A Cross-Language Acoustic-Perceptual Study of the Effects of Simulated Hearing Loss on Speech Intonation

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Audiology

in the Department of Communication Disorders

at the University of Canterbury

By

Paul Daniell

University of Canterbury

2012

Table of Content

Acknowledgement iv

Abstract v

List of Tables vii

List of Figures viii

1. **Introduction** 1
   - 1.1 Intonation 1
     - 1.1.1 Features of Intonation 2
     - 1.1.2 Perception of Pitch 3
     - 1.1.3 Rising and Falling Boundary Tones 4
   - 1.2 Prosodic Language Typology 5
     - 1.2.1 English Intonation 7
       - 1.2.1.1 Dialectal Variation in English 7
       - 1.2.1.2 Stress and Intonation in English 9
     - 1.2.2 Mandarin Intonation 10
       - 1.2.2.1 Mandarin Lexical Tones 10
       - 1.2.2.2 Perception of Mandarin Intonation 11
       - 1.2.2.3 Acoustic Markers of Declarative Questions in Mandarin 14
     - 1.2.3 Taiwanese Intonation 16
       - 1.2.3.1 Taiwanese Lexical Tones 16
       - 1.2.3.2 Studies of Taiwanese Intonation 19
   - 1.3 Factors Affecting Question/Statement Perception 20
     - 1.3.1 Language 20
     - 1.3.2 Language Experience 22
   - 1.4 Hearing Loss and Intonation Perception 23
     - 1.4.1 Hearing Loss and Pitch Perception 24
     - 1.4.2 Hearing Loss and Temporal Fine Structure 26
   - 1.5 Cochlear Implants and the Encoding of Pitch 29
     - 1.5.1 Pitch Contour Perception of Cochlear Implant Users 30
     - 1.5.2 Electric and Acoustic Hearing Combined 32

1.6 Simulated Hearing Loss 33

1.6.1 Cochlear Hearing Loss 34

1.6.2 Temporal Jittering 36

1.7 Research Questions and Importance 37

1.8 Hypotheses and Rationales 38

2. **Method** 39

2.1 Participants 39

2.2 Participants' Tasks 40

2.3 Instrumentation 41

2.4 Signal Manipulation 42

2.5 Procedure 44

2.5.1 Recording of Sentences 44

2.5.2 Perceptual Experiment 44

2.6 Acoustic Analysis 46

2.6.1 Original Signals Only 46

2.6.2 Seven Types of Signals 47

2.7 Statistical Analysis 48

2.7.1 Acoustic Data 48

2.7.1.1 Original Signals 48

2.7.1.2 Seven Types of Signals 49

2.7.2 Perceptual Data 49

2.8 Reliability 51

3. **Results** 55

3.1 Acoustic Markers of Declarative Questions 55

3.2 Effects of Simulated Hearing Loss 57

3.2.1 Acoustic Effects 57

3.2.1.1 Percent Jitter 59

3.2.1.2 MomentCOV 60

3.2.2 Perceptual Effects 67

3.2.2.1 Percent Correct 67

3.2.2.2 Proportion of the Area (PA) 70

3.2.2.3 Error Analysis 72

3.2.2.3.1 Stress and Lexical Tone 72

3.2.2.3.2 Speaker Gender 75

4. **Discussion** 78
4.1 Study Findings in Relation to Research Questions 78
4.2 Study Findings in Relation to Previous Research 81
4.2.1 Acoustic Measures of Intonation 81
4.2.2 Perception of Vocoder-Processed Tokens 82
4.2.3 Perception of Temporally-Jittered Tokens 83
4.2.4 The Effect of Language 85
4.2.5 The Effect of Tone 86
4.2.6 The Effect of Stress 87
4.2.7 The Effect of Speaker Gender 88
4.3 Clinical Implications 89
4.4 Limitations and Future Directions 90
4.5 Conclusion 92
5. **References** 93

Appendix 1: Rules of Tone Sandhi in Mandarin 100

Appendix 2: Rules of Tone Sandhi in Taiwanese 101

Appendix 3: Instructions to Speakers 106

Appendix 4: Instructions to Listeners 107

Appendix 5: Consent Form 108

Appendix 6: List of 28 Phrases in English, Mandarin, and Taiwanese 109

Appendix 7: Categorisation of the last syllable for the 28 Phrases in English, Mandarin, and Taiwanese 110

Appendix 8: Predicting the F0ratio of a declarative question from the F0ratio or RMSratio of a statement 111

Acknowledgments

I would firstly like to express my sincere thanks to my supervisor Dr Emily Lin for her ongoing support and guidance over the last year, and for the many hours she spent on this project. Her expertise in statistical analysis and Mandarin and Taiwanese linguistics were indispensable. I would also like to express my sincere appreciation to my co-supervisor Dr Donal Sinex for his advice throughout the project and for his help in creating the hearing loss simulations. Finally, many thanks to all of the participants involved for giving their time to this study.

