Frequency Importance Functions for Words and Sentences in Mandarin Chinese: Implications for Hearing Aid Prescriptions in Tonal Languages

Ming-Wen Kuo

Thesis Submitted for the Degree of Doctor of Philosophy
MARCS Institute
University of Western Sydney
March, 2013
Acknowledgements

First, I would like to thank to my supervisory panel, Dr Christine Kitamura and Dr. Gary Walker from MARCS, UWS and Dr. Teresa Ching from the National Acoustics Laboratories (NAL) for giving me the opportunity to undertake this project. I would also like to thank Professor Harvey Dillon from NAL for his expertise and input in this project. I especially thank my principal supervisor, Christine, for her input, support, and assistance in the research and writing up of this project. I am grateful for her encouragement and for steadfastly refusing to accept anything less than my best.

Thanks to everyone at MARCS and NAL which together, created a fabulous research atmosphere. Special thanks to Mark Seeto at NAL, for his mathematical advice, and the implementation of smoothing procedures for the data; and to Colin Schoknecht and Johnson Chen, and other technical staff at MARCS, Steven Fazio, Donovan Govan, and Lei Ching.

Sincere thanks must also go to all the Mandarin Chinese participants from University of Western Sydney, and at Macquarie University. Without their time, patience and assistance, there would be no thesis.

Special thanks to my family, especially Mum and Dad, and my siblings, Wei-Kai, Meng-Fang and Lan-Chen, for supporting me from afar.

Finally, heartfelt thanks to my housemate Benjawan Kasisopa, and my close friends, Yvonne Leung, Leo Chong and Wei-Chien Huang for supporting me in so many ways.
The work presented in this thesis is, to the best of knowledge and belief, original except as acknowledged in the text. I hereby declare that I have not submitted this material, either in full or in part, for a degree at this or any other institution.

(Signature)
# Table of contents

Acknowledgements ............................................................................. 2

List of tables ......................................................................................... 8

List of figures ....................................................................................... 9

Abbreviations ....................................................................................... 12

Abstract ............................................................................................... 13

Chapter 1 Overview ............................................................................. 1

1.1 The research problem ................................................................. 1

1.2 Structure of the thesis ............................................................... 3

Chapter 2 Physics of sound ................................................................. 5

2.1 Sound waves ............................................................................... 5

2.2 Modifying sound waves in this thesis ........................................ 9

2.2.1 Filtering speech ....................................................................... 9

2.1.2 Signal to noise ratios ............................................................... 10

2.2 Summary ...................................................................................... 11

Chapter 3 Human speech ................................................................. 12

3.1 Speech production ....................................................................... 12

3.2 The speech hierarchy: from phonemes to continuous speech .... 13

3.2.1 Phones, phonemes and tonemes ............................................. 14

3.2.2 Syllables .................................................................................. 15

3.2.3 Morpheme ............................................................................... 15

3.2.4 Continuous speech ................................................................. 15

3.3 Vowels, consonants, lexical tone and prosody ......................... 16

3.3.1 Vowels ................................................................................... 17

3.3.2 Consonants ............................................................................. 18

3.3.3 Speech Prosody ....................................................................... 19

3.3.4 Lexical tone ............................................................................ 20

3.4 Mandarin Chinese ....................................................................... 22

3.4.1 Phonemic system ................................................................... 22

3.4.2 Syllable structures ................................................................. 23
3.4.3 Lexical tone auditory cues ........................................... 25
3.4.4 Concatenated tonal variations ..................................... 26
3.4.5 Lexical tone and prosody ............................................. 27
3.5 Summary ........................................................................... 28

Chapter 4 The auditory system and hearing loss ................................................. 30

4.1 Anatomy and physiology of hearing ............................................. 30
4.1.1 The outer ear ............................................................. 31
4.1.2 The middle ear ........................................................... 31
4.1.3 The inner ear ............................................................. 32
4.2 Hearing loss ..................................................................... 34
4.2.1 Assessment of hearing loss ............................................. 35
4.2.2 Types of hearing loss ..................................................... 36
4.2.3 Degrees of hearing loss ................................................ 37
4.2.4 The impact of hearing loss on understanding speech ......................... 38
4.3 Hearing loss and hearing aids ............................................. 39
4.3.1 Amplification formula .................................................... 41
4.4 Summary ........................................................................... 43

Chapter 5 Hearing aid prescription for better speech intelligibility ......................... 44

5.1 Hearing aid prescription for different languages ............................. 44
5.1.1 Differences in phonemes ................................................. 45
5.1.2 Differences in syllable structure ..................................... 45
5.1.3 Differences in syntax ..................................................... 46
5.2 Types of speech test materials .............................................. 47
5.2.1 Monosyllabic words ....................................................... 47
5.2.2 Sentences ................................................................. 48
5.2.3 Non-meaningful word and sentence tests ............................. 49
5.2.4 Mandarin speech tests .................................................. 49
5.2.5 Speech recognition with competing noise ......................... 51
5.3 Measuring speech intelligibility using ANSI standards .................... 52
5.3.1 The speech intelligibility index ...................................... 53
5.3.2 The transfer function .................................................... 56
5.4 The frequency importance function ................................................. 57
5.4.1 Filter slope. ........................................................................... 58
5.4.2 Noise. .................................................................................... 59
5.4.3 Transducer bandwidth ............................................................ 59
5.4.4 Scoring methods ................................................................. 60
5.4.5 Data smoothing ................................................................. 61
5.5 Factors which affect FIFs .......................................................... 62
5.5.1 Gender of the talker ............................................................. 62
5.5.2 Phonetic content in speech materials ................................. 63
5.5.3 Context and redundancy ..................................................... 64
5.5.4 Response format ............................................................... 65
5.5.5 Mandarin Chinese FIF ......................................................... 65

Chapter 6 Overview of this study ..................................................... 68
6.1 Experimental plan .................................................................... 68
6.2. Predictions .............................................................................. 70
6.2.1. Predictions for performance-intensity functions ............ 70
6.2.2 Predictions for frequency importance function ................... 71
6.2.3 Predictions for SII transfer function ..................................... 73
6.3 Conclusion ................................................................................ 74

Chapter 7 Method .......................................................................... 75
7.1 Method ..................................................................................... 76
7.1.1 Mandarin Chinese speech materials .................................. 76
7.1.2 Recording procedure ......................................................... 76
7.1.3 Token selection and editing .............................................. 78
7.1.4 Speech filtering and noise additive procedures ............... 78
7.1.5 Test conditions ................................................................. 79
7.1.6 Pilot study: method ............................................................ 81
7.1.6.1 Participants .................................................................. 81
7.1.6.2 Speech materials ....................................................... 81
7.1.6.3 Procedure ................................................................. 81
7.1.7 Pilot study: results ............................................................. 82
7.2 Method for speech recognition test ........................................ 82
7.2.1 Participants ......................................................................... 82
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.2.2 Hearing evaluation</td>
<td>84</td>
</tr>
<tr>
<td>7.2.3 Apparatus</td>
<td>84</td>
</tr>
<tr>
<td>7.2.4 Speech recognition test materials</td>
<td>85</td>
</tr>
<tr>
<td>7.2.5 Procedure</td>
<td>87</td>
</tr>
<tr>
<td>7.3 Scoring and reliability</td>
<td>88</td>
</tr>
<tr>
<td>7.4 Overview of FIF development</td>
<td>88</td>
</tr>
</tbody>
</table>

**Chapter 8 Results** ................................................................. 90

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 Performance intensity functions</td>
<td>91</td>
</tr>
<tr>
<td>8.2 Developing Mandarin frequency importance functions</td>
<td>92</td>
</tr>
<tr>
<td>8.2.1 Speech recognition scores</td>
<td>92</td>
</tr>
<tr>
<td>8.2.2 Cross-over frequencies</td>
<td>95</td>
</tr>
<tr>
<td>8.2.3 Relative transfer function</td>
<td>99</td>
</tr>
<tr>
<td>8.2.4 Frequency importance functions</td>
<td>105</td>
</tr>
<tr>
<td>8.2.5 Absolute transfer function</td>
<td>109</td>
</tr>
<tr>
<td>8.3 Summary</td>
<td>113</td>
</tr>
</tbody>
</table>

**Chapter 9 Discussion** .................................................................. 115

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1 Overview of findings</td>
<td>115</td>
</tr>
<tr>
<td>9.1.1 Overview</td>
<td>115</td>
</tr>
<tr>
<td>9.2 Performance intensity functions</td>
<td>117</td>
</tr>
<tr>
<td>9.3 Cross-over frequencies</td>
<td>119</td>
</tr>
<tr>
<td>9.4 Frequency importance functions</td>
<td>121</td>
</tr>
<tr>
<td>9.4.1 Mandarin FIFs in this study</td>
<td>121</td>
</tr>
<tr>
<td>9.4.2 Word FIFs: comparing Mandarin Chinese and English</td>
<td>123</td>
</tr>
<tr>
<td>9.4.3 Mandarin versus English and Cantonese sentences</td>
<td>127</td>
</tr>
<tr>
<td>9.5 Transfer functions</td>
<td>130</td>
</tr>
<tr>
<td>9.6 Implications for Mandarin hearing aid prescription formulas</td>
<td>133</td>
</tr>
<tr>
<td>9.7 Future research directions</td>
<td>134</td>
</tr>
<tr>
<td>9.8 Summary and conclusions</td>
<td>135</td>
</tr>
</tbody>
</table>

**References** .................................................................................. 138

**Appendices** .................................................................................. 153
List of tables

Table 4.1. Classification of hearing impairment levels .................................................................37
Table 5.1. List of CFs (Hz) for different phonemic features ....................................................63
Table 7.1. Average hearing threshold of participants for left and right ears ..................................83
Table 7.2. Shows the word and sentence subtest lists allocated to HP (top) and LP (bottom) frequency bandwidth and SNR conditions .................................................................86
Table 8.1. Mean percentage recognition scores (%) for sentences for HP (left) and LP (right) filtered conditions ........................................................................................................93
Table 8.2. Mean percentage recognition scores (%) for words for HP (left) and LP (right) filtered conditions ........................................................................................................93
Table 8.3. Mean percentage recognition scores (%) for segments for HP (left) and LP (right) filtered conditions .........................................................................................................94
Table 8.4. Mean percentage recognition scores (%) for tones for HP (left) and LP (right) filtered conditions ...............................................................................................................94
Table 8.5. CFs (Hz), for sentences, words, segments and tones under different SNR conditions ...............................................................................................................................98
Table 8.6. Relative transfer functions for predicting Score and SII and fitting constants for different speech materials .................................................................................................102
Table 8.7. Sentence SII as a function of SNR and frequency bandwidth. The final weight or importance (%) for each band was calculated from the average SII of all SNRs for each frequency bandwidth .................................................................106
Table 8.8. Importance weighting of each octave band for sentences, words, segments and tones .........................................................................................................................107
Table 8.9. Fitting constants for the best fit TFs ..............................................................................111
Table 9.1. Comparison of TFs. TF slopes are based on observed scores between 20% - 80%. The TF slope is calculated as percentage (%) increase per .033 SII ................................132
List of figures

Figure 2.1. A tuning fork produces a sinusoidal sound wave which transfers vibrating motions through air particles. The sound wave (longitudinal and transverse views) is characterized by the movement of individual particles to produce wave propagation (Gelfand, 2004). .................................................................6
Figure 2.2. The root-mean-square, peak, and peak-to-peak amplitudes (Gelfand, 2004). .7
Figure 2.3. Examples of four commonly used filters (Haughton, 2002). .........................10
Figure 3.1. The resonance cavities along the vocal tract serve as a filter for the vocal folds. Retrieved from Speech Science (Borden, Raphael, & Harris, 2003)..........................13
Figure 3.2. An English speech spectrum (Haughton, 2002). .............................................16
Figure 3.3. The criteria to describe vowels: the tongue in the direction of the palate, the extent direction to the tongue rises towards the palate, generally, high, mid or low, the high/low raising of the soft palate, and the type of opening, rounded/unrounded/spread, made by the lips (MacNeilage & Davis, 2000) .........................................................17
Figure 3.4. The source-filter theory of speech production: The source spectrum represents (a) the spectrum of typical glottal air flow with air passing through it and (b) with a transfer function or vocal tract filter output to (c) provide an energy spectrum which represents characteristics of vocal tract shaped by articulators for F1 and F2 (Gelfand, 2004). ......18
Figure 3.5. Mean F0 contours of the four Mandarin tones on monosyllable /ma/. The duration of tone1, 2, and 4 are normalized against tone 3 (Xu, 1997).............................24
Figure 3.6. The third-tone sandhi rule using /ma ma/ sequence in Mandarin where each tone shape is the average of 48 contours from 12 speakers. The F0 contours of the first instance of tone 3 syllable /ma/, are typically modified at common point 25% into the segment (Xu, 1997). .....................................................................................................................26
Figure 3.7. Effects of first tone on F0 contour of the following syllable in /ma ma/ sequences in Mandarin where the tone in the second syllable is held constant, and the tone of the first syllable is varied. The large variations of the onset F0 (50% of the segment) of the second syllable is presented in each panel. ..................................................................................................................................27
Figure 4.1. Cross-sectional view of the human ear showing the divisions of the outer, middle and inner ear.................................................................30
Figure 4.2. Diagram of an uncoiled cochlea (Martin & Clark, 2006).................................32
Figure 4.3. The flow of sound waves of three different frequencies (Yost, 2007). ....33
Figure 4.4. Cross-sectional view of the cochlea (Martin & Clark, 2006)............................34
Figure 4.5. Schematic of a hearing aid (Katz, Burkard, & Medwetsky, 2002). .................40
Figure 4.7. A diagram of a hearing aid receiver (Dillon, 2001a)......................................41
Figure 4.6. Cross-sectional view of a hearing aid microphone (Dillon, 2001a)..............40
Figure 5.1. Performance intensity functions showing the effects of redundancy and contextual clues on digits, words in sentences and nonsense syllables (Miller, Heise, & Lichten, 1951). ...........................................................................................................52
Figure 5.2. shows how the three factors (W, D, I) are related for calculation of a SII. In panel (A) the range of LTASS are plotted on the audiogram to compare with the listener’s threshold, and are shown as black bars for each frequency threshold at the audible dynamic range or W. In panel (B) the distortion factor is shown as a function of loudness or D; and (C) shows the relative weighting of different frequencies bands. Retrieved from ANSI – S3.5 (1997).

Figure 5.3. Comparison of transfer functions derived from three different speech materials: monosyllabic words from French and Steinberg (1947); nonsense syllables from Black (1959); continuous discourse from Studebaker (1987).

Figure 5.4. Examples of English language FIFs for monosyllabic words (Black, 1959) and continuous discourse (Studebaker, Pavlovic, & Sherbecoe, 1987). The relative percentage of speech intelligibility importance is shown as a function of each frequency cut-off.

Figure 5.5. The FIFs for monosyllabic words, sentences, and continuous discourse. The FIF for continuous discourse reveals a higher broader span compared to monosyllabic words; whereas, the FIF for sentences is intermediate (DePaolis, Janota, & Frank, 1996).

Figure 7.1 Shure SM10a Headset Dynamic Microphone.

Figure 7.2. A display of waveforms of a HINT sentence in Channel 1 (top), and noise profile in Channel 2 (bottom).

Figure 8.1. Performance intensity functions showing the effects of SNR and sentence, word, segment and tone as a function of percentage correct.

Figure 8.2. Sentence recognition scores plotted as a function of cut-off frequency at different SNRs. The solid lines indicate the smoothed values; the dashed lines indicate the unsmoothed values. The left panel displays SNRs +9, +3 and -3 dB; the right panel displays SNRs +6, derived +1, 0, and -6 dB. The location of CF is marked on the left panel. The procedures used to derive points on the SII relative transfer function are plotted at the right panel.

Figure 8.3. Word recognition scores plotted as a function of cut-off frequency at different SNRs. The solid lines indicate the smoothed values; the dashed lines indicate the unsmoothed values. The left panel displays SNRs +12, +6, 0 and -6 dB; the right panel displays SNRs +15, +9, +3 and -3 dB. The location of CF is marked in the left panel. The procedures used to derive points on the speech intelligibility transfer function are plotted at the right panel.

Figure 8.4. Segment recognition scores, plotted as a function of cut-off frequency at different SNRs. The solid lines indicate the smoothed values; the dashed lines indicate the unsmoothed values. The left panel displays SNRs +12, +6, 0 and -6 dB; the right panel displays SNRs +15, +9, +3 and -3 dB. The location of CF is marked on the left panel. The procedures used to derive points on the speech intelligibility transfer function are plotted at the right panel.

Figure 8.5. Tone recognition scores plotted as a function of cut-off frequency at different SNRs. The solid lines indicate the smoothed values; the dashed lines indicate the unsmoothed values. The left panel displays SNRs +15, +9, +3 and -6 dB; the right panel displays SNRs +12, +6, 0 and -3 dB. The location of CF is marked on the left panel. The procedures used to derive points on the speech intelligibility transfer function are plotted at the right panel.

Figure 8.6. Relative transfer function for sentences. The left panel transfer function predicts the % correct score from SII, is described by Equation 1 and has Q and N values of 0.53 and 3.89, respectively; the right panel transfer function predicts the SII from % correct score, is described by Equation 2 and has Q and N values of 0.44 and 5.74, respectively.
**Figure 8.7.** Relative transfer function for words. The left panel transfer function predicts the % correct score from SII, is described by Equation 1 with Q and N values of 0.86 and 2.23, respectively; the right panel transfer function predicts the SII from % correct score, is described by Equation 2, and has Q and N values of 0.68 and 3.20, respectively. 103

**Figure 8.8.** Relative transfer function for segments. The left panel transfer function predicts the % correct score from SII, is described by Equation 1 and has Q and N values of 0.82 and 2.43, respectively; the right panel transfer function predicts the SII from % correct score, is described by Equation 2 and has Q and N values of 0.69 and 3.18, respectively. 104

**Figure 8.9.** Relative transfer function for tones. The left panel transfer function predicts the % correct score from SII, is described by Equation 1, and has Q and N values of 0.42 and 0.88, respectively; the right panel transfer function predicts the SII from % correct score, is described by Equation 2, and has Q and N values of 0.46 and 0.73, respectively. 104

**Figure 8.10.** The frequency importance function for sentences, words, segments and tones. 108

**Figure 8.11.** Comparison of cumulative FIFs for sentences, words, segments and tones. 109

**Figure 8.15.** Absolute TF for lexical tones using constant, K=9, is best described by Equation 2, with values for Q and N of 0.30 and 1.51, respectively. 112

**Figure 8.14.** Absolute TF for segments using constant, K=8, is best described by Equation 2, with values for Q and N of 0.55 and 2.92, respectively. 112

**Figure 8.13.** Absolute TF for words using constant, K=7, is best described by Equation 2, with values for Q and N of 0.52 and 2.90, respectively. 112

**Figure 8.12.** Absolute TF for sentences using constant, K=6.8, is best described by Equation 2, with values for Q and N of 0.23 and 4.06, respectively. 112

**Figure 8.16.** Comparison of absolute TFs for the four speech materials. 113

**Figure 9.1.** Word FIFs for the current study and four other studies: (1) English, W-22 (Studebaker & Sherbecoe, 1991) (2) English, NU-6 (Studebaker, et al., 1993) (3) English, PB50 (DePaolis et al., 1996) (4) Mandarin, 50 words (own lists) (Yeh, 2005). 126

**Figure 9.2.** Cumulative word FIF comparing the current study to four similar studies: (1) English, W-22 (Studebaker & Sherbecoe, 1991) (2) English, NU-6 (Studebaker, et al., 1993) (3) English, PB50 (DePaolis et al., 1996) (4) Mandarin, 50 words (own lists) (Yeh, 2005). 126

**Figure 9.3.** Comparison of the current study MHINT sentence frequency importance functions to other two similar studies: English-SPIN (DePaolis, et al., 1996) and Cantonese HINT (Wong, Ho, et al., 2007). 129

**Figure 9.4.** Cumulative sentence FIFs for the current study compared to other studies, (1) English, SPIN (DePaolis, et al., 1996) and (2) Cantonese HINT (Wong, Ho, et al., 2007). 130

**Figure 9.5.** The TF of Mandarin words from MSTMs in this study is shown in comparison with other reported TFs: (1) English words from W-22 (Studebaker & Sherbecoe, 1991) and (2) NU-6 (Studebaker, et al., 1993); and (3) English, using the PB50 test (DePaolis, et al., 1996); and (4) Mandarin words (own lists) (Yeh, 2005). 131

**Figure 9.6.** The TF of Mandarin sentences from MHINT in this study is shown in comparison with other reported sentence TFs: (1) English sentences from SPIN (DePaolis, et al., 1996) and (2) HINT (Eisenberg, et al., 1998); and (3) CHINT sentences (L. L. Wong, Ho, et al., 2007). 132
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI</td>
<td>Articulation Index</td>
</tr>
<tr>
<td>ANSI</td>
<td>American National Standard Institute</td>
</tr>
<tr>
<td>ASHA</td>
<td>American Speech-Language-Hearing Association</td>
</tr>
<tr>
<td>CHINT</td>
<td>Cantonese Hearing In Noise Test</td>
</tr>
<tr>
<td>CF</td>
<td>Crossover Frequency</td>
</tr>
<tr>
<td>CID</td>
<td>Central Institute for the Deaf (Everyday Speech Sentences)</td>
</tr>
<tr>
<td>CVC</td>
<td>Consonant-Vowel-Consonant</td>
</tr>
<tr>
<td>CV</td>
<td>Consonant-Vowel</td>
</tr>
<tr>
<td>dB</td>
<td>Decibel</td>
</tr>
<tr>
<td>DSL [i/o]</td>
<td>Desired Sensation Level input/output formula</td>
</tr>
<tr>
<td>F0</td>
<td>Fundamental Frequency</td>
</tr>
<tr>
<td>FIF</td>
<td>Frequency Importance Function</td>
</tr>
<tr>
<td>HI</td>
<td>Hearing Impaired</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>HP</td>
<td>High Pass</td>
</tr>
<tr>
<td>LTASS</td>
<td>Long Term Average Speech Spectrum</td>
</tr>
<tr>
<td>LP</td>
<td>Low Pass</td>
</tr>
<tr>
<td>MHINT</td>
<td>Mandarin Hearing In Noise Test</td>
</tr>
<tr>
<td>MMRT</td>
<td>Mandarin Monosyllable Recognition Test</td>
</tr>
<tr>
<td>MSTMs</td>
<td>Mandarin Speech Test Materials</td>
</tr>
<tr>
<td>NAL-NL1</td>
<td>National Acoustic Laboratories- Nonlinear, version 1 formula</td>
</tr>
<tr>
<td>NU-6</td>
<td>Northwestern University test no.6</td>
</tr>
<tr>
<td>PB-50</td>
<td>Phonetically Balanced test</td>
</tr>
<tr>
<td>RMS</td>
<td>Root-Mean-Square</td>
</tr>
<tr>
<td>RTF</td>
<td>Relative Transfer Function</td>
</tr>
<tr>
<td>SBS</td>
<td>Special Broadcasting Service</td>
</tr>
<tr>
<td>SHM</td>
<td>Simple Harmonic Motion</td>
</tr>
<tr>
<td>SII</td>
<td>Speech Intelligibility Index</td>
</tr>
<tr>
<td>SNR</td>
<td>Signal-to-Noise Ratio</td>
</tr>
<tr>
<td>SOV</td>
<td>Subject-Object-Verb</td>
</tr>
<tr>
<td>SVO</td>
<td>Subject-Verb-Object</td>
</tr>
<tr>
<td>SPIN</td>
<td>Speech Perception In Noise Tests</td>
</tr>
<tr>
<td>SRT</td>
<td>Speech Recognition Test</td>
</tr>
<tr>
<td>SSI</td>
<td>Synthetic Sentence Identification test</td>
</tr>
<tr>
<td>TF</td>
<td>absolute Transfer Function</td>
</tr>
<tr>
<td>WDP</td>
<td>Word Discrimination Performance</td>
</tr>
<tr>
<td>W-22</td>
<td>CID W-22 word lists</td>
</tr>
</tbody>
</table>
Abstract

Hearing aids increase the amount of speech perceived by the listener by amplifying the speech signal using a prescription that takes into account the type of hearing loss, and the frequency importance function (FIF) for the listeners’ language. An FIF describes how each frequency band contributes to speech intelligibility and is a reflection of the structure of the input language. Therefore, for Mandarin, the FIF may be affected not just by its phonemic structure, but also by the inclusion of low-frequency lexical tone. To date, most FIFs have been developed for English language background listeners. This project aims to fill a knowledge gap by providing a systematic evaluation of how the language features of Mandarin Chinese contribute to FIF for hearing aid prescription. Hence, this project uses series of speech recognition scores to develop Mandarin FIFs to examine weightings for lexical tone, phonetic content in words, and sentences, together with other functions to determine how the language factor (Mandarin Chinese) affects the estimation of speech intelligibility index (SII).

Speech stimuli were recorded by a male broadcaster using monosyllabic words from Mandarin Speech Test Materials (MSTMs) and ten-word sentences from the Chinese Mandarin Hearing in Noise Test (CHINT). The speech was passed through a broadband digital filter to separate the speech signal into high- and low-pass conditions for discrete frequency bands (141, 224, 355, 562, 891, 1413, 2239, 3548, 5623, and 8913 Hz) and speech noise added to create 9 conditions for words and 7 conditions for sentences signal-to-noise ratios (SNRs: 15, 12, 9, 6, 3, 0, -3,-6,-9 dB). In the speech recognition test, normal hearing Mandarin speaking participants were asked to orally repeat words and sentences; and in the word session, to indicate the tone used on the screen. These data are used to determine performance intensity
functions, the FIFs of tones, segments, words, and sentences. Lastly, the derived FIFs were incorporated with the published Long Term Average Speech Spectrum (LTASS) to evaluate SII transfer functions.

Results of this project suggest that language specific inventories of phonemes and tones, in addition to context and redundancy effects, influence performance intensity functions, CFs, FIFs and TF slopes in Mandarin. The Mandarin CFs for the four speech types are remarkably similar for words (1807 Hz) and segments (1813 Hz), but lower for sentences (1570 Hz) and tones (743 Hz). FIF data showed a primary peak centered at ~1800 Hz and secondary peak at ~112 Hz for word and sentence materials. The Mandarin TFs indicate that tones, with high redundancy had the steepest slope, followed by sentences, then by words and segments, which both having similar shallow slopes.

This study is the first tonal language study to separate the effects of words, segments and tones on performance intensity functions, FIFs, and TFs. The reported findings provide important new directions for developing hearing aid prescriptions for tonal language speakers, but specifically for Mandarin Chinese.
Chapter 1 Overview

1.1 The research problem

Most patients with hearing loss are introduced to amplification devices in order to help understand speech. Among the types of amplification devices, hearing aids are the most common and acceptable intervention strategy prescribed by audiologists. A hearing aid plays the role of altering the input level of signals in order to increase the amount of speech perceived. Over the past two decades, the rapid growth of technology in English speaking countries has made hearing aids increasingly more efficient and sophisticated for hearing loss patients (Levitt, 1987, 2007; Levitt, Neuman, & Sullivan, 1990). However, satisfaction research in non-English speaking countries, such as China, has reported that Cantonese speaking Chinese with hearing loss only use hearing aids approximately four hours per day, in contrast those with hearing loss from English speaking countries who average seven or more hours use per day (Bentler, Niebuhr, Getta, & Anderson, 1993; Humes, Garner, Wilson, & Barlow, 2001; Munro & Lutman, 2004; Wong, Hickson, & McPherson, 2009). Recent data suggest that, for Mandarin speakers, speech perception might be improved if the prescription included more low frequency information to allow access to lexical tone (Fu, Zeng, Shannon, & Soli, 1998; Nilsson, Soli, & Sullivan, 1994; Zhang & McPherson, 2008). It is clear that hearing aid prescriptions are necessary for tonal language speakers in order to improve their level of satisfaction and their speech perception. This project will examine how the tonal nature of Mandarin Chinese affects the composition of the frequency importance function, an important hearing aid prescription element.
The loss of hearing varies greatly in its degree and type. For example, hearing loss can be categorized as mild, moderate, severe, profound, high-frequency or low-frequency loss. Consequently, an audiologist needs to adjust the hearing aid to the appropriate level amplification for each hearing impaired person. Clinically, the common strategy for reaching such a goal is the use of a prescriptive formula through reference to the language type, plus the characteristics of the hearing loss. Several prescriptive formulas are widely used and have been effectively tested on English speakers (Byrne, 1986; Byrne, Dillon, Ching, Katsch, & Keidser, 2001; Ching, 2002; Harvey Dillon, 2001b; Scollie, Ching, Seewald, Dillon, Britton, Steinberg, & Corcoran, 2010; S. Scollie, Ching, Seewald, Dillon, Britton, Steinberg, & King, 2010). However, the efficacy of prescriptive formulas which have been developed with reference to speakers of English do not translate well to speakers of tonal languages, such as Mandarin Chinese, where lexical tone is an integral part of the language. Unlike English, tonal languages require access to low-frequency information contained in fundamental frequency. Thus, there is the need to investigate whether a different type of hearing aid prescription is required for Mandarin speakers.

The English hearing aid prescriptive formula is based upon the English speech intelligibility index (SII) and its frequency importance function (FIF). The SII was first developed as the articulation index, and was used for estimating which frequencies were most important and should be amplified for efficient communication using the telephone (French & Steinberg, 1947). Later, the SII was also applied to evaluate the efficacy of hearing aid fitting, and then incorporated into hearing aid prescriptive formulas. One of the key components of the SII is the FIF. A FIF indicates the relative importance of each frequency band.
in the speech spectrum, and the degree to which it contributes to speech intelligibility. For example in English, the frequency importance curve peaks at 1800 Hz. Therefore, the FIF for a specific language is a reflection of the nature of the language. The English-based FIF is incorporated into the prescriptive formula in hearing aids worn by non-tonal and tonal language speakers. This occurs despite the fact that tonal languages use F0 to distinguish the meaning of lexical items. Thus, it might be assumed that the FIF for non-tonal languages such as English and tonal languages such as, Mandarin Chinese might differ. This project will focus on developing FIFs for Mandarin Chinese in two contexts: monosyllabic words and sentences.

1.2 Structure of the thesis

The introductory chapters provide background literature necessary to establish the research context. These chapters provide an overview of how human speech is related to current hearing aid technology.

Chapter 2 provides brief overview of the literature concerning the physics of sound, including details of the physical properties of sound, and the underlying principles of speech filtering and adding noise to speech signals as they relate to this project.

Chapter 3 describes the source-filter model of speech production, including the levels of speech used in this thesis, and their components, vowels, consonants and tones. It concludes with details of the lexical tone system used in Mandarin Chinese, with particular focus on research in respect to tonal languages.

Chapter 4 reviews the physiology of the human auditory system, its pathologies, together with an overview of commonly used methods of assessing hearing, different types of hearing loss, and the resulting difficulties in
understanding speech. It also discusses the physical characteristics of hearing aids, and provides an overview of different amplification formulas.

Chapter 5 reviews the development of hearing aid technology, describes how language specific information affects hearing aid prescription, and the types of tests used to test speech intelligibility. In addition, there is an overview of the SII equation and its key components, the FIF. Finally, published FIFs are reviewed, together with current unresolved issues in tonal language FIFs.

Chapter 6 outlines the approach taken in this study, including the experimental plan and the predictions associated with the performance intensity functions, the FIFs and SII.

Chapter 7 describes the method for recording and editing words and sentences for use in the speech recognition test, and the procedure for collecting scores for the purpose of developing Mandarin word and sentence FIFs. This includes collecting three different recognition scores for words: (i) word score (correct repetition of segment + tone); (ii) tone score (correct identification of tone); (iii) segment score (correct repetition of word irrespective of the tone used), and for the sentence test: the recognition scores consist of the number of words correctly identified in a 10-word sentence.

Chapter 8 presents the development of the FIF and TF in respect to the effects of words, segments and tones on their characteristics.

