helping them to take advantage of more available temporal dips in two-talker babble and thus obtaining significant better SRTs of enhanced speech, even though Chinese HI listeners had poorer SRTs of original speech. The amount of masking release from temporal modulation in noise could not be calculated directly in this study. However, SII values indirectly confirmed that for both English and Chinese groups with NH and HI, F2 enhancement brought in more available acoustic cues for enhanced speech compared with original speech, particularly in two-talker babble (see Figure 17 for English group; see Figure 18 for Chinese group).

5.2.2 Information masking interference and F2 enhancement

Informational masking has rarely been evaluated by current clinical tests to assess various cochlear and retro-cochlear functions. The amount of information masking depends on several stimulus factors, such as the similarity between signal and masker (i.e., speech signals and competing speech are produced by talkers of the same gender), listeners' familiarity to the stimuli (e.g., native or non-native language); and stimulus uncertainty. Li et al. (2004) found that older listeners underperformed younger listeners when they recognized meaningless sentences in background noise, whereas younger and older listeners had similar informational masking. Similarly, Helfer and Freyman (2005) observed that older listeners, especially those with hearing loss, showed greater difficulty to selectively attend to the target speech, even when the masker voice was from a talker with an opposite gender. However, there was no difference in the amount of informational masking on sentence recognition in competing speech maskers between younger and older listeners. A later study by Agus, Akeroyd, Gatehouse, and Warden (2009) reported similar findings that in a task of sentence recognition in competing speech maskers and speech-shaped noise, younger and older listeners showed comparable informational masking.

By using three types of maskers (i.e. a single sentence, a reversed sentence, and a steady state speech spectrum noise), Summers et al. (2004) investigated whether increasing the level of target speech and noise masker could improve performance for hearing-impaired listeners. Results revealed that only two of six HI listeners benefited from simple amplification of speech and noise together where the noise was a single sentence. Two other HI listeners even did worse. Large individual differences on informational masking were observed in the HI group. In addition, overall increasing the levels of speech and noise was not efficient for listeners to overcome the informational masking of speech maskers.

Given the comparable informational masking for older and young listeners from the previous studies described above, it was expected that the effect of informational masking was similar between elderly NH and HI listeners in the present study. The perceptual benefits with the F2 enhancement in the two-talker babble, but not the six-talker babble, may not be connected to informational masking. That is, the presentation of F2-enhanced speech signals in multi-talker babble may have little impact on the informational masking of babble.

In order to quantify the amount of informational masking when benefits from the technology is evaluated, two types of masker could be used in the future study: multi-talker babble and babble-modulated noise that had a long-term spectrum and temporal envelope matched with the multi-talker babble. The difference in the identification scores between the two maskers is considered to be the quantitative measure of the informational masking (Li et al., 2016).

5.2.3 Language background and F2 enhancement

As shown in Figure 15, the F2-enhancement benefits were presented for two language groups with normal hearing or hearing loss in the two-talker babble. Overall, English group obtained slightly larger F2-enhancement benefits than Chinese group in two-talker babble. Particularly, English listeners with normal hearing had higher benefits from F2 enhancement than Chinese NH listeners, while English HI listeners had similar benefits as Chinese HI listeners. One of reasons why English group had more F2-enhancement benefits may be that F2 enhancement helped English listeners improve the capacity to use temporal modulations in two-talker babble to perceive target speech.

English and Chinese are two language systems. Chinese differs from English in many aspects, such as phonetics, phonology, prosody, and tone. Therefore, temporal features of Chinese speech (e.g., temporal modulation frequency and depth) may be significantly different from those of English speech. Calandruccio et al (2010) observed that English two-talker babble was more temporally modulated than Chinese two-talker babble, implying that the ability of using temporal dips in noise may differ between English- and Chinese-native listeners. In this study, as shown in Fig. 19, more available acoustic cues for speech recognition in English two-talker babble with higher SIIs than in Chinese two-talker babble. Thus, there might be more speech information in English two-talker babble than in Chinese two-talker babble, possibly due to the different temporal structures of the two speech babble.

On the other hand, previous research have showed that native language background influenced listeners' ability of temporal dips listening for their speech recognition in noise (Guan et al., 2015; Li et al., 2016). In particular, English-native listeners received greater benefits from the temporal fluctuation of noise than Chinese-native listeners when identifying English vowels (Guan et al., 2015). Moreover, extensive native English

experiences for Chinese-native listeners who lived in the US for 1-3 years may help them improve the capacity to use temporal cues in noise to perceive English vowels (Guan et al., 2015), and Chinese vowels (Li et al., 2016) better than their peers who had no residency history in English-speaking countries. These studies suggest that the capacity of using temporal dips in noise to perceive speech signals may differ between English- and Chinese-native speakers. Thus, in this study, Chinese older listeners had smaller benefits from temporal modulation in the two-talker babble than English older listeners.