Abstract

Aim : The purpose of this study was to examine the impact of simulated hearing loss on the acoustic contrasts between declarative questions and declarative statements and on the perception of speech intonation. A further purpose of the study was to investigate whether any such effects are universal or language specific. **Method**: Speakers included four native speakers of English and four native speakers of Mandarin and Taiwanese, with two female and two male adults in each group. Listeners included ten native English and ten native speakers of Mandarin and Taiwanese, with five female and five male adults in each group. All participants were aged between 19 and 55 years old. The speaker groups were asked to read a list of 28 phrases, with each phrase expressed as a declarative statement or a declarative question separately. These phrases were then filtered through six types of simulated hearing loss configurations, including three levels of temporal jittering for simulating a loss in neural synchrony, a high level of temporal jittering in combination with a high-pass or a low-pass filter that simulate falling and rising audiometric hearing loss configurations, and a vocoder processing procedure to simulate cochlear implant processing. A selection of acoustic measures was derived from the sentences and from some embedded vowels, including /i/, /a/, and /u/. The listener groups were asked to listen to the tokens in their native language and indicate if they heard a statement or a question.

Results: The maximum fundamental frequency (F0) of the last syllable (MaxF0-last) and the maximum F0 of the remaining sentence segment (MaxF0-rest) were found to be consistently higher in declarative questions than in declarative statements. The percent jitter measure was found to worsen with simulated hearing loss as the level of temporal jittering increased. The vocoder-processed signals showed the highest percent jitter measure and the spread of spectral energy around the dominant pitch. Results from the perceptual data showed that participants in all three groups performed significantly worse with vocoder-processed tokens compared to the original tokens. Tokens with temporal jitter alone did not result in significantly worse perceptual results. Perceptual results from the Taiwanese group were significantly worse than the English group under the two filtered conditions. Mandarin listeners performed significantly worse with the neutral tone on the last syllable, and Taiwanese listeners performed significantly worse with the rising tone on the last syllable. Perception of male intonation was worse than female intonation with temporal jitter and high-pass filtering, and perception of female intonation was worse than male intonation with most temporal jittering conditions, including the temporal jitter and low-pass filtering condition.

Conclusion: A rise in pitch for the whole sentence, as well as that in the final syllable, was identified as the main acoustic marker of declarative questions in all of the three languages tested. Perception of intonation was significantly reduced by vocoder processing, but not by temporal jitter alone. Under certain simulated hearing loss conditions, perception of intonation was found to be significantly affected by language, lexical tone, and speaker gender.

List of Tables

Table 1. The four Mandarin tones, with examples of minimally contrastive monosyllabic words. 10

Table 2. The seven Taiwanese tones, with examples of minimally contrastive monosyllabic words. 17

Table 3. Descriptive statistics for the six acoustic measures with data from native English, Mandarin, and Taiwanese and non-native English productions combined. 56

Table 4. Results from the four-way (2 sentence types X 7 signal conditions X 3 languages X 2 genders) Mixed Model MANOVA conducted on the two vowel-based measures (%Jit and MomentCOV). 58

Table 5. Results from the four-way (7 signal conditions X 3 languages X 2 sentence types X 2 genders) Mixed Model ANOVA conducted on the percent correct scores. 61

Table 6. Results from the four-way (2 speaker genders X 2 signal conditions X 3 languages X 2 listener genders) Mixed Model ANOVA conducted on the error rates. 62

List of Figures

Figure 1. The computer interface for the perceptual experiment. 45

Figure 2. The means and standard errors of the measures of percent jitter (%Jit) for a 500 Hz pure tone processed with different levels of temporal jittering. 52

Figure 3. The means and standard errors of the measures of MomentCOV for a 500 Hz pure tone processed with different levels of temporal jittering. 54

Figure 4. The average %Jit values for each of the seven signal conditions with all languages, sentence types, and vowels combined. 59

Figure 5. The average MomentCOV values obtained from native English productions in each of the seven signal conditions (“Original”, “Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) with all sentence types and vowels combined. 61