In conclusion, Chapter 9 provides a general discussion of the experimental findings, their comparisons to other languages and their implications for hearing aid prescription.
Chapter 2 Physics of sound

When sound waves reach human ears, it produces a physiological and psychological response. The study of psychoacoustics focuses on the relationship between physical stimuli and psychological responses, and includes, among other things the perception of pitch or the experience of frequency of sound waves, (ii) loudness or the experience of intensity/amplitude of sound waves, (iii) durational qualities experienced as rhythm or tempo and (iv) timbre or the experience of the quality of the emitting object, e.g., piano versus violin. The physical characteristics of sound waves that determine the perception of timbre include the frequency components and the envelope composition of the amplitude of single and complex sound waves.

A basic understanding of the acoustics of sound and its fundamental physical properties is a prerequisite for understanding speech and hearing. In the following sections, a number of physical principles will be reviewed including basic definitions and concepts that are used in this thesis.

2.1 Sound waves

Sound maybe generated by an object, that has the properties of elasticity and mass, which in turn creates vibrating patterns in the particles of medium that are in contact with the object. These take the object the form of pressure waves in the medium. Thus, when an external force is applied to an object, such as a tuning fork, or vocal folds, the resulting to-and-fro motions enter the surrounding air medium to create sound waves. For example, in Figure 2.1, when the striking of a tuning fork initiates a vibration, the surrounding air particles are alternately
pushed together and pulled apart to form sound waves. When these waves reach the ears, they move the eardrum producing the sensation of hearing a sound.

The motion produced by an object’s vibration that results in the compression and rarefaction of air particles is described as simple harmonic motion (SHM) because each vibrating particle continues repeating the to-and-fro motion around its own resting points. SHMs produce waves with sinusoidal characteristics. In other words, sine waves are commonly used to describe waves resulting from SHMs. These waves are categorized as longitudinal waves or waves moving along the same axis as medium particles. In contrast to longitudinal waves, transverse waves are often what is shown schematically. As shown on the top of Figure 2.1, the high points of the curves above the resting line depict the compression area or the pushing force, in contrast to the pulling force (decrease in intensity or pressure) which represents the rarefaction area.

![Diagram of wave propagation](image)

Figure 2.1. A tuning fork produces a sinusoidal sound wave which transfers vibrating motions through air particles. The sound wave (longitudinal and transverse views) is characterized by the movement of individual particles to produce wave propagation (Gelfand, 2004).

The length of one to-and-fro cycle of a wave is called wavelength (λ). It is calculated as the distance between the point of maximum displacement of one wave and the next in the same direction. A period is defined as the time it takes
each participle to complete a cycle. In addition, the number of cycles that occur within a time period (usually a second) is called frequency \(1/t=\) frequency). Frequency is described as the number of cycles per second and expressed as hertz (Hz). If frequency is doubled, it is defined as an increase of one octave.

Instantaneous amplitude is the magnitude of displacement of the medium’s particles as they change from moment to moment. The general amplitude characteristics of a sine wave are shown in Figure 2.2 and include: (1) peak to peak amplitude or overall displacement from one extreme point to the other; (2) peak amplitude or the magnitude from the resting point to the extreme displacement in one direction and (3) root-mean-square (RMS) amplitude or square root of the mean of the squared instantaneous amplitudes over one cycle. The rms amplitude of a sinusoid wave is equal to 0.707 times the peak amplitude.

![Figure 2.2. The root-mean-square, peak, and peak-to-peak amplitudes (Gelfand, 2004).](image)

A single sinusoidal sound wave seldom appears in nature. Most sounds, speech, music, and noises, are composed of waves with different frequencies, amplitudes and timing relationships. When combining two or more sinusoid
waves, a complex wave is formed. Depending on the frequency relationships between the component waves, the outcome sound wave will vary. The general rules are: (1) Combining waves of vibrating particles moving in the same direction (i.e. phase) provides reinforcement and increases amplitude; whereas waves made up of particles in opposite directions (i.e. out phase) results in cancellation and reduces amplitude. (2) If complex waves are composed of waves with a multiple frequency relationship, that is, if they contain harmonics, this will result in a periodic sound wave. The lowest frequency component is known as fundamental frequency (F0). The harmonics are component frequencies that are integer multiples of the fundamental frequency, for example, if F0 (the fundamental frequency) is 100 Hz, the harmonics will be at 200, 300, 400 Hz etc. On the other hand, if sound waves are combined that are not harmonically related, the resultant waveform will not repeat over time and is known as an aperiodic waveform. Noises are examples of aperiodic sound waves.

Humans hear when sound pressure variations are applied to the ear drum. The greater the pressure the louder is the sound that is perceived. The most intense sound that humans can tolerate is on the order of 10 million times more greater sound pressure than that of the softest audible sound. Accurate for pressure range overall 120 dB. Thus, a logarithmic (log) not linear scale is used to describe the difference in amplitude between two sounds. Typically the decibel scale expresses the sound pressure level as shown in the formula (ANSI S1.1-1994):

$$\text{SPL}_{dB} = 20 \log_{10} \left( \frac{P_{\text{rms}}}{P_{\text{ref}}} \right) \text{ dB}.$$  

In the above formula, $P_{\text{ref}}$ is the reference sound pressure unit and $P_{\text{rms}}$ is the rms sound pressure being measured. The reference unit is $P_{\text{ref}} = 20 \mu \text{Pa}$ (rms),
which is usually considered to be near the minimum intensity perceived by humans, depending upon frequency.

2.2 Modifying sound waves in this thesis

In this thesis, participants are asked to recognize words and sentences that have been high- and low-pass filtered and have different signal-to-noise ratios. These are described as follows:

2.2.1 Filtering speech

Filtering a sound wave involves modifying or subtracting the amplitude of frequency components of complex sound waves. There are at least four types of filters: (1) High pass (HP)/low cut filters are designed to allow frequencies above a certain frequency point to remain unmodified, and to attenuate frequencies below that frequency point. (2) Low pass (LP)/high cut filters are the opposite of (1), that is, they leave lower frequencies intact, and attenuate higher frequencies. (3) Band pass filters allow a range of frequencies to remain intact while frequencies above and below the defined band are attenuated. (4) Band reject filters remove frequencies within a frequency range and leave the frequencies outside that range intact. Examples are shown in Figure 2.3. An ideal filter should allow all designated frequencies to pass equally without affecting the original intensities, that is, have minimum ripple (<1 dB). Filter slope is also important, e.g., when a high pass filter is set at 1000 Hz, frequencies below the cut-off frequency are attenuated. Generally it is better to have a steeper slope to attenuate neighboring frequencies just outside the designated filter cutoff frequency.
2.1.2 Signal to noise ratios

Signal-to-noise ratio (SNR) is a term that describes the relationship between the intensity level of the speech signal and the level of background noise. It is demonstrated that speech becomes more intelligible when either the intensity of the speech increases in relation to the background noise or when the level of noise decreases in relationship to that of the speech signal (French & Steinberg, 1947). A ratio higher than 1:1 or an SNR of 0 indicates more speech signal than noise. A SNR of 30dB is deemed to be effectively a clean signal. The current project involves measuring Mandarin speech recognition performance under different SNRs that range from -9 dB to 15 dB. These data will then be served as a
basis for understanding whether the Mandarin hearing aid prescription can be modified to achieve better speech intelligibility.

2.2 Summary

Sound is generated by creating vibrating patterns in the particles of a medium, usually the medium is air. These vibrating patterns or pressure waves in the air reach the ears and are psychologically interpreted as pitch, loudness and timbre. In speech, vocal fold constrictions cause different vibration patterns that are modified by the mouth, lips and tongue to produce sounds that are recognized as speech. In the next chapter, different aspects of the speech production process will be discussed, including important details involved in the production of Mandarin Chinese, the language to be used in this study.
Chapter 3 Human speech

In oral communication, human beings use speech, a series of sequential sounds, in order to exchange ideas, feelings, and information. The production of human speech relies on physical organs, such as the lips, tongue, vocal folds, and lungs to articulate and produce speech, and on auditory organs to hear and perceive, and understand the meaning of the speech signal. The term generally used to describe the finer details of spoken speech sounds is articulatory or acoustic phonetics. Speech sounds are only described phonologically when referring to speech sounds or phonemes of a particular language. This chapter will focus on the acoustic aspects of speech production, which are discussed below. In addition, the discussion will turn to the speech hierarchy, including speech segments, syllables, morphemes, and continuous speech, and then to speech at the level of consonants, vowels, tones, and prosody. This is followed by current research findings on the study of tonal languages, and on Mandarin Chinese, the language investigated in this thesis.

3.1 Speech production

The source-filter model of speech production is commonly accepted as how the vocal tract produces the components (vowels, consonants, tones) of speech (Kent & Read, 2002). At its base, this model proposes that during the production of speech, pulmonic pressure produced by the lungs serves as the energy source for vibrations in the vocal folds (see Figure 3.1). The vibrating frequency of the vocal folds is commonly called the fundamental frequency (F0), and is determined by the elasticity, tension, and mass of the folds. Therefore, an individual can manipulate the tension and stiffness of their vocal folds by using
the laryngeal muscles to change the resulting F0. Furthermore, the acoustic signal generated in the larynx will be filtered by the vocal tract cavities above the glottis. The vocal tract muscles and articulators (lips, teeth and the tongue etc.) then shape the resonating cavities to create the various consonants and vowels.

![Image](image.png)

Figure 3.1. The resonance cavities along the vocal tract serve as a filter for the vocal folds. Retrieved from Speech Science (Borden, Raphael, & Harris, 2003).

The quality of speech output is highly related to the physiologic structure which varies by individual characteristics, such as sex, and age. Children, men, and women have differing speech production systems in terms of the size of the vocal folds and the volume of the lungs. In general, male vocal folds have higher mass and longer length (17 - 25 mm) to produce F0 ranging from about 100Hz to 150Hz. The female vocal tract (12.5 mm to 17.5 mm) produces F0 ranging from 170Hz to 220Hz. Children have even higher F0, but as a child progresses to adulthood, the vocal folds increase in length to gradually lower the produced F0 (Abitbol, Abitbol, & Abitbol, 1999; Keating & Buhr, 1978).

### 3.2 The speech hierarchy: from phonemes to continuous speech

In 1886, the International Phonetic Association was formed to promote the scientific study of phonetics. From this long tradition, has come a speech hierarchy which means speech can be analyzed at many different levels. For the purposes of this thesis, the following levels of the speech hierarchy are pertinent
and are described below: phones/phoneme/segment, syllables, morphemes, and continuous speech.

3.2.1 Phones, phonemes and tonemes

When humans produce sounds through the organs involved in phonation, the smallest distinctive units that can be classified by its articulatory or acoustic features are called phones. All phones can be analysed non-linguistically at a psychoacoustic level or linguistically at phonemic level. The subset phones, such as the vowels and consonants of the native language, are called phonemes.

Segments or phones are individual consonant and vowel sounds (e.g., /b/ as in big and /i/ as in beat) and each language creates meaning by using only a subset of all possible speech segments. These are known as phonemic segments (or phonemes) and they can be recognised by virtue of their ability to alter meaning. For example, the word bag is comprised of three phonemes – /b/, /æ/, and /ɡ/. Each can be replaced with another to change the meaning of the word. For instance, /b/ could be replaced with /n/ to form the word nag or /æ/ could be replaced with /ɪ/ forming the word, big and so on. Phonemes in a particular language are idealised categories that may be realised by one or more phones (or allophones). For example, in English the phoneme /p/ has two allophones because it is produced as an aspirated bilabial stop /ph/ as in pool, or as an unaspirated bilabial stop /p/ as in spoon. Furthermore, tonal languages use tones to distinguish lexical meaning of words that are analogous to consonants and vowels. Such tonal phonemes are called tonemes.
3.2.2 Syllables

A syllable typically comprises a syllable nucleus (most often a vowel) with a consonantal onset and/or coda and is separated by a vocalic peak from another syllable. For example, the word water has two syllables, and the vocalic peak between the syllables offers a clear distinction for the auditory system. Words can be monosyllabic, disyllabic, trisyllabic, or polysyllabic. Furthermore, syllables can be imposed with different pitch and stress characteristics to express different communication purpose. Lexical tone can be imposed on syllables change their meaning in tonal languages.

3.2.3 Morpheme

A morpheme is the smallest semantically meaningful unit in a language. However, a morpheme is not identical to a word. Some single morphemes are words while other words have two or more morphemes within them. Therefore words can have two or more syllables but be a single morpheme, e.g. elephant, banana. On the other hand, words can have two morphemes and only one syllable; examples include *dogs, barked*.

3.2.4 Continuous speech

The long-term average speech spectra (LTASS) of continuous speech is of interest here because it is an important component in the equation that is used to derive frequency importance function and, subsequently, the speech intelligibility index, which is the main purpose of this study (French & Steinberg, 1947). Speakers typically produce about 4 syllables per second. In order to understand the acoustic components of speech, speech samples can be displayed as LTASS to show how frequency varies in relation to intensity (see Figure 3.2). Byrne et al.
(1994) measured the LTASS of 12 languages using both male and female speech, and found that the dynamic values were similar across all 12 different languages: English (several dialects), Swedish, Danish, German, French (Canadian), Japanese, Cantonese, Mandarin, Russian, Welsh, Singhalese, and Vietnamese. The average intensity of all LTASS was 70 dB SPL (sound pressure level) with LTASS similar for both sexes between 250 to 5000 Hz. The findings suggest that the LTASS is very similar across all the 12 different languages and that no single language or group of languages was very different from another. Furthermore, they found no systematic differences between the English versus the non-English languages or between non-tonal versus the tonal languages (Cantonese, Mandarin, and Vietnamese). Male intensity was significantly higher at frequencies below 160 Hz and female slightly higher above 6300 Hz. The finding that LTASS is the same in English and Mandarin supports the finding of a previous work by McCullough et al. (1993), using similar techniques to record and analyze Mandarin and English.

![Figure 3.2. An English speech spectrum (Haughton, 2002).](image)

**3.3 Vowels, consonants, lexical tone and prosody**

In the following sections, the physical and acoustic manifestations relating to the production of vowels, consonants, lexical tone and prosody are described. Human speech is produced by combining vowels and consonants in non-tonal
languages; and consonants, vowels and lexical tone in tonal languages; and all languages use prosody across the segments of an utterance. While lexical tone is used in tonal languages to change the meaning of the words, prosody allows perception of stress and rhythm in addition to the perception of emotion and intent of the speaker.

### 3.3.1 Vowels

Vowels are produced by shaping the vocal tract through the manipulation of different articulators. Vowel sounds are made without constriction in the vocal tract, and air escapes relatively unimpeded through the mouth and nose, but is filtered or shaped by movements of the tongue and lips. Vowels are normally described with reference to four criteria using the four articulators, lips, tongue, palate, and soft palate as shown in Figure 3.3. These are: (1) the part of the tongue that is raised (front, centre or back); (2) the extent to which the tongue rises in the direction of the palate, generally, high, mid or low; (3) the position of the soft palate, whether it is raised or lowered; and (4) the type of opening made by the lips (rounded, unrounded, spread).

![Figure 3.3: The criteria to describe vowels](image)

Figure 3.3. The criteria to describe vowels: the tongue in the direction of the palate, the extent direction to the tongue rises towards the palate, generally, high, mid or low, the high/low raising of the soft palate, and the type of opening, rounded/unrounded/spread, made by the lips (MacNeilage & Davis, 2000)
The constriction of air flow by the lip and tongue causes the formation of format resonances. Figure 3.4 shows the effect of this, such that the articulators form a filter to shape the sounds emanating from the vocal folds to form resonances. These resonances of vowels are called formats and labeled as first format (F1), second format (F2), third format (F3), etc. Combinations of formant values (especially, F1, F2, and F3) serve to distinguish between vowels (Fant, 1970). Furthermore, the center frequency of each vowel format tends to be higher for men than women than children due to the differences in vocal tract length. It is also important to realize that formats may vary slightly between talkers and the change of physical conditions, brought on by growth, aging, and disease.

Figure 3.4. The source-filter theory of speech production: The source spectrum represents (a) the spectrum of typical glottal air flow with air passing through it and (b) with a transfer function or vocal tract filter output to (c) provide an energy spectrum which represents characteristics of vocal tract shaped by articulators for F1 and F2 (Gelfand, 2004).

3.3.2 Consonants

Consonants are produced with either total or partial constriction of the vocal folds and are typically aperiodic—and are represented by noise acoustically. Consonants can be described with reference to five criteria: (1) whether ingressive (clicks, ejectives, implosives) or egressive as is the vast majority of consonants; (2) whether there is vibration in the vocal folds (voiced) or no vibration (voiceless); (3) whether the soft palate is raised or lowered; (4) whether the place of articulation in the vocal tract involves the lips, the teeth, the velar, etc. and can be thus described as, e.g., bilabial, dental, alveolar, retroflex, or palatal; and (5) the
manner of articulation, that is, whether they are plosives, nasals, or affricates. For example fricatives are characterised by the high-frequency aperiodic noise that is produced by air being pushed through two closely held articulators. Typically, the concentration of energy for both /s/ and /f/ is above 4000 Hz, with higher frequency values obtained for /s/ than /f/ (J. Clark & Yallop, 1990). For adults, fricatives are among the most confusable contrasts in the English repertoire (Miller & Nicely, 1955). High frequency consonants, such as fricatives and sibilants can cause problems for individuals with high frequency hearing loss because the frequency of distinguishing characteristics is above their frequency threshold (see section 4.2).

3.3.3 Speech Prosody.

The prosodic aspects of language require sequences of speech sounds longer than single segments for their realization. The acoustic correlates of prosody are fundamental frequency, duration, and intensity. Fundamental frequency (F0), which is directly related to the vibration of the vocal cords, is perceived as pitch (Wightman & Green, 1974). Pitch is involved in the production of stress, rhythm, and intonation (Atkinson-King, 1980), and may be superimposed over syllables, words, phrases or sentences (as can loudness and rate). Stress is the emphasis or accent produced by increased pitch, duration, and to a lesser extent loudness, at the syllable, word or sentence level (Cruttenden, 1986). In English, stress is used at the syllabic level as a rhythmic timing device; at a word level, to emphasize key words in a sentence; and at the sentence level, to convey affective intent, such as anger. Finally, intonation is the rising and falling pattern of the pitch of the voice over a phrase or sentence, and can function linguistically, e.g., to distinguish a question or statement, or affectively, e.g., to
convey attitude and intent. Speakers around the world commonly modulate F0 to express intentions, such as questions or declaratives at the sentence level. For example, “Are you fine?” has a rising intonation pattern, while the declarative “I’m fine” has a falling pattern.

In English and other Germanic languages, such as Dutch, rhythm occurs in connected speech because the stress patterns on disyllabic words tend to have strong weak or weak-strong patterns. Romance languages, such as Italian, Spanish and French, are syllable-timed, and in disyllabic words the syllables tend to occur at equal time intervals. Languages such as Japanese are mora-timed. Morae are subsyllabic units that are based on the time taken to produce each sound, with each mora having the same duration when used in fluent speech.

3.3.4 Lexical tone.

Some languages change pitch at the word level in order to change its lexical meaning. This production relies on the speaker manipulating F0 of monosyllables. These languages are categorized as containing lexical tones. Yip (2002) indicates that around 60-70 percent of the world’s languages are tonal and over half of the world’s populations speak a tone language (Fromkin, 1978). They are located in three major language areas, Africa (northern Nigeria, Ethiopia etc.), South-East Asia (China, Vietnam and Thailand etc.), the Pacific (Hupa), and the Americas (Mexico). Amongst these localities, Asia is the richest in tonal languages and includes the Chinese language family, Vietnamese, and Thai (Downing, 2001). In this project, Mandarin Chinese, the most spoken language in the world, is studied.

Although the main feature of lexical tones is F0 (pitch), other features such as duration (timing), amplitude (loudness) and vocal quality (register) also
Contribute to the characteristics of tone (Burnham & Mattock, 2007). Conventionally, tones are classified as dynamic (or contour) tones, in which there are movements of F0; or classified as static (or level) tones, in which there is little movements of F0 tones but they differ in the level of pitch. Within a tone language, it is possible to contain both types of tones, for example, Thai contains three static and two dynamic tones (Abramson, 1978). In many tone languages, tones are governed by tone sandhi rules. Sandhi is a Sanskrit word, meaning “putting together”, and sandhi rules refer to post-lexical tonal changes that are conditioned by the preceding word (Gussenhoven, 2004). Thus tone sandhi places constraints on the tone assigned to individual words based on the pronunciation of the words that surround them in.

Lexical tone perception by native tonal language speakers occurs at the phonemic level in contrast to non-tonal speakers who perceive tones at the psychoacoustic level. Burnham and Jones (2002) asked native monolingual Australian English and native Thai speakers to listen and identify tones in four types of stimuli: full spectrum speech, speech low-pass filtered at 400Hz, speech with synthesized F0 to sound like music and sine wave (F0) speech. All mimicked Thai tonal contrasts. The results indicated that identification functions occurred more at the phonemic level for Thai than for English listeners in the full speech condition, but when speech information was reduced to non-speech in low-pass filtered, music, and sinewave conditions, the two groups perceived tone equivalently.

The development of phoneme perception by native tonal speakers is linked with linguistic experiences in the ambient environment. A testing method developed by Werker and Logan (1985) that uses a 500 ms inter-stimulus interval
(ISI) between two sounds is shown to induce a phonetic mode of perception such that even a non native speaker can identify contrast differences. An ISI of 1500 ms, on the other hand, tends to induce a phonological mode of perception such that only native speakers can identify tone differences. This method was adopted by Burnham and Francis (1997), with native Thai and English speakers, including adults, 8-, 6-, and 4 year-olds. Participants were asked to listen and compare a tonal contrast imposed on the same syllable with two ISIs (500 ms and 1500 ms). It was found that regardless of the ISI difference, Thai talkers’ performance was significantly better than that of English talkers. Furthermore, for both language groups, the 4 year-olds were significant better at identifying the tone difference at 1500 msec than 500 msec. The authors suggest that for children, the development of tone perception involves the increasing experience with tonal language, as a language-specific mode of tone perception appears to gain precedence between age of 4 and 6 years.

3.4 Mandarin Chinese

Among tonal languages, Mandarin Chinese, also known as Putonghua, is the most commonly used and spoken by 885 million people around the world (M. J. W. Yip, 2002). In Mandarin, tonal information plays a role as fundamental as the role of its sequences of consonants and vowels. In the following section, an overview of Mandarin phonetic system is provided in respect to its syllable structure, lexical tone characteristics, and related perception studies.

3.4.1 Phonemic system

The Mandarin vowel system comprises 6 monophthongs (single vowels), 9 diphthongs (2 vowel combinations), and 4 triphthongs (3 vowel combinations),
and the consonant system contains 23 consonants (Ma & Shen, 2004). However, it is necessary to point out that due to the large area of Mandarin speaking in the world, the retroflex consonants, /zh/, /ch/ and /sh/, are pronounced as their non-retroflex counterparts /z/, /c/, /s/. This is commonly observed in the south parts of China or Taiwan (P. C. Yip, 2000). The foundational unit or morpheme structure in Mandarin starts with a monosyllable, and each Mandarin monosyllable is assigned a specific tone to provide it with a specific lexical meaning. Each unique syllable can also be combined to become a polysyllabic or compound word (Zhou & Marslen-wilson, 1995). In Eric Shen Liu’s Frequency Dictionary of Chinese (1973), it lists 3,000 commonly used lexemes, and of these 68% are disyllabic, and 30% multisyllabic words.

3.4.2 Syllable structures

The structure of a Mandarin monosyllable has 4 possibilities: (1) V, (2) CV, (3) VC, or (4) CVC (V=vowel; C=consonant). In addition, each monosyllable can be assigned one of five tones to carry different lexical meaning. Unlike English, Mandarin is not written alphabetically, but with a different character to represent each syllable. In addition, syllables with different lexical tones are written as having different characteristics. Therefore, each monosyllable can be represented by many words (Chao, 1968; Zhou & Marslen-wilson, 1995).

Generally, there are five tones in Mandarin. These tones are commonly known as Tone 1, Tone 2, Tone 3, Tone 4 and Tone 5. However, because the fifth tone is a neutral tone, it is usually not considered in most research. Based on the spectral analyses, Xu (1997) shows F0 contours for the four tones. Figure 3 shows the mean F0 contours of the four tones of a CV sequence, /ma/. Depend on the tone used, /ma/ can have 4 different meanings as follows:
In Mandarin, Tone 1 is the only static tone, and has no drastic pitch movement, whereas Tones 2, 3, and 4 are dynamic (see Figure 3.5). The name of each tone is based on its pattern of pitch change (Speer, Shih, & Slowiaczek, 1989). Thus, Tone 1 is called the high level tone and is produced with a high and relatively constant pitch; Tone 2, the rising tone, is produced with a pitch contour that rises from a average value to a relatively high level. Tone 4, the high falling tone, on the other hand, starts high, and decreases in pitch. The pitch pattern associated with Tone 3 varies depending on the context, that is, whether spoken in isolation or phrase-finally, (Chao, 1948; Cutler, Dahan, & Van Donselaar, 1997; Gandour & Harshman, 1978; Shih, 1997; Speer, et al., 1989; Xu, 1994, 1997). As such there is no clear consensus on its name: it has been referred to as low-falling (e.g., Shih, 1986; Speer, Shih, & Slowiaczek, 1989), whereas others call it low-dipping or low-falling rising (Blicher, Diehl, & Cohen, 1990; Wise & Chong, 1957). Here, the term low-dipping will be used following Xu (1997).
3.4.3 Lexical tone auditory cues

The auditory cues for Mandarin lexical tones depend mainly on changes to F0, but can also depend on other acoustic cues, such as duration and the size of the amplitude envelope for a given syllable. As shown in Figure 3.5 the F0 contours, pitch height, and duration are important in cueing Mandarin tone perception. Lin and Repp (1989) report that for the perception of static tones like Tone 1, the only cue required is pitch height; whereas for dynamic tone perception, height and F0 movement is used by listeners. Moreover, not only do the four tones have different contour shapes, but duration also facilitates tone differentiation, for example, tone 3 is longest in duration (Howie, 1976; Massaro, 1985; Whalen & Xu, 1992).

Liu and Samuel (2004) examined how well listeners perceive Mandarin monosyllables when the F0 information is unavailable. Signal processing techniques were used to selectively neutralize portions of the F0 contour hypothesized to cue the identity for words in isolation from 100% information till all F0 information was removed. Perceptual judgments by Mandarin Chinese listeners showed they could identify tones correctly when 70 percent of the information was available. This suggested that the duration cues of F0 acted as a secondary cue to identify the difference between Tone 1, 4 and 3. They have concluded that Mandarin speakers promote the utility of secondary cues when the primary cue of F0 is obscured, and that the flexible use of cues to tone in Mandarin is similar to the flexibility that has been found in the production and perception of cues to phonetic identity in non-tonal languages.
3.4.4 Concatenated tonal variations

When Mandarin syllables are concatenated into words, phrases and sentences, several pitch alterations are superimposed to maintain the fluency of speech. In other words, when a Mandarin monosyllable is produced in isolation, the F0 contours are well defined and stable. However, when produced in sentence context, tonal contours will vary depending on the preceding and following tones (Chao, 1968; Howie, 1976; Lin, 1988; Lin, 1965; Wu, 1984, 1988). One frequently used sandhi rule is related to the third tone. As shown in Figure 3.6, the F0 contour of the first syllable is affected by the following syllable. When two instances of monosyllables using tone 3 appear in succession, the first instance changes from falling-rising to rising, but if a tone 3 follows a tone 1, 2 or 4 that tone 3 changes from mid-falling-rising to mid-falling (Speer, et al., 1989). In addition, by analyzing Beijing Mandarin citation speech, Kratochvil (1984) observed that in sequences of Tone 2+Tone 4 and Tone 4+Tone 4, the frequency range of F0 contour of the first syllable was narrowed, whereas for the rest of the combinations, the range were expanded.

Figure 3.6. The third-tone sandhi rule using /ma ma/ sequence in Mandarin where each tone shape is the average of 48 contours from 12 speakers. The F0 contours of the first instance of tone 3 syllable /ma/, are typically modified at common point 25% into the segment (Xu, 1997).
As shown in Figure 3.7 the influence of tonal change is more apparent on the first tone than the second tone. The offset F0 value appears to determine the initial F0 of the following tone, and this influence can be seen at least two thirds of the way into the vowel of the following syllable, and sometimes the influence spreads to the end of the second syllable. Xu (1997) argued that the determination F0 level by influence of offset value of the first tone would potentially avoid conflict at the neuro-muscular level, and unnecessary muscle strain.

![Figure 3.7. Effects of first tone on F0 contour of the following syllable in /ma ma/ sequences in Mandarin where the tone in the second syllable is held constant, and the tone of the first syllable is varied. The large variations of the onset F0 (50% of the segment) of the second syllable is presented in each panel.]

### 3.4.5 Lexical tone and prosody

Like all languages, tonal languages change surface pitch for communicative functions, such as emphasis and questions. In order to impose prosodic considerations of attitude, intent and emotion on a sentence, the lexical tones can be distorted to achieve this purpose. To address how Mandarin lexical tones are implemented phonetically when uttered with emphasis, Chen and
Gussenhoven (2008) had 5 native speakers read 32 sentences that contained four lexical tone target syllables, with different pre- and post-tonal contexts, and under two conditions of emphasis, medium emphasis and high emphasis, together with a no emphasis base-line condition. The findings suggest that there is a gradual increase in syllable duration from the no emphasis to medium and high emphasis conditions. Moreover, there was a robust increase of F0 range from the no emphasis to the high emphasis condition, although little difference between medium and high emphasis conditions. This supports the previous finding showing that in Mandarin Chinese, emphasis does induce a durational increase and some F0 range expansion (Y. Chen, 2003; Hu, Dong, Tao, & Huang, 2005; Jin, 1996; Xu, 1999). Therefore, in order to achieve emphasis, speakers of Mandarin rely more consistently on duration rather than F0 expansion. In English, emphasis also induces durational increase and F0 expansion, but the magnitude of the duration increase is small. In other words, English speakers rely more on F0 range expansion and less on duration to indicate emphasis (Arvaniti & Garding, 2007; Lehiste, 1970; Sluijter, 1995; Turk & Sawusch, 1997). Chen and Gussenhoven (2008) argue that in Mandarin the minor expansion of F0 range may be due to the speakers’ physiological limits. According to Berstein (1979) weighting of different acoustic cues for a linguistic function may be dependent on the functional load of the cues in conveying other linguistic information. In Mandarin, F0 variation indicates lexical tonal contrasts and therefore is required to be somewhat more restricted in conveying degrees of emphasis.

3.5 Summary

The production of human speech relies on a number of physical organs to produce its components (vowels, consonants, tones) and on other auditory organs
to hear and perceive the speech signal. In tonal languages, the meaning of a word can be changed by manipulating F0 using the vocal folds. Mandarin, used in this thesis, contains four tones for each monosyllabic word. Therefore, successful perception of Mandarin will rely on not only phonemic information, but also Mandarin tonal cues. The thesis aims to understand how tonal cues affect weighting in Mandarin speech recognition tasks with the end objective, appropriate weighting for Mandarin hearing aid prescriptions. The next chapter overview hearing loss, assessment methods, types of hearing loss and how they relate to speech perception.
Chapter 4 The auditory system and hearing loss

The acoustic signals of speech are first received and amplified by the outer and middle ear. Then, the auditory signal is translated into neural input in the inner ear for the nervous system to propagate the information to the brain. In order to have clear neural signals for the brain, sometimes the ear will filter, amplify or compress the input signal if speech occurs in undesired conditions, such as extreme soft/loud intensity or a noisy background. However, these abilities could be lost if there is hearing loss in the ear/s. This chapter begins by reviewing the anatomy and physiology of the ear followed by an introduction to the commonly used methods of assessing hearing, different types of hearing loss, and the resulting difficulties in understanding speech.

4.1 Anatomy and physiology of hearing

The ear comprises three parts: the outer, the middle and the inner parts of the ear as shown in Figure 4.1.