5.2.4 F2 enhancement and other enhancement technology

The purpose of this study was to evaluate perpetual effects of F2 enhancement in adverse listening conditions with three challenge levels for older listeners with normal hearing and hearing loss. Speech with F2 enhancement was processed before multi-talker babble was added rather than the F2 enhancement on the mixture of speech and noise. Compared with other studies of enhancement technology, F2 enhancement had the following advantages and limitations.

First, thresholds of speech recognition significantly improved by about 1-1.5 dB as a result of F2 enhancement for older listeners in both language groups, especially in the two-talker babble. This improvement was relatively robust in comparison with previous attempts at speech enhancement. Similarly, Boers (1980) applied formant enhancement only on sentence lists with unenhanced speech-shaped noise for NH and HI listeners. However, he reported that speech reception scores even worsened with formant enhancement for both groups and unexpected distortion was introduced. Some other studies applied speech enhancement technique on both speech signals and noise simultaneously, which was more feasible with naturally occurring speech signals that were usually contaminated with noise in our daily life (Simpson et al., 1990; Stone & Moore,

1992; Baer et al., 1993; Franck et al., 1999; Lyzenga et al., 2002). Nevertheless, speech intelligibility tests showed there was no improvement (Baer et al., 1993; Stone & Moore, 1992; Lyzenga et al., 2002), or small improvement (around 6%) in speech intelligibility in the enhanced noise (Simpson et al., 1990). Significant improvement usually needed extra processing algorithms in addition to spectral enhancement (e.g., moderate compression, Baer et al., 1993; noise reduction and high-frequency amplification, Lyzenga et al., 2002). Therefore, with no extra processing algorithms, simply enhancing F2 in this study, which has made significant improvement in speech reception thresholds for older listeners, seems very promising.

Second, the F2-enhancement algorithm in the current study had detectors for both spectral peaks and spectral valleys. In this way, instead of only targeting speech peaks, spectral contrasts between peaks and valleys of speech were enhanced so as to restore robust spectral representation for HI listeners. Bunnell (1990) applied spectral envelope estimator that mainly changed the amplitudes of spectral peaks and did not find larger increase in syllable identification in quiet. Scores of syllables with /d/ in CV (i.e. consonant-vowel) tokens even reduced. Syllable identification with speech enhancement in noise was not tested in that study. In the current study, results of speech identification test in noise demonstrated that F2 enhancement by enlarging amplitudes between the F2 peaks and adjacent valleys improved speech intelligibility for both NH and HI listeners in two-talker babble. Given the findings of this study, it may speculate that the F2 enhancement could also work for speech-plus-noise situations when SNR is at a favorable level, e.g., formant peaks are observable from noise background. However, when SNR is quite low, e.g., formant peaks are engulfed by noise, F2 enhancement may not apply or even enhance noise spectrum, possibly resulting in no perceptual benefit. Such speculation regarding the F2 enhancement on speech-plus-noise needs future investigations.

On the other hand, the current F2-enhancement technology has some limitations. Firstly, only speech signals were enhanced with no enhancement processing for background noise. The effectiveness of F2 enhancement on speech-plus-noise to simulate hearing-aid processing was needed for further examination. Secondly, the enhancement was applied for the segment that contained F2 peaks, possibly resulting in no enhancement for consonant segments like stops and fricatives. It was uncertain about which part of sentences (e.g. vowels, consonants, or formant transitions between phonemes) gained more benefits from F2 enhancement. Alcantara et al. (1994) employed dynamic band-pass filters centered at F1 and F2 with varied bandwidths to enhance speech signals, which was similar as this study. Prior to signal processing and in the absence of noise, acoustic analyses of the stimuli were completed to identify F1 and F2 characteristics. They found that in a background of multi-talker babble, speech enhancement successfully improved vowel identification in a /hVd/ context by 10-15% but not consonant intelligibility in /aCa/ context or sentence recognition in noise. Franck et al. (1999) also observed that vowels perception scores improved with spectral enhancement, while consonant scores dropped with enhancement. In this study, because the F2 enhancement was applied on the speech segments with F2 peaks detected, spectral enhancement was primarily presented in vowels and formant transitions rather than consonants. It should be noted that formant transitions also provided information for consonant place and manner (Alcantara et al., 1994), thus, the F2 enhancement may improve the identification of both vowels and consonants. Further study with a wider range of stimuli, such as vowels, consonants and syllables, need to be conducted to evaluate the effectiveness of F2 enhancement on different phonemic and syllabic levels.