Figure 6. The average MomentCOV values obtained from native Mandarin /i/ productions in each of the seven signal conditions (“Original”, “Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) with all sentence types and vowels combined. 62

Figure 7. The average MomentCOV values obtained from native Taiwanese /i/ productions in each of the seven signal conditions (“Original”, “Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) with all sentence types and vowels combined. 64

Figure 8. The average MomentCOV values obtained from native Taiwanese /a/ productions in each of the seven signal conditions (“Original”, “Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) with all sentence types and vowels combined. 65

Figure 9. The average MomentCOV values obtained from native Taiwanese /u/ productions in each of the seven signal conditions (“Original”, “Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) with all sentence types and vowels combined. 66

Figure 10. The average Percent Correct scores obtained from the perceptual task for question and statement stimuli presented in each of the seven signal conditions. 69

Figure 11. Mean and standard error of the PA measure across seven signal conditions for the English stimuli. 71

Figure 12. Mean and standard error of the PA measure across seven signal conditions for the Mandarin stimuli. 71

Figure 13. Mean and standard error of the PA measure across seven signal conditions for the Taiwanese stimuli. 72

Figure 14. The mean error rate for each of the seven tone types in Mandarin. 73

Figure 15. The mean error rate for each of the seven tone types in Taiwanese. ..74

Figure 16. The means and standard errors of the error rate measure for female and male voices across seven signal conditions. 77

1 Introduction

The major goal of hearing aid fitting and cochlear implantation is to improve the speech perception of individuals with hearing loss. Some aspects of the speech signal may vary across languages; therefore, amplification needs for different language users may be different. For example, the usage of pitch-related features in speech perception may differ between non-tonal and tonal language users. The current study aimed to evaluate the impact of different types of simulated hearing loss on the perception of speech intonation. Furthermore, as prosodic systems vary across languages, this study also aimed to determine whether the effect of simulated hearing loss on the perception of speech intonation is language specific or universal.

This chapter provides a review of the literature regarding speech intonation in general and in the languages investigated in this study, which included one non-tonal language, English, and two tonal languages, Mandarin and Taiwanese. A literature review on the perception of speech intonation in individuals with hearing loss is also provided to support the formulation and importance of the proposed research questions and methodology.

1.1 Intonation

Intonation can be defined as “the use of suprasegmental phonetic features to convey ‘post-lexical’ or sentence-level pragmatic meanings in a linguistically-structured way” (Ladd, 2008, p. 4). According to Ladd’s definition, intonational meaning applies to whole sentences or utterances, thereby excluding word level features such as stress, lexical tone, or pitch-accent, which are considered a part of the lexicon. By “linguistically structured”, it is suggested that intonation can convey a phonological code shared among speakers of a common language. This qualifier distinguishes intonation from more general paralinguistic features of the vocal code which may be understood by those who do not know the language, such as those used to perceive gender or age.

According to Crystal (2008), intonation has two functions. Firstly, it performs a grammatical role similar to that of punctuation in written language. This includes the marking of sentence and clause boundaries and the differentiation between questions and statements. Secondly, it can have the pragmatic role of expressing emotion and attitude, such as signalling certainty or uncertainty, sarcasm, or anger. The current study focuses on the grammatical role of intonation in distinguishing between interrogative and declarative utterances.

1.1.1 Features of Intonation

According to Cruttenden (1997), intonation involves variations of three main features: time length, loudness, and most importantly, pitch. Acoustically, pitch pattern is mainly associated with a global variation of the fundamental frequency (F0) of vocal fold vibration. A faster rate of vocal fold vibration results in a higher F0 and thus a higher perceived pitch. A higher pitch normally results from the contraction of the cricothyroid muscles and thus the lengthening and stiffening of vocal folds, whereas a reduction in pitch results from a decline in vocal fold tension or subglottal pressure (Snow, 2007). Due to differences in the mass and length of vocal folds, male speakers tend to have a lower F0 than female speakers. The average F0 of a male speaker is 125Hz while the average of a female speaker is 225Hz (Gussenhoven, 2004). The other two features of intonation, time length and intensity, also play a role in speech intonation. Time length refers to the relative duration of a linguistic unit, such as a syllable. Intensity, which is the main acoustic correlate of loudness, is produced physiologically mainly by breath-force (Cruttenden, 1997). Time length, pitch, and intensity appear to be interrelated. For example, a stressed syllable is normally longer than an unstressed syllable in English. In addition, a local pitch contour (i.e., pitch changes within a syllable) is often found in English to be associated with a longer syllable duration (Gussenhoven, 2004). Intensity has been found to vary with pitch in many languages. A falling pitch contour is usually accompanied by a drop in intensity, whereas a rising pitch contour is either associated with an increase in intensity or with no change in intensity (Oller, 1972).