![Cross-sectional view of the human ear showing the divisions of the outer, middle and inner ear.](image)
4.1.1 The outer ear

As shown in Figure 4.1, the outer ear comprises the pinna or auricle and the external auditory meatus, a canal leading to the tympanic membrane. The ear canal is an S-shaped tube about 2.5 cm which connects the pinna and the middle ear. The pinna is an irregular plate of elastic cartilage covered with skin, and its rim is somewhat fleshier, and shows individual variation. The pinna helps to collect sound waves travelling through air and directs them into the auditory meatus. In sum, the outer ear is responsible for collecting and directing the sounds through the auditory meatus to the middle ear, and for altering the spectral shape to provide information about directional location of sound.

4.1.2 The middle ear

Sound waves entering the external auditory canal eventually hit the tympanic membrane, or eardrum, the boundary between the outer and middle ears. The middle ear is a cavity also connected to the oral cavity through the Eustachian tube (see Figure 4.1) which allows air to pass between the tympanic cavity and the outside of the body by way of the throat and mouth. It is important for maintaining equal air pressure on both sides of the eardrum, and important for normal hearing.

The tympanic membrane consists a very thin layer of translucent connective tissue, averaging about 0.7 mm thick with a surface area about 63 mm² (Martin & Clark, 2006). Due to its fineness, the tympanic membrane is an extremely efficient vibrating surface for conducting sound, and the eardrum, in turn, transfers the sound energy to the tiny bones of the ossicular chain: the malleus, incus and stapes. The stapes at the end of ossicles transmits the vibratory
motion of the eardrum to the oval window, which in turn sets the fluids of the inner ear into motion, eventually exciting the hearing receptors.

![Diagram of an uncoiled cochlea](image)

Figure 4.2. Diagram of an uncoiled cochlea (Martin & Clark, 2006).

### 4.1.3 The inner ear

The inner ear is called the labyrinth and comprised of the osseous and membranous labyrinths. The osseous labyrinth is a bony canal in the temporal bone; while the membranous labyrinth lies within the osseous one and has a similar shape. Between the osseous and membranous labyrinths is a fluid called perilymph that is secreted by cells in the wall of the bony canal. The membranous labyrinth contains endolymph, whose composition is slightly different. There are three structural cavities within the bony labyrinth, the vestibule, the semicircular canals (anterior, posterior and lateral), and the cochlea. The necessary organs for hearing and maintaining balance are embedded and supported within these three regions.

The cochlea is a snail shaped organ that serves to convert mechanical sound waves into neural signals for auditory nerve and brain processing. The human cochlea (Figure 4.1) is about 35 mm long, and forms a cone-shaped spiral with about 2 ¾ turns. In order to understand the physiological function of the cochlea, it is better to imagine the spiral uncoiled as shown in
Figure 4.2. Inside the cochlea, as shown in Figure 4.4, there are three fluid-filled spaces: the scala vestibuli, scala media, and scala tympani, separated by a vestibular membrane (or Reissner’s membrane) and the basilar membrane. The cochlear duct is separated from the scala vestibuli by a vestibular membrane and from the scala tympani by a basilar membrane. The upper compartments, called the scala vestibuli, leads from the oval window to the apex of the spiral. The lower compartment, the scala tympani, extends from the apex of the cochlea to a membrane-covered opening in the wall of the inner ear called the round window. Sound waves travel from the ossicular chain through the oval window and enter the scala vestibuli (see Figure 4.3) and always flow from the base to the apical end. Furthermore, waves of low frequency tones propagate all the way to the apical end, whereas waves of high frequency tones remain around the base.

Figure 4.3. The flow of sound waves of three different frequencies (Yost, 2007).

The hearing sensory unit located on the basilar membrane is called the organ of Corti. As shown in Figure 4.4, the organ of Corti runs along the basilar membrane longitudinally and is made up of three rows of outer hair cells and a single row of inner hair cells. The base of these hair cells touch the hearing nerves such that when sound waves pass through the basilar membrane, the tip cilia of the hair cells move back and forth and are stimulated to change acoustic
inputs into neural impulses. Various receptor cells, however, have slightly different sensitivities to such deformation of the hairs. Thus, a sound that produces a particular frequency of vibration will excite certain receptor cells, while a sound involving another frequency will stimulate a different set of cells.

![Cross-sectional view of the cochlea](image)

**Figure 4.4. Cross-sectional view of the cochlea (Martin & Clark, 2006).**

### 4.2 Hearing loss

Hearing loss can be genetic, or acquired from a virus or other illness, ototoxic (ear-damaging) drugs, exposure to loud noises, tumors, head injuries, or the aging process. Any lesion formed from trauma, a virus infection or a foreign body may locate at the peripheral section of the auditory system between inner and outer ear and cause physical blockage of sound processing. Much impairment to hearing sensitivity, e.g., middle ear infection or breaks in the ossicular chain can be repaired with drug treatment or surgery. However, if there is a problem inside the cochlea, such as the loss of hair cells, hearing loss will be permanent, and amplification devices are necessary to improve hearing sensitivity. Clinically, audiologists and physicians separate the hearing pathways into the conductive portion - the outer and middle ear; and the sensorineural portion - the inner ear and the auditory nerve.
4.2.1 Assessment of hearing loss

Two of the most commonly used hearing assessment techniques are pure tone audiometry and speech audiometry, which use instruments tuned specifically for testing hearing sensitivities. The term hearing level (HL) is used to reference the average hearing sensitivity of a group of normal hearing individuals on a scale where 0 dB HL is defined as the lowest intensity that can just stimulate hearing in a normal individual. Statistically, a person with hearing threshold above 25 dB HL is considered as having hearing loss. Detailed referencing information can be obtained from ANSI S3.6 (1996). On completion of a hearing test, the hearing thresholds obtained signify the softest frequency (tone) a patient can identify 50% of the time (Sivian & White, 1933) which is then plotted on an audiogram.

**Pure Tone Audiometry.** Pure tone audiometry is used to test the ability to perceive tones presented at different frequencies (250 Hz, 500Hz, 1kHz, 2kHz, 4kHz, and 8kHz). The participant sits in a sound–isolated chamber with standard ear phones and the audiologist operates the audiometer to test the participant’s responses to pure tone stimuli. Detailed testing guidelines can be reviewed from the American Speech and Hearing Association guidelines (ASHA, 1978). At the end, an audiogram is obtained with both conductive and sensorineural testing results for further diagnosis.

**Speech Audiometry.** Speech audiometry is used when a patient has difficulty understanding speech. Clinically, there are two procedures involved. The first procedure is used to obtain a speech recognition threshold which defines the softest level at which the participant can perceive speech 50% of the time (Carhart, 1946). In the second, the speech recognition test, patients are asked to
repeat a list of standard words which are presented at his/her comfort loudness level. The percentage of correct responses will be then recorded as a speech recognition score. This score is used to predict speech understanding for the patient. The types of material used in speech recognition tests are detailed in Chapter 5.

4.2.2 Types of hearing loss

Hearing loss can be affected along auditory pathways that (1) use air conduction to pass sounds through the outer, middle and inner ear; and/or (2) use bone conduction to bypass the outer and middle ear by vibrating the skull and stimulating the inner ear directly. By assessing hearing in these two pathways, audiologists can differentiate hearing loss into three major types for further management.

Conductive Hearing Loss. Whenever there is a pathology located at the outer or middle ear, the individual will experience a loss of hearing sensitivity due to the blockage of sounds because sounds lose their intensity before reaching the inner ear. A hearing test will reveal impaired air conduction thresholds with normal bone conduction thresholds.

Sensorineural Hearing Loss. If hearing loss occurs beyond the air conduction mechanisms in the inner ear or there is hearing nerve pathology, then impaired air and bone conduction thresholds will present. The obtained thresholds of both air and bone conduction pathways will be fairly close, since there is no blockage of sound conduction before the inner ear.

Mixed Hearing Loss. Some individuals can have lesions occurring simultaneously in both air and conductive mechanisms. Hearing thresholds of such individuals will show better bone conduction than air conduction thresholds,
since the sounds stimulated the inner ear will be attenuated by the outer or middle ear before the inner ear.

4.2.3 Degrees of hearing loss

The degree of hearing loss is based on the threshold average of the three frequencies most related to understanding speech: 500, 1000 and 2000 Hz. According to the threshold average, the severity of hearing loss can be classified into different categories as shown in Table 4.1. Normal hearing is classified as below 15 dB HL and borderline normal hearing at 25 dB HL. Although communication may be impacted slightly in borderline cases, the intervention for amplification is reserved. If thresholds are located between 26 to 40 dB HL, it is considered mild hearing loss and audiological treatment is introduced. Hearing thresholds between 41 to 55 dB HL are considered moderate loss; between 56 to 70 dB HL, moderately severe loss; between 71 to 90 dB HL, severe loss; and greater than 90 dB HL, profound loss.

<table>
<thead>
<tr>
<th>Hearing level (dB)</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10 to 15</td>
<td>Normal hearing</td>
</tr>
<tr>
<td>16 to 25</td>
<td>Slight hearing loss</td>
</tr>
<tr>
<td>26 to 40</td>
<td>Mild hearing loss</td>
</tr>
<tr>
<td>41 to 55</td>
<td>Moderate hearing loss</td>
</tr>
<tr>
<td>56 to 70</td>
<td>Moderately severe hearing loss</td>
</tr>
<tr>
<td>71 to 90</td>
<td>Severe hearing loss</td>
</tr>
<tr>
<td>&gt;90</td>
<td>Profound hearing loss</td>
</tr>
</tbody>
</table>

Table 4.1. Classification of hearing impairment levels.
4.2.4 The impact of hearing loss on understanding speech

People with hearing loss have different types of difficulties with communication and understanding speech, and thus may experience one or more of the following.

**Decreased Audibility.** As the degree of hearing loss increases, sensitivity to sounds gradually decreases. In other words, a person with greater hearing loss in a certain frequency region will require higher auditory intensity in this than other regions. As people aging, they tend to lose hearing sensitivity to high frequencies (4 kHz~6 kHz), required to correctly perceive consonants such as /f/, /s/, /th/.

**Abnormal Loudness Growth.** Abnormal loudness growth or reduced dynamic range is an issue for patients with cochlear pathology (Fowler, 1928). If a person has a normal 1000 Hz hearing threshold in one ear at 5 dB HL, and a threshold at 45 dB HL in the hearing loss ear, then a 5 dB stimulus presented in the normal ear will sound as loud as a 45 dB stimulus presented in the hearing loss ear. If the auditory stimulus is increased to 70 dB HL, both ears will also judge these stimuli equally loud. However, the impaired ear allows 25 dB increases (45-70 dB), while the good ear allows 65 dB increases (5-70 dB). The abnormal loudness growth or reduced dynamic range can lead to intolerance to loud amplified sounds.

**Difficult Hearing in Background Noise.** Most patients with sensorineural hearing loss report difficulty hearing in noise. As shown in Figure 4.3, the cochlea has place codes for specific frequencies but any lesion on the basilar membrane will make the cochlea less frequency specific. For example, a person with a lesion around the 1 kHz coded area may develop difficulties differentiating 1 kHz and 1.2 kHz. Therefore, when speech is present with background noise,
individuals with hearing loss have difficulty sorting out the speech for the brain to process.

4.3 Hearing loss and hearing aids

Hearing aids are essentially sound processors that transform and amplify wave forms. There are two types of techniques used in hearing aids: (1) analog in which continuous sound waves are transformed into electrical signals and (2) digital in which continuous sound waves are sampled for transformation into representative numbers which are processed mathematically according to the hearing aid microchip. Analog hearing aids continue to decline in favor of their digital counterparts with a recent American survey reporting that 75% of fitted hearing aids are digital (Kirkwood, 2004). However, digital hearing aids still require an analog microphone and receiver. In the following sections, the important components of a hearing aid will be described. The hearing aid derives from the early invention of Bell Telephone Laboratories in the 1950s. There are different hearing aid styles, for example, a hearing aid can be either worn behind the ear or be inserted into the ear canal but no matter how it is worn, the main components remain similar. As shown in Figure 4.5, there are three major components inside the conventional hearing aid, a microphone, amplifier, and receiver.
Microphone. A hearing aid microphone is an analog input transducer responsible for collecting sound waves and transforming them into electrical signals. The transduction function of a microphone relies on the vibration of a thin diaphragm above an electrical back plate. As showed in Figure 4.6, sound waves enter the inlet port and vibrate the thin diaphragm. This results in voltage changes on the electrical back plate which transform sound waves from acoustic into electrical signals.

Amplifier. The function of an amplifier, whether analog or digital, is to increase the voltage of the electrical signals received from the microphone. In an analog hearing aid, the heart of the amplifier is an integrated circuit which contains a
transistor, diodes, and capacitors; whereas a digital hearing aid has a microcomputer chip that replaces all the parts of an analog amplifier.

**Receiver.** A receiver, also called a speaker, is an output transducer that converts electrical signals into acoustic output. Like a microphone, the transforming function also relies on a thin diaphragm. As shown in Figure 4.7, the electrical current passes through the center coil to create a magnetic field, and as the passing current alternates direction, the polar direction also alternates. The alternating of the polar direction results in a moving force on the diaphragm that turns the electrical signal into sound waves.

![Figure 4.7. A diagram of a hearing aid receiver (Dillon, 2001a).](image)

### 4.3.1 Amplification formula

The amount of loudness increase for various degrees of hearing loss depends on the prescriptive formula. The formula is used to adjust the amount of amplification to different frequencies and different input levels. In general, the amount of amplification is based on increasing the loudness of speech to the most comfortable level. In the past, the half-gain rule was used which specifies that for each 1 dB of hearing loss, approximately 0.5 dB amplification is required (Lybarger, 1944, 1963). Moreover, research has shown that half-gain rule is ideal for mild and moderate degrees of hearing loss (Pascoe, 1988; Schwartz, Lybarger, & Lundh, 1988) but for hearing loss greater than 60 dB HL, the half gain increase will make the speech insufficiently loud for the listener. In 1990s, two major
prescriptive approaches were distributed for use in hearing instrument fitting. (1) The Desired Sensation Level (DSL) input/output formula (DSL [i/o]) version 4.1 for pediatric hearing instrument fitting (Cornelisse, Seewald, & Jamieson, 1995), and (2) the National Acoustic Laboratories- Nonlinear, version 1 (NAL-NL1) formula for adults (Dillon, 1999). Both formulas calculate the desired amplification values based upon the degrees of hearing loss and the client’s age (Bagatto et al., 2005; Seewald, 1995), so hearing aids would deliver audible, comfortable, and undistorted amplified speech (Ross & Seewald, 1988; Seewald, Ross, & Spiro, 1985; Seewald, Ross, & Stelmachowicz, 1987). Although the goals of both formulas are to provide audible and comfortable amplified speech, for children, the DSL [i/o] aims to provide the most possible amplification that across the full bandwidth and envelope of conversation-level speech so that all phonemic patterns can be learned (Ling, 1989; Scollie, 2005); for adults, the NAL-NL1 aims to achieve the best speech intelligibility under the assumption that the input speech is similar to the international LTASS (Byrne, et al., 1994). Compared to other formulas, NAL-NL1 provides the greatest amplification to those frequencies that make the greatest contribution to intelligibility as suggested by the SII.

The validity of the nonlinear amplification formulas is under constant revision. In particular the relationship between degrees of hearing loss and the prediction of speech intelligibility, and user’s experience has been assessed and has lead to a revision in its formula called NAL-NL2. For the purpose of making speech more intelligible, both NAL formulae incorporate English SII data and has shown clinical benefits (Byrne, et al., 2001; Dillon, 2006; Keidser & Dillon, 2006; Scollie, Ching, Seewald, Dillon, Britton, Steinberg, & Corcoran, 2010; S. Scollie, Ching, Seewald, Dillon, Britton, Steinberg, & King, 2010). However, the benefit
of NAL formulas on other language speakers, such as Mandarin, has not been examined.

4.4 Summary

There are two primary intervention methods for dealing with the hearing loss. For some with outer and middle ear pathology, surgery can remedy the problem but for others amplification devices are introduced to compensate for the loss of hearing. Among these devices, hearing aids remain the most popular choice, and their design has improved dramatically over the last 20 years due to digital technology. The focus of this project is to collect Mandarin speech recognition data to serve as a basis for deriving speech intelligibility information for Mandarin speakers. In next chapter, hearing aid prescription will be explained in detail, including the use of the SII and its key component the FIF.
Chapter 5 Hearing aid prescription for better speech intelligibility

Hearing loss affects how much information in speech is audible over the spread of frequency bands and concomitant intensities. Hearing aid technology is designed to optimize the speech signal taking into consideration the adjustment at different frequency bands. This chapter details much of the information used in relation to hearing aid prescription that can contribute to speech intelligibility. The intelligibility of speech is assumed to relate to the sum of the contributions from different frequency bands, with some frequency bands being more important than others. The articulation index (AI), renamed the speech intelligibility index (SII), is an attempt to quantify the role of audibility in speech perception. SII theory allows for the calculation of a hearing aid response for a given audiogram in order to maximize intelligibility. However, before describing this in more detail, the language specific features which affect hearing aid prescription and the types of tests used to test audibility/intelligibility are discussed. This will then lead the discussion to the SII equation and its key component, the FIF. Finally, published FIFs will be reviewed which introduces the current unresolved issues concerning tonal language FIFs, and some of the technical issues that can affect the derivation of a FIF.

5.1 Hearing aid prescription for different languages

Due to the increasing use of hearing aids in non-English speaking populations, there is a growing imperative to investigate how amplification formulas should be tuned based on the similarities and differences among
commonly spoken languages. Chasin (2008) suggests that speech should be examined at three levels:

5.1.1 Differences in phonemes

Each language has its own specific set of phonemes. Thus, when any two languages are compared, each has a different inventory of vowels and consonants. The specific inventory of a language has the potential to influence how each frequency band contributes to total speech understanding. For example, in English, the high frequency bands are involved in the production of some consonants and contribute to speech perception; but in language that include lexical tone, fundamental frequency and harmonics (< 2000 Hz) also play an important role because these features require perception in the lower frequency bands. Therefore, amplification formulas that reflect the phonetic characteristics of non-English tonal languages is of interest. However, the SII has been only fully investigated in English.

5.1.2 Differences in syllable structure

An efficient hearing aid often has to respond quickly in order to provide the appropriate amount of amplification when the intensity of speech input fluctuates. The most obvious shift in intensity lies in the difference between vowels and consonants. Because of vocal tract constriction and the partitioning of vocal fold energy, consonants are usually produced with less loudness than vowels (Fant, 1973; Ladefoged, 1972; Stewart, 1922). Therefore, a language rich in complex syllables, e.g., the consonant-vowel-consonant (CVC) structure of Japanese, requires a rapid response time for the second (quieter) consonant that follows the middle (intense) vowel in order to be audible. However, for languages
that contain consonant-vowel (CV) syllable structure, such as English, response
time can be slower, because there is more time between first syllable vowel, and
the initial consonant of the second syllable.

5.1.3 Differences in syntax

Generally, the intensity of a sentence gradually declines over the course of
a spoken sentence due to expiration and decrease in lung volume. However, the
syntactic structure of a sentence can also affect intensity levels. For example,
there are differences between syntactic structures of subject-verb-object (SVO)
and subject-object-verb (SOV) structures indicating that differences in word order
can affect intensity in the sentence final position. If an object is in the sentence
final position, it will have more intensity, especially if it is a noun; whereas,
sentence final verbs are produced with less intensity (Sagisaka, 1990). It has been
shown that languages with a SOV word order may need hearing aid adjustment to
increase amplification for soft levels in sentence final position. Chasin (2011)
tested participants with mild to moderate, or mild to moderately-severe sensory
neural hearing loss. In this study they were fitted with appropriate amplification
based on English, and asked to listen to speech recordings of their other language
that contained SVO syntactic structures (Hindi-Urdu, Turkish, Japanese, or
Korean) and to self-adjust the amount of amplification required for the sentences
to be intelligible. Participants did the same task in English (SVO) also with the
ability to self-adjust amplification. The results suggest that languages with a SVO
word order, such as English, require less amplification than languages with the
word order, SOV, such as Hindi, Urdu, Japanese, Korean, Iranian, or Turkish.
5.2 Types of speech test materials

Speech test materials used to test speech intelligibility typically consist of an open- or closed-set approach. In an open-set approach, the listener is asked to repeat (or write) what was heard without the knowledge of the corpus of test items. For closed-set materials, the listener is required to select the correct item from a list of items including the correct response, or the individual is familiarized with the test content prior the test. Because the resulting scores are influenced by the subject’s short term acoustic memory and linguistic ability, reducing the number of response items and increasing familiarity of the words in the test can significantly improve performance scores (Rosenzweig & Postman, 1957). This also applies to speech intelligibility in noise: the average frequency of occurrence of a word in a language is an important determiner of how well a speech item is perceived under challenging noise competition (Howes, 1957; Owens, 1961).

5.2.1 Monosyllabic words

The aim of test materials is to determine how well patients understand everyday conversation. Hence, speech test materials need to be developed based on the criteria of familiarity, homogeneity, and phonetic balance in the speech materials. Phonetically balanced items contain a proportional representation of phonemes that occur in everyday oral communication. Likewise equivalence of familiarity or difficulty of test words and homogeneity across subtests is taken into account in commonly used tests. Speech tests should contain multiple subtests of different items in order to assess progress and ensure scores are not biased by memory effects. For example, an individual with hearing loss will need to be tested before and after hearing aid fitting to assess the degree of benefit from
their device, and later to adjust hearing settings. Each sub-list should also contain enough test items to represent conversational phonemes, but not be too lengthy. Early studies indicated that a minimum of 50 monosyllabic words was required (Egan, 1948). One commonly used set of monosyllabic word lists was designed by Egan (1948) who introduced the idea of phonetically balanced lists. In this test, each list contains 50 monosyllabic words, and many tests adopt this design including lists in the W-22 (Ross & Huntington, 1962), NU-6 (Lehiste & Peterson, 1959), CNC (Causey, Hood, Hermanson, & Bowling, 1984), PB-50 (Hirsh et al., 1952). Penrod (1994) compared the number of test words using several commonly used word lists (W-22, NU-6, PB-50) and concluded that as the number of test items decreases, so does test-retest reliability. In English, there are several monosyllabic word and sentence tests available.

5.2.2 Sentences

Sentence test materials need to contain phonemically balanced sentence lists; equivalence of sentence difficulty across sentences, in addition to the lists being homogeneous across test. The effect of sentence context is commonly used to test the effects of speech recognition thresholds, the accuracy of the pure tone thresholds, and generally to be an index of speech sensitivity (Brandy, 2002; Hudgins, Hawkins, Kaklin, & Stevens, 1947). The use of sentences as speech test materials was introduced by Jerger et al. (1968) to assess the degree to which context effects speech understanding. When single words are used, it tends to test the listeners’ ability to perceive acoustic cues as words lack redundancy. But sentences and continuous discourse are rich in contextual information, and allow individuals to use these cues to help perceive meaning. Sentence material varies in respect to the length of each sentence, the number of sentences per list, and the
number of words in sentences. For example, the Central Institute for the Deaf (CID) everyday speech sentences (Davis & Silverman, 1970) are made up of 10 sets of 10 sentences with the length of sentences in each set varying between 2 to 12 words; the Speech Perception in Noise (SPIN) tests (Beattie, Edgerton, & Svihovec, 1977) contains 8 sets of 50 sentences with 5 to 8 words; the Hearing in Noise Test (Nilsson, et al., 1994) is comprised of 25 sets of 10 sentences with 5 to 7 words.

5.2.3 Non-meaningful word and sentence tests

Non-speech test materials, such as nonsense words and sentences are also used in order to understand how well individual phonemes are recognized without contextual cues. Edgerton and Danhauer (1979) developed two lists of 25 nonsense words, each composed of two syllables with each syllable created by a consonant followed by a vowel. Typically, these words are abstract and difficult for listeners to recognize because language experience fails to provide predictive cues (Carhart, 1965); but even with sentences there are memory and familiarity effects (Jerger, et al., 1968). The Synthetic Sentence Identification (SSI) test has also been developed and contains 24 sets of 10 sentences, in which each sentence contains 7 words, with a noun, predicate, object, but carries no meaning.

5.2.4 Mandarin speech tests

In Mandarin, or Putongha, there are very few speech tests. It should also be noted that communication between Mainland China and Taiwan was prohibited from 1945 to 1997, and thus there are now differences in the accent and vocabulary of Mandarin spoken in Mainland China and in Taiwan. Typically, therefore, Mandarin speech test materials are developed for either Mainland
Chinese or Taiwan speakers. In general, the Mandarin speech materials adopt the format of English test materials, that is, there are several subtests with each list containing 50 items. There are three commonly known Mandarin monosyllabic word tests available.

For native speakers of Taiwan Mandarin, there is word recognition test which contains four 50-word lists, and in which each list can be tested as a half-list (Nissen, Harris, & Dukes, 2008). While the first 25 or the second 25 items can be used as a half test list and still offer enough audiological diagnostic information, the full lists are still preferable due to their improved reliability. Digital recordings of this material have both male and female versions. There is also the (Taiwan) Mandarin Monosyllable Recognition test (MMRT) (Tsai, Tseng, Wu, & Young, 2009). The MMRT is software based material which contains a male voice repeating 4733 monosyllabic words. It also contains software to select monosyllabic words to form unlimited numbers of 25 word lists or 50 word lists under the criteria of phonetic balance. Finally, there is the monosyllabic Mandarin Speech Test Materials (MSTMs) which have high familiarity, are phonetically balanced and found to have sufficient reliability and validity to be used in speech recognition tests (Han et al., 2009). It contains seven 50 word lists, and in general is the most popular choice of speech test material, and will be used in this study.

In respect to sentence material, the Mandarin Hearing in Noise Test (MHINT) is the only available sentence material for Mandarin Chinese speakers. There are two versions, one for Mainland Mandarin speakers (MHINT-M) and the other for Taiwanese Mandarin speakers (MHINT-T). Both versions have the same format, each sentence in the list has 10 syllables represented by 10 Chinese
characters and the complete test contains twelve lists of 20 sentence (Wong, Soli, Liu, Han, & Huang, 2007). The MHINT-M will be used in this study.

5.2.5 Speech recognition with competing noise

In order to assess the types of communication problems that patients encounter in daily life, speech recognition tests are often conducted under noisy conditions. When presenting speech items with background noise, the relative intensity difference between the speech signal and the noise is identified as the signal to noise ratio (SNR). For example, if speech with the constant intensity of 65 dB SPL is presented under the three levels of noise: 55, 65, 75 dB SPL (re 20 uPa), the SNRs will be recorded as +10, 0, -10 dB. Speech recognition improves as speech signal becomes more audible (SNR increases) and decreases as the noise level is raised (SNR decreases) (Fletcher, Allen, & Ward, 1996). The correlation between speech recognition scores and changing of SNR is known as the performance intensity function.

Several factors, such as redundancy/context cues, word familiarity, and response set have been shown to have effects on speech intelligibility. A classic study conducted by Miller and colleagues (1951) asked subjects to identify (1) words in sentences; (2) digits from 0 to 9; and (3) nonsense syllables. Performance intensity functions of speech recognition scores as a function of SNRs were generated for these three materials in order to understand the differences in intelligibility between words with extremely high redundancy (digits), low redundancy (nonsense syllable) and with high contextual cues (words in sentence). As shown in Figure 5.1, recognition scores were best for digits (high redundancy), lower for words in sentences (high contextual cues) and poorest for the nonsense syllables (low redundancy and contextual cues) at each SNR. It

51
seems that even a small increase in SNR resulted in a substantial increase in digit recognition, a smaller improvement for the word perception, and the least increase for nonsense syllables. Thus, they concluded that materials with the most redundancy are more intelligible than the less redundant ones.

![Psychometric functions showing the effects of redundancy and contextual clues on digits, words in sentences and nonsense syllables (Miller, Heise, & Lichten, 1951).](image)

**Figure 5.1.** Psychometric functions showing the effects of redundancy and contextual clues on digits, words in sentences and nonsense syllables (Miller, Heise, & Lichten, 1951).

### 5.3 Measuring speech intelligibility using ANSI standards

The American National Standard Institute (ANSI) offers guidelines to ensure that the characteristics and performance of auditory/acoustical products are consistent. The standards used in this study are those based on an early study by the Bell Telephone Laboratories on human-to-human communication under a variety of conditions, and a related investigation by French and Steinberg (1947) investigating factors governing the intelligibility of speech sounds. In 1969, ANSI published the *American National Standard Methods for the Calculation of the Articulation Index* (ANSI S3.5-1969) based on these and other such studies, e.g. Kryter (1962a, 1962b). French and Steinberg (1947) identified four factors that affect the listener’s ability to perceive speech: (1) the basic characteristics of the speech signal, and the listener’s hearing sensitivity; (2) the electrical and
acoustical characteristics of the intermediate instruments/circuits between talker and listener; (3) the environmental conditions for communication; and (4) communication-related behavior changes in the talker and listener. Thus, the AI was developed as an acoustical index that could be used to predict the speech recognition ability of normal hearing adults listening to speech under a variety of conditions, e.g., the telephone, or radio communication between pilots. Application of this index to the hearing impaired listeners did not occur until the early 1980s (Dugal, Braida, & Durlach, 1980; Kamm, Dirks, & Bell, 1985; C.V. Pavlovic, 1984). More recently, the AI was renamed the Speech Intelligibility Index or SII (ANSI S3.5-1997), and while similar to the earlier standard in its basic concept there were some procedural updates that came with the renaming. One of the more fundamental changes concerned the FIF, a change which implied that frequency weighting is somewhat dependent on or is influenced by the characteristics of the speech material and not solely on the auditory system. The new standard included updated methods for computing a FIF that were governed by factors such as speech clarity, listeners’ hearing sensitivity, noise level in the communicating environment and the nature of the spoken language. The following sections introduce the speech intelligibility formula and its key components as they relate to hearing aid prescriptions for everyday use.

5.3.1 The speech intelligibility index

The formula for the SII can be used to derive a speech recognition score for audiological purposes (Halpin, Thornton, & Hous, 1996). The SII is expressed as a factor ranging from 0.0 to 1.0 which indicates the importance weighted proportion of the speech signal that is audible to the listener. Based on ANSI S.3.5 (1997), the equation is as follows:
\[ \text{SII} = P \sum_{i=1}^{n} I_i W_i D_i \]

In the above equation, \( P \) is a proficiency factor defined by the efficiency of communication between speaker and listener (Fletcher, 1948). For example, if both speakers and listeners have the same mother tongue, the average proficiency factor will have a value of 1.0. However, if one of the talkers uses a second language during the conversation, the proficiency factor will decrease due to unfamiliar accent, or grammar, etc.

\( I_i \) represents the relative importance of the \( i \)th frequency band in the speech spectrum, and is related to the degree to which this frequency band contributes to speech intelligibility, that is, it represents the FIF for that language.

\( W_i \) represents the proportion of the speech signal that is audible to the listener above the noise or absolute threshold at the \( i \)th frequency band.

\( D_i \) is intensity distortion factor that occurs when the human auditory system receives speech signals at relatively high intensity levels, typically more than 80 dB HL. It is well established that as speech signals become more audible, speech intelligibility increases and its most optimal is a factor of \( D = 1 \). However, at higher levels, after full audibility has been received, speech may become less intelligible due to factors such as distortion and spread of masking. At those levels, the value of \( D \) becomes less than 1. Therefore, the SII is the product of a speech recognition\( (P) \) language proficiency between talkers and listeners; \( I \) the relative importance of each frequency band to speech intelligibility; \( W \) proportion of the speech signal that is audible to the listener in each frequency band; and \( D \) the intensity distortion factor. The SII is the total summed across frequency bands. The following is an example of estimating a value at 125 Hz. In Figure 5.2 the speech spectrum’s dynamic range has values ranging from 15 to 45
dB HL at 125 Hz. It is shown as a black bar in the audiogram in panel A. If the listener’s hearing threshold is 30 dB HL, then the listeners’ (W) at 125 Hz is 0.5 \[\frac{(45-30)}{(45-15)}\]; the FIF weighting (I) at 125 Hz is .04 in English; and the distortion factor (D) is assumed to be 1. Therefore, the contribution to the SII at 125 Hz will be as follows:

\[ I_{125} \times W_{125} \times D_{125} = (0.04) (0.5) (1) = 0.02 \]

Figure 5.2 shows how the three factors (W, D, I) are related for calculation of a SII. In panel (A) the range of LTASS are plotted on the audiogram to compare with the listener’s threshold, and are shown as black bars for each frequency threshold at the audible dynamic range or W. In panel (B) the distortion factor is shown as a function of intensity or D; and (C) shows the relative importance weighting of different frequencies bands. Retrieved from ANSI – S3.5 (1997).
5.3.2 The transfer function

Speech recognition scores are usually collected under several signal-to-noise ratios and fitted to the SII formula. This involves developing a transfer function to show the relationship between speech recognition scores and the values for the SII, and use the transfer function to predict speech recognition performance (Kryter, 1962a, 1962b). Figure 5.3 shows transfer functions for different speech test materials – monosyllabic words, continuous discourse and nonsense syllables. For example, an SII of 0.5 corresponds to speech recognition scores of about 70% for nonsense syllables, 75% for phonetically balanced monosyllabic words, and 100% for continuous discourse. Beranek (1947) proposed that for speech to be intelligible the SII should be >0.6, and when < 0.3, is much less intelligible. According to Figure 5.3, an SII of 0.6 is associated with speech recognition scores approximately 100% for continuous discourse and 85% for words; whereas when the SII is only 0.3, speech intelligibility falls to about 80% for continuous discourse, 45% for words and <30% for nonsense syllables.