5.3 Summary of results

In summary, according to the three goals of this study, some results in this study consistently met the hypotheses, while others did not. First, older listeners with hearing loss needed slightly higher SNR to recognize speech in noise than their peers with normal hearing. Regardless of hearing status, F2 enhancement helped both NH and HI listeners improve their performance mainly in two-talker babble, but not in six-talker babble. These results indicate that F2 enhancement technology produced general benefits across different languages.

Second, speech recognition performance and the improvement of speech recognition by F2 enhancement were not found to be significantly correlated with listeners' hearing thresholds and memory abilities in most listening conditions against the hypothesis. This is possibly because speech recognition was measured in noise and also because of the relative simplicity of speech materials in the present study.

Third, for both English and Chinese listeners, two-talker babble had significantly higher SII values than six-talker babble and the SII values significantly dropped with F2 enhancement. The perceptual benefits from F2 enhancement in two-talker babble across language groups may be associated with that more acoustic cues were available in the two-talker babble because of temporal variations.

5.4 Future directions

Elderly listeners with normal and impaired hearing were evaluated with the effect of spectral enhancement on speech perception. In the future's study, younger listeners with normal hearing and hearing loss could be recruited to reveal the aging effect on SRTs and F2-enhancement benefits and whether F2 enhancement could improve speech perception in noise for younger listeners.

In order to separately investigate the relationship between temporal masking release/informational masking and F2-enhancement benefits, other types of maskers could be employed in the future, such as spectral-matched stationary noise, and babble-modulated noise that was matched with the spectral and/or temporal features of multi-talker babble.

Moreover, instead of only enhancing the speech signals, speech-plus-noise as input signals with different SNRs could be applied with F2 enhancement in the future so as to simulate hearing-aid processing.

Additionally, a broader range of speech intelligibility tests including vowels, consonants and syllables, could be employed to testify the effectiveness of the F2 enhancement algorithm on different levels of speech in the future.

5.4 Clinical application of F2 enhancement

According to the findings of the present study, elderly listeners with different language backgrounds may gain benefits from F2 enhancement in their speech recognition in noise. As a signal-processing algorithm, F2 enhancement appears auspicious to be implemented in hearing devices for elderly listeners, such as the listening scenario in which speech signals are enhanced under an unchanged noise background.

Speech intelligibility is one of essential aspects that are needed to be considered when a signal-processing algorithm is judged. Other aspects, such as speech clarity, listening efforts, and cognitive demand are also important. Only when listeners are pleased to keep the hearing devices with the new signal-processing algorithm and wear them constantly, they can obtain improvements from the amplification. For example, compared with nonlinear digital hearing aids that aim to restore normal loudness growth by specifically amplify soft input sounds more than intense ones for HI listeners with a

smaller-than-normal dynamic range, they may not tolerate the hearing aids that simply amplify overall levels and overshoot listeners' upper uncomfortable boundary. When Stone and Moore (1992) did not find any significant benefit in speech intelligibility with their spectral enhancement, the algorithm was still worthwhile when the majority of participants reported enhanced speech had overall better quality than unprocessed speech. Success with amplification in the real daily life should take care of many aspects of listening.

In addition, another implication based on the findings in this study is that cognitive function may associate with the effectiveness of F2 enhancement for HI listeners. Previous research indicated that those listeners with worse cognitive capacity (e.g. working memory) could have less success with complex signal-processing algorithms (Lunner et al., 2003). All in all, many factors might need to be incorporated into future signal-processing algorithms and related audiological rehabilitation: audibility, spectral and temporal resolution, cognitive function, and also the effectiveness of signal-processing algorithms on speech intelligibility, speech clarity, perceived comfort and so on.