1.1.2 Perception of Pitch

Two theories have traditionally been postulated to account for the perception of pitch: the place theory and the temporal theory (Oxenham, 2008). Both theories attempt to explain how pitch is represented in the auditory nerve. The place theory holds that pitch is determined by which auditory nerve fibres are excited the most for a particular signal. The temporal theory holds that pitch is determined by the timing of action potentials in the auditory nerve. These action potentials are determined by the incoming sound signal, as they are directly related to the sinusoidal phase-locking of the inner hair cells. Although research continues into both theories, it is now generally thought that pitch perception draws on both place and temporal mechanisms (Moore & Carlyon, 2005). Temporal cues are thought to be most important for frequencies below 4 to 5 kHz where phase-locking in the cochlea occurs (Moore, 1973a, 1973b, 2003). Above 4 to 5 kHz, the place mechanism is thought to be of most importance (Moore, 1973b; Sek & Moore, 1995)

As a complex sound, speech consists of an F0 and many harmonics. The harmonics of a complex sound provide information about the F0. In fact, even if the F0 is deleted from the complex sound, the perceived pitch will remain the same (Hartmann, 1996). This phenomenon is called the “pitch of the missing fundamental", and it has the effect of guarding the integrity of sounds even if partial masking occurs (Oxenham, 2008). It is thought the frequency information of individual resolved harmonics is cross-correlated to deduce information about the overall F0 (Oxenham, 2008).

The lower numbered resolved harmonics are more important in accurate pitch recognition when compared to higher numbered unresolved harmonics. In an early study, Plomp (1967) found that the dominant harmonics for F0s up to 350 Hz were the fourth harmonic and higher, for F0s up to 700 Hz, the third and higher, for F0s up to 1,400 Hz the second and higher, and for F0s above 1400 Hz the first. Moore, Glasbery, and Peters (1985) found that for F0s of 100, 200, and 400 Hz, the dominant harmonic was always within the lowest six harmonics, and was usually the second, third, or fourth harmonic. Dai (2000) found that harmonics around the frequency region of 600 Hz were the most dominant for determining pitch, and if the F0 was above 600 Hz, the fundamental itself dominated. Although differences exist among different studies, there is general agreement that harmonics between the first and fifth tend to dominate pitch perception for complex tones, and as F0 increases the dominant harmonic number decreases (Plack & Oxenham, 2005).

Perception of intonation relies largely on recognising the overall pitch pattern. The F0 is not a completely constant sequence as only voiced phonemes involve vocal fold vibration. Therefore, perceptual allowances are made for the gaps in the F0 contour as well as for the F0 fluctuation resulting from the inherent F0 differences between vowels (Cruttenden, 1997).

1.1.3 Rising and Falling Boundary Tones

When intonational tones are at the end of a prosodic constituent, they are said to be boundary tones (Pierrehumbert, 1980). Rising and falling intonation boundary tones display a common tendency in their presence across languages. Bolinger (1978) reported that several studies have found most languages (about 70%) use a rising terminal pitch for questions, whereas the remaining languages use a higher overall pitch for questions when compared to statements. Yes-no questions are more commonly associated with a rising pitch toward the end of a sentence, whereas wh-questions are often associated with a falling pitch (Cruttenden, 1997). There are, however, exceptions to this tendency in some languages, such as Chickasaw, Belfast English, and Bengali. Chickasaw and Belfast English speakers prefer a rising tone for declaratives while Bengali speakers employ a complex low-high-low tone for interrogatives (Gussenhoven, 2004).

In some languages (e.g. Portuguese), yes-no questions and statements may have identical syntax and morphology and, therefore, intonation is the only distinguishing feature (Cruttenden, 1997). Cruttenden (1997) terms these types of yes-no questions "declarative questions". Languages that do mark yes-no questions grammatically may also employ declarative questions as a form of questioning. For example, in English, declarative questions are commonly used with multiple functions, including requesting confirmation, initiating a conversational turn, and expressing surprise (Weber, 1993).