Figure 5.3. Comparison of transfer functions derived from three different speech materials: monosyllabic words from French and Steinberg (1947); nonsense syllables from Black (1959); continuous discourse from Studebaker (1987).
5.4 The frequency importance function

A key component of the SII for a specific language is the FIF for that language. The FIF indicates the relative importance of each speech frequency band as it contributes to speech intelligibility. Due to the differences between languages, it is accepted that certain frequency bands may be more important to intelligibility in one language over another. In this section FIFs are discussed in relation to some specific methodological issues that affect the development of FIFs - issues, such as the steepness of the filter slope, type of noise spectrum, transducer bandwidth of microphone, scoring participant’s responses, and data smoothing. Other factors such as gender of the talker, phonemic content, and context can also affect FIF outcomes and findings relating to these factors are reviewed. Also reviewed are English language findings and the three studies that have examined FIFs in Mandarin and Cantonese Chinese.

The technique to derive an FIF involves collecting speech recognition scores while participants listen to LP and HP filtered speech, at different frequency cut-off regions and SNRs. Initially speech recognition scores for LP and HP filtered speech bands are plotted as a function of cut-off frequencies to produce cross-over frequencies, or the point at which LP and HP filtering yields equivalent performance. Using this information, FIFs are plotted to show frequency cut-off regions relative to speech intelligibility importance. As shown in Figure 5.4, the peaks indicate the higher importance of certain cut-off frequencies and demonstrate where greater speech recognition is shown relative to other frequency bands. It should also be noted that the peaks differ depending on the type of test material. Although ANSI S3.5-1997 describes the protocol for deriving FIFs, there are several methodological issues that can affect its
development. These include the steepness of the filter slope, noise standards for SNRs, transducer bandwidth, scoring, and data smoothing. These are discussed below.

5.4.1 Filter slope.

It is essential that both high and low filters have very steep slopes when determining the importance of each frequency band. Speaks (1967) presented English sentences to normal hearing subjects under a series of filtering conditions. He found that speech low-pass filtered at 125 Hz can range from being completely unintelligible to be somewhat intelligible when the difference between the filter slopes ranged from a gentle slope at 24 dB/octave to a steep slope at 60 dB/octave. With the advent of computer based digital software for filtering, it is possible to achieve slopes of greater than 100 dB/octave. In general, a filter with slope greater than 60 dB/octave is considered appropriate for FIF data collection (Studebaker & Sherbecoe, 1993).

Figure 5.4. Examples of English language FIFs for monosyllabic words (Black, 1959) and continuous discourse (Studebaker, Pavlovic, & Sherbecoe, 1987). The relative percentage of speech intelligibility importance is shown as a function of each frequency cut-off.
5.4.2 Noise.

Different noise a spectra can also shift FIF weightings. Webster and Klumpp (1963) compared the effect of 16 noise profiles on crossover frequencies and found that crossover frequencies can vary by up to an octave depending on the type of noise used. Most noise profiles have flatter spectra (equal intensity across the frequency range) than speech which has higher intensities between 0.5 to 4 kHz. Therefore, when speech is masked by noise with a flat spectrum, the high frequency end would be masked more than other regions; and when the SNR decreases, speech recognition becomes more dependent on low frequency speech cues which can result in inaccuracies. Hence, the ideal noise profile should have a spectrum that parallels the speech spectrum. Under such conditions, the SNR remains constant across the frequency range (Bell, Dirks, & Trine, 1992).

5.4.3 Transducer bandwidth.

When the speech material for deriving the FIF is recorded, the frequency range of the microphone or the transducer bandwidth which records acoustic signals can be limited and not optimal. While normal hearing individuals can perceive sounds when the frequency range is quite narrow, e.g., between 20 Hz to 2000 Hz, the frequency range should be between 200 Hz and 7000 Hz according to ANSI S 3.5 (1969). However, this frequency range is not optimal and more likely due to the limitations on microphone technology of the time, as it would typically cause some loss of low fundamental frequency and high frequency fricative information impeding speech clarity. Moreover, it has been shown that frequencies around 8912 Hz are important to English word recognition, and it seems that some individuals with severe to profound hearing loss can retain
hearing between 8000 to 14000 Hz for perceiving speech information (Berlin, Wexler, Jerger, Halperin, & Smith, 1978; Collins, Cullen, & Berlin, 1981; Duggirala, Studebaker, Pavlovic, & Sherbecoe, 1988). Ideally, the microphone used to record speech materials for deriving FIFs should have a bandwidth that covers the whole speech frequency range. Contemporary microphones and software typically record frequencies between 110Hz to 44000Hz.

5.4.4 Scoring methods

The response scoring method can also affect a FIF, as its calculation relies on speech recognition data to be collected under a multitude of conditions at different frequency bands and different SNRs. Most reported FIFs, including Mandarin FIFs (J. Chen et al., 2008; Yeh, 2005) have experimenters monitoring responses, that is, listening to the responses and marking them right or wrong. It has been shown that the discrepancy between two experimenters can increase up to 20% when aural monitoring is used instead of a written response format (Merrell, 1964; Nelson & Chaiklin, 1970). Owens and Schubert (1977) found that it is especially difficult for examiners to differentiate the voiced sounds from unvoiced sounds, such as /f/ and /v/. Monitoring a Mandarin speech test responses this way could have similar issues not just for phonemes but more particularly for lexical tone information. For example, tone 1 and 3 can be difficult to distinguish due similarities in duration and F0. Collecting the written responses of listeners seems to be one way to resolve this issue but does extend testing time (Brandy, 1966).
5.4.5 Data smoothing

FIF derivation depends on measuring speech recognition under multiple frequency bands and SNRs, and despite using a relatively large pool of participants’ speech recognition scores do not always track smooth curves across frequency bands and SNRs. At present, there is no consensus on a method that should be followed to smooth speech recognition data. Most researchers deal with this issue by using the mathematical concept of interpolating or extrapolating from the speech recognition scores. Studebaker and Sherbecoe (1991) report general guidelines for smoothing data, and most studies follow these guidelines (Bell, et al., 1992; J. Chen, et al., 2008; DePaolis, et al., 1996; Sherbecoe & Studebaker, 2002; Studebaker & Sherbecoe, 1993; Studebaker, Sherbecoe, & Gilmore, 1993; Wong, Ho, Chua, & Soli, 2007; Yeh, 2005). In their method, scores are smoothed by eye using independent judgments and the following criteria: (1) A curve at one SNR cannot intersect a curve at another SNR unless one of the curves is a HP curve and the other is a LP curve. (2) Both the HP and LP curve for each SNR should terminate at near the same recognition score (corresponding to the full frequency bandwidth). (3) Scores should increase or remain constant when either the frequency bandwidth or SNR benefit recognition. (4) Data is smoothed by independent experimenters. The independent estimates are then averaged and become the curve values. However, it is of some concern that the raw data were smoothed and fitted by eye and that this may still cause artifacts due to individual differences in experimenters’ judgments.
5.5 Factors which affect FIFs

Language is just one factor that can affect FIFs. Findings from the most extensively studied language, English, suggest that the frequency bands that range from 500 Hz to 2 kHz make greater contributions to intelligibility than other frequency bands (Schum, Matthews, & Lee, 1991; Wang, Reed, & Bilger, 1978) (see Figure 5.4 & 5.5), and explain why hearing loss in this region has a major effect on speech perception (Martin & Clark, 2006). However, there are effects of other speech variables on FIF, for example, talker gender, phonemic content, and context have also been examined (DePaolis, et al., 1996; Chaslav V. Pavlovic, 1994; Sherbecoe & Studebaker, 2002; Sherbecoe, Studebaker, & Crawford, 1993; Studebaker, et al., 1987; Studebaker & Sherbecoe, 1991, 1993; Studebaker, et al., 1993). Findings related to these factors in English language speech materials will be discussed in more detail below. Finally FIF results from three studies of Cantonese and Mandarin Chinese will be presented.

5.5.1 Gender of the talker

A review by Studebaker (1987) which included nonsense syllables, monosyllabic words and disyllabic words, suggests that FIFs for male talkers have a lower relative importance region than those for female talkers. When gender differences are examined with continuous discourse, there are no differences between male and female talkers (Studebaker, et al., 1987). The differences between the findings of these two studies appear to lie in the nature of the speech material, that is, whether single words or continuous speech with greater contextual cues was used. Although it is possible that the gender of the talker affects relative weighting in low frequency regions, others have suggested that it
is due to chance, and the fact that most speech recognition data is collected using speech materials recorded by one or two speakers (Duggirala, et al., 1988; Sherbecoe, et al., 1993; Studebaker, et al., 1987).

5.5.2 Phonetic content in speech materials

There is also data suggesting that FIFs can be altered by the speech materials used, especially if phonemic contrasts are used with different articulatory features. Duggirala et al. (1988) summarizes the results of three English language studies (see Table 5.1) and draws the conclusion that the differences lie in the different contrasts used, and that even the use of different voicing contrasts, can produce different CFs (Duggirala, et al., 1988; Miller & Nicely, 1955; Wang, et al., 1978). Other languages may have an even greater potential to produce different CFs/FIFs. For example, Palva (1965) reported that Finnish has a CF slightly above 1000 Hz, which is lower than English (1500 Hz). He concluded that the lower CF of Finnish is probably because Finnish contains more phonemes with low frequency phonetic characteristics than English.

Table 5.1. List of CFs (Hz) for different phonemic features.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Duggirala et al. CF</th>
<th>Miller and Nicely Feature CF</th>
<th>Wang et al. Feature CF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasality</td>
<td>472</td>
<td>Nasality</td>
<td>450</td>
</tr>
<tr>
<td>Voicing</td>
<td>758</td>
<td>Voicing</td>
<td>500</td>
</tr>
<tr>
<td>Place</td>
<td>1290</td>
<td>Place</td>
<td>1900</td>
</tr>
<tr>
<td>Graveness</td>
<td>1618</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compactness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manner</td>
<td>1800</td>
<td>Manner</td>
<td>750</td>
</tr>
<tr>
<td>Sustention</td>
<td>2521</td>
<td>Duration</td>
<td>2200</td>
</tr>
<tr>
<td>Sibilation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5.5.3 Context and redundancy

FIF findings show that any increase in context and redundancy such as provided by sentences and continuous discourse shift the weighting of the FIF into the low-frequency region and broaden its shape (DePaolis, et al., 1996; Sherbecoe & Studebaker, 2002). FIFs for nonsense syllables and monosyllabic words, on the other hand, have relatively more importance at higher frequencies than continuous stretches of speech (Black, 1959; Dirks, Dubno, Ahlstrom, & Schaefer, 1990; Studebaker, et al., 1987; Studebaker & Sherbecoe, 1991). Figure 5.5 shows the differences between the shapes of FIFs obtained from words, sentences and continuous discourse. DePaolis (1996) found that increasing context shifts the relative importance of frequency cut-off regions away from the central peak of the FIF (shown for monosyllabic words) and broadens its span (DePaolis, et al., 1996).

![Figure 5.5. The FIFs for monosyllabic words, sentences, and continuous discourse. The FIF for continuous discourse reveals a higher broader span compared to monosyllabic words; whereas, the FIF for sentences is intermediate (DePaolis, Janota, & Frank, 1996).](image-url)
5.5.4 Response format

The response format of speech recognition materials has also been found to affect FIFs. Speech materials with limited response options (closed-response format) are reported to shift the FIF down to the low-frequency regions more than speech materials with more response options (open-response format) (Speaks, 1967; Studebaker & Sherbecoe, 1993). Speaks (1967) used the 10-item closed-response format of the Synthetic Sentence Identification test to obtain a cross-over frequency of 725 Hz, which is much lower than typical cross-over frequencies in English (Dirks, et al., 1990; Studebaker, et al., 1987). These findings suggest that the low cross-over frequencies may be due to the limited response options of its closed-response format and the relatively low number of items in the test (Braida et al., 1979). Moreover, it is shown that for normal hearing listeners using a closed-set response format, it is possible to correctly identify items even when only vowels are heard (Braida, et al., 1979).

5.5.5 Mandarin Chinese FIF

Unlike English, Mandarin Chinese contains lexical tones which are used to convey lexical meaning in this language. Thus, in tonal languages critical speech information is conveyed by the shape and height of the F0 of different lexical tones, and this low-frequency information affects intelligibility. Currently, only two studies have examined FIFs for Mandarin Chinese monosyllabic words. Both are yet to be published. Chen and colleagues (2008) reported a FIF with a peak at 2240 Hz in a conference poster of a study that used male and female recordings of 50 phonetically balanced words. However, the use of 50 words in 144 filter frequency x SNR conditions means that there is the potential to significantly bias
the results due to memory/familiarity effects of words being repeated. It is also not specified how many participants were used in the recognition task. Likewise, in a Masters’ thesis with two participants, Yeh (2005) shows greater importance above 2000 kHz for words. However, he also used only 50 words in 320 filter frequency x SNR conditions.

Both the previous Mandarin word studies exhibit shortcomings in terms of the number words used for the number of conditions, and the number of participants. Although both studies indicate that compared to English (see Figure 5.5), Mandarin has more importance above 2 kHz, these results should be viewed with caution. It could be expected that low frequency information would contribute more to speech intelligibility in a tonal than non-tonal languages, such as English. However, the Mandarin word findings suggest that frequencies above 2 kHz contribute more in Mandarin than in English (Chen, 2008; Yeh et al., 2005). Conversely, published findings for Cantonese Chinese reveal a FIF CF that is much lower (~1600 Hz) than English (~1800 Hz) and that there is a second low frequency importance peak around 160 Hz (Wong, Ho, Chua, & Soli, 2007). However, this study used sentence lists from the Cantonese Hearing in Noise Test (CHINT), and generally sentences tend to have lower frequency peaks than monosyllabic words (see Figure 5.5).

It is important to point out that among the available studies with Cantonese sentence and Mandarin monosyllabic words, only the Cantonese sentence data show the low frequency peak around 160 Hz. In addition, the FIF peak obtained using Cantonese sentences was lower than the typical English sentence FIF peak (see Figure 5.5) and the findings with Mandarin Chinese (Chen, 2008; Yeh, 2005). The peak at 160 Hz is suggested to derive from the tonal characteristics of
the speech materials, but also could be affected by the use of a male speaker. The differences in these data sets certainly require further investigation and raise more questions than answers.

In summary, the Cantonese Chinese sentence FIF has a similar shape to the English sentence FIF, but the main importance peak is in a lower frequency region than reported for English. It also has a secondary peak around 160 Hz. The Mandarin Chinese word FIF, in contrast, tends to be higher than reported for English. Nonetheless, it is difficult to make accurate judgments because there is no Cantonese monosyllabic word data available, and no sentence data available for Mandarin. This study will attempt to remedy this situation as it aims to develop FIFs for both Mandarin words and sentences. The current study will be outlined in the next chapter.
Chapter 6 Overview of this study

Given the large population of Mandarin Chinese speakers across the globe, it is essential to examine whether current hearing aid prescriptions for Mandarin reflect its language-specific characteristics, especially lexical tone. This is the first time a large scale study has been conducted in Mandarin Chinese. The current project aims to fill a knowledge gap by providing a systematic evaluation of how the language features of Mandarin Chinese contribute to frequency importance weightings for hearing aid prescription. Hence, the primary aim of this project is to use series of speech recognition scores to develop Mandarin FIFs to examine weightings for lexical tone, phonetic content in words, and sentences, together with other functions to determine how the language factor, in the case of Mandarin, affects SII estimation. This chapter provides an overview of the experimental plan using different speech materials to determine performance intensity functions, FIFs, and related SII transfer function calculation. In addition, it includes predictions associated with each of these functions.

6.1 Experimental plan

Established FIF study indicated that speech recognition data to be collected under different noise conditions (SNRs) and different LP and HP filtered frequency bands (ANSI 1969, 1997). In the first instance, these data are used to determine performance intensity functions which describe the relationship between speech recognition scores and different SNRs. Secondly, the FIFs of lexical tones, segments, words, and sentences will be developed to examine how low-frequency lexical tones, language-specific phonemes, and sentence context play a role in the importance weightings for frequency bands in Mandarin
Chinese. Lastly, the derived FIFs will be incorporated with the published LTASS (D. Byrne, et al., 1994) to evaluate SII transfer functions which provide the relationship between the speech recognition scores and SII values in daily communication.

The speech materials for the recognition tests were recorded by a Mandarin speaker using two different Mandarin speech materials: monosyllabic words and 10-word sentences. In monosyllabic word recognition test, listeners were required to (1) select the lexical tone used, and (2) orally repeat the word. Results were then scored correct or incorrect for the following: (i) tone recognition, (ii) segment recognition (segment only) (iii) word recognition (segment and tone). In the sentence recognition test, participants were asked to orally repeat the sentence, and scored on the number of words that were correctly recognized.

This is the first time tone and segment response data have been collected and analyzed separately from word responses. The Mandarin speech recognition data collected by Yeh and colleagues (2005) and Chen (2008) only scored overall word responses. It should be recalled also that these studies had several shortcomings (see Section 5.5.5). The reason for developing separate functions of tones, segments and words is to allow for differentiation between the effects of phonetic content and lexical tone in Mandarin Chinese. In addition, technical aspects, such as filter slope, noise, transducer bandwidth and data smoothing etc utilized modern techniques for better accuracy. Thus, some procedures were modified, most particularly the use of curve-fitting techniques to smooth curve functions.
Finally, the derived Mandarin data are compared with English and Cantonese FIFs in Chapter 9. Comparison with English, which is the most studied language in this area, should provide some evidence on how lexical tone, phonemic content and context/redundancy affect these functions. The sentence data can also be compared with Cantonese results (Wong, Ho, et al., 2007), also a language of the Chinese tonal family, should be similar in how lexical tone affects the characteristics of these functions in a sentence context. At this point, though, the following predictions are made associated with Mandarin FIFs.

6.2. Predictions

The current project plans to explore how phonemic content, lexical tone, and sentence context affects the shape of performance intensity functions, FIFs and SII transfer functions by comparison of the types of Mandarin Chinese materials used and other published data with English and Cantonese Chinese. In the following sections, predictions are made for the three functions (performance intensity function, frequency importance function and transfer function of SII) in comparison to published evidence from English and Cantonese Chinese.

6.2.1. Predictions for performance-intensity functions

It is shown that speech recognition improves as the speech signal becomes more audible (SNR increases) and worsens as the noise level is raised (SNR decreases). Furthermore, for the same increase in SNR, there is a faster increase in recognition of speech with increasing SNR for materials containing higher contextual/redundant information than the speech materials containing less redundant material (Fletcher, et al., 1996; Miller, et al., 1951). In this respect, the speech materials used in this project can be ranked in respect to their level of
redundancy. Tonal information is the most redundant, and easiest to correctly 'guess' especially in Mandarin as it only has four tones. Tones are followed by sentences which contain considerable contextual cues, with segments in third place, and words, in which both segmental and tonal information must be recognized, being the least redundant. Therefore, it is predicted that the context/redundancy effect will be of greater benefit to Mandarin tone and sentence recognition scores at high SNRs than segments and words at the same SNRs. In other words, a small increase in SNR should result in a greater improvement in tone and sentence recognition than segment and word recognition.

6.2.2 Predictions for frequency importance function

Findings for English words and sentences show that the peak percentage of importance weighting tends to center on 1800 Hz; and that the increase in contextual information in sentences tends to shift the relative importance to a lower percentage, and broaden its frequency span compared to words (DePaolis et al., 1996). Thus, it is expected that highest percentage importance weighting for Mandarin sentences will be smaller than for words. It is also reported that Cantonese Chinese sentence data contain a secondary peak of importance at frequencies below 200 Hz (Wong, Ho, et al., 2007), English does not have this secondary peak. The Cantonese results suggest that these frequencies contribute some importance weighting for intelligibility of tonal languages but the question remains as to whether this secondary peak under 200 Hz is a reflection of tonal language characteristics or differences in the phonetic characteristics between English and Cantonese Chinese (DePaolis et al., 1996; Wong, Ho et al., 2007). If the derived Mandarin FIF for lexical tone in the current study contain noticeable low frequency weighting under 200 Hz, it would suggest that it can be attributed
to the tonal nature of the language. In addition, comparison of word (segment +
tone) and segment only (minus tone) has the potential to determine whether the
secondary peak is due to lexical tone or acoustic-phonetic differences. This is
discussed below.

Previous studies of frequency importance functions using Mandarin words
(Chen, 2008; Yeh, 2005) reveal that greater weighting should be given to
frequency bands above 2 kHz. While these findings could be due to the
conditions used in these studies (Chen, 2008; Yeh, 2005), and the results of the
current study might not concur with those findings, we are assuming that there
may be differences between English and Mandarin FIFs, that may be due to
acoustic differences in native language phonemes, and/or the effect of the tonal
characteristics of Mandarin Chinese. Thus, the following alternate predictions are
made based on the increase in importance weighting > 2 kHz for word data in
Mandarin compared to English (~1800 Hz) (DePaolis, et al., 1996; Studebaker &
Sherbecoe, 1991; Studebaker, et al., 1993). Firstly, if the frequency weighting for
Mandarin segment-only responses derived in this study are similar to the
Mandarin FIF in other studies using words (segment + tone) (Chen, 2008; Yeh,
2005), this would indicate that it is phonetic (not tonal) features that play the
major role in contributing to the higher weighting above 2 kHz. Alternatively, if
the frequency weighting for Mandarin segment-only data in the current study
shows decreased importance weighting compared with the Mandarin word FIF
(segment + tone), this would indicate that tonal (not phonetic) characteristics of
Mandarin words make a major contribution to the importance weighting above 2
kHz. Similarly, the same predictions can be made for any secondary peaks in low
frequency regions using sentences.
6.2.3 Predictions for SII transfer function

A transfer function converts speech recognition scores into SII values, which are an estimate of communication efficiency in everyday life (Clark, 1992; Killion, Mueller, Pavlovic, & Humes, 1993). To compare transfer functions, and indeed, all other functions from different studies and languages, it is important that studies use the same derivation procedures; and that studies compare both sentences and words at the same time. Details of transfer function creation were not provided in the conference abstract by Chen and colleagues, (2008); and are only available for Mandarin words in the Master’s thesis by Yeh (2005) and the paper by Wong and colleagues, (2007). In this study the derived Mandarin TFs are compared with the transfer functions for English words and sentences, and use the same FIF to TF procedures (DePaolis, et al., 1996; Miller, et al., 1951; Wong, Ho, et al., 2007). It is expected that the greater the context cues and redundant information, the higher will be the percentage of speech recognized for a given SII. Thus, Mandarin tones and sentences rich in redundant/contextual information may be linked with higher speech recognition scores than words and segments at the same SII value as found in English (Hudgins, et al., 1947; Miller, et al., 1951). It is also reported that Cantonese Chinese provides greater redundancy than similar English sentence material (Wong, Ho, et al., 2007), and it, therefore, might be expected than the Mandarin data collected here will have higher speech recognition scores than English at the same SII value.
6.3 Conclusion

Optimal intelligibility for hearing impaired Mandarin listeners can only be achieved though the development of FIFs for Mandarin Chinese. This will allow the relative importance of each frequency band as it contributes to the total speech intelligibility score for Mandarin to be specified, and enable correct adjustment of hearing aid prescription for hearing impaired speakers of this language. Furthermore, once the FIF is determined, a series of transfer functions can be computed to predict appropriate SII values for listeners under different conversational conditions. The next chapter details the method and procedures used to recording and editing speech materials, and to collect speech recognition data for the development of Mandarin performance intensity functions, frequency importance functions and transfer functions.
Chapter 7 Method

This chapter describes the method for recording and editing words and sentences for the speech recognition tests (SRT), and the procedure for collecting SRT scores. The several steps began with the recording of a Mandarin male speaker repeating the speech materials to be used. In this study it was words from the monosyllabic Mandarin Speech Test Materials (MSTMs) (Han, et al., 2009) and sentences from the Mandarin Hearing in Noise Test (MHINT) (Wong, Soli, et al., 2007) that were used. The words and sentences were then high-pass (HP) and low-pass (LP) filtered into specific bandwidths, and noise was added to each filtered frequency band to create different SNRs. In obtaining the SRT, participants were required to listen and repeat the words and sentences for the HP and LP filtered frequency bands at different SNRs, and to indicate the lexical tone used. The percentage of correct responses was recorded as a recognition score. For the word test, three sub-responses are recorded as three different recognition scores: (i) word recognition score for correct repetition of segment + tone; (ii) tone recognition scores for correct identification of tone used irrespective of phonemic structure; (iii) segment recognition scores: the correct responses for segment was used irrespective of the tone used. For the sentence test, the recognition scores consist of the number of words correctly identified in a 10-word sentence. Finally, The FIF was calculated by analysis of the speech recognition scores based on mathematic functions stated in the ANSI S3.5 standard (1997). Separate FIFs for words, tones, segments and sentences were derived. This, chapter provides details of the methods for recording and editing the speech materials, and describes the procedures used for collecting the SRT
scores. Finally an overview of the protocols and steps involved for deriving FIFs is provided.

7.1 Method

7.1.1 Mandarin Chinese speech materials

According to the requirements of the ANSI (ANSI S3.5-1997) and described in Chapter 5, SRT materials need to be developed based on the criteria of familiarity, homogeneity, and phonetic balance of the speech materials. Two clinically popular and robust speech tests were used in the study:

(1) The monosyllabic words were from the MSTMs developed by Beijing Institute of Otolaryngology (Han, et al., 2009). There are seven subtests in the test and each subtest is a 50-word list. In this study the four lists used were recommended as the most homogenous on evaluation by the authors (Han et al., 2009).

(2) The 10-word sentences were taken from the MHINT (Wong, Soli, et al., 2007) and comprise 12 sets of 20 ten-word sentences. This study used four of the 12 lists based on similar criteria to the words (MSTMs).

7.1.2 Recording procedure

The purpose of this study is to compare different derived functions for different speech types with a consistent methodology. Therefore, it is essential to use the same talker to record both sets of speech materials. The Mandarin speaker was a male professional broadcaster from special broadcasting service (SBS), a multicultural Australian television station. The news broadcaster was
recruited to record the speech materials. Original recording were not used, because word and sentence lists were recorded by different talkers. Prior to the recording session, the broadcaster was asked to listen to the original materials recorded by Han et al. (2009) and Wong et al. (2007) in order to produce similar quality of speech in terms of pronunciation and rate. For sentences, speech rate was approximately 4 syllables per second, resulting in each 10 word sentence having a duration of 2.5 to 3 seconds.

The Mandarin word and sentence materials were recorded in an anechoic chamber at MARCS Institute which is designed to attenuate all sound reverberation. The recording instruments consisted of a high-sensitivity dynamic microphone headset (Shure SM10A), as shown in Figure 7.1. This was connected to an Audio/MIDI interface (EDIROL UA-25) which provides 24-bit/96kHz recording and playback. The input was recorded directly to computer software, CoolEdit, installed on a Lenovo T61p laptop. The recordings used a sampling frequency at 44.1 kHz with 16-bit amplitude resolution. During the recording session, the speaker wore the headset microphone with a fixed distance of 13 mm, and was instructed to speak at a normal loudness level, which corresponded to about 65 dB SPL as monitored by the A weighting sound level meter. Each monosyllabic word and each sentence was recorded 3 times.

Figure 7.1 Shure SM10a Headset Dynamic Microphone.
7.1.3 Token selection and editing

The experimenter (a licensed audiologist) selected the best exemplar from the three recordings based on judgments of vocal quality and the integrity of the waveform. The recorded speech was edited from the waveform into monosyllabic words and 10-word sentence units. The silent period at the beginning and end of each speech unit was always less than 1ms to eliminate intensity differences that might arise if too much silence was attached to the tokens when noise was added to the speech signals. In total, 200 monosyllabic words and 80 sentence .wav files were created. The average root-mean-square (rms) of these units used a level of a 1000 Hz tone for normalizing intensity following ANSI standards.

7.1.4 Speech filtering and noise additive procedures

For this study, interfering noises were generated by shaping white noise to match the individual short spectra of each of the speech files. First, the rms intensity level of each speech word and sentence file was measured. Then the noise files were generated by shaping the white noise to match the averaged spectrum of each word / sentence file using software designed by the technicians at the National Acoustic Laboratories. Thus, matched noise files were created for each of the monosyllabic words and each of the sentences. Figure 7.2 shows a matched noise spectrum in Channel 2 for the speech waveform in Channel 1.

Speech editing software, CoolEdit 2000 was used to filter the sound files and mix the filtered speech with the noise files. As seen in Figure 7.2, each .wav file contained a speech waveform in Channel 1 and noise waveform in Channel 2. The words and sentences in Channel 1 were HP and LP filtered into specific frequency bandwidths. For example, the 20-band method uses 112, 141, 178, 224,
282, 355, 447, 562, 708, 891, 1122, 1413, 1778, 2239, 2818, 3548, 4467, 5623, 7080, 8913 Hz. To separate the speech signal into HP and LP frequency bandwidths the speech was passed through a broadband digital filter (110-11,000 Hz) using a Hanning windowing function, and a sharp cut-off of 90 dB/octave to ensure as little overlap as possible between frequency bands. The target frequency band was also set, e.g., HP at 8913 Hz. The intensity of the filtered speech was then amplified or reduced with reference to the noise level in Channel 2, and finally, the speech and noise channels were mixed to create a single waveform in Channel 1 and 2.

![Figure 7.2](image_url)

Figure 7.2. A display of waveforms of a HINT sentence in Channel 1 (top), and noise profile in Channel 2 (bottom).

### 7.1.5 Test conditions

According to the current established study for the development of FIFs, speech recognition data requires each participant to listen to a standard list of 50 words or 20 sentences with 20 filtered frequency bandwidths (112, 141, 178, 224, 282, 355, 447, 562, 708, 891, 1122, 1413, 1778, 2239, 2818, 3548, 4467, 5623,
7080, 8913 Hz) at each 2 step SNR (8, 6, 4, 2, 0, -2, -4, -6, -8, -10 dB) in HP and LP conditions. For a single word or sentence, this would require each participant to complete 400 conditions [20 bands x 2 filter (HP+LP) x 10 SNRs] but when multiplied by 50 words, would result in 20,000 trials for words; and for 20 sentences, 8000 trials. Undoubtedly this is far too onerous and fatiguing for most participants. Therefore, two strategies were applied to reduce the number of conditions for collecting speech recognition data collection: (1) Reducing the numbers of frequency bands and SNRS by widening the frequency bandwidths and differences between SNRs, and (2) reducing the numbers of words and sentences to be completed by each participant.