Chapter 6: Conclusion

In conclusion, results of this study suggested that regardless of hearing status and native language background, older listeners performed significantly better in speech perception with F2 enhancement than original speech, particularly in the two-talker babble, but not in the six-talker babble. Findings supported that the manipulation of F2 enhancement may help older listeners improve speech perception in noise across languages: American English and Mandarin Chinese. However, age, cognitive memory abilities, and hearing thresholds were not strong predictors to estimate listeners' speech recognition in noise independently. Instead, the SII model could successfully interpret the SRTs in multi-talker babble and the benefits associated with F2 enhancement. The perceptual improvement with F2 enhancement for both English and Chinese listeners in the two-talker babble was probably due to that more acoustic cues were available for them to perceive speech in temporally dynamic noise. Still, other aspects were needed to be incorporated into the future development of signal-processing algorithms, such as more speech materials and F2-enhancement in speech-plus-noise (e.g. hearing aid listening condition).

References

- Agus, T. R., Akeroyd, M. A., Gatehouse, S., & Warden, D. (2009). Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise. *The Journal of the Acoustical Society of America*, 126(4), 1926-1940.
- Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing impaired adults. *International Journal of Audiology*, 47(Suppl. 2), S53–S71.
- Alcántara, J. I., Dooley, G. J., Blamey, P. J., & Seligman, P. M. (1994). Preliminary evaluation of a formant enhancement algorithm on the perception of speech in noise for normally hearing listeners. *Audiology*, 33(1), 15-27.
- ANSI. (1997). American National Standard for methods for calculation of the speech intelligibility index. *ANSI S3.5-1997*. New York.
- Bacon, S. P., Opie, J. M., & Montoya, D. Y. (1998). The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds, *Journal of Speech, Language, and Hearing Research*, 41, 549-563.
- Baer, T., Moore, B. C. J., & Gatehouse, S. (1993). Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: effects on intelligibility, quality and response times. *Journal of Rehabilitation Research and Development*, 30(1), 49-72.

- Bernstein, J. G., & Grant, K. W. (2009). Auditory and auditory-visual intelligibility of speech in fluctuating maskers for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 125(5), 3358-3372.
- Boers, P. M. (1980). Formant enhancement of speech for listeners with sensorineural hearing loss. *IPO annual progress report*, 15, 21.
- Bruce, I.C. (2004). Physiological assessment of contrast-enhancing frequency shaping and multiband compression in hearing aids. *Physiological Measurement*, 25(4), 945-56.
- Bunnell, H. T. (1990). On enhancement of spectral contrast in speech for hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 88(6), 2546-56.
- Buus, S., & Florentine, M. (2002). Growth of loudness in listeners with cochlear hearing losses: Recruitment reconsidered. *JARO-Journal of the Association for Research in Otolaryngology*, 3(2), 120-139.
- Calandruccio, L., Dhar, S., & Bradlow, A. R. (2010). Speech-on-speech masking with variable access to the linguistic content of the masker speech. *The Journal of the Acoustical Society of America*, 128(2), 860-869.
- CHABA (Committee on Hearing, Bioacoustics and Biomechanics). (1988). Speech understanding and aging. *The Journal of the Acoustical Society of America*, 83, 859–895.

- Chen, F., Wong, L. L., & Hu, Y. (2014). Effects of lexical tone contour on Mandarin sentence intelligibility. *Journal of Speech, Language, and Hearing Research*, 57(1), 338-345.
- Chen, J., Baer, T., & Moore, B. C. (2012). Effect of enhancement of spectral changes on speech intelligibility and clarity preferences for the hearing impaired. *The Journal of the Acoustical Society of America*, 131(4), 2987-2998.
- Ching, T. Y., Dillon, H., & Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *The Journal of the Acoustical Society of America*, 103(2), 1128-1140.
- Chung, K. (2004). Challenges and recent developments in hearing aids: Part I. Speech understanding in noise, microphone technologies and noise reduction algorithms. *Trends in Amplification*, 8(3), 83-124.
- Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, 119(3), 1562-1573.
- Coughlin, M., Kewley-Port, D., & Humes, L. E. (1998). The relation between identification and discrimination of vowels in young and elderly listeners. *Journal of the Acoustical Society of America*, 104, 3597-3607.
- Dubno J.R., Dirks D.D., & Morgan D.E. (1984). Effects of age and mild hearing loss and