1.2 Prosodic Language Typology

Although pitch is the principle exponent of intonation, it also has other functions which may vary across languages. Traditionally, languages have been classified according to their use of pitch at the lexical level (Jun, 2005). According to this typology, languages have been categorised into tone languages, stress languages, and pitch-accent languages. Tone languages, such as Mandarin, use different pitch patterns to make lexical distinctions. Therefore, a number of individual words may be identical except for a difference in the pitch pattern. Stress languages and pitch-accent languages are similar in that both place emphasis on a certain syllable or syllables in a word. An example of a pitch-accent language is Japanese, where an accented word has a high pitch on the accented syllable followed by a low pitch on the following unaccented syllables (Jun, 2005). This pattern is predictable and cannot be changed by intonation. In stress languages, such as English, syllable emphasis may be achieved through intensity, and also through pitch-accent and duration. However, unlike pitch-accent languages, the type of pitch-accent given to the syllable is determined by the intonation of the utterance. The same stressed syllable in one word may receive a high, low, or other tone depending on the context (Cruttenden, 1997; Fang Liu, 2009). Jun (2005) notes that the categories of tone, stress, and pitch accent languages are not mutually exclusive. For example, tone languages and pitch accent languages can also have stress.

A tone language is characterised by the use of tonemes, which are phonologically distinctive units showing unique pitch patterns at the word level to convey differences in lexical meaning. Different tonemes (or tone types) can be differentiated from one another based on three main features: register, contour, and length (Duanmu, 2007). Tone register refers to different levels of pitch height, such as low, mid, and high. Contour refers to the rising or falling pattern of the pitch. Tone length is more related to how a tone is terminated than the actual length of the tone and thus can be considered as consisting of two main categories, checked and unchecked. A checked tone, also known as an entering tone, refers to a tone ending with an unaspirated voiceless stop (e.g., /p/, /t/, and /k/) or glottal stop. The finding from the study of tone languages, such as Taiwanese, that an unchecked tone is longer than a checked tone suggests that there is a relationship between the tone and time length of a tone-bearing unit (Pan, 2008).

Studies of the use of speech intonation for differentiating between questions and statements in one non-tonal language, English, and two tonal languages, Mandarin, and Taiwanese are reviewed respectively in the following section.

1.2.1 English Intonation

English is an example of a stress language, which has no lexical tone (Jun, 2005). As with most languages, English uses intonation in differentiating between questions and statements. In general, English uses a falling tone for declaratives and wh-questions, and a rising tone for yes-no questions (Wells, 2006). The dialectal variation and the control mechanism concerning the use of intonation to differentiate between questions and statements are discussed in the following sections.

1.2.1.1 Dialectal Variation in English

A study of American English telephone conversations (Hedberg, Sosa, & Fadden, 2004) revealed that a rising pitch pattern (rising intonation) was used in most cases of positive yes-no questions (81%), negative yes-no questions (76%), positive declarative questions (82%), and negative declarative questions (82%). Falling intonation was used with wh-questions in 82% of all occurrences. According to Bolinger (1978), the type of question which most consistently uses rising intonation in American English is an echo question, whereby the speaker repeats something they have just heard to ask for confirmation.

In a study of four British English dialects, Kochanski, Grabe, and Coleman (2004) found that the average F0 of questions was higher than the average F0 of statements. It was also found that declarative questions showed rising intonation in

78% of declarative questions, 61% of yes-no questions, and 38% of wh-questions. It was theorised that use of rising intonation was related to the relative number of syntactic cues available to distinguish the question from a statement. Since declarative questions do not have any syntactic cues, yes-no questions have a change in word order, and wh-questions have a change in word order and a question word as cues, it appears that the lower the number of syntactic cues, the more likely that rising intonation would be used to signal questioning.

Amongst the four British English dialects studied in the Kochanski et al. (2004) study, only Belfast English used rising intonation for declarative statements. This pattern is known as a high rise terminal (HRT), and it is also found in New Zealand and Australian English (Fletcher, Grabe, & Warren, 2007). The occurrence rate of HRTs in declarative utterances in New Zealand English, has been found to be about 1.5% for older speakers, and about 7.9% for younger speakers (Britain, 1992). In New Zealand English, the HRT is associated with the speech of young people, females, and Maori males (Britain, 1992; Warren & Britain, 2000). The HRTs in English were once thought to mark tentativeness or uncertainty; however, based on more recent discourse analysis, Warren and Britain (2000, p. 169) propose that "HRTs function as positive politeness markers, serving to overcome interspeaker hurdles, and to build and maintain speaker-hearer solidarity". Acoustically, HRTs in statements can be difficult to separate from the rising intonation in questions. However, a tendency has been noted for the rising intonation in questions to begin earlier than statement HRTs, especially in the speech of younger people, and these earlier rises are also more likely to be perceived as questions when compared to later rises (Warren, 2005).