Several researchers have used wider bandwidths and these appear to provide sufficient data for deriving FIFs (DePaolis, et al., 1996; Studebaker, et al., 1987; Studebaker & Sherbecoe, 1991; Studebaker, et al., 1993; Wong, Ho, et al., 2007). For example, Wong et al (2007) selected 136 conditions (88 LP and 48 HP) from the original 400 conditions. In the current study, to achieve a reduction in the number of conditions, the bandwidth was doubled, resulting 10 cut-off frequencies instead of the mostly used 20 cut-off frequency bands. Thus, the cut-offs used were 141, 224, 355, 562, 891, 1413, 2239, 3548, 5623, and 8913 Hz. Also broadened were the original SNRs from 2 dB steps to 3 dB steps to include seven (not 10) SNRs at 9, 6, 3, 0, -3, -6, and -9 dB. This resulted in a reduction to 140 test conditions [10 bands x 2 filter (HP+LP) x 7 SNRs] for each word and sentence. To ensure this would provide suitable data for developing Mandarin FIFs, a pilot study was conducted with four participants. The aim of this study was to determine whether recognition percentages would provide a full range of scores, that is, from 0 to 100%.
7.1.6 Pilot study: method

7.1.6.1 Participants

Four Mandarin Chinese speaking participants were recruited using flyers (see Appendix A) posted on bulletin boards at the University of Western Sydney and Macquarie University. All participants were required to be between 20 and 45, years of age originate from China less than five years ago, use Mandarin at least three days a week or 25 hours a week, and report normal hearing and no history of neurologic disease. For more participant detail, the questionnaire is in Appendix C and there are more specific demographic details in Section 7.2.1. The hearing of the participants in the pilot test was also evaluated as described in section 7.2.2 below. A travel reimbursement of $30 was given to each participant at the end of each session.

7.1.6.2 Speech materials

The speech materials were a single word list (50 words) and single sentence list (20 sentences) presented with ten cut-off frequencies (141, 224, 355, 562, 891, 1413, 2239, 3548, 5623, and 8913 Hz) and at seven SNRs (9, 6, 3, 0, -3, -6, and -9 dB) outlined in Section 7.1.5 above. Thus, the number of trials each participant completed in the word session was 7000 [50 words x 10 bands x 2 filter x 7 SNRs] and in the sentence condition was 2800 [20 sentences x 10 bands x 2 filter x 7 SNRs].

7.1.6.3 Procedure

The speech recognition test for words and sentences was conducted in two sessions on different days. In the word session, participants were required to listen
and repeat each word, and indicate the tone used (tone 1, tone 2, tone 3, tone 4) using the mouse to click their selection on the computer screen. For the word test, three recognition scores were recorded: (i) word recognition score (segment + tone correct); (ii) tone recognition score (tone correctly identified); and (iii) segment recognition scores (segment correct). For the sentence test, the recognition scores consisted of the number of words correctly identified in a 10-word sentence. The percentage of correct responses was recorded as a recognition score. Participants were able to take breaks whenever they felt fatigued.

7.1.7 Pilot study: results

Recognition scores for the sentence test ranged from 0 to 100 % suggesting that reduction to 10 frequency bands and seven SNRs were sufficient to provide a full range sentence recognition scores. However, in the word test, the range of scores was not ideal as they ranged between 0 to 80%. To overcome this, two additional quiet SNR conditions (+15 and +12 dB) were added to the word list conditions resulting an increase to 180 conditions [10 bands x 2 filters x 9 SNRs] for words. It is also important to note that hearing evaluation took 1.5 hours, and the sentence and word sessions each took approximately two hours (~5.5 hours in total).

7.2 Method for speech recognition test

7.2.1 Participants

Section 7.1.6.1 contains details of how participants were recruited and an overview of the demographics of the participant groups used in the study. A total 50 adults (25 males; 25 females) aged between 25 and 35 years participated in the
main SRT study. The questionnaire in Appendix F revealed that all participants were born and raised in China; had received mandatory Mandarin Chinese education; had arrived in Australia from China within the five previous years; and came from different Chinese provinces, ensuring there were speakers/listeners with a variety of accent backgrounds. All participants were undergraduate or graduate students in Sydney, Australia, and over the past three months typically spoke Mandarin from four to seven hours daily. Appendix F contains more detailed information in respect to the participants.

Hearing evaluation (see Section 7.2.2) showed all participants had normal hearing sensitivity with pure-tone thresholds around 15 dB HL between 250 and 8000 Hz in both ears (see Table 7.1). Another four participants were excluded based on their hearing evaluation and referred for further audiology assessment. None of the participants had formal musical education of more than five years, and none had received any musical training in the preceding three years. All participants were tested with the monosyllabic words, and 20 of these participants returned to complete the SRT for sentence materials. This group of participants also received travel reimbursements of $30 at the end of each session.

Table 7.1. Average hearing threshold of participants for left and right ears.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Right (dB HL)</th>
<th>SD Error (dB)</th>
<th>Left (dB HL)</th>
<th>SD Error (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>500 Hz</td>
<td>14.7</td>
<td>0.54</td>
<td>13.4</td>
<td>0.71</td>
</tr>
<tr>
<td>1000 Hz</td>
<td>15.9</td>
<td>0.62</td>
<td>12.7</td>
<td>0.64</td>
</tr>
<tr>
<td>2000 Hz</td>
<td>13.4</td>
<td>0.87</td>
<td>13.5</td>
<td>1.14</td>
</tr>
<tr>
<td>4000 Hz</td>
<td>9.9</td>
<td>0.88</td>
<td>9.3</td>
<td>0.86</td>
</tr>
<tr>
<td>8000 Hz</td>
<td>8.2</td>
<td>0.87</td>
<td>7</td>
<td>0.88</td>
</tr>
</tbody>
</table>
7.2.2 Hearing evaluation

On arrival, participants were given an information sheet about the study, and a consent form to sign (see Appendix B). To evaluate hearing, several hearing tests were administered to each participant prior to the speech recognition test. Firstly, an otoscopy was used to examine participants’ ear canals and ear drums to ensure there was no blockage of cerumen, or redness on the ear drums. Secondly, Interacoustic tympanometry equipment was used to test middle-ear function. Thirdly, a pure tone Interacoustic audiometer with earphones (Model TDH-39) was used to identify each participant’s hearing threshold levels based on their responses to pure tone stimuli. Both latter equipments were calibrated weekly as per ANSI S3.6 standards. Only subjects with bilateral pure tone threshold of 15dB HL or better at the frequency range between 125 Hz and 8000 Hz and with no middle ear dysfunction were recruited for the study. Tympanometry was repeated for the 20 participants who returned to complete the SRT for sentences. Based on the hearing evaluation guidelines of the American Hearing and Speech Association, pure tone hearing test is only required every six months, but tympanometry was conducted on the second visit, and participants were asked about noise exposure and any changes in hearing sensitivity since the last visit (see Appendix C).

7.2.3 Apparatus

The speech recognition tests were conducted in a sound-proof booth with noise level attenuation appropriate for diagnostic audio testing, according to the ANSI S3.1 (1991), relating to the maximum permissible ambient noise levels for audiomeric test rooms. A high-sensitivity Audio-Technica stereo headset (Model
BPHS1) was used to record participant’s responses. This was connected to an Edirol UA-25 audio interface with participant word or sentence repetition input recorded directly to CoolEdit, installed on a Lenovo T61p laptop. Also installed on the laptop was testing software, DMDX (http://www.u.arizona.edu/~kforster/dmdx/ dmdx.htm) which was used to sequence the speech recognition test presentations, collect oral responses and tone identification responses in .csv worksheets.

7.2.4 Speech recognition test materials

The SRT materials for monosyllabic words were four subtests from the MSTMs with each subtest being a 50 word list (N=200); and for sentences there were four subtests of 10-word sentences from the MHINT with each subtest containing 20 sentences (N=80) (Wong, Soli, et al., 2007). Developing FIFs typically involves each participant completing a word and sentence list under all filter frequency band and SNR conditions, but as reported elsewhere (Studebaker & Sherbecoe, 1993) and found in the pilot study (Section 7.1.6) participants report the task of completing a whole list of 50 words or 20 sentences in all conditions is extremely fatiguing. Because fatigue has the potential to impinge on the accuracy of the scores (DePaolis, et al., 1996), some researchers have increased the number of participants and reduced the number of words and/or sentences each participant is required to complete (Sherbecoe & Studebaker, 2003). The current study used a similar strategy by having each participant complete 180 conditions for words by randomly assigning a word from one of the four lists allocated to each condition; and similarly the same strategy was used for four sentence lists. Thus, there are two parts to this strategy. Firstly, each of the four lists of monosyllabic
MSTMs (or MHINT sentences) was assigned to each of the filter bandwidth and SNR conditions. Table 7.2 shows the conditions. The numbers one to four refer to which of the four lists of words (or sentences) is allocated to each condition. As each of the word and sentence lists is developed based on the criteria of familiarity, homogeneity, and phonetic balance of the speech materials, this suggests that there is no potential bias associated with one set of words or sentences over another. In addition, the word and sentence lists used were those lists which revealed the minimum subtest differences during evaluations by Han, and colleagues, 2009 and Wong, Soli, et al., 2007. Secondly, for each participant, the words (or sentences) from each of these lists in each condition was randomly assigned to each participant for the SRT. Excel was used to generate randomly assigned words from the lists to each condition for each participant. This strategy also negates the effects of learning and familiarity that can potentially occur when the same words (and sentences) are repeated under different filter bandwidth and SNR conditions.

Table 7.2. shows the word and sentence subtest lists allocated to HP (top) and LP (bottom) frequency bandwidth and SNR conditions.

<table>
<thead>
<tr>
<th>HP / SNR (dB)</th>
<th>15</th>
<th>12</th>
<th>9</th>
<th>6</th>
<th>3</th>
<th>0</th>
<th>-3</th>
<th>-6</th>
<th>-9</th>
</tr>
</thead>
<tbody>
<tr>
<td>141 Easier</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>224</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>355</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>562</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>891</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1413</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2239</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3548</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5623</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>8913 Harder</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>LP / SNR (dB)</td>
<td>15</td>
<td>12</td>
<td>9</td>
<td>6</td>
<td>3</td>
<td>0</td>
<td>-3</td>
<td>-6</td>
<td>-9</td>
</tr>
<tr>
<td>---------------</td>
<td>----</td>
<td>----</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>141 Harder</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>224</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>355</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>562</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>891</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>1413</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>2239</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3548</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>5623</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>8913 Easier</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

7.2.5 Procedure

Participants sat at a computer during testing and were instructed to listen to and repeat each word or sentence they heard. In the monosyllabic word session, they were also required to select the tone used on the word, by using the mouse to indicate on screen their choice of tone 1, tone 2, tone 3, or tone 4. DMDX software was used to present word and sentence trials, and to record and collect the speech recognition data. For each participant, the ID assigned by DMDX would randomly assign group of words, one for each of the 180 conditions, and likewise for the 140 sentence conditions. Responses were categorized into three subsets: (a) word recognition scores: correct only if the participant used the correct tone and phonemes of a given word; (b) tone recognition scores: correct only if the participant indicated the correct tone used on a word; (c) segment recognition scores: correct only if the participant used the correct phonemes in a word regardless the tone. For the sentence test session, participants were required to repeat each of the sentences they heard. Sentences were scored by counting, how many words were correct in a 10 word sentence.
7.3 Scoring and reliability

The approach to scoring depended on whether the materials were word segments or sentences. For word segments, the recorded speech tokens were scored correct or incorrect by the experimenter, a native Mandarin speaker. For ten-word sentences, the experimenter counted and recorded the number of words correct in each sentence recording. The reliability of the experimenter’s judgments was checked by having a second native Mandarin speaker who listened to at least 10% of the recordings for both speech material types. For word segments, there were a total of 9000 recorded responses (180 conditions x 50 participants). Of these ~14.5% (n=1310) of the recordings were selected for a reliability check. For the ten-word sentences, there were a total of 7000 recorded responses (140 conditions x 50 participants). Of these more than 11% (n=800) were used for the reliability check. Pearson correlation coefficients were produced to examine the relationship between the examiners’ scores. These showed significant correlations for word segments (r=0.99; p<0.0001), and for sentences (r=0.98; p<0.0001), an indication that there was high reliability between scorers (see Appendix G).

7.4 Overview of FIF development

Development of the FIFs based on the speech recognition data involved a number of steps. First, the SRT data were plotted and smoothed to ensure that performance decreases steadily as the speech material became less audible. Second, the relative transfer function (RTF) was derived using the bisection method to define the function of the relationship between SII values and SRT scores. In this procedure, the point of 100% speech recognition was identified as
having a SII with a value of 1.0; 0% speech recognition has a SII value of 0.0; and 50% speech recognition has a SII value of 0.5. These bisection procedures were continued until a total 13 reference points were collected, that is, SII values of 0.25, 0.125 etc. Thus, the importance of each frequency band was calculated using the RTF function to convert the SII values of each frequency band into a relative percentage scale. Thirdly, curve fitting software was used with these reference points as input to obtain the fitting constants Q and N (French & Steinberg, 1947). Finally, this fitting equation was used to calculate the weighted importance of each frequency band to create a FIF. The RTF was then modified by the outcome FIF and long-term average speech spectrum (LTASS) to produce the absolute transfer function (TF) which best describes the correlation between SII s and SRT scores.
Chapter 8 Results

This chapter details the outcomes of the project which, in the first instance, includes the performance intensity functions that demonstrate speech recognition performance at different SNRs for different materials; and the steps involved in deriving the Mandarin Chinese FIF. The technique for deriving Mandarin FIFs uses participants' speech recognition scores for HP and LP filtered frequency cutoffs, and SNRs to develop relative transfer function (RTF) data to depict the relationship between speech recognition scores and their expected speech intelligibility index (SII). Finally, the FIF data incorporates long term average spectrum (LTASS) to derive the absolute transfer function (TF) which depicts the relationship between recognition scores and the SII. In the next stage the data are smoothed to ensure the SRT performances decrease steadily as the speech stimuli become less audible. Second, the point where 100% speech recognition scores is identified as having a SII with a value of 1.0; 0% speech recognition, as having a SII value of 0.0; and the central point (50%), a SII value of 0.5. These bisection procedures are continued until a total 13 reference points are collected, that is, SII value of 0.25, 0.125 etc. Thirdly, curve fitting software uses these reference points to obtain the fitting constants Q and N (French & Steinberg, 1947). An SII and relative transfer function was then computed. Finally, this outcome equation was used to calculate the importance weight of each frequency band in the FIF. The relative transfer function used to derive the FIF can then be adjusted to take accounts into factors of LTASS to derive absolute TF, and is useful for clinical hearing evaluation.
8.1 Performance intensity functions

Performance intensity functions were generated by plotting the percentage correct in the SRT of LP 8913Hz as a function of SNRs. As shown in Figure 8.1, speech recognition performance decreases steadily as each of the speech materials becomes less intelligible and form asymmetric S-shaped curves with a range of 24 dB for scores from 0 to 100%. The figure shows that recognition scores for sentences are below 10% intelligibility from -9 to -3 dB, show a steep rise between -3 to +3 dB, and a tail with greater than 90% intelligibility from 3 to 15 dB. It is also shown that lexical tones are more recognized than sentences, words or segments at the lowest SNRs. Segments and words are almost equally intelligible, and more intelligible than sentences at low SNRs (-9 to -3 dB). However, sentences became more intelligible than words and segments above -3 dB. The SNR values necessary for 50% correct responses were approximately -5 dB for tones, -1.5 dB for sentences, and +3 dB for words and segments.

![Performance intensity functions](image)

Figure 8.1. Performance intensity functions showing the effects of SNR and sentence, word, segment and tone as a function of percentage correct.
8.2 Developing Mandarin frequency importance functions

In order to develop Mandarin FIFs, first the speech recognition scores are converted to percentages, then plotted as a function of HP and LP filtered frequency bands at each SNR. This allows observation of the CFs for each speech type or the points where the half of total speech intelligibility is located. The CF is calculated from smoothed data using the geometric mean of the intersect frequencies of all SNRs. The next step involves calculating the RTF using the curve bisection method to define SII references point at 0.5, 0.25 etc. which are then to identify the best fit curve (maximum $r^2$) for the constants $Q$ and $N$. From this, FIFs, cumulative FIF, and the absolute TF can be developed.

8.2.1 Speech recognition scores

Initially, the speech recognition scores for each of the speech materials were converted to percentages, and are shown in Tables 8.1 to 8.4. In all, there were 140 sentence conditions, and 180 word, segment, and tone conditions. The tables report the mean percentage scores for sentence (Table 8.1), word (Table 8.2), segment (Table 8.3) and lexical tone (Table 8.4) recognition as a function of HP and LP frequency cut-off conditions. These tables show that increasing amounts of HP or LP filtering reduced intelligibility and that sentence, word and segment recognition decreased to < 10% when the available frequencies were <500 Hz or >5000 Hz. However, for lexical tone at most SNRs, the recognition scores remain relatively high even in the more extreme frequency bands.
Table 8.1. Mean percentage recognition scores (%) for **sentences** for HP (left) and LP (right) filtered conditions.

<table>
<thead>
<tr>
<th>filter cut-off (Hz)</th>
<th>High Pass (SNR)</th>
<th>Low Pass (SNR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9  6  3  0 -3 -6 -9</td>
<td>9  6  3  0 -3 -6 -9</td>
</tr>
<tr>
<td>141</td>
<td>100 99 89.5 53 32.5 2 1</td>
<td>1 0 1 0 0 0 0</td>
</tr>
<tr>
<td>224</td>
<td>100 88.5 90 51 13.5 0 0</td>
<td>2 0 0 0 0 0 0</td>
</tr>
<tr>
<td>355</td>
<td>100 96.5 79 70.5 10 2 0</td>
<td>8.5 4 1 1 0 0 0</td>
</tr>
<tr>
<td>562</td>
<td>99 94 79.5 52 10.5 1.5 0</td>
<td>29 15.5 7.5 0 2 0</td>
</tr>
<tr>
<td>891</td>
<td>86 99 65.5 11.5 34.5 1.5 0.5</td>
<td>90.5 82 60.5 18.5 2 0 0</td>
</tr>
<tr>
<td>1413</td>
<td>44 31.5 14 20 4.5 0 0.5</td>
<td>95 98 79 27 14 0 1</td>
</tr>
<tr>
<td>2239</td>
<td>43 18.5 6.5 17 2.5 0 0</td>
<td>94.5 93.5 87 24 15 2.5 0</td>
</tr>
<tr>
<td>5623</td>
<td>1 0.5 0 0 1.5 0 0</td>
<td>94 100 87 57 9.5 0.5 0</td>
</tr>
<tr>
<td>8913</td>
<td>0.5 0 3.5 5 0 0.5 0.5</td>
<td>97.5 99.5 97 72.5 7.5 3.5 0</td>
</tr>
</tbody>
</table>

Table 8.2. Mean percentage recognition scores (%) for **words** for HP (left) and LP (right) filtered conditions.

<table>
<thead>
<tr>
<th>filter cut-off (Hz)</th>
<th>High Pass (SNR)</th>
<th>Low Pass (SNR)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15 12 9 6 3 0 -3 -6 -9</td>
<td>15 12 9 6 3 0 -3 -6 -9</td>
</tr>
<tr>
<td>141</td>
<td>98 86 82 64 40 40 14 12 2</td>
<td>2 2 4 0 2 0 0 0 0</td>
</tr>
<tr>
<td>224</td>
<td>94 90 70 60 48 24 22 12 2</td>
<td>0 0 0 0 0 0 2 2 0 0</td>
</tr>
<tr>
<td>355</td>
<td>86 82 58 52 50 26 16 2 0</td>
<td>10 0 2 6 0 2 0 0 0</td>
</tr>
<tr>
<td>562</td>
<td>88 76 72 52 48 24 10 6 0</td>
<td>16 10 10 4 4 2 0 0 2</td>
</tr>
<tr>
<td>891</td>
<td>70 52 50 38 34 40 8 2 0</td>
<td>8 26 12 4 10 4 0 0 0</td>
</tr>
<tr>
<td>1413</td>
<td>58 58 46 52 20 14 14 2 0</td>
<td>30 36 28 26 12 10 10 0 0</td>
</tr>
<tr>
<td>2239</td>
<td>38 30 30 20 16 10 4 2 0</td>
<td>62 68 46 34 24 26 4 2 0</td>
</tr>
<tr>
<td>3548</td>
<td>14 10 14 4 6 10 4 0 4</td>
<td>62 70 52 54 32 24 10 2 2</td>
</tr>
<tr>
<td>5623</td>
<td>4 8 2 0 0 2 0 0 0</td>
<td>92 80 86 42 48 30 14 8 2</td>
</tr>
<tr>
<td>8913</td>
<td>0 2 2 2 2 0 0 0 0</td>
<td>96 86 82 72 48 40 22 6 0</td>
</tr>
<tr>
<td>filter cut-off (Hz)</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>-------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>141</td>
<td>98</td>
<td>86</td>
</tr>
<tr>
<td>224</td>
<td>94</td>
<td>90</td>
</tr>
<tr>
<td>355</td>
<td>86</td>
<td>82</td>
</tr>
<tr>
<td>562</td>
<td>88</td>
<td>76</td>
</tr>
<tr>
<td>891</td>
<td>70</td>
<td>54</td>
</tr>
<tr>
<td>1413</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>2239</td>
<td>42</td>
<td>32</td>
</tr>
<tr>
<td>3548</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>5623</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>8913</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 8.4. Mean percentage recognition scores (%) for lexical tones for HP (left) and LP (right) filtered conditions.
8.2.2 Cross-over frequencies

CFs are the intersection points of plotted speech recognition scores for HP and LP filtered frequency bands at each SNR, and set the point of equal intelligibility for HP and LP filtered materials. Figures 8.2 to 8.5 show the speech recognition scores plotted as a function of HP and LP cut-off frequencies. Results from HP filtering conditions start from the left side of each graph and those from LP filtering conditions start from the right side. Unsmoothed data is shown as dotted lines, and smoothed data is shown as solid lines for sentences in Figure 8.2, for words in Figure 8.3, for segments in Figure 8.4 and for tones in Figure 8.5. At present, there is no consensus on a method for smoothing data. The ANSI S3.5 (1997) states smoothing should be conducted using interpolation and extrapolation of the speech recognition scores. The Studebaker and Sherbecoe (1991) method (detailed in Section 5.4.5), however, remains the most popular approach to data smoothing. It used four expert judges to visually smooth the data to fit curves and reported the mean of the four judges as the final smoothed functions. Also, it is important to note that each curve was fitted individually. The current study differed as it, firstly, applied a frequency (Hz) logarithmic scale to the speech recognition data; and secondly, the smoothing approach also took into account each curve’s relationship to neighboring curves to further reduce artifacts (see Appendix H). In the left panels of Figures 8.2 to 8.5, vertical solid lines were plotted to represent the location of the CFs of sentences, words, segments and tones.
Figure 8.2. Sentence recognition scores plotted as a function of cut-off frequency at different SNRs. The solid lines indicate the smoothed values; the dashed lines indicate the unsmoothed values. The left panel displays SNRs +9, +3 and -3 dB; the right panel displays SNRs +6, derived +1, 0, and -6 dB. The location of CF is marked on the left panel. The procedures used to derive points on the SII relative transfer function are plotted at the right panel.

Figure 8.3. Word recognition scores plotted as a function of cut-off frequency at different SNRs. The solid lines indicate the smoothed values; the dashed lines indicate the unsmoothed values. The left panel displays SNRs +12, +6, 0 and -6 dB; the right panel displays SNRs +15, +9, +3 and -3 dB. The location of CF is marked in the left panel. The procedures used to derive points on the speech intelligibility transfer function are plotted at the right panel.
Figure 8.4. Segment recognition scores, plotted as a function of cut-off frequency at different SNRs. The solid lines indicate the smoothed values; the dashed lines indicate the unsmoothed values. The left panel displays SNRs +12, +6, 0 and -6 dB; the right panel displays SNRs +15, +9, +3 and -3 dB. The location of CF is marked on the left panel. The procedures used to derive points on the speech intelligibility transfer function are plotted at the right panel.

Figure 8.5. Tone recognition scores plotted as a function of cut-off frequency at different SNRs. The solid lines indicate the smoothed values; the dashed lines indicate the unsmoothed values. The left panel displays SNRs +15, +9, +3 and -6 dB; the right panel displays SNRs +12, +6, 0 and -3 dB. The location of CF is marked on the left panel. The procedures used to derive points on the speech intelligibility transfer function are plotted at the right panel.
The CF of each speech type is calculated by using the geometric mean of the intersect frequencies of all SNRs using smoothed data. Table 8.5 report the frequencies of the intersect points of the SNRs for four speech types. The smoothed geometric average of the CFs for sentences, words, segments, tones are 1570 Hz, 1807 Hz, 1813 Hz, 743 Hz, respectively. It was not possible to calculate the CFs at the SNR of -9 dB for sentences, words and segments due to the recognition scores being too low across all filtering conditions. Furthermore, the CF of lexical tone at SNR -6 dB is very different from the other SNRs in tone conditions, and thus this score (1853 Hz) was excluded from the geometric average for tones to avoid skewing the data. It is also important to note that the smoothed geometric CF has to be close to the geometric CF of the unsmoothed data (see Appendix I) otherwise it would suggest that the data was biased by smoothing procedures. The CF differences between smoothed and unsmoothed data were relatively small (the maximum difference is 74 Hz for the tones).

Table 8.5. CFs (Hz), for sentences, words, segments and lexical tones under different SNR conditions.

<table>
<thead>
<tr>
<th>SNR</th>
<th>Sentence</th>
<th>Word</th>
<th>Segment</th>
<th>Lexical Tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>1826</td>
<td>1826</td>
<td>726</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1800</td>
<td>1826</td>
<td>700</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1309</td>
<td>1800</td>
<td>1800</td>
<td>726</td>
</tr>
<tr>
<td>6</td>
<td>1400</td>
<td>1800</td>
<td>1800</td>
<td>750</td>
</tr>
<tr>
<td>3</td>
<td>1413</td>
<td>1810</td>
<td>1800</td>
<td>700</td>
</tr>
<tr>
<td>0</td>
<td>1648</td>
<td>1826</td>
<td>1826</td>
<td>726</td>
</tr>
<tr>
<td>-3</td>
<td>2239</td>
<td>1800</td>
<td>1826</td>
<td>891</td>
</tr>
<tr>
<td>-6</td>
<td>1800</td>
<td>1800</td>
<td>1853</td>
<td></td>
</tr>
</tbody>
</table>

Geometric Average: 1570.5 1807.7 1813.0 743.3
SD: 377.6 11.8 13.9 66.4
8.2.3 Relative transfer function

The RTF is derived through fitting relative data between SII values and speech recognition scores using the curve bisection method (Duggirala, Studebaker, Palovic, 1988, Studebaker, 1987). The RTF assumes that the maximum SII is 1.0 and equals the unfiltered condition with the highest speech recognition scores. Other SNR and cut-off frequency conditions in turn have SII relative to the maximum SII in relation to recognition scores. Thus, the curve bisection procedure (Studebaker, et al., 1987; Studebaker & Sherbecoe, 1991) used to prepare the RTF is shown in the right panels of Figures 8.2 to 8.5. As shown in these panels, the smoothed HP and LP scores for the highest SNR were plotted as a function of filter cut-off frequency. For example, in the sentence data for curves of +9 dB (Figure 8.2), the cut-off frequency conditions below 3548 Hz and above 562 Hz should be attributed maximum recognition. However, according to Wong, and colleagues (2007) that developed the Cantonese FIF, the SNR +9 dB recognition scores of cut-off frequency LP 3548 Hz and HP 562 Hz is likely due to a ceiling effect and not appropriate for deriving the RTF. Similarly, in this study, SNR +9 dB curves were not used for developing the RTF. Instead, the frequency where the +6 dB SNR curves intersect (CF=1400 Hz), was used. Therefore, the cut-off frequency conditions below 5623 Hz and above 244 Hz were attributed maximum recognition. The SII at the CF (1400Hz) equals 0.50 because half of the total auditory area is available to the listener above or below this point, and the total area for the SNR +6 dB is assumed to have a SII of 1.00.

The next point to be established was 0.25 SII on the transfer function. This is determined by using the score at 0.5 SII as the maximum score, identifying a pair of HP and LP curves terminating at the score at that CF (74% shown in
right panel Figure 8.2). However, none of the tested SNRs produced scores in that
matched the 0.50 SII score (74%), therefore curves were derived from the data set
which terminated at approximately 74%, that is, at +1 dB SNR. The curves of the
derived +1 dB SNR shown have scores for LP at 75.5%; and for HP for 74%. As
previously, the 0.25 SII at the CF was equal to one-half of the previous SII value
of 0.5. In this case, the SII value of 0.25 is related to score, 30%, of the CF. The
process of halving the SII continues until scores of successively smaller SII
fractions had been determined.

Scores for SII values above 0.50 were obtained by identifying locations on
the curves for the +6 dB condition that complemented those below 0.50. For
example, as shown in the right panels of Figures 8.2 to 8.5, the 0.75 SII point was
produced by extending a horizontal line from the score for 0.25 SII CF until it
intersected the HP and LP curve for SNR +6 dB condition. A vertical line was
then drawn from the point on the LP curve until it touched the HP curve at +6 dB
SNR. This produced a HP curve estimate for cutoff frequencies and score for 0.75
SII. The LP curve estimate was obtained similarly, that is, by drawing a vertical
line from the HP curve until it touched the LP curve at +6 dB SNR. The final
score for 0.75 SII was the average of the HP (94.4 %) and LP (93.5%) estimates.
Other points were found in a similar fashion. The above bisection procedures
were applied to word, segment and tone data, and plotted in the right panels of
Figure 8.3 to 8.5. For each speech material, 13 sets of SII values were calculated
from the corresponding speech recognition scores for each stimulus type. SPSS
was used to fit the scores and the corresponding SIIIs using the equation:
Equation 1: \[ S = \left( 1 - 10^{-PA/Q} \right)^N \]

Equation 2: \[ A = -\frac{Q}{P} \log \left( 1 - s^{1/N} \right) \]

It should be noted that Equation 2 is a transformation of Equation 1. Equation 1 is used to predict percentage correct of speech recognition scores from the SII, whereas equation 2 is used to predict the SII from percentage correct speech recognition scores. In each equation, \( S \) is the speech recognition score, \( A \) is the SII value; \( P \) stands for a proficiency factor that accounts for listener’s competence and practice and is assumed to be 1 for native speakers. Both \( Q \) and \( N \) are fitting constants and depend on the characteristics of the speech materials (Fletcher, 1948). More specifically, \( Q \) is a correction factor to compensate for changes in proficiency with the test materials under experimental conditions; and \( N \) represents the number of independent sounds in a test item or a constant that controls the shape of the function curve (Studebaker & Sherbecoe, 1991).

To obtain the fitting constants \( Q \) and \( N \), the scores and the corresponding SII obtained were substituted into both equations in order to identify the best fit curve (maximum \( r^2 \)) for \( Q \) and \( N \) using SPSS. Table 8.6 reports fitting constants and \( r^2 \) of RTF. Figures 8.6 to 8.9 show the fitting curves for Equation 1 in the left panel and for Equation 2 in the right panel for each sentences, words, segments and tones. Among all four speech types, the \( r^2 \) values are above 0.9, except for tones, for the prediction of either SRT scores or SII values, indicating that the curve fitting model provided a good fit to the data. The comparison of constants and \( r^2 \) values of the speech types revealed that the \( Q \) and \( N \) values (Table 8.6) of the words and segments are similar and have similar RTF slopes as shown in Figure 8.7 and 8.8.
Table 8.6. Relative transfer functions for predicting Score and SII and fitting constants for different speech materials.

<table>
<thead>
<tr>
<th>Prediction</th>
<th>% Score</th>
<th>Q</th>
<th>N</th>
<th>R²</th>
<th>Prediction</th>
<th>SII</th>
<th>Q</th>
<th>N</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence</td>
<td>0.53</td>
<td>3.89</td>
<td>0.94</td>
<td></td>
<td>Sentence</td>
<td>0.44</td>
<td>5.74</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Word</td>
<td>0.86</td>
<td>2.23</td>
<td>0.92</td>
<td></td>
<td>Word</td>
<td>0.68</td>
<td>3.20</td>
<td>0.94</td>
<td></td>
</tr>
<tr>
<td>Segment</td>
<td>0.82</td>
<td>2.43</td>
<td>0.93</td>
<td></td>
<td>Segment</td>
<td>0.69</td>
<td>3.18</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Tone</td>
<td>0.42</td>
<td>0.88</td>
<td>0.80</td>
<td></td>
<td>Tone</td>
<td>0.46</td>
<td>0.73</td>
<td>0.82</td>
<td></td>
</tr>
</tbody>
</table>
Figure 8.6. Relative transfer function for sentences. The left panel transfer function predicts the % correct score from SII, is described by Equation 1 and has Q and N values of 0.53 and 3.89, respectively; the right panel transfer function predicts the SII from % correct score, is described by Equation 2 and has Q and N values of 0.44 and 5.74, respectively.