- age on speech recognition in noise. *The Journal of the Acoustical Society of America*, 76, 87-96.
- Dubno, J. R., Horwitz, A. R., & Ahlstrom, J. B. (2002). Benefit of modulated maskers for speech recognition by younger and older adults with normal hearing. *Journal of Acoustical Society of America*, 111, 2897-2907.
- Eddins, D.A., & Hayes, D. (2007). Evaluation of the antiShock Algorithm for Impulse Correction of Transient Signals by the Unitron Element Hearing Aid, *Association for Research in Otolaryngology* Abstr. 30.
- Emery, L., Myerson, J., & Hale, S. (2007). Age differences in item manipulation span: the case of letter-number sequencing. *Psychology and aging*, 22(1), 75.
- Fitzgibbons, P. J., & Gordon Salant, S. (1995). Age effects on duration discrimination with simple and complex stimuli. *The Journal of the Acoustical Society of America*, 98(6), 3140-3145.
- Franck, B.A.M., Van Kreveld-Bos, C.S.G.M., Dreschler, W.A., & Verschuure, H. (1999). Evaluation of spectral enhancement in hearing aids, combined with phonemic compression. *The Journal of the Acoustical Society of America*, 106(3), 1452-64.
- Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *The Journal of the Acoustical Society of America*, 115(5), 2246-2256.

Fogerty, D. & Kewley-Port, D. (2009). Perceptual contributions of the consonant-vowel boundary to sentence intelligibility. *Journal of the Acoustical Society of America*, 126, 847-857.

Fogerty, D., Kewley-Port, D. & Humes, L.E. (2012). The relative importance of consonant and vowel segments to the recognition of words and sentences: Effects of age and hearing loss. *Journal of the Acoustical Society of America*, 132, 1667-1678.

Füllgrabe, C., Meyer, B., & Lorenzi, C. (2003). Effect of cochlear damage on the detection of complex temporal envelopes. *Hearing research*, 178(1), 35-43.

Gelfand, S. A., Ross, L., & Miller, S. (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.

Gordon-Salant, S., & Fitzgibbons, P. (1999). Profile of auditory temporal processing in older listeners. *Journal of Speech, Language, and Hearing Research*, 42, 300-311.

Gordon-Salant, S., & Fitzgibbons, P. (2001). Sources of age related recognition difficulty for time-compressed speech. *Journal of Speech, Language, and Hearing Research*. 44, 709-719.

Gordon-Salant, S., Yeni-Komshian, G. H., & Fitzgibbons, P. J. (2010). Recognition of accented English in quiet and noise by younger and older listeners. *The Journal of the Acoustical Society of America*, 128(5), 3152-3160.

- Grose, J. H., Mamo, S. K., & Hall III, J. W. (2009). Age effects in temporal envelope processing: speech unmasking and auditory steady state responses. *Ear and hearing*, 30(5), 568
- Guan, J., Liu, C., Tao, S., Mi, L., Wang, W., & Dong, Q. (2015). Vowel identification in temporal-modulated noise for native and non-native listeners: Effect of language experience. *The Journal of the Acoustical Society of America*, 138(3), 1670-1677.
- Helfer, K. S., & Freyman, R. L. (2005). The role of visual speech cues in reducing energetic and informational masking. *The Journal of the Acoustical Society of America*, 117(2), 842-849.
- Hornsby, B. W. (2004). The Speech Intelligibility Index: What is it and what's it good for? *The Hearing Journal*, 57(10), 10-17.
- Houtgast, T., & Festen, J. M. (2008). On the auditory and cognitive functions that may explain an individual's elevation of the speech reception threshold in noise. *International journal of audiology*. 47(6):287-95.
- Humes, L. E. 2002. Factors underlying the speech-recognition performance of elderly hearing-aid wearers. *The Journal of the Acoustical Society of America*, 112, 1112–1132.
- Humes, L. E., & Young, L. A. (2016). Sensory–Cognitive Interactions in Older Adults. *Ear and Hearing*, 37, 52S-61S.

- Humes, L. E., & Young, L. A. (2016). Sensory-cognitive interactions in older adults. *Ear Hearing*, 37, 52S–61S.
- Jenkins, L., Myerson, J., Joerding, J. A., & Hale, S. (2000). Converging evidence that visuospatial cognition is more age-sensitive than verbal cognition. *Psychology and aging*, 15(1), 157.
- Jin, S. H., Liu, C., & Sladen, D. P. (2014). The effects of aging on speech perception in noise: comparison between normal-hearing and cochlear-implant listeners. *Journal of the American Academy of Audiology*, 25(7), 656-665.
- Jin S-H., & Nelson P. (2006). Speech perception in gated noise: the effects of temporal resolution, *Journal of the American Academy of Audiology*, 119 (5): 3097-3108.
- Jin S-H., & Nelson P. (2010). Interrupted speech perception: the effects of hearing sensitivity and frequency resolution, *The Journal of the Acoustical Society of America*, 128 (2): 881-889.
- Johnson, C. E., Danhauer, J. L., Ellis, B. B., & Jilla, A. M. (2016). Hearing aid benefit in patients with mild sensorineural hearing loss: A systematic review. *Journal of the American Academy of Audiology*, 27(4), 293-310.
- Kale, S., & Heinz, M. G. (2010). Envelope coding in auditory nerve fibers following noise-induced hearing loss. *Journal of the Association for Research in Otolaryngology*, 11(4), 657-673.