Figure 8.7. Relative transfer function for words. The left panel transfer function predicts the % correct score from SII, is described by Equation 1 with Q and N values of 0.86 and 2.23, respectively; the right panel transfer function predicts the SII from % correct score, is described by Equation 2, and has Q and N values of 0.68 and 3.20, respectively.
Figure 8.8. Relative transfer function for segments. The left panel transfer function predicts the % correct score from SII, is described by Equation 1 and has Q and N values of 0.82 and 2.43, respectively; the right panel transfer function predicts the SII from % correct score, is described by Equation 2 and has Q and N values of 0.69 and 3.18, respectively.

Figure 8.9. Relative transfer function for tones. The left panel transfer function predicts the % correct score from SII, is described by Equation 1, and has Q and N values of 0.42 and 0.88, respectively; the right panel transfer function predicts the SII from % correct score, is described by Equation 2, and has Q and N values of 0.46 and 0.73, respectively.
8.2.4 Frequency importance functions

The RTF was used to determine FIFs. The procedures used were comparable to those used in previous studies (Duggirala et al., 1988; Studebaker et al, 1987), and involved the following calculations.

1) Mean HP and LP recognition scores for each SNR were transformed into SII values by using Equation 2: 
\[ A = -\frac{Q}{p} \log \left(1 - \frac{1}{S^\frac{1}{N}}\right) \]
and inserting the recognition scores and the fitting constants Q and N from Table 8.6 for each of the speech types. For example, the equation for HINT sentences is \( A = 0.44 \log (1-S^{1.74}) \). For the LP condition at 891 Hz, SNR +6 dB, the converted SII is 0.4678, obtained by substituting the recognition score of 59.5 from Table 8.1.

2) SII values for each frequency band were calculated. For the LP data, this was done by subtracting the SII value of the lower cut-off frequency from the SII value for the higher cut-off frequency. For example in Table 8.1 for sentences, for SNR +6 dB, the LP SII value between 562 Hz and 891 Hz was obtained by subtracting the SII value at 562 Hz, corresponding to a percent correct score of 15.5% from the SII value at 891 Hz, corresponding to a percent correct score of 59.5%. For the HP data, the reverse procedure was used, that is, by subtracting the SII value at 891 Hz from the SII value at 562 Hz. This results in two estimates for SII, one HP estimate and one LP estimate, for each band at each SNR. These estimates were then averaged to provide a single estimate for each frequency bandwidth and SNR. Table 8.7 provides example data for sentences.
3) Frequency importance values can now be calculated by reference to Table 8.7. Table 8.7 shows example data for sentences. It contains the mean SII values of HP and LP for each bandwidth x SNRs [as discussed in (2) above], and also shows the average SIIs at the SNR for each frequency band. More importantly, it shows the percentage importance (%) at each frequency band. The percentage importance or FIF is the average SII converted to a percentage, and was achieved by dividing each average SII by the sum (0.54) and multiplying by 100. Importance values of FIF for each material type are reported in Table 8.8, and plotted in Figure 8.10.

Table 8.7. Sentence SII as a function of SNR and frequency bandwidth. The final weight or importance (%) for each band was calculated from the average SII of all SNRs for each frequency bandwidth.

<table>
<thead>
<tr>
<th>Centre (Hz)</th>
<th>Band (Hz)</th>
<th>Signal to Noise Ratio</th>
<th>Avg.</th>
<th>Importance (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>+6 dB</td>
<td>+3 dB</td>
<td>0 dB</td>
</tr>
<tr>
<td>125</td>
<td>110-141</td>
<td>0.050</td>
<td>0.0384</td>
<td>0.0273</td>
</tr>
<tr>
<td>178</td>
<td>141-224</td>
<td>0.021</td>
<td>0.0137</td>
<td>0.0087</td>
</tr>
<tr>
<td>282</td>
<td>224-355</td>
<td>0.039</td>
<td>0.0226</td>
<td>0.0132</td>
</tr>
<tr>
<td>447</td>
<td>355-562</td>
<td>0.080</td>
<td>0.0436</td>
<td>0.0230</td>
</tr>
<tr>
<td>708</td>
<td>562-891</td>
<td>0.144</td>
<td>0.0871</td>
<td>0.0451</td>
</tr>
<tr>
<td>1122</td>
<td>891-1413</td>
<td>0.194</td>
<td>0.1390</td>
<td>0.0767</td>
</tr>
<tr>
<td>1778</td>
<td>1413-2239</td>
<td>0.200</td>
<td>0.1578</td>
<td>0.0922</td>
</tr>
<tr>
<td>2818</td>
<td>2239-3548</td>
<td>0.174</td>
<td>0.1365</td>
<td>0.0797</td>
</tr>
<tr>
<td>4467</td>
<td>3548-5623</td>
<td>0.137</td>
<td>0.0931</td>
<td>0.0501</td>
</tr>
<tr>
<td>7080</td>
<td>5623-8913</td>
<td>0.092</td>
<td>0.049</td>
<td>0.0257</td>
</tr>
</tbody>
</table>

**SUM** | **0.54** | **100** |
Table 8.8. Importance weighting of each octave band for sentences, words, segments and tones.

<table>
<thead>
<tr>
<th>Centre (Hz)</th>
<th>Band (Hz)</th>
<th>Sentence (%)</th>
<th>Word (%)</th>
<th>Segment (%)</th>
<th>Tone (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>125</td>
<td>110-141</td>
<td>5.13</td>
<td>11.28</td>
<td>11.25</td>
<td>6.52</td>
</tr>
<tr>
<td>178</td>
<td>141-224</td>
<td>1.86</td>
<td>2.98</td>
<td>2.75</td>
<td>9.15</td>
</tr>
<tr>
<td>282</td>
<td>224-355</td>
<td>3.15</td>
<td>4.47</td>
<td>4.07</td>
<td>10.27</td>
</tr>
<tr>
<td>447</td>
<td>355-562</td>
<td>6.09</td>
<td>6.89</td>
<td>6.31</td>
<td>11.14</td>
</tr>
<tr>
<td>708</td>
<td>562-891</td>
<td>11.52</td>
<td>10.17</td>
<td>9.61</td>
<td>11.69</td>
</tr>
<tr>
<td>1122</td>
<td>891-1413</td>
<td>17.25</td>
<td>13.28</td>
<td>13.13</td>
<td>11.75</td>
</tr>
<tr>
<td>1778</td>
<td>1413-2239</td>
<td>19.02</td>
<td>14.79</td>
<td>15.21</td>
<td>11.35</td>
</tr>
<tr>
<td>2818</td>
<td>2239-3548</td>
<td>16.61</td>
<td>14.25</td>
<td>14.95</td>
<td>10.59</td>
</tr>
<tr>
<td>4467</td>
<td>3548-5623</td>
<td>12.06</td>
<td>12.26</td>
<td>12.83</td>
<td>9.49</td>
</tr>
<tr>
<td>7080</td>
<td>5623-8913</td>
<td>7.31</td>
<td>9.67</td>
<td>9.89</td>
<td>8.04</td>
</tr>
</tbody>
</table>

Sentences, words and segment FIFs tend to peak in the bandwidth 1413-2239 Hz, (centered around 1800 Hz). Word and segment FIFs are similar, but sentence FIF has 10% more importance across a broader span (between center frequencies 708 and 4467 Hz) than words and segments. The FIF findings reported in Figure 8.10 also indicate secondary peaks in the 110-141 Hz bandwidth in three of the speech materials, sentences (~5%), words (~11%) and segments (~11%). Here words and segments contain twice the importance of sentence materials.

Of all the FIFs, however, the FIF for lexical tone has the most distinctive shape. The importance distribution of tone FIF is curved slightly across frequency bands. Between 355 Hz to 3548 Hz, importance weighting is ~11%, but there is less than 2% difference in importance weighting between 141 Hz and 5623 Hz, and even at extreme frequencies of <141 and >5623 the differences are not dramatic [<141 (6.5%) and >5623 (8%)]. Thus the importance distribution for the FIF for lexical tone shows relative equivalence across frequency bands.
The cumulative FIF values of the four speech types across frequencies are plotted in Figure 8.11. Cumulative FIFs are calculated by summing all importance weightings (%) that occur at each of the center frequencies, with all the importance weightings (%) that occur at center frequencies below the target frequency. For example, for sentences, the cumulative FIF value at 708 Hz, would be 11.5 (708 Hz) plus the values at each centre frequency below 708 Hz, that is, + 6.1 (447 Hz) + 3.2 (282 Hz) + 1.9 (178 Hz) + 5.1 (112 Hz). Cumulative FIFs plot center frequencies as a function of percentage correct and allow the technician/clinician to predict frequency bandwidths necessary to obtain a score of, for example, 50% correct.

Figure 8.10. The frequency importance functions for sentences, words, segments and tones.
8.2.5 Absolute transfer function

A transfer function shows the relationship between speech recognition scores and SII values, and can be used, for example, to predict intelligibility from speech recognition performance or to provide information for the calculation of a hearing aid response for a given audiogram in order to maximize intelligibility. The RTF derived in section 8.2.3 depicts the same relationship between speech recognition scores and SII values, but is developed based on listening to the speech of a single male talker. In order to take into account the considerable speech variability among talkers, the RTF is adjusted using LTASS (see Section 3.2.3) (D. Byrne, et al., 1994). This new adjusted function aims to predict the best SII values from SRT scores for everyday communication.
The adjustment from RTF to TF was performed using the following equation:

**Equation 3:** \[ \text{SII} = P \sum_{i=1}^{n} \left[ \frac{(\text{SNR} + K)}{\text{LTASS}} \right] \times \text{FIF}_i \]

It should be noted that Equation 3 is an expansion of the SII equation from Section 5.3.1: \[ \text{SII} = P \sum_{i=1}^{n} \text{LiWiDi} . \] In Equation 3, the SII is expressed as a factor ranging from 0.0 to 1.0 from a speech audibility which represents a importance weighted audibility. In order to find the best function for converting the speech recognition scores into SII values, the Q and N values that best describe average recognition performance were determined. First, an SII value for each listening condition (SNR x filter cut-off) was calculated using Equation 3. Each SII was the product of frequency weight (FIF) and the audibility weight \[ \left[ \frac{(\text{SNR} + K)}{\text{LTASS}} \right] \] for the condition. To determine audibility weight, the unknown is the constant (K). The starting estimate for determining K was 12 (as suggested by ANSI S3.5 (1997), and once calculated, a new K is selected moving to higher and lower numbers until the maximum \( r^2 \) is obtained.

Following this, unsmoothed mean scores of each condition are plotted as a function of their SII values, and Q and N (Equation 1) varied until the best fit curve was obtained. The Ks that produced maximum \( r^2 \) are shown in Table 8.9, together with associated Q and N values, and \( r^2 \) for each speech material type. The absolute TFs for the four sets of speech materials are shown in Figures 8.12 to 8.15. Among the speech types, \( r^2 = -0.9 \) for the prediction of scores for sentences, words and segments. For lexical tones \( r^2 = 0.79 \). These values indicate the K value that is appropriate for fitting speech recognition data into the curve model (Equation 3). For the further comparison, the TFs of sentences, words, segments, and tones, are shown in Figure 8.16. This shows that lexical tones have the
steepest slope followed by sentences, and those words and segments both have similar shallower slopes. For example, the SII of 0.5 corresponds to speech recognition scores of ~70% for segments and words, and ~97 % for tone and sentences. It is possible to further adjust the Q value to compensate for changes in proficiency, such as when participants have been familiarized with speech materials prior to SRT test (Studebaker & Sherbecoe, 1991), but participants in this study were not familiarized with the test content, and no further adjustment was conducted to the Q values.

Table 8.9. Fitting constants for the best fit TFs.

<table>
<thead>
<tr>
<th>Prediction Score</th>
<th>K</th>
<th>Q</th>
<th>N</th>
<th>$R^2$</th>
<th>Prediction Score</th>
<th>K</th>
<th>Q</th>
<th>N</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentence</td>
<td>6.8</td>
<td>0.23</td>
<td>4.06</td>
<td>0.90</td>
<td>Sentence</td>
<td>6.8</td>
<td>0.16</td>
<td>7.38</td>
<td>0.86</td>
</tr>
<tr>
<td>Word</td>
<td>7.0</td>
<td>0.52</td>
<td>2.90</td>
<td>0.93</td>
<td>Word</td>
<td>7.0</td>
<td>0.41</td>
<td>3.89</td>
<td>0.90</td>
</tr>
<tr>
<td>Segment</td>
<td>8.0</td>
<td>0.55</td>
<td>2.92</td>
<td>0.94</td>
<td>Segment</td>
<td>8.0</td>
<td>0.43</td>
<td>3.92</td>
<td>0.90</td>
</tr>
<tr>
<td>Tone</td>
<td>9.0</td>
<td>0.30</td>
<td>1.51</td>
<td>0.85</td>
<td>Tone</td>
<td>9.0</td>
<td>0.23</td>
<td>2.08</td>
<td>0.79</td>
</tr>
</tbody>
</table>
Figure 8.15. Absolute TF for sentences using constant, K=6.8, is best described by Equation 2, with values for Q and N of 0.23 and 4.06, respectively.

Figure 8.14. Absolute TF for words using constant, K=7, is best described by Equation 2, with values for Q and N of 0.52 and 2.90, respectively.

Figure 8.13. Absolute TF for segments using constant, K=8, is best described by Equation 2, with values for Q and N of 0.55 and 2.92, respectively.

Figure 8.12. Absolute TF for lexical tones using constant, K=9, is best described by Equation 2, with values for Q and N of 0.30 and 1.51, respectively.
8.3 Summary

The FIF shapes across frequency bands are similar for sentences, word and segments, and are weighted most heavily around 1800 Hz (1413 to 2239Hz). There are also secondary peaks below 140 Hz for words and segments, and a similar but smaller peak for sentences. Lexical tones, on the other hand, tend to be weighted equivalently across most frequency bandwidths with somewhat less weight at extreme frequency bands, resulting in a curvilinear shape importance function. This study is the first tonal language study to separate the effects of words, segments and lexical tones on the FIF, and the results show substantial differences between those reported for English (DePaolis, et al., 1996; Sherbecoe & Studebaker, 2002; Studebaker, et al., 1987; Studebaker & Sherbecoe, 1991; Studebaker, et al., 1993), and those reported for Cantonese Chinese (Wong, Ho, et al., 2007). The FIF peak below 141 Hz for Mandarin sentences, words, and
segments (not lexical tone) is interesting, as it indicates that lexical tone per se is not influencing this weighting but that the phonemic characteristics of the Mandarin Chinese maybe. The results, their comparisons to FIF in other languages and their implications will be discussed in the next chapter.
Chapter 9 Discussion

9.1 Overview of findings

Current research on Mandarin FIFs is far from adequate in providing speech intelligibility values for clinical hearing evaluations in order to make appropriate language-specific adjustments to hearing aids. In this study, speech recognition scores were collected under different noise conditions (SNRs), and HP and LP filtered frequency bandwidths, all of which exert an influence on speech intelligibility. From the word and sentence materials used in the test, four different speech types: sentences, words, segments and tones were scored to determine the frequency importance weightings for Mandarin lexical tone, phonemic information, and sentence context. Development of FIFs in this way, not only provides an understanding of the frequency band weightings that affects different speech types, but in addition provides performance intensity functions to aid understanding of how speech recognition changes as a function of SNR and type of speech material; and finally, it provides a series of CFs and absolute TFs that enable prediction of SII values for listeners in everyday communication. Performance intensity functions, CFs, FIFs and TFs are discussed in turn below after a brief overview of the results at the word and sentence level.

9.1.1 Overview

The findings for Mandarin FIFs are depicted in Figure 8.10 and Table 8.8. Findings at the monosyllabic word level were separated into three categories, word (segment + tone), segment (segment only) and tone (tone only). The findings are remarkably similar for word and segment and show a primary peak in
the bandwidth centered at ~1800 Hz, whereas the FIF for lexical tone has its own unique slightly curvilinear shape. At the sentence level, the FIF results reveal a narrow peak in percentage weighting which also occurs at center frequency ~1800 Hz. Essentially, the span is similar to that for words and segments, but shows a stronger narrower peak for sentences (~19%) compared to words and segments (~14%). Results are similar to those in English language studies, which also show a greater percentage weighting for sentences than words with a frequency centre located at ~1800 Hz peak. Albeit, it should be noted that the percentage weighting differs between English language studies, depending on the speech materials used in the recognition test, the size and numbers of the bandwidths used, etc. For instance, (DePaolis et al., 1996) found the peak percentage weighting for sentences (using SPIN test) was ~29% and for words (using PB-50 test) was ~26% whereas separate studies with different word recognition tests by Studebaker and colleagues found a percentage weighting of ~22% using the W22 test (Studebaker & Sherbecoe, 1991) and ~20% using the NU6 test (Studebaker, et al., 1993). Thus, it can be seen that the nature of the test materials, irrespective of the language used, can affect the importance weighting distribution to some extent.

Secondary low-frequency peaks in the 110-141 Hz bandwidth were also found in three of the speech materials, sentences (~5%), words (~11%) and segments (~11%). This type of low frequency weighting has also been found in Cantonese Chinese sentence data (Wong, Ho, et al., 2007) but not Mandarin word data (J. Chen, et al., 2008; Yeh, 2005). This will be discussed further below (Section 9.4.2). Of all the FIFs, however, the FIF for lexical tone has the most distinctive shape as the importance distribution is slightly curved across all frequency bands. Thus, the importance weight distribution for lexical tone shows
relative equivalence across frequency bands. In addition, the psychometric function showed that tone recognition was more resistant to noise than words, segments, and sentences, that is, lexical tones were more easily recognized under noisier conditions, e.g., -9 dB (see Figure 8.1).

9.2 Performance intensity functions

Speech recognition performance is always better when a listener has fewer choices, e.g., Miller et al. (1951) found participants recognized digits better than words in sentences. Nonsense syllables were understood least well (see Section 5.2.5). Thus, for lexical tones, the listener only requires an impression of relative F0 height, pitch contour shape, and duration to identify the differences between the tones. Because there are only four possible alternatives for tones in Mandarin Chinese, it is easier to identify them correctly under poor SNRs. And as noted earlier, the pitch information is spread widely across frequencies. With segments and words, however, there are many more possibilities, and the listener must perceive each phoneme correctly, as it is not enough to recognize the vowel and/or tone to identify a test word. In sentences, there is a more restricted range of possibilities for each word given neighboring words provides a context. These restrictions should make sentence recognition easier under unfavorable SNR conditions but this was not the case (see below).

The performance intensity function shows percentage correct as a function of SNR for Mandarin tones, segments, words, and sentences in Figure 8.1, and reveals that lexical tones were more audible than sentences, words or segments at all SNRs. In addition, segments and words are almost equally intelligible, and less intelligible than sentences at SNRs 0 dB to 15 dB. This suggests that speech recognition scores benefit from context/redundancy effects of lexical tone and
sentences. For example, to obtain 50% correct score, lexical tones can be presented at -6 dB SNR, sentences at -1.5 dB (between -3 dB and 0 dB SNR), while words and segments need to be presented at +3 dB SNR or quieter. The context/redundancy benefit in Mandarin speech recognition scores for sentences and words at a given SNR is consistent with English data (DePaolis, et al., 1996; Miller, et al., 1951).

Typically English language studies find steep curves in percentage correct for sentences across the range of SNRs. In the current study, Mandarin sentence recognition scores at SNRs between -9 and -3 dB are poorer than scores for tones, words and segments, but rise steeply for SNRs at 0 dB to 15 dB. However, there are factors that should be taken in to account when considering the low sentence recognition scores in the noisier conditions. It should be recalled that Mandarin is constructed of monosyllabic units that each have their own meaning and that these units can be put together to make up polysyllabic words, phrases and sentences. To perceive Mandarin monosyllabic words and sentences successfully the listener must perceive both phonemic and tonemic information correctly. According to the literature (Section 3.4.4), when Mandarin words are presented in continuous speech, the tonal features can be distorted depending on the preceding and following tones (Chao, 1968; Howie, 1976; M. C. Lin, 1988; Lin, 1965; Wu, 1984, 1988). Thus, when a Mandarin monosyllable is produced in isolation, the F0 contours will be well defined and stable, but when produced in a sentence the tones will vary depending on neighboring tones. Therefore, in Mandarin sentences, phonemic information is more important than tonal information but for words, phonemes and tones play an equal role in correct recognition.
As discussed above, in very noisy conditions between -9 dB and -3 dB, word recognition can be better than sentence recognition because words have both phonemes and tones correctly pronounced whereas sentences do not. The contextual benefit usually associated with sentence recognition can only occur if enough phonemic units are correctly recognized. However, in very noisy conditions when it is difficult to recognize enough word units (due to less precise tonal information) word recognition is better than sentence recognition scores. It is only when the noise level lowers and crosses that critical SNR 0 dB barrier, that sentence recognition improves dramatically (Figure 8.1).

9.3 Cross-over frequencies

CFs are the intersection points of plotted speech recognition scores for HP and LP filtered frequency bands at each SNR, and set the frequency point of equal intelligibility for HP and LP filtered materials. Thus, CFs are defined as regions that divide the whole FIF frequency span into two parts of equal importance. The Mandarin CFs for the four speech types, as reported in Table 8.5, indicate that CFs are remarkably similar for words (1807 Hz) and segments (1813 Hz), while sentences (1570 Hz) have a lower CF and tones (743 Hz) have a much lower CF.

The Mandarin CFs of sentences (1570 Hz) is lower in frequency than for words (1807 Hz), an indication that the addition of sentence context and tonal redundancy can shift the CF towards a lower frequency region which is similar to findings in English which compare words and sentences (DePaolis et al., 1996). The very low CF (743 Hz) for Mandarin tones indicates the relative importance of weighting in low frequency regions for tonal languages as compared to nontonal languages. This is consistent with other findings showing that important auditory cues for lexical tone perception are fundamental frequency and harmonics under
1kHz (see Section 3.4.3) (Liu & Samuel, 2004). Similarly, Zhang and McPherson (2008) found that a decrease of amplification in low frequency regions (<250 Hz) reduces the ability of HI participants to discriminate tones. Thus, it is important to include weighting in low frequency regions for tonal language speaking hearing impaired listeners.

When the word CF of Mandarin (1807 Hz) is compared to English word data, it is higher than for English: W-22 test (1314 Hz); NU-6 test (1454 Hz) (Studebaker & Sherbecoe, 1991; Studebaker, et al., 1993). That is, in order for 50% of the speech information to be audible, the cutoff bandwidth needs to be around 1800 Hz. When the sentence CF of Mandarin (1570 Hz) is compared with English: SPIN test (1599 Hz), HINT test (1550 Hz) (DePaolis, et al., 1996; Eisenberg, Dirks, Takayanagi, & Martinez, 1998), the difference is within 50 Hz range, in other words, there is a negligible difference for sentence CFs between English and Mandarin. On the other hand, the sentence CFs for Cantonese is much lower (1075 Hz) (Wong, Ho, et al., 2007) than both English and Mandarin. From this, it can be concluded that for sentences, Cantonese contains more speech information in low frequency regions than English and Mandarin. It would be interesting to see where the Cantonese CF for words is located, but there is only sentence data available for this tonal language. Nonetheless, the greater weighting in the lower frequency region in Cantonese compared to Mandarin may be related to the fact that Cantonese has nine lexical tones, compared four tones in Mandarin Chinese. Thus, low frequency regions carry more weight for Cantonese than Mandarin as it appears to show that the greater the amount of tonal differentiation required, the greater low frequency weighting required.
9.4 Frequency importance functions

The Mandarin FIFs derived in this study indicate that phonemic content, lexical tone and sentence context affect the shapes of the FIFs. In addition, there are differences between the current findings and other studies using English, Cantonese and Mandarin (J. Chen, et al., 2008; DePaolis, et al., 1996; Studebaker & Sherbecoe, 1991; Studebaker, et al., 1993; Wong, Ho, et al., 2007; Yeh, 2005).

The following sections, firstly, examine FIFs by comparing overall shapes and the distribution of importance weightings for the four derived FIFs; and secondly, compare the current findings for word materials to other Mandarin (J. Chen, et al., 2008; Yeh, 2005) and English studies (DePaolis, et al., 1996; Studebaker & Sherbecoe, 1991; Studebaker, et al., 1993); and finally compare the current findings for sentence materials to Cantonese (Wong, Ho, et al., 2007) and English (DePaolis, et al., 1996) studies.

9.4.1 Mandarin FIFs in this study

As shown in Table 8.8, the findings for monosyllabic materials show peak importance weighting occurred in the bandwidth 1413-2239 Hz, centered at ~1800 Hz for words (14.8%) and segments (15.2%), and with neighboring importance weightings that were >10% for frequencies 708-4467 Hz. At the sentence level, the FIF results revealed a narrower peak in importance weighting also occurring at the centre frequency ~1800 Hz (19%) with percentage weightings over the same frequency span that were >11% (see Table 8.8). Essentially, the span is similar to that for word data, but there is a stronger narrower peak for sentences (~19%) around 1800 Hz compared to words and segments (~15%). In addition, the FIF for sentences has at least 10% more
importance weighting across this span (708 Hz - 4467 Hz) than words and segments.

In accord with sentence data from Wong et al. (2007), secondary low-frequency peaks were also found in the current study in the 110-141 Hz bandwidth for sentences accounting for ~5% importance weighting, and for words and segments (~11%). This provides strong support for a secondary peak occurring in tonal language FIFs (Wong et al., 2007). Here the word FIF (~11%) has almost twice the importance weighting of sentences (~5%) at 112 Hz whereas in the case of the primary peak at ~1800 Hz, importance weighting is greater for sentences than words/segments. The relative low weighting centered at 112 Hz of sentences than words could result from the distortion of lexical tone that occurs when words are concatenated in sentences. In other words, when Mandarin produced continuously, the F0 contours of words are distorted in order to achieve fluency (see Section 3.4.4) (Chao, 1968; Howie, 1976; Lin, 1988; Lin, 1965; Wu, 1984, 1988). As for the higher peak at 1800 Hz for sentences than words, it could be a contribution of the richer phonemic information contained in sentences than words.

It could be suggested that both sentence context and the associated tonal effects have shifted the distribution of importance weightings in Mandarin FIFs compared to the English FIFs. To perceive Mandarin monosyllabic words and sentences successfully the listener must perceive phonemic and tonemic information correctly. As reviewed in Section 3.4.4), when Mandarin words are presented in continuous speech, the tonal features can be distorted (Chao, 1968; Howie, 1976; Lin, 1988; Lin, 1965; Wu, 1984, 1988) because the production of tonal information in continuous speech depends to a large degree on neighboring
tones (Speer, et al., 1989). Thus, as seen in our Mandarin FIFs, the importance weighting for words (~11%) centered at 112 Hz is because tonemic information is more readily accessible on words than sentences (~5%) where there is a greater reliance on phonemic information.

The results forlexical tone differed from words and segments primarily in the shape of the importance weightings across frequency bands as they revealed almost equivalent weighting across bandwidths, albeit comprising a curvilinear shape with the greatest percentage weighting on central frequency bands, and lower percentage weightings at the more extreme frequencies. As shown in Figure 8.10, the importance weightings were 10% to 11% between 224 Hz and 3548 Hz, and reduced in a stepwise manner at the extremities. This is contrary to what might be expected for a language containing lexical tone, in which it might be expected that you would find lower frequency bands more heavily weighted than higher frequency bands. Nonetheless, the recognition data for words, segments and sentences shows secondary low-frequency importance weightings centered at ~112 Hz implying that F0 does contribute significant weight for both word and sentence speech understanding.

9.4.2 Word FIFs: comparing Mandarin Chinese and English

In this section, FIFs and cumulative FIFs for words from this study are compared to those from other studies. The other studies are (1) three English studies using the W-22 test (Studebaker & Sherbecoe, 1991), the NU-6 test (Studebaker, et al., 1993) and the PB-50 test (DePaolis, et al., 1996); and (2) two other Mandarin studies using 50 phonetically balanced words (J. Chen, et al., 2008; Yeh, 2005). It should be noted that these two Mandarin studies are unpublished (see Section 5.5.5) and that Chen (2008) is a conference proceedings
that does not contain details of importance weighting for each frequency. Therefore, Chen’s data are not included in Figure 9.1 and 9.2. Figure 9.1 depicts comparative data for word FIFs and Figure 9.2 plots the comparative cumulative FIFs across the five studies.

The two Mandarin monosyllabic word FIFs reported by Yeh (2005) and Chen (2008) are very different from the FIFs derived in this study. First, there are no secondary peaks in the 110-141 Hz bandwidth. Second, both Yeh’s (2005) and Chen’s (2008) word FIFs exhibit a higher importance weighting at ~2500 Hz than the word and segment FIFs in the current study (~1800 Hz). There are methodological differences, however, between these studies and the current one. Firstly, there were only two participants in the Yeh study, and no mention of participant numbers in the Chen study. Secondly, both these studies used 50 monosyllabic words for total of 320 test conditions, compared to 200 words for total of 180 conditions in the current study. Using 50 words for 320 conditions would mean that all words are repeated at least four times, and this could lead to a familiarity or learning effects. When learning is involved in a speech recognition test, the midpoint FIF (50 percentile of the cumulative FIF) as shown in Figure 9.2 tends to be in a higher frequency band than one obtained from a large word pool (Studebaker & Sherbecoe, 1993), and indeed, the peak in both the Chen and Yeh studies is higher than the current study. In Figure 9.2, the 50% correct point of Yeh’s FIF is centered at 1778 Hz, much higher than the current study at 1122 Hz. In addition, the familiarity of test items benefits participants and results in better consonant recognition (Studebaker & Sherbecoe, 1993).

As shown in Figure 9.1 current Mandarin and English FIF data have primary peaks centered around 1800 Hz, while Mandarin data by Yeh et al.(2005)
and Chen et al. (2008) reported a peak around 2500 Hz. For the span between 224 Hz to 1413 Hz, English FIFs derived using the W-22 and NU-6 (Studebaker & Sherbecoe, 1991; Studebaker, et al., 1993) have 13-15 % more importance than our Mandarin word FIF, but the FIF of PB-50 is only ~5% more (DePaolis, et al., 1996). When compared the three English FIFs to current Mandarin, English have about 13 % more importance within the range between 891 Hz and 8913 Hz.

Figure 9.1 also shows that, although there is a consistent pattern whereby the three English word FIFs peak at ~1800 Hz and have greater importance than current Mandarin in the range of 891-8913Hz, there is also considerable variability among the percentage importance weighting centered on 1800 Hz for the English studies. This variability can be due to the use of different filtering bands, different numbers of conditions, different word lists, and/or different methods of smoothing the data. For instance, Depaolis et al. (1996) used only 9 filtering conditions using the PB-50 test compared Studebaker’s use of 20 filtering conditions using the NU-6 test (Studebaker, et al., 1993) and the W-22 test (Studebaker & Sherbecoe, 1991). Furthermore, there is no consensus on a method that should be followed to smooth speech recognition data (see Section 5.4.5). The SRT data smoothing procedure involved in FIF derivation in the above studies with English were smoothed by eye using independent judgments; therefore, here we are focusing on the general pattern rather than the fine details of the differences.
Figure 9.1. Word FIFs for the current study and four other studies: (1) English, W-22 (Studebaker & Sherbecoe, 1991) (2) English, NU-6 (Studebaker, et al., 1993) (3) English, PB50 (DePaolis et al., 1996) 4) Mandarin, 50 words (own lists) (Yeh, 2005).

Figure 9.2. Cumulative word FIF comparing the current study to four similar studies: (1) English, W-22 (Studebaker & Sherbecoe, 1991) (2) English, NU-6 (Studebaker, et al., 1993) (3) English, PB50 (DePaolis et al., 1996) (4) Mandarin, 50 words (own lists) (Yeh, 2005).
9.4.3 Mandarin versus English and Cantonese sentences

For sentences, the Mandarin FIF in this study is compared to studies in English using the SPIN sentence test (DePaolis, et al., 1996) and Cantonese using the CHINT (Wong, Ho, et al., 2007) and these are plotted in Figure 9.3. Both Cantonese and current Mandarin FIFs in Figure 9.3 have used HINT as sentence material. It would be ideal to plot English HINT values for comparison but the only study using the English HINT focused on deriving of a transfer function (Eisenberg, et al., 1998). This data is discussed in section 9.5, but not here.

Figure 9.3 reveals two differences that set tonal language findings apart from the typical English FIF findings. Firstly, both Mandarin and Cantonese show a secondary peak (~5%) in the 110-141 Hz bandwidth (Wong, Ho, et al., 2007), that is not found in English studies. These findings imply that tonal language listeners require greater low frequency weighting to differentiate tonal information. While duration is a strong cue for tone recognition, the other critical cues are F0 contour shape and F0 height which allow the listeners to differentiate between lexical tones (Garding, Kratochvil, Svantesson, & Zhang, 1986; Sagart, 1986 Howie, 1976; Massaro, 1985; Whalen & Xu, 1992; Xu, 1997). Thus, the secondary peak indicates that a secondary weighting is required in low-frequency regions which contain F0 information.

The second difference lies in the shape of the primary peak. Although Mandarin, Cantonese and English sentence FIFs all show a primary peak at ~1800 Hz, the English FIF has a narrower span than the FIF for Mandarin and Cantonese Chinese. Specifically, the peak importance weighting for English sentences at ~29% (DePaolis et al., 1996) is stronger and narrower than the peak at ~1800 Hz for Cantonese (Wong, Ho, et al., 2007) and Mandarin at ~19%. These findings
may be attributed to differences between the phonemic inventories of English and Mandarin. For example, English has more high frequency fricatives than Mandarin.

There are also differences in the cumulative sentence FIFs shown in Figure 9.4. The FIF midpoint (50 percentile of the cumulative FIF) locates at ~1100 Hz for Cantonese HINT sentences (Wong et al., 2007), which is lower than for the current Mandarin study at ~1500 Hz, and also for the English SPIN test (~1700 Hz). (De Paolis et al., 1996). The much lower midpoint location of Cantonese than English and Mandarin could be influenced by differences in its phonemic and tonemic inventories. Research shows that Cantonese has fewer consonants and more vowels than English (So & Dodd, 1995). In addition, Cantonese has more tones (9) (Browning, 1974; Dodd & So, 1994; Fok Chan, 1974) than Mandarin (4). Thus, Cantonese and Mandarin, as dialects of the Chinese, might share the similar phonemes but Cantonese contains more lexical tones than Mandarin (Bauer & Benedict, 1997; Howie, 1976), and both dialects contain more vowels than English. Taken together these language features seem to contribute to the lower FIF midpoint in Cantonese than Mandarin and English.

A comparison of word and sentence FIFs in English (DePaolis et al., 1996) and the current Mandarin study, reveals that English has greater importance weighting at ~1800 Hz than Mandarin (>10% difference for words, >15% difference for sentences) within the range of 708-4467Hz. These differences can be accounted for by phonemic differences between the two languages. When the phonetic features of English are manipulated, a comparison of FIFs across different phonetic features reports that dental, alveolar, and labial consonants produce a CF >1700 Hz (section 5.5.2) (Duggirala, et al., 1988; Miller & Nicely,
1955; Wang, et al., 1978). It is also the case that fronted speech sounds (dental, alveolar and labial) comprises 80% of English consonants compared to 60% of Mandarin consonants (Denes, 1963; Ma & Shen, 2004; Mines, Hanson, & Shoup, 1978). Thus, again, it is shown that the differences in the phonemic inventories of the Mandarin and English may contribute to differences between English and Mandarin FIFs for words and sentences.

Figure 9.3. Comparison of the current study MHINT sentence frequency importance functions to other two similar studies: English-SPIN (DePaolis, et al., 1996) and Cantonese HINT (Wong, Ho, et al., 2007).
9.5 Transfer functions

In this study, TFs for converting the speech recognition scores to the SII values were derived for words, tones, segments and sentences. The results indicate that tones had the steepest slope, followed by sentences with words and segments which both having similar shallow slopes (see Figures 8.12 to 8.15). These findings are consistent with English data which also show contextual cues in sentences produce steeper slopes than for words (Hudgins, et al., 1947; Miller, et al., 1951). Thus, similarly in Mandarin Chinese, sentences with high context, and tones with high redundancy show steep slopes which are linked with higher speech recognition scores than scores for words and segments at the same SII value. For example, the SII value of 0.5 corresponds to scores of ~97% for tones and sentences, and~70% for words and segments.

In Figure 9.5, the TF for words is plotted along with three other reported TFs: (1) English using the W-22 test (Studebaker & Sherbecoe, 1991), (2)
English, using the NU-6 test (Studebaker, et al., 1993), (3) English, using the PB50 test (DePaolis, et al., 1996) and (4) Mandarin words (Yeh, 2005). The calculation of TF slopes is based on observed scores between 20% - 80% and presented as percentage increase per .0333 SII. As shown in Figure 9.5, the TFs for English words (W-22, NU-6 and PB50) are steeper than those for Mandarin words in the current study, and in the Yeh et al. (2005) study. Thus, the slope of the Mandarin TF is approximately half that of the English slope, for example, the SII value of 0.3 corresponds to scores of ~35-40% for Mandarin words, and ~50-60% for English words. Typically, redundant material types yield better speech intelligibility than non-redundant materials for a given SII (Miller, et al., 1951; Wong, Ho, et al., 2007). Here the results indicate Mandarin words are more difficult to recognize than English words because the listeners need to correctly recognize not only individual phonemes but also the associated tone, thus making Mandarin word lists less predictable than English ones.

Figure 9.5. The TF of Mandarin words from MSTMs in this study is shown in comparison with other reported TFs: (1) English words from W-22 (Studebaker & Sherbecoe, 1991) and (2) NU-6 (Studebaker, et al., 1993); and (3) English, using the PB50 test (DePaolis, et al., 1996); and (4) Mandarin words (own lists) (Yeh, 2005)
Figure 9.6. The TF of Mandarin sentences from MHINT in this study is shown in comparison with other reported sentence TFs: (1) English sentences from SPIN (DePaolis, et al., 1996) and (2) HINT (Eisenberg, et al., 1998); and (3) CHINT sentences (Wong, Ho, et al., 2007).

Table 9.1. Comparison of TFs. TF slopes are based on observed scores between 20% - 80%. The TF slope is calculated as percentage (%) increase per .033 SII.

<table>
<thead>
<tr>
<th>Study</th>
<th>Language</th>
<th>Speech Materials</th>
<th>Q</th>
<th>N</th>
<th>TF Slope (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Study</td>
<td>Mandarin</td>
<td>MTSM words</td>
<td>0.52</td>
<td>2.9</td>
<td>4.4</td>
</tr>
<tr>
<td>Yeh et al. (2005)</td>
<td>Mandarin</td>
<td>Words</td>
<td>0.50</td>
<td>3.8</td>
<td>~3.6</td>
</tr>
<tr>
<td>Studebaker et al. (1991)</td>
<td>English</td>
<td>W-22 words</td>
<td>0.28</td>
<td>4.1</td>
<td>10.2</td>
</tr>
<tr>
<td>Studebaker et al. (1993)</td>
<td>English</td>
<td>NU-6 words</td>
<td>0.40</td>
<td>3.3</td>
<td>6.4</td>
</tr>
<tr>
<td>Depaolis et al. (1996)</td>
<td>English</td>
<td>PB50 words</td>
<td>0.64</td>
<td>2.4</td>
<td>~4.7</td>
</tr>
</tbody>
</table>

| Current Study          | Mandarin  | HINT Sentences   | 0.23| 4.1 | 10.9         |
| Wong et al. (2007)     | Cantonese | HINT sentences   | 0.18| 12.6| 11.0         |
| Eisenberg et al. (1998)| English   | HINT Sentence    | 0.24| 15.1| ~10.0        |
| Depaolis et al. (1996) | English   | SPIN sentences   | 0.33| 4.5 | ~8.0         |
In Figure 9.6, the sentence TF for the current study is compared with (1) English sentences (Eisenberg, et al., 1998); (2) English using SPIN sentences (DePaolis, et al., 1996); (3) Cantonese HINT (Wong, Ho, et al., 2007). Table 9.1 presents the TF slope (%)s from these studies, in addition to associated Q and N values. The sentence data shows that the Mandarin TF has the steepest slope; that the Cantonese slope is shallower, and that English slope is the shallowest. This is the reverse of what was found for word TFs where Mandarin displayed the shallowest slope, and English the steepest. Thus, it appears that Mandarin and Cantonese sentences are more intelligible, and by extension contain more redundant cues than English sentences. However, according psychometric functions (Figure 8.1 and Section 9.2) under high noise conditions (SNR -9 to -3 dB), listeners have great difficulty recognizing Mandarin sentences. It can be argued when noise levels are high, listeners struggle to recognize the combination of tone and phoneme cues in sentences, but when the noise level decreases sufficiently, there is a dramatic increase in intelligibility which compensates in the overall assessment for these poor recognition scores at low SNRs.

9.6 Implications for Mandarin hearing aid prescription formulas

The data obtained in this study provide important implications for developing hearing aid prescriptions for tonal language speakers, and specifically Mandarin Chinese. Comparison with English data indicates there are issues with using English data to estimate SII values for Mandarin speaking hearing impaired listeners. Although, clinicians are aware that prescription formulas should provide amplification in the area containing fundamental frequency, until now, this has been done by guess-work. Now, the findings from the current Mandarin study reveal more exact areas that require amplification which, to a large extent,
but certainly not in all aspects, accord with Cantonese Chinese (Wong, Ho, et al., 2007). Furthermore, when tonal words are concatenated as typically happens in day to day conversations, low frequency amplification is more complicated, as the perception of words requires more low frequency amplification than for tonal sentences. This is different to fitting hearing aids for English speakers, because the low frequency region typically contains unwanted noise components, and more relative weight is provided in high frequency regions to aid speech perception.

It is clear that using the TFs derived from Cantonese and English data to predict SII values of Mandarin will not be accurate. For words, the use of English word TFs would under-estimate actual SII values, and for sentences, TFs derived from English or Cantonese would over-estimate actual SII values. These issues reinforce the point that only a TF derived specifically from the listener’s language can provide appropriate and accurate SII predictions. To date, this is the first Mandarin Chinese study to systematically investigate the effects of this language on FIFs, and the first tonal language study to provide individual weighting data for words/segments, tones, and sentences. Furthermore, deriving FIFs in multiple contextual formats provides good validity to the TFs for estimating accurate SII value when evaluating the communication efficacy.

9.7 Future research directions

The FIFs obtained in this study have several important implications for how hearing aids should be adjusted for Mandarin speakers. However, this is only the first step in understanding how hearing aids can be more beneficially prescribed for Mandarin speakers. In order to modify the current English based prescription for lexical tonal language users, further research needs to be done. In
the next step, the aim would be to compare English FIF of the NAL-NL2 to and a
tonal version of NAL-NL2 with Mandarin speakers. Mandarin speakers with
different degrees and types of hearing loss should be used in order to examine the
interaction between different types of hearing loss, whether Mandarin speakers do
benefit from the modified amplification formula. Specifically native Mandarin
speakers would participate in (1) an intelligibility preference test using continuous
discourse recorded by a native Mandarin speaker; and (2) an objective sentence
recognition test using the MHINT. This would provide information on the
efficacy of tonal versus non-tonal prescriptions for Mandarin speakers, but also
which of the two prescriptions is preferred by Mandarin speakers, and whether
these preferences and judgments are related to the configuration of hearing loss,
recognition scores, etc. It would also be of interest to develop FIF for other tonal
languages, such as Thai and Vietnamese in order to examine differences in the
patterns of findings between tonal languages.

9.8 Summary and conclusions

Results of this project suggest that language specific inventories of
phonemes and tones, in addition to context and redundancy effects, influence
psychometric functions, CFs, FIFs and TF slopes in Mandarin. Psychometric
functions show these effects under various noise conditions, with tones the easiest
to identify, and containing the most redundancy of all the speech material types.
Moreover, the context/redundancy benefit for Mandarin word and sentence
recognition is consistent with English data (DePaolis, et al., 1996; Miller, et al.,
1951), except for Mandarin sentence recognition scores between -9 and -3 dB
SNR, where the performance slope remains relatively flat, and then rises steeply
at 0 dB to 15 dB. The Mandarin CFs are remarkably similar for words (1807 Hz)
and segments (1813 Hz), but lower for sentences (1570 Hz) and tones (743 Hz), and are in general agreement with English studies that include sentences and word materials (DePaolis et al., 1996). The current data also highlight how sentence context and tonal redundancy can shift the CF towards a lower frequency region, and the importance of auditory cues under 1 KHz (S. Liu & Samuel, 2004).

The very steep slope for tones in the psychometric functions shows their resistance to noise. More importantly, the very low CF (743 Hz) for Mandarin tones reveals that important auditory cues for tone perception are fundamental frequency and harmonics at less than 1 kHz, and accords with the secondary peaks found for word and sentence FIFs, which also indicate the role of importance weighting within the low frequency region. The psychometric function for words was also steeper than that for sentences in very noisy conditions because words have tones correctly pronounced whereas sentences may not. Taken together, the findings reveal the influence of Mandarin tones in speech recognition tasks, and that tonal language words contain less predictability than non-tonal ones.

The FIF data showed a primary peak centered at ~1800 Hz and secondary peak at ~112 Hz for word and sentence materials in accord with Cantonese sentence FIFs (Wong et al., 2007). The secondary peak carries more weighting for words/segments (11%) than sentences (5%) but generally confirms that low frequencies are more important in tonal than non-tonal languages for speech intelligibility. The Mandarin TFs concur, and indicate that tones, with high redundancy had the steepest slope, followed by sentences, then by words and segments, which both having similar shallow slopes. These findings are consistent with English materials where redundancy and contextual cues in sentences produce steeper slopes than for words (Hudgins, et al., 1947; Miller, et al., 1951).
However, it should also be noted that the word data for non-tonal English has a steeper slope than that for tonal Mandarin and Cantonese Chinese; whereas, for sentences, it is reversed, Mandarin and Cantonese have steeper TF slopes than those for English.

Typically, redundant material yields steeper TF slopes than non-redundant material (Miller, et al., 1951; Wong, Ho, et al., 2007). Thus, it is argued that for word recognition tests in tonal languages, not only do the phonemes need to be correctly recognized, but also the associated lexical tone. Thus, tonal language words contain less predictability than non-tonal ones. Nonetheless, for sentences, in which tones are less fully realized than in words, phonemes play a central role, and phoneme recognition is easier due to the highly redundant nature of the structure of Mandarin Chinese. Nonetheless, it is clear from the TF data that English word TFs under-estimate SII values for Mandarin speakers, and that TFs derived from English or Cantonese sentences would over-estimate SII values for Mandarin speakers. The results of this investigation suggest that language specific derivations of CFs, FIFs and TFs are needed for appropriate amplification of hearing aids for Mandarin speakers, and that generalization from other tonal languages such as Cantonese Chinese may not be appropriate.

The FIF comparison between Mandarin, Cantonese and English indicates that due to language specific phoneme and toneme inventories, certain frequency bands are more important to intelligibility in one or another language. Furthermore, word and sentence materials shifts FIF intelligibility weight differently between any two languages. Therefore, it is clear that an appropriate amplification formula for Mandarin speakers will be different than for English or even Cantonese speakers.


Palva. (1965). Filtered speech audiometry: I. Basic studies with finnish speech towards the creation of a method for the diagnosis of central hearing disorders. acta oto-laryngologica. supplement(210), 7-86.


Scollie, S., Ching, T., Seewald, R., Dillon, H., Britton, L., Steinberg, J., et al. (2010). Children's speech perception and loudness ratings when fitted with hearing aids using the DSL v.4.1 and the NAL-NL1 prescriptions. International Journal of Audiology, 49(s1), S26-s34.


Appendices

Appendix A  Advertisement for Participants - English Version
Advertisement for Participants - Mandarin Version
Appendix B  Participant Information and Consent Form
Appendix C  Participant Language Information
Appendix D  Audiogram
Appendix E  Audiometer Calibration Certificate
Appendix F  Language Background of Participants
Appendix G  SPSS Output Tables for Reliability
Appendix H  NAL Smoothed Algorithm - Intelligibility of Filtered Mandarin
Appendix I  Comparison of Unsmoothed and Smoothed Crossover Frequencies
Appendix J  Lists of Words and Sentences
Appendix K  A Sample DMDX Script
Appendix L  CD Extras

1. Word audio files used for testing
2. Sentence audio files used for testing
Appendix A
Advertisement for Participants - English Version

Mandarin Chinese Speakers Wanted

UWS Ethics Approval #H8609

We are looking for adults to participate in a study examining how to improve hearing aids for Mandarin Chinese people.

REQUIREMENTS

1. Aged between 18-45 years
2. Have no history of hearing disorder or brain injury
3. Speak Mandarin as a native language

The study takes about two hours.

You will be paid a travel reimbursement of $60

If you are interested in participate this study, please contact:

Ming-Wen Kuo on (02) 9772 6573 or email: m.kuo@uws.edu.au
徵求“說國語”受試者

我們正在尋找說國語成年人參與如何改善助聽器研究。

要求條件

1. 年齡介於 18-45 歲
2. 沒有聽力障礙或腦損傷的歷史
3. 日常生活以國語為主要溝通語言

這項研究需時約兩個小時。

您將獲得 1200 元 (約 60 元澳幣)的交通補助費。

如果你有興趣參與這項研究，請聯繫:
郭小姐 (02) 9772 6573 or email: m.kuo@uws.edu.au
Appendix B

Participant Information and Consent form

MARCS Auditory Laboratories

UWS Ethics Approval No. H8609

DERIVING THE FREQUENCY IMPORTANCE FUNCTION OF MANDARIN CHINESE

PARTICIPANT INFORMATION SHEET

Who is carrying out the study?

You are invited to participate in a study conducted by Ming-Wen Kuo, a PhD Candidate at MARCS Auditory Laboratories at the University Of Western Sydney. The study will form the basis for the PhD degree at the University Of Western Sydney under the supervision of Dr. Christine Kitamura.

What is the study about?

The purpose of the study is to investigate the frequency importance function for Mandarin Chinese. A frequency importance function is used for setting hearing aids and these setting depend on the nature of the language, e.g. tonal or non-tonal.

What does the study involve?

You will be asked to repeat sentences and words back after you listen to them. Some may be harder to hear than others because they are mixed with noise.

How much time will the study take?

It will take approximately 2 hours.

Will the study benefit me?

The study will not benefit you directly but will help with deriving the most suitable amplification prescription for use in hearing aids for Mandarin Chinese, and other tonal language speakers.

Will the study involve any discomfort for me?
The study involves no discomfort but should you feel discomfort in any way, please let the researcher know, and you can take a break, or withdraw from participating.

**How is this study being paid for?**

The study is being sponsored by the University Of Western Sydney.

**Will anyone else know the results? How will the results be disseminated?**

All aspects of the study, including results, will be confidential, and only the researchers will have access to information on participants. The group results will be published in journal articles, and a PhD thesis. No individual results will be reported.

**Can I withdraw from the study?**

Participation is entirely voluntary. You are not obliged to be involved and – if you do participate – you can withdraw at any time without giving any reason and without any consequences.

**Can I tell other people about the study?**

Yes, you can tell other people about the study by providing them with the chief investigator’s contact details. They can contact the chief investigator to discuss their participation in the research project and obtain an information sheet.

**What if I require further information?**

When you have read this information, Ming-Wen will discuss it with you further and answer any questions you may have. If you would like to know more at any stage, please feel free to contact Ming-Wen at (02) 9772 6573.

**What if I have a complaint?**

This study has been approved by the University Of Western Sydney Human Research Ethics Committee. The Approval number is [enter approval number]. If you have any complaints or reservations about the ethical conduct of this research, you may contact the Ethics Committee through the Office of Research Services on (02) 4736 0083, or by email at humanethics@uws.edu.au. Any issues you raise will be treated in confidence and investigated fully, and you will be informed of the outcome.

If you agree to participate in this study, you will be asked to sign a consent form.
DERIVING THE FREQUENCY IMPORTANCE FUNCTION OF MANDARIN CHINESE

CONSENT FORM

I have been asked to participate in the research DERIVING THE FREQUENCY IMPORTANCE FUNCTION OF MANDARIN CHINESE conducted by Ming-Wen Kuo and give my free consent by signing this form. I acknowledge that:

1. The research project will be carried out as described in the information sheet, a copy of which I have retained. I have read and understood the Information Sheet and have had the opportunity to discuss the information and my involvement in the project with the researcher(s) to my satisfaction.

2. My involvement is confidential and the information gained during the study may be published, but no personal information will be used in any way that reveals my identity.

3. My consent to participate is voluntary. I may withdraw from the study at any time, without affecting my relationship with UWS. I do not have to give a reason for the withdrawal of my consent.

Signed: ………………………………………………………………………

Name: ………………………………………………………………………

Date: ………………………………………………………………………
Appendix C

Participant Language Information

Name: ____________________________ (Mandarin Name) _________________________________ (English Name)

Contact Phone Number (optional):

Email Address:

Mother Tongue:

Dialect (Province):

How long have you been in Australia?

How many hours do you speak Mandarin weekly (for the past 3 months)?

Music Training Background (years)

Occupation

Second Visit

Date:

Do you notice any hearing sensitivity changes since your last visit as the date provided above?

Have you been exposed to loud noises for the past three days?
# Appendix D

Audiogram

## Audiometry Report

<table>
<thead>
<tr>
<th>1. Name</th>
<th>2. CAA Client No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. Postal Address</th>
<th>4. Date of Birth</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. Certificate(s) applied for</th>
<th>6. Applicant’s Signature: To be signed in front of examiner.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 1 □ Class 2 □ Class 3 □</td>
<td>Date: / /</td>
</tr>
</tbody>
</table>

### 7. Pure Tone Audiometry (all applicants)

#### Right Ear

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Intensity (dBnHL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>10</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>10</td>
</tr>
<tr>
<td>1500</td>
<td>10</td>
</tr>
<tr>
<td>2000</td>
<td>10</td>
</tr>
<tr>
<td>3000</td>
<td>10</td>
</tr>
<tr>
<td>4000</td>
<td>10</td>
</tr>
<tr>
<td>6000</td>
<td>10</td>
</tr>
<tr>
<td>8000</td>
<td>10</td>
</tr>
</tbody>
</table>

#### Left Ear

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Intensity (dBnHL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>10</td>
</tr>
<tr>
<td>500</td>
<td>10</td>
</tr>
<tr>
<td>1000</td>
<td>10</td>
</tr>
<tr>
<td>1500</td>
<td>10</td>
</tr>
<tr>
<td>2000</td>
<td>10</td>
</tr>
<tr>
<td>3000</td>
<td>10</td>
</tr>
<tr>
<td>4000</td>
<td>10</td>
</tr>
<tr>
<td>6000</td>
<td>10</td>
</tr>
<tr>
<td>8000</td>
<td>10</td>
</tr>
</tbody>
</table>

It is mandatory to record at 500, 1000, 2000 and 3000 Hz. Other frequencies up to 8000 Hz are desirable.

**SYMBOLS**

- **Right**
  - ○ Air
  - ● Air Masked
  - < Bone
  - ▲ Bone Masked

- **Left**
  - ○ Speech
  - ● Speech Masked

**Audiometer:**

**Calibration Date:**

### 8. Speech Audiometry (as indicated)

#### Right Ear

<table>
<thead>
<tr>
<th>Percentage</th>
<th>dBnHL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

#### Left Ear

<table>
<thead>
<tr>
<th>Percentage</th>
<th>dBnHL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

**Earphones:**

- Insert 3A
- TDH Headsets

**SYMBOLS**

- **Right**
  - ○ Speech
  - ● Speech Masked

- **Left**
  - ○ Speech
  - ● Speech Masked

### 9. Impittance Audiometry (as indicated)

<table>
<thead>
<tr>
<th>Right</th>
<th>Left</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td></td>
</tr>
<tr>
<td>MEP</td>
<td></td>
</tr>
<tr>
<td>Immit</td>
<td></td>
</tr>
<tr>
<td>Vol</td>
<td></td>
</tr>
</tbody>
</table>

**Contralateral Acoustic Reflex**

**Ipsilateral Acoustic Reflex**

**Normal**

**Elevated**

**Absent**

### 10. Diagnosis/Comments

### 11. Print Examiner’s Name and Address

(Practise Stamp Preferred)

### 12. Client’s ID:

Indicate the type of photographic ID sighted, serial number and expiry date.

- Client’s photographic ID sighted at the medical examination.

### 13. Examiner’s Declaration:

I hereby certify that I personally identified and examined the applicant named on this medical report and that this report, with any attached notes, embodies my examination completely and correctly.

**Telephone Number:**

**Facsimile Number:**

**Examiner signature:**

**Date:** / /
### Audiometer Calibration Certificate

**Date:** 7/10/2013

**Model:** 8208

**Serial No.:** 102400

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Right</th>
<th>Left</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>61</td>
<td>57</td>
</tr>
<tr>
<td>2000</td>
<td>63</td>
<td>61</td>
</tr>
<tr>
<td>3000</td>
<td>65</td>
<td>63</td>
</tr>
<tr>
<td>4000</td>
<td>67</td>
<td>65</td>
</tr>
<tr>
<td>5000</td>
<td>69</td>
<td>67</td>
</tr>
<tr>
<td>6000</td>
<td>71</td>
<td>69</td>
</tr>
<tr>
<td>7000</td>
<td>73</td>
<td>71</td>
</tr>
<tr>
<td>8000</td>
<td>75</td>
<td>73</td>
</tr>
<tr>
<td>9000</td>
<td>77</td>
<td>75</td>
</tr>
<tr>
<td>10000</td>
<td>79</td>
<td>77</td>
</tr>
</tbody>
</table>

**Coverage:** 6.7 to 6.8 kHz

**Test Equipment:** PAXTON - BARROW ELECTROMEDICS

**Manufacturer:** PBE

**Quality Control:**

---

**Appendix E**

**Test Procedure 1000 Hz**

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Right</th>
<th>Left</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>61</td>
<td>57</td>
</tr>
<tr>
<td>2000</td>
<td>63</td>
<td>61</td>
</tr>
<tr>
<td>3000</td>
<td>65</td>
<td>63</td>
</tr>
<tr>
<td>4000</td>
<td>67</td>
<td>65</td>
</tr>
<tr>
<td>5000</td>
<td>69</td>
<td>67</td>
</tr>
<tr>
<td>6000</td>
<td>71</td>
<td>69</td>
</tr>
<tr>
<td>7000</td>
<td>73</td>
<td>71</td>
</tr>
<tr>
<td>8000</td>
<td>75</td>
<td>73</td>
</tr>
<tr>
<td>9000</td>
<td>77</td>
<td>75</td>
</tr>
<tr>
<td>10000</td>
<td>79</td>
<td>77</td>
</tr>
</tbody>
</table>

**Coverage:** 6.7 to 6.8 kHz

**Test Equipment:** PAXTON - BARROW ELECTROMEDICS

**Manufacturer:** PBE

**Quality Control:**

---

**Note:** This document includes a table and a diagram related to audiometer calibration. The table provides test results for frequencies from 1000 to 10000 Hz, with measurements for right and left ears. The diagram illustrates the frequency coverage and test equipment used.
Appendix F

Language Background of Participants

Total fifty (25 males and 25 females) participated in the experiment. All participants participated in the monosyllabic word recognition test; and twenty of these participants also completed the sentence recognition test in a separate session. Participants originated from different Chinese provinces and ensured some variety of accent backgrounds in our recognition test. Table F.1 lists the number of participants each of the 22 provinces (at the highest level), four municipalities, five autonomous regions, and two special administrative regions (e.g., Taiwan and Hong Kong) that divide China geographically and administratively. All of the participants had arrived in Australia from China within the five previous years (Figure F.1), and for the largest cohort (54%) the time spent speaking Mandarin was between 4 to 7 hours daily (Figure F.2). All participants were undergraduate or graduate students at Sydney, Australia. None of the listeners had formal musical education of more than 5 years and no listeners had received any musical training in the preceding 3 years.

![Residential Period in Australia](image1)

![Mandarin Speaking Hours Daily](image2)

Figure F.1. Residential period in Australia.  
Figure F.2. Average hours speaking Mandarin
### Table F.1. Participant numbers for each of the Chinese provinces, regions, etc.

<table>
<thead>
<tr>
<th>Number of Participants</th>
<th>English Province Name</th>
<th>Chinese Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Anhui Province</td>
<td>安徽省</td>
</tr>
<tr>
<td>2</td>
<td>Beijing Municipality</td>
<td>北京市</td>
</tr>
<tr>
<td>2</td>
<td>Chongqing Municipality</td>
<td>重庆市</td>
</tr>
<tr>
<td>2</td>
<td>Fujian Province</td>
<td>福建省</td>
</tr>
<tr>
<td>0</td>
<td>Gansu Province</td>
<td>甘肃省</td>
</tr>
<tr>
<td>2</td>
<td>Guangdong Province</td>
<td>广东省</td>
</tr>
<tr>
<td>1</td>
<td>Guangxi Zhuang Autonomous Region</td>
<td>广西壮族自治区</td>
</tr>
<tr>
<td>1</td>
<td>Guizhou Province</td>
<td>贵州省</td>
</tr>
<tr>
<td>0</td>
<td>Hainan Province</td>
<td>海南省</td>
</tr>
<tr>
<td>5</td>
<td>Hebei Province</td>
<td>河北</td>
</tr>
<tr>
<td>1</td>
<td>Heilongjiang Province</td>
<td>黑龙江省</td>
</tr>
<tr>
<td>3</td>
<td>Henan Province</td>
<td>河南省</td>
</tr>
<tr>
<td>0</td>
<td>Hong Kong Special Admin Region</td>
<td>香港特别行政区</td>
</tr>
<tr>
<td>3</td>
<td>Hubei Province</td>
<td>湖北省</td>
</tr>
<tr>
<td>3</td>
<td>Hunan Province</td>
<td>湖南省</td>
</tr>
<tr>
<td>0</td>
<td>Inner Mongolia Autonomous Region</td>
<td>内蒙古自治区</td>
</tr>
<tr>
<td>2</td>
<td>Jiangsu Province</td>
<td>江苏省</td>
</tr>
<tr>
<td>2</td>
<td>Jiangxi Province</td>
<td>江西省</td>
</tr>
<tr>
<td>1</td>
<td>Jilin Province</td>
<td>吉林省</td>
</tr>
<tr>
<td>2</td>
<td>Liaoning Province</td>
<td>辽宁省</td>
</tr>
<tr>
<td>0</td>
<td>Macau Special Administrative Region</td>
<td>澳门特别行政区</td>
</tr>
<tr>
<td>0</td>
<td>Ningxia Hui Autonomous Region</td>
<td>宁夏回族自治区</td>
</tr>
<tr>
<td>0</td>
<td>Qinghai Province</td>
<td>青海省</td>
</tr>
<tr>
<td>2</td>
<td>Shaanxi Province</td>
<td>陕西省</td>
</tr>
<tr>
<td>1</td>
<td>Shandong Province</td>
<td>山东省</td>
</tr>
<tr>
<td>3</td>
<td>Shanghai Municipality</td>
<td>上海市</td>
</tr>
<tr>
<td>2</td>
<td>Shanxi Province</td>
<td>山西省</td>
</tr>
<tr>
<td>3</td>
<td>Sichuan Province</td>
<td>四川省</td>
</tr>
<tr>
<td>0</td>
<td>Taiwan Province †</td>
<td>台湾省</td>
</tr>
<tr>
<td>2</td>
<td>Tianjin Municipality</td>
<td>天津市</td>
</tr>
<tr>
<td>0</td>
<td>Tibet Autonomous Region</td>
<td>西藏自治区</td>
</tr>
<tr>
<td>1</td>
<td>Xinjiang Uyghur Autonomous Region</td>
<td>新疆维吾尔自治区</td>
</tr>
<tr>
<td>2</td>
<td>Yunnan Province</td>
<td>云南省</td>
</tr>
<tr>
<td>1</td>
<td>Zhejiang Province</td>
<td>浙江省</td>
</tr>
</tbody>
</table>
### Appendix G

**SPSS Output Tables for Reliability**

Table G.1. Reliability coefficients and descriptive statistics for two judges scoring words/segments as correct or incorrect

#### Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td>.28</td>
<td>.449</td>
<td>1310</td>
</tr>
<tr>
<td>Examiner 2</td>
<td>.28</td>
<td>.448</td>
<td>1310</td>
</tr>
</tbody>
</table>

#### Correlations

<table>
<thead>
<tr>
<th></th>
<th>Wei</th>
<th>Ming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td>1</td>
<td>.994**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>1310</td>
</tr>
<tr>
<td>Examiner 2</td>
<td>.994**</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>1310</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

Table G.2. Reliability coefficients and descriptive statistics for two judges scoring the numbers of words correct in sentences.