- Kalluri, S., & Humes, L. E. (2012). Hearing technology and cognition. *American journal of audiology*, 21(2), 338-343.
- Keidser, G. (1991). Computerized Measurement of Speech Intelligibility I : Development of System and Procedures. *Scandinavian audiology*, 20(3), 147-152.
- Kewley-Port, D., Burkle, T. Z., & Lee, J. H. (2007). Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 122(4), 2365-2375.
- Klein A., Mills J., & Adkins W. (1990). Upward spread of masking, hearing loss, and speech recognition in young and elderly listeners. *The Journal of the Acoustical Society of America*, 87(3), 1266-71.
- Kryter, K. D. (1962). Validation of the articulation index. *The Journal of the Acoustical Society of America*, 34(11), 1698-1702.
- Leek, M. R., Dorman, M. F., & Summerfield, Q. (1987). Minimum spectral contrast for vowel identification by normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, 81(1), 148-54.
- Leek, M. R., & Summers, V. (1993). Auditory filter shapes of normal-hearing and hearing-impaired listeners in continuous broadband noise. *The Journal of the Acoustical Society of America*, 94(6), 3127-3137.

- Leek, M. R., & Summers, V. (1996). Reduced frequency selectivity and the preservation of spectral contrast in noise. *The Journal of the Acoustical Society of America*, 100(3), 1796-1806.
- Leeuw, A. R., & Dreschler, W. A. (1991). Advantages of directional hearing aid microphones related to room acoustics. *Audiology*, 30, 330-344.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467-477.
- Lewis, H. D., Benignus, V. A., Muller, K. E., Malott, C. M., & Barton, C. N. (1988). Babble and random-noise masking of speech in high and low context cue conditions. *Journal of Speech, Language, and Hearing Research*, 31(1), 108-114.
- Li, L., Daneman, M., Qi, J. G., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *Journal of Experimental Psychology: Human Perception and Performance*, 30(6), 1077.
- Li, M., Wang, W., Tao, S., Dong, Q., Guan, J., & Liu, C. (2016). Mandarin Chinese vowel-plus-tone identification in noise: Effects of language experience. *Hearing research*, 331, 109-118.
- Liu, C & Kewley-Port, D. (2004). Formant discrimination in noise for isolated vowels. *Journal of the Acoustical Society of America*, 116, 3119-3129.

- Liu, C. & Eddins, D. (2008). Effects of spectral modulation filtering on vowel identification. *Journal of the Acoustical Society of America*, 124, 1704-1715.
- Liu, C., & Kewley-Port, D. (2007). Factors affecting vowel formant discrimination for hearing-impaired listeners. *Journal of the Acoustical Society of America*, 122, 2855–2864.
- Lock, D., 2009. The New Children's Encyclopedia. *DK Publishing*, Inc, New York, NY, pp. 8-78.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., & Moore, B. C. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proceedings of the National Academy of Sciences*, 103(49), 18866-18869.
- Lyzenga, J., Festen, J.M., & Houtgast, T. (2002). A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity. *The Journal of the Acoustical Society of America*, 112(3), 1145-1157.
- Lunner, T. (2003). Cognitive function in relation to hearing aid use. *International journal of audiology*, 42, S49-S58.
- Luts, H., Maj, J. B., Soede, W., & Wouters, J. (2004). Better speech perception in noise with an assistive multimicrophone array for hearing aids. *Ear and hearing*, 25(5), 411-420.

- Martin E., & Pickett, J.M. (1970). Sensorineural hearing loss and upward spread of masking. *Journal of Speech and Hearing Research*, 13, 426-437.
- Miller, R.L., Calhoun, B.M., & Young E.D. (1999). Contrast enhancement improves the representation of /e/-like vowels in the hearing-impaired auditory nerve. *The Journal of the Acoustical Society of America*, 106(5), 2693-2708.
- Moore, B. C. J. (1996). Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear Hearing*, 17, 133–161.
- Moore, B. C. J. (2007). Cochlear Hearing Loss: Physiological, Psychological and Technical Issues, 2nd ed. (Wiley, Chichester), pp. 332.
- Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, 9, 399–406.
- Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1992). Detection of temporal gaps in sinusoids by elderly subjects with and without hearing loss. *Journal of the Acoustical Society of America*, 92, 1923–1932.
- Moore, T. (1981). Voice communication jamming research, in AGARD Conference Proceedings 331: *Aural Communication in Aviation*, pp. 2:1–2:6, Neuilly-Sur-Seine, France.