#### Descriptive Statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td>3.47</td>
<td>4.401</td>
<td>800</td>
</tr>
<tr>
<td>Examiner 2</td>
<td>3.48</td>
<td>4.307</td>
<td>800</td>
</tr>
</tbody>
</table>

#### Correlations

<table>
<thead>
<tr>
<th></th>
<th>Ming</th>
<th>Wei</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examiner 1</td>
<td>1</td>
<td>.981**</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>800</td>
</tr>
<tr>
<td>Examiner 2</td>
<td>.981**</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>800</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
Appendix H

NAL Smoothed Algorithm-Intelligibility of Filtered Mandarin

By Mark Seeto

2 May 2012

1. Introduction

The four scores of interest were segment, tone, word, and sentence intelligibility scores, expressed as percentages. There were two sets of data. The first set, containing the segment, tone and word scores, had 180 data points: 2 filter types (LP, HP) × 9 SNRs × 10 cut-off frequencies. In the first data set, each data point came from the same 50 subjects. The second data set, containing the sentence scores, had 140 data points: 2 filter types (LP, HP) × 7 SNRs × 10 cut-off frequencies. In the second data set, each data point came from the same 20 subjects.

The aim was to find, for each of the eight combinations of score type and filter type, a function of SNR and cut-off frequency which fits the data reasonably well, is a monotonic function of SNR for each cut-off frequency, and is a monotonic function of cut-off frequency for each SNR.

2. General approach

Each of the eight combinations of score (segment, tone, word, sentence) and filter type (LP, HP) is considered separately. To work with frequency on a logarithmic scale, define

\[ x = \log_2 \left( \frac{\text{cut-off frequency}}{250 \, \text{Hz}} \right) \]

For each of the eight combinations of score and filter type, these steps are followed:

1. For each SNR, fit a logistic function

\[ \text{Intelligibility}(x) = a \left(1 - \frac{1}{1 + e^{-b(x-c)}}\right) \]

2. Use the estimates of \( a \) from Step 1 (possibly with modifications) to fit \( a \) as a function of SNR.
3. Use the estimates of \( c \) from Step 1 (possibly with modifications) to fit \( c \) as a function of SNR.
4. Substitute the functions obtained from Steps 2 and 3 into

\[ \text{Intelligibility}(x, \text{SNR}) = a(\text{SNR}) \left(1 - \frac{1}{1 + e^{-b(x-c(\text{SNR}))}}\right) \] (1)

and then fit this function on the full data set (for the current combination of score and filter type), that is, use the full data set to estimate \( b \).

The functions for \( a \) in Step 2 were logistic functions with upper asymptote 100 and lower asymptote 0. I chose to assume that \( c \) varies monotonically with SNR, although I don’t have any real justification for this. For LP filtered
speech, I assumed that $c$ decreases asymptotically with increasing SNR. On the other hand, for decreasing SNR, it wasn’t clear to me whether $c$ should increase to an asymptote or increase unboundedly. However, because $a$ decreases to 0 as SNR decreases, the limiting behavior of $c$ as SNR decreases probably doesn’t matter much for practical purposes. Similarly, for HP filtered speech, I assumed that $c$ increases asymptotically with increasing SNR but I didn’t know whether $c$ should decrease to an asymptote or decrease unboundedly with decreasing SNR. I used the unbounded option, with the function being

$$c(\text{SNR}) = k_1 e^{-k_2 \text{SNR}} + k_3.$$  

The procedures used for Steps 2 and 3 were not exactly the same for each combination of score and filter type, and will be described below.

(1) **Segment, LP**

The following plot shows the data for the segment score with low-pass filtering.

The following plots show the estimates and fitted curves for $a$ and $c$. The estimate of $a$ for SNR = −3 seemed wildly incorrect, so it was not used in the fitting. The circles represent the estimated $a$ and $c$ values which were used in the fitting, while the crosses represent the estimated $a$ and $c$ values which were not used. The squares represent modified values which were used in the fitting. These modified values were ones I chose by looking at plots of the original data. I modified the $c$ value for SNR = −9, −6, −3 because the estimates seemed incorrect. I modified the $c$ value for SNR = 15 because the corresponding $a$ estimate was greater than 100.
Upper asymptote for segment, low pass

Location of inflection for segment, low pass
The following plot shows the fitted curves.

The following plot shows the fitted curves together with the data.
The estimate for $\text{SNR} = -6$ was left out. The $c$ estimate for $\text{SNR} = -6$ was modified.
For a, I modified the estimate for SNR = −9 and left out the estimate for SNR = −3. I used a modified estimate for SNR = −9 because with no data at lower SNRs, it seemed beneficial to provide guidance for the curve, whereas at SNR = −3 this seemed unnecessary. For c, I modified the estimates for SNR = −9, −6, −3, 15. The modification for SNR = 15 was made because the $a$ estimate for SNR = 15 was greater than 100.
The $a$ estimate for SNR = −6 was left out. The $c$ estimate for SNR = −6 was modified.
(5) Tone, LP

For both $a$ and $c$, I modified the estimate for $\text{SNR} = -9$. For $c$, I fixed the asymptote by using $c(\text{SNR}) = k_1 e^{-k_2 \text{SNR}} - 8$. Having the asymptote as a variable gave a good fit to the data, but the estimated value of $-26$ seemed unrealistically low, corresponding to a cut-off frequency of $10^{-6}$ Hz. On reflection, the value of $-8$ also seems too low, corresponding to a cut-off frequency of about 1 Hz.
I found this data set more difficult to fit than the other data sets. The $c$ parameter estimates varied “very” non-monotonically with SNR, but I don’t know whether this reflected the true relationship or if it was just a result of random error. I tried fitting a monotonic curve for $c$ as well as a non-monotonic (quadratic) curve, but neither gave a good result. I also tried fitting the slope at the point of inflection as a function of SNR, and using this with $c$ assumed to be constant across SNR, and with $c$ as a function of SNR, but this did not produce good results either.

I then swapped the roles of SNR and cut-off frequency. I fitted a logistic function for each cut-off frequency and then fitted the parameter estimates as
functions of cut-off frequency. I first tried letting $b$ be constant across cut-off frequency, but this seemed inadequate, so I ended up fitting functions for $a$, $b$ and $c$. The final function was

$$\text{Intelligibility}(x, \text{SNR}) = a(x) \left( 1 - \frac{1}{1 + e^{-b(x)(\text{SNR} - c(x))}} \right),$$

with $a$ being a logistic function with asymptotes 0 and 100, $b$ being a logistic function with upper asymptote 0 and free lower asymptote, and $c$ being an exponential function as before.

The estimates of $a$ and $c$ were modified for the two highest cut-off frequencies. The fitted function for $b$ gave a final function which was not monotonic, so I modified one of the parameters for the $b$ function, reducing the slope slightly.

The final function appears to satisfy the monotonicity conditions in the range of the data, but not outside that range.
I modified the \( a \) estimates for SNRs of \(-6\) and \(0\). For \(c\), I did not modify any estimates, but I weighted the points differently, according to how reliable the estimates seemed to be. I later changed the weights to try to improve the fit.
I modified the $a$ estimate for SNR = $-9$. I used different weightings for the $c$ estimates.
4. Summary of fitted functions

For all combinations of score and filter type except for (tone, high pass), the fitted intelligibility function of SNR and cut-off frequency \( f \) (in Hz) is

\[
\text{Intelligibility} = a(\text{SNR}) \left( 1 - \frac{1}{1 + \exp(-b|x - c(\text{SNR})|)} \right),
\]

in which

\[
x = \log_2(f/250)
\]

\[
a(\text{SNR}) = 100 \left( 1 - \frac{1}{1 + \exp(-k_{a1}(\text{SNR} - k_{a2}))} \right)
\]

\[
c(\text{SNR}) = k_{c1} \exp(-k_{c2}\text{SNR}) + k_{c3}.
\]

The values of the parameters are given in the table below.

| Parameter | Segment |  |  | Syllable |  |  | Tone |  | Sentence |  |  |
|-----------|---------|-----------|-----------|---------|-----------|-----------|---------|-----------|-----------|-----------|
|           | LP      | HP        | LP        | HP      | LP        | LP        | HP      | LP        | LP        | HP        |
| \(k_{a1}\) | -0.265  | -0.215    | -0.301    | -0.222  | -0.502    | -0.865    | -0.548  |           |           |           |
| \(k_{a2}\) | 1.344   | 3.946     | 1.772     | 4.057   | -4.073    | -0.880    | -0.454  |           |           |           |
| \(k_{c1}\) | 0.995   | -0.079    | 5.479     | -0.081  | 8.764     | 7.873     | -17.072 |           |           |           |
| \(k_{c2}\) | 0.049   | 0.327     | 0.005     | 0.386   | 0.032     | 0.030     | 0.005   |           |           |           |
| \(k_{c3}\) | 2.421   | 2.886     | -2.296    | 2.752   | -8.000    | -4.717    | 19.531  |           |           |           |
| \(b\)     | -1.220  | 1.464     | -1.196    | 1.428   | -0.868    | -1.909    | 2.088   |           |           |           |
For the tone score with high-pass filtering, the fitted intelligibility function of SNR and cut-off frequency $f$ (in Hz) is

$$\text{Intelligibility} = a(x) \left( 1 - \frac{1}{1 + \exp(-b(x)\text{SNR} - c(x))} \right),$$

in which

$$x = \log_2(f/250)$$

$$a(x) = 100 \left( 1 - \frac{1}{1 + \exp(-k_{a1}(x - k_{a2}))} \right)$$

$$c(x) = k_{c1}\exp(-k_{c2}x) + k_{c3}$$

$$b(x) = \frac{k_{b3}}{1 + \exp(-k_{b1}(x - k_{b2}))}.$$

The values of the parameters are given in the table below.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$k_{a1}$</td>
<td>1.417</td>
</tr>
<tr>
<td>$k_{a2}$</td>
<td>4.482</td>
</tr>
<tr>
<td>$k_{c1}$</td>
<td>2.618</td>
</tr>
<tr>
<td>$k_{c2}$</td>
<td>-0.354</td>
</tr>
<tr>
<td>$k_{c3}$</td>
<td>-5.871</td>
</tr>
<tr>
<td>$k_{b1}$</td>
<td>-0.500</td>
</tr>
<tr>
<td>$k_{b2}$</td>
<td>2.959</td>
</tr>
<tr>
<td>$k_{b3}$</td>
<td>-0.420</td>
</tr>
</tbody>
</table>
Appendix I

Comparison of Unsmoothed and Smoothed Crossover Frequencies

Based on the requirements of ANSI standard, the way to examine whether the SRT smoothing operation has affected the accuracy of outcome FIF is by examining the crossover frequencies pre- and post- smoothing. The crossover frequencies for both unsmoothed and smoothed data at each SNR are shown in Table I.1. As can be seen, the differences between unsmoothed and smoothed data are relatively small (< 100 Hz). This suggests that the smoothing operation did not bias FIF outcomes.

Table I.1. Both unsmoothed and smoothed crossover frequencies at each SNR for the four speech materials.

<table>
<thead>
<tr>
<th>SNR</th>
<th>Unsmoothed</th>
<th>Smoothed</th>
<th>Geometric Avg.</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tone</td>
<td>Segment</td>
<td>Word</td>
<td>Sentence</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>562</td>
<td>2033</td>
<td>2033</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>600</td>
<td>1800</td>
<td>1826</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>700</td>
<td>2000</td>
<td>1826</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>726</td>
<td>1826</td>
<td>1826</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>700</td>
<td>1826</td>
<td>1826</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>562</td>
<td>1619</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>-3</td>
<td>891</td>
<td>2000</td>
<td>1826</td>
</tr>
<tr>
<td></td>
<td>-6</td>
<td>1283</td>
<td>1826</td>
<td>1826</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>726</td>
<td>1826</td>
<td>1826</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>700</td>
<td>1826</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>726</td>
<td>1800</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>750</td>
<td>1800</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>700</td>
<td>1800</td>
<td>1810</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>726</td>
<td>1826</td>
<td>1826</td>
</tr>
<tr>
<td></td>
<td>-3</td>
<td>891</td>
<td>1826</td>
<td>1800</td>
</tr>
<tr>
<td></td>
<td>-6</td>
<td>1853</td>
<td>1800</td>
<td>1800</td>
</tr>
<tr>
<td>Geometric Avg.</td>
<td>669.14</td>
<td>1861.67</td>
<td>1847.36</td>
<td>1597.94</td>
</tr>
<tr>
<td>SD</td>
<td>116.64</td>
<td>138.34</td>
<td>75.05</td>
<td>294.64</td>
</tr>
<tr>
<td></td>
<td>743.25</td>
<td>1812.95</td>
<td>1807.72</td>
<td>1570.52</td>
</tr>
<tr>
<td></td>
<td>66.42</td>
<td>13.90</td>
<td>11.78</td>
<td>377.58</td>
</tr>
</tbody>
</table>
## Appendix J

Lists of Words and Sentences

Table J.1. The four of seven lists from MSTMs which have the better reliability and validity are used in this project. The content of each list are presented in the following four tables with the Mandarin character, and its’ pinyin transcription.

### Word list 1

<table>
<thead>
<tr>
<th>Word list 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>念 niàn</td>
</tr>
<tr>
<td>透 tòu</td>
</tr>
<tr>
<td>哀 āi</td>
</tr>
<tr>
<td>贺 hè</td>
</tr>
<tr>
<td>星 xīng</td>
</tr>
<tr>
<td>毫 háo</td>
</tr>
<tr>
<td>肥 féi</td>
</tr>
<tr>
<td>打 dǎ</td>
</tr>
<tr>
<td>做 zuò</td>
</tr>
<tr>
<td>龙 lóng</td>
</tr>
</tbody>
</table>

### Word list 2

<table>
<thead>
<tr>
<th>Word list 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>科 kē</td>
</tr>
<tr>
<td>翘 qiào</td>
</tr>
<tr>
<td>谢 xiè</td>
</tr>
<tr>
<td>锄 gē</td>
</tr>
<tr>
<td>吴 wú</td>
</tr>
<tr>
<td>挤 jǐ</td>
</tr>
<tr>
<td>兰 lái</td>
</tr>
<tr>
<td>色 sè</td>
</tr>
<tr>
<td>顶 dǐng</td>
</tr>
<tr>
<td>词 cí</td>
</tr>
</tbody>
</table>
### Word list 3

<table>
<thead>
<tr>
<th>物</th>
<th>代 dì</th>
<th>枭 zhāo</th>
<th>菱 yíng</th>
<th>仔 zǐ</th>
</tr>
</thead>
<tbody>
<tr>
<td>赶 gǎn</td>
<td>例 lì</td>
<td>拱 gǒng</td>
<td>冼 xiǎn</td>
<td>耐 nài</td>
</tr>
<tr>
<td>鼎 dǐng</td>
<td>闰 rùn</td>
<td>饶 ráo</td>
<td>耘 yún</td>
<td>蚯 qiū</td>
</tr>
<tr>
<td>聚 jù</td>
<td>握 wǔ</td>
<td>拥 yōng</td>
<td>燃 rán</td>
<td>燃 rán</td>
</tr>
</tbody>
</table>

### Word list 4

<table>
<thead>
<tr>
<th>物</th>
<th>命 mìng</th>
<th>为 wéi</th>
<th>点 diǎn</th>
<th>罚 fá</th>
</tr>
</thead>
<tbody>
<tr>
<td>比 bǐ</td>
<td>员 yuán</td>
<td>情 qíng</td>
<td>拧 nǐng</td>
<td>拧 nǐng</td>
</tr>
<tr>
<td>奴 nú</td>
<td>儿 ér</td>
<td>痛 tòng</td>
<td>忿 fèn</td>
<td>忿 fèn</td>
</tr>
<tr>
<td>咽 yān</td>
<td>鲁 lǔ</td>
<td>和 hé</td>
<td>臭 chòu</td>
<td>等 děng</td>
</tr>
<tr>
<td>仇 chóu</td>
<td>低 dī</td>
<td>窄 zǎi</td>
<td>低 dī</td>
<td>低 dī</td>
</tr>
<tr>
<td>借 jiè</td>
<td>醋 chī</td>
<td>酱 jiàng</td>
<td>醋 chī</td>
<td>醋 chī</td>
</tr>
<tr>
<td>真 zhēn</td>
<td>是 shì</td>
<td>脚 jiǎo</td>
<td>党 dǎng</td>
<td>烟 yān</td>
</tr>
<tr>
<td>汕 shān</td>
<td>民 mín</td>
<td>悦 yuè</td>
<td>体 tǐ</td>
<td>饥 jī</td>
</tr>
<tr>
<td>筹 chóu</td>
<td>甄 zhēn</td>
<td>又 yòu</td>
<td>烤 kǎo</td>
<td>烤 kǎo</td>
</tr>
</tbody>
</table>
Table J.2. Four sentence lists with better reliability and validity are selected for the SRT test from the total of 12 CHINT lists. Each list contains 20 sentences. The content of each used lists are presented in Mandarin in the following.

<table>
<thead>
<tr>
<th>Sentence List1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 他吃完晚饭 (经常 / 常常) (去) 散步</td>
</tr>
<tr>
<td>2. 他第二天就把钱送来 (了)</td>
</tr>
<tr>
<td>3. 我大年初一给爸爸拜年</td>
</tr>
<tr>
<td>4. 他家今天来 (了) (很多 / 许多 / 好多) 客人</td>
</tr>
<tr>
<td>5. 工地上有 (很多 / 许多 / 好多) 建筑工 (人)</td>
</tr>
<tr>
<td>6. 我在桌 (子) 下发现 (一) 支笔</td>
</tr>
<tr>
<td>7. 我十分钟后在门口等你</td>
</tr>
<tr>
<td>8. 他们请了 (个) 保姆 (看 / 带) 小孩</td>
</tr>
<tr>
<td>9. 这 [个 / 家] 大公司今年要裁员</td>
</tr>
<tr>
<td>10. 这 [个 / 位 / 名] 小女孩长得 (很 / 好) 秀气</td>
</tr>
<tr>
<td>11. 那 [个 / 家] 饭馆 (的) 包子 (很 / 好) 出名</td>
</tr>
<tr>
<td>12. 上 (个) 月 (很多 / 许多 / 好多) 人得了 (了) 流感</td>
</tr>
<tr>
<td>13. 他 (的) 作文在比赛中获奖</td>
</tr>
<tr>
<td>14. 姐夫修好了那 [个 / 台] 收音机</td>
</tr>
<tr>
<td>15. 小朋友用铅笔画向日葵</td>
</tr>
<tr>
<td>16. 她 (和 / 跟) 同学们失去 (了) 联系</td>
</tr>
<tr>
<td>17. 公司接到 (一) 份国外订单</td>
</tr>
<tr>
<td>18. 她今天下午 (突然 / 忽然) 肚子 (痛 / 疼)</td>
</tr>
<tr>
<td>19. 那 [个] 牌子 (的) 电器非常好</td>
</tr>
<tr>
<td>20. 爸爸准备去考 (驾驶执照 / 驾照)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sentence List2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. 农村过年 (的) 气氛 (很 / 好) 热闹</td>
</tr>
<tr>
<td>2. 元诞他收到 (一) 大 (堆 / 擎) 贺卡</td>
</tr>
<tr>
<td>3. (许多 / 很多 / 好多) 人业余时间学 (英文 / 英语)</td>
</tr>
<tr>
<td>4. 这个学习班 (很多 / 许多 / 好多) 人报名</td>
</tr>
<tr>
<td>5. 爸爸上街买回几斤橙子</td>
</tr>
<tr>
<td>6. 她今天 (戴了 / 戴着) (一) 条金项链</td>
</tr>
<tr>
<td>7. 今晚有奥运 (会) 实况转播</td>
</tr>
<tr>
<td>8. 看完演出 (后) 大家 (很 / 好) 兴奋</td>
</tr>
<tr>
<td>9. 我们都在学习电脑打字</td>
</tr>
<tr>
<td>10. 我有想法 (要) (和 / 跟) 经理谈谈</td>
</tr>
<tr>
<td>11. 他们后天 (要) 回老家探亲</td>
</tr>
<tr>
<td>12. 哥哥的脚磨起 (了 / 个) 大水泡</td>
</tr>
<tr>
<td>13. 爸爸 (每天 / 天天) 晚上 (要) 看新闻</td>
</tr>
<tr>
<td>14. 年纪大的人 (很 / 好) 容易骨折</td>
</tr>
<tr>
<td>15. 我明天上午九点有约会</td>
</tr>
<tr>
<td>16. 学生 (每天 / 天天) 上午 (要) 做早操</td>
</tr>
<tr>
<td>17. 她从来不吃刺激性食物</td>
</tr>
<tr>
<td>18. (很多 / 许多 / 好多) 人准备考律师执照</td>
</tr>
<tr>
<td>19. 他 (星期一 / 周一) (要) 参加会议</td>
</tr>
<tr>
<td>20. 他已经被 (送往 / 送去) 儿童医院</td>
</tr>
</tbody>
</table>
1. 我们(今天晚上/今晚)要 看球赛
2. 今天出门(要) (看/记得) 带雨伞
3. 运动员在操场练习投篮
4. 小朋友们在看蚂蚁搬家
5. 孩子都喜欢吃麦当劳
6. 这个(个/位/名)大夫手术做得(很/好) 好
7. 他(的) 表演吸引(很多/许多/好多) 观众
8. 天气冷(要) 注意多穿衣服
9. 他穿了一件灰格子上衣
10. 火车站(就) 在大楼(的) 右边
11. 叔叔脸上(的) 表情(很/好) 奇怪
12. 她一吃酸(的) 东西就牙(痛/疼)
13. 我们学校要举办运动会
14. 因为过敏他不能吃海鲜
15. 她们宿舍(的) 人都有蚊帐
16. 她(每天/天天)都(要) 工作到深夜
17. 我得承认他(很/好) 会作广告
18. 她们能写出(很/好) 抒情的诗
19. 我买了(一) 包观赏龟饲料
20. 我昨天没(能) 参加招待会

21. 他出差忘(记) 带身份证(了)
22. 他(每天/天天)上班都(要) 打领带
23. 这(棵) 茶树有三百年树龄
24. 今天(外面/外边) (起/下) 了(很好) 大(的) 雾
25. 下(星期一/周一) 我们(要) 交作业
26. 她们全家寒假(要) 去旅遊
27. 老师(很好) 关心我们的(的) 生活
28. 他在这方面有丰富经验
29. 我(很好) 相信她的决策能力
30. 我们(在) 看法上是一致的

Sentence List 4

1. 他们单位春节(放假五天/放五天假)
2. 护士(正在/在) 准备给她打针
3. 他们几个围在一起聊天
4. (昨天晚上/昨晚) 聚会她喝醉了
5. 看电影时(很多/许多/好多) 观众哭(了)
6. 橘黄色的灯显得(很/好) 温馨
7. 我早上在汽车站(遇到/遇见) 他
8. 旅行团(去/到) 北京参观故宫
9. 他没留神打破(一) (只/个) 花瓶
10. 他出差忘(记) 带身份证(了)
11. 我周末爱(到/去) 图书馆看书
12. 这(棵) 茶树有三百年树龄
13. 今天(外面/外边) (起/下) 了(很好) 大(的) 雾
14. 他(每天/天天) 上班都(要) 打领带
15. 下(星期一/周一) 我们(要) 交作业
16. 她们全家寒假(要) 去旅遊
17. 老师(很好) 关心我们的(的) 生活
18. 他在这方面有丰富经验
19. 我(很好) 相信她的决策能力
20. 我们(在) 看法上是一致的
Appendix K

A Sample DMDX Script

$ 0 "You will be listening to Mandarin words in noise",
  <ln 1> "press Spacebar to continue";

0 "Press SPACE for the 4 tones you will be hearing"

0 "+/" <jpg>"tone1Demo.JPG" / <wav 2>"ai1.wav";
0 "+/" <jpg>"tone2 Demo.JPG" / <wav 2>"ai2.wav" ;
0 "+/" <jpg>"tone3 Demo.JPG" / <wav 2>"e3.wav" ;
0 "+/" <jpg>"tone4 Demo.JPG" / <wav 2>"e4.wav";

0 "Now you will need to repeat each word you hear out loud",
  <ln 1>" and then click the icon that matches the tone ",

  <ln 5> "press Spacebar to continue";

0 "Sometimes it will be very hard",
  <ln 1>" please make your best judgment ",

  <ln 5> "press Spacebar to continue";

0 "The following are practice trials",
  <ln 1> “press Spacebar to start”;

+8 "+/<wav 2>"TW-HP-141-SN+9-a-1.wav"/*<jpg>"responseS_a_test.JPG";
+9"+/< wav 2>"TW-HP-891-SN+3-duan-4.wav"/*<jpg>"responseS_a_test.JPG";
+10"+/< wav 2>"TW-HP-8913-SN+6-re-4.wav"/*<jpg>"responseS_a_test.JPG";

0 "This is the end of the practice trials",
  <ln 1> "Please press the spacebar to start the experiment";
$
$0 "Please take a break and call the experimenter";$

+2346"+"<wav 2> "W2-HP-2239-SN-0-yu4.wav"/*<jpg>"responseS_a.JPG";
+2374"+"<wav 2> "W2-HP-3548-SN+3-lan2.wav"/*<jpg>"responseS_a.JPG";
+2403"+"<wav 2> "W2-HP-3548-SN-9-bing3.wav"/*<jpg>"responseS_a.JPG";
+2478"+"<wav 2> "W2-HP-5623-SN-3-mian2.wav"/*<jpg>"responseS_a.JPG";
+2525"+"<wav 2> "W2-HP-5623-SN-9-lin2.wav"/*<jpg>"responseS_a.JPG";
+2567"+"<wav 2> "W2-HP-8913-SN+6-ge3.wav"/*<jpg>"responseS_a.JPG";
+2645"+"<wav 2> "W2-HP-8913-SN-6-yang4.wav"/*<jpg>"responseS_a.JPG";
+2659"+"<wav 2> "W2-HP-141-SN+6-di3.wav"/*<jpg>"responseS_a.JPG";
+2737"+"<wav 2> "W2-HP-141-SN-6-ti1.wav"/*<jpg>"responseS_a.JPG";
+2753"+"<wav 2> "W2-HP-224-SN-3-bing3.wav"/*<jpg>"responseS_a.JPG";
+2804"+"<wav 2> "W2-HP-224-SN+9-chi3.wav"/*<jpg>"responseS_a.JPG";
+2874"+"<wav 2> "W2-HP-355-SN+3-lan2.wav"/*<jpg>"responseS_a.JPG";
+2926"+"<wav 2> "W2-HP-355-SN-9-lun4.wav"/*<jpg>"responseS_a.JPG";
+3001"+"<wav 2> "W2-HP-891-SN+6ben1.wav"/*<jpg>"responseS_a.JPG";
+3003"+"<wav 2> "W2-HP-891-SN-6bing3.wav"/*<jpg>"responseS_a.JPG";
+3093"+"<wav 2> "W2-HP-891-SN-6-xiu4.wav"/*<jpg>"responseS_a.JPG";
+3144"+"<wav 2> "W2-HP-1413-SN-3-yan3.wav"/*<jpg>"responseS_a.JPG";
+3197"+"<wav 2> "W2-HP-1413-SN+9hao4.wav"/*<jpg>"responseS_a.JPG";
+3233"+"<wav 2> "W2-HP-2239-SN+3-se4.wav"/*<jpg>"responseS_a.JPG";
+3301"+"<wav 2> "W2-HP-3548-SN+0-ben1.wav"/*<jpg>"responseS_a.JPG";
+3308"+"<wav 2> "W2-HP-3548-SN+0-dan1.wav"/*<jpg>"responseS_a.JPG";
+3365"+"<wav 2> "W2-HP-5623-SN+6-gai4.wav"/*<jpg>"responseS_a.JPG";
+3439"+"<wav 2> "W2-HP-5623-SN-6-wan4.wav"/*<jpg>"responseS_a.JPG";
+3455"+"<wav 2> "W2-HP-8913-SN-3-chu1.wav"/*<jpg>"responseS_a.JPG";
" $0 "Please take a break and call the experimenter";$

+3513"+"<wav 2> "W2-HP-8913-SN+9e2.wav"/*<jpg>"responseS_a.JPG";
+3596"+"<wav 2> "W3-LP-141-SN+3-ying3.wav"/*<jpg>"responseS_a.JPG";
+3614"+"<wav 2> "W3-LP-141-SN-9-geng4.wav"/*<jpg>"responseS_a.JPG";
+3687"+"<wav 2> "W3-LP-224-SN-3-ta1.wav"/*<jpg>"responseS_a.JPG";
+3709"+"<wav 2> "W3-LP-224-SN+9-du4.wav"/*<jpg>"responseS_a.JPG";
+3789"+"<wav 2> "W3-LP-355-SN+6-w01.wav"/*<jpg>"responseS_a.JPG";
+3829"+"<wav 2> "W3-LP-355-SN-6-qiao3.wav"/*<jpg>"responseS_a.JPG";
+3859"+"<wav 2> "W3-LP-562-SN-0-du4.wav"/*<jpg>"responseS_a.JPG";
+3949"+"<wav 2> "W3-LP-891-SN+3-z13.wav"/*<jpg>"responseS_a.JPG";
+3965"+"<wav 2> "W3-LP891-SN-9-guo2.wav"/*<jpg>"responseS_a.JPG";
+4023"+"<wav 2> "W3-LP-1413-SN-3-lang4.wav"/*<jpg>"responseS_a.JPG";
+4092"+"<wav 2> "W3-LP-1413-SN+9-xue3.wav"/*<jpg>"responseS_a.JPG";
+4151"+"<wav 2> "W3-LP-2239-SN-6-bo2.wav"/*<jpg>"responseS_a.JPG";
+4195"+"<wav 2> "W3-LP-2239-SN-6-ying3.wav"/*<jpg>"responseS_a.JPG";
+4222"+"<wav 2> "W3-LP-3548-SN-0-ke3.wav"/*<jpg>"responseS_a.JPG";
+4267"+"<wav 2> "W3-LP-5623-SN+3-hu4.wav"/*<jpg>"responseS_a.JPG";
+4307"+"<wav 2> "W3-LP-5623-SN-9-de2.wav"/*<jpg>"responseS_a.JPG";
+4376"+"<wav 2> "W3-LP-8913-SN-3-mi2.wav"/*<jpg>"responseS_a.JPG";
+4411"+"<wav 2> "W3-LP-8913-SN+9-e2.wav"/*<jpg>"responseS_a.JPG";
+4478"+"<wav 2> "W3-HP-141-SN+0-ni4.wav"/*<jpg>"responseS_a.JPG";
$0 "End of the block. Please call the experimenter";$

+6812"+"<wav 2> "W4-HP-3548-SN+9-e4.wav"/*<jpg>"responseS_a.JPG";
+6865"+"<wav 2> "W4-HP-5623-SN+3-gan4.wav"/*<jpg>"responseS_a.JPG";
+6908"+"<wav 2> "W4-HP-5623-SN-9dian3.wav"/*<jpg>"responseS_a.JPG";
+6983"+"<wav 2> "W4-HP-8913-SN+0-she3.wav"/*<jpg>"responseS_a.JPG";
Appendix L

CD Extras
Attached to this document is a CD.

List of contents:
1. Word audio files used for testing
2. Sentence audio files used for testing.