- Moore, B. C., Peters, R. W., & Stone, M. A. (1999). Benefits of linear amplification and multichannel compression for speech comprehension in backgrounds with spectral and temporal dips. *The Journal of the Acoustical Society of America*, 105(1), 400-411.
- Myerson, J., Emery, L., White, D. A., & Hale, S. (2003). Effects of age, domain, and processing demands on memory span: Evidence for differential decline. *Aging, Neuropsychology, and Cognition*, 10(1), 20-27.
- Nelson, D. A., & Pavlov, R. (1989). Auditory time constants for off-frequency forward masking in normal-hearing and hearing-impaired listeners. *Journal of Speech, Language, and Hearing Research*, 32(2), 298-306.
- Nie, K., Stickney, G., & Zeng, F. G. (2005). Encoding frequency modulation to improve cochlear implant performance in noise. *IEEE Transactions on Biomedical Engineering*, 52(1), 64-73.
- Oxenham AJ, & Bacon SP. (2003). Cochlear compression: Perceptual measures and implications for normal and impaired hearing. *Ear and Hearing*, 24,352-366.
- Oxenham, A. J., Rosengard, P. S., & Braida, L. D. (2004). Perceptual consequences of normal and abnormal peripheral compression: Potential links between psychoacoustics and speech perception. *The Journal of the Acoustical Society of America*, 115(5), 2421-2421.

- Parikh, G., & Loizou, P. 2005. The influence of noise on vowel and consonant cues. *The Journal of the Acoustical Society of America*, 118, 3874–3888.
- Patuzzi, R. B., Yates, G. K., & Johnstone, B. M. (1989). Outer hair cell receptor current and sensorineural hearing loss. *Hearing Research*, 42, 47-72.
- Peterson G. E. and Barney H.L. (1952). Control methods used in a study of the vowels. *Journal of Acoustic Society of America*, 24(2), 175-184.
- Phatak, S., & Allen, J. B. (2007). Consonant and vowel confusions in speech weighted noise. *Journal of the Acoustical Society of America*, 121, 2312–2326.
- Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hearing research*, 223(1), 114-121.
- Pichora - Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America*, 97(1), 593-608.
- Plomp R. (1986). A signal-to-noise ratio model for the speech-reception threshold of the hearing impaired. *Journal of Speech Language and Hearing Research*, 29, 14654.
- Plomp, R. & Mimpen (1979). Speech reception threshold for sentences as a function of age and noise level. *Journal of the Acoustical Society of America*, 66, 1333-1342.

- Plomp, R. (1978). Auditory handicap of hearing impairment and the limited benefit of hearing aids. *Journal of the Acoustical Society of America*, 63(2), 533-49.
- Purcell, D. W., John, S. M., Schneider, B. A., & Picton, T. W. (2004). Human temporal auditory acuity as assessed by envelope following responses. *The Journal of the Acoustical Society of America*, 116(6), 3581-3593.
- Rhebergen, K. S., & Versfeld, N. J. (2005). A speech intelligibility index-based approach to predict the speech reception threshold for sentences in fluctuating noise for normal-hearing listeners. *The Journal of the Acoustical Society of America*, 117(4), 2181-2192.
- Ronnberg, J. (2003). Cognition in the hearing impaired and deaf as a bridge between signal and dialogue: A framework and a model. *International Journal of Audiology*, 42, S68-S76.
- Rosen S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society B*, 367-373.
- Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, 133(4), 2431-2443.
- Rout, A. (2006). The effect of spectral enhancement on speech recognition performance of normal-hearing and hearing-impaired individuals. *ProQuest*.

- Salthouse, T. A. (2001). General and specific age-related influences on neuropsychological variables. In F. Boller & S. Cappa (Eds.), *Handbook of neuropsychology* (2nd ed.). Amsterdam: Elsevier.
- Schneider, B. A., Daneman, M., & Pichora-Fuller, M. K. (2002). Listening in aging adults: from discourse comprehension to psychoacoustics. *Canadian Journal of Experimental Psychology*, 56(3), 139.
- Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, 18(7), 559-572.
- Shannon R.V., Zeng F-G., Kamath V., Wygonski J., and Ekelid M. (1995). Speech recognition with primarily temporal cues, *Science*, 270(5234), 303-304.
- Shrivastav MN, Humes LE, & Kewley-Port D. (2006) Individual differences in auditory discrimination of spectral shape and speech-identification performance among elderly listeners. *The Journal of the Acoustical Society of America*, 119, 1131–1142.
- Simpson, S. A., & Cooke, M. (2005). Consonant identification in N-talker babble is a nonmonotonic function of N. *The Journal of the Acoustical Society of America*, 118(5), 2775-2778.
- Simpson, A. M., Moore, B. C. J., & Glasberg, B. R. (1990). Spectral enhancement to

- improve the intelligibility of speech in noise for hearing-impaired listeners. *Acta Oto-Laryngologica. Supplementum*, 469:101-107.
- Smits, C., & Festen, J. M. (2013). The interpretation of speech reception threshold data in normal-hearing and hearing-impaired listeners: II. Fluctuating noise. *The Journal of the Acoustical Society of America*, 133(5), 3004-3015.
- Smooorenburg, G. F. (1992). Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *The Journal of the Acoustical Society of America*, 91(1), 421-437.
- Souza, P. E., & Turner, C. W. (1994). Masking of speech in young and elderly listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*, 37(3), 655-661.
- Stiles, D. J. (2010). Influences of working memory and audibility on word learning and vocabulary acquisition in children with hearing loss. Doctoral dissertation, University of Iowa.
- Stone, M. A. & Moore, B. C. J. (1992). Spectral feature enhancement for people with sensorineural hearing impairment: effects on speech intelligibility and quality. *Journal of Rehabilitation Research and Development*, 29(2), 39-56.
- Strelcyk, O., & Dau, T. (2009). Relations between frequency selectivity, temporal fine-

- structure processing, and speech reception in impaired hearing. *The Journal of the Acoustical Society of America*, 125(5), 3328-3345.
- Summers V, & Molis MR. (2004) Speech recognition in fluctuating and continuous maskers: effects of hearing loss and presentation level. *Journal of Speech, Language, and Hearing Research*, 47:245–256
- Summerfield, Q., Foster, J., & Tyler, R. (1985). Influences of formant bandwidth and auditory frequency selectivity on identification of place of articulation in stop consonants. *Speech Communication*, 4, 213-229.
- Takahashi, G. A., & Bacon, S. P. (1992). Modulation detection, modulation masking, and speech understanding in noise in the elderly. *Journal of Speech, Language, and Hearing Research*, 35(6), 1410-1421.
- Ter Keurs, M., Festen, J. M., & Plomp, R. (1992). Effects of spectral envelope smearing on speech reception. I. *The Journal of the Acoustical Society of America*, 91(5), 2872-2880.
- Tyler, R. S. (1986). Frequency resolution in hearing-impaired listeners. Frequency Selectivity in Hearing, edited by B.C. J. Moore (Academic, London), 309-371.
- Van Engen, K. J., & Bradlow, A. R. (2007). Sentence recognition in native-and foreign-language multi-talker background noise. *The Journal of the Acoustical Society of America*, 121(1), 519-526.

- Verhaeghen, P., Steitz, D. W., Sliwinski, M. J., & Cerella, J. (2003). Aging and dual-task performance: A metaanalysis. *Psychology and Aging*, 18, 443–460.
- Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index. *Journal of the American Academy of Audiology*, 23(10), 779-788.
- Viemeister, N. F. (1979). Temporal modulation transfer functions based upon modulation thresholds. *The Journal of the Acoustical Society of America*, 66(5), 1364-1380.
- Wechsler, D. (1997). Wechsler adult intelligence scale, 3rd Edition (WAIS-III). San Antonio, TX: The Psychological Corporation.
- Wojtczak, M., & Viemeister, N. F. (2005). Forward masking of amplitude modulation: Basic characteristics. *The Journal of the Acoustical Society of America*, 118(5), 3198-3210.
- Woodall, A., & Liu, C. (2013). Effects of signal level and spectral contrast on vowel formant discrimination for normal-hearing and hearing-impaired listeners. *The American Journal of Audiology*, 22, 94-104.
- Yang, J., Luo, F-L., and Nehorai, A. (2002). Spectral contrast enhancement: Algorithms and comparisons. *Speech Communication*, 39:33-46.