1.2.1.2 Stress and Intonation in English

In a stress language, the mechanisms used in the production of intonation are also involved in syllable stress. Gussehhoven (2004) outlines the three measures by which the difference between stressed and unstressed syllables is achieved. Firstly, stressed syllables have an even intensity spread across the frequency spectrum, whereas unstressed syllables tend to have lower intensities for higher frequencies. Secondly, vowels in unstressed syllables are more schwa-like in that they are more centralised and more rounded. Thirdly, the length of both consonants and vowels in stressed syllables is longer than in unstressed syllables. It should be noted that the distinction between stressed and unstressed syllables in a stress language does not appear to be achieved either by the stressed syllable being louder overall or by the difference in F0 pattern.

Pitch does interact with stress as it is the stressed syllables which have the potential for receiving pitch accent (Bolinger, 1958). A pitch accent in English is given to syllables with the highest degree of stress. These types of syllables may be used in distinguishing between verb-noun pairs such as “permit” (noun), which has the stress in the first syllable, and “permit” (verb), which has the stress in the second syllable (Gussenhoven, 2004). However, even though a particular syllable may be pitch accented in a particular context, an intonational tone (e.g. rising question tone) may change the pitch accent. Liu (2009) found that the pitch accents of all stressed syllables were found to change from being either high or falling in statements to a rising in questions. For this reason, intonational tones are thought of as being independent from the pitch accents used in syllables with a high degree of stress (Gussenhoven, 2004).

1.2.2 Mandarin Intonation

In a tone language, pitch is used in differentiating between lexical items as well as in varying the overall intonation. This section includes a description of the Mandarin lexical tone system and a review of the acoustic and perceptual studies of question/statement intonation in Mandarin.

1.2.2.1 Mandarin Lexical Tones

Mandarin is a tone language with four lexical tones, which are all unchecked tones. The four Mandarin tones have traditionally been labelled from one to four, with Tone 1 being a high-level tone, Tone 2 a high-rising tone, Tone 3 a low-dip-rise tone, and Tone 4 a high-falling tone (Chao, 1965). Table 1 shows the four Mandarin tones with the qualitative descriptions and the traditional five-point numerical scales often used to indicate the three main features of a toneme (Chao, 1930). The four Mandarin tones can be described as consisting of two tone registers, high and low, and four contours, level, rising, falling, and falling-and-rising.

Table 1. The four Mandarin tones, with examples of minimally contrastive mono-syllabic words.

Tone type	Description	Numerical scale	Example
Tone 1	High-level	55	/pa^{55}/ "eight"
Tone 2	High-rising	35	/pa^{35}/ "to pull"
Tone 3	Low-dip-rise	21(4)	/pa^{21}/ "handle"
Tone 4	High-falling	51	/pa^{51}/ "father"

Chao (1965) also described a neutral tone (Tone 0), which is realised on weakly stressed syllables and is short in duration. As a neutral tone is rarely present in a single monosyllabic word, it is not classified as a toneme. In Mandarin, the preceding tone determines the pitch of the neutral tone so that the neutral tone is high after Tone 3 and low after the other tones. The Mandarin rules of tone change (i.e., tone sandhi) involve mainly a change of Tone 3 to either a falling or a rising tone depending on the following tone (see Appendix 1).

1.2.2.2 Perception of Mandarin Intonation

Due to the fact that Mandarin is a lexical tone language, there has been some controversy regarding the role of intonation in the formation of yes-no questions in both perceptual and acoustic studies. It has been suggested that identification of questions in Mandarin mainly relies on morphology or context. For example, Gao (2000, p. 138) states that "Generally speaking, as a tonal language, Chinese does not use intonation in asking questions." However, studies have found that Mandarin listeners are able to distinguish statements from yes-no questions using intonation cues alone. In a study employing one male adult native speaker and one female adult native speaker of Mandarin as listeners, Liu (2009) found that using intonation cues alone, participants were able to correctly identify yes-no questions or statements in 89.12% of 3,520 trials [11 listeners x 2 speakers (1 male, 1 female) x 4 tones x 4 focuses x 2 sentence types x 5 repetitions]. Perception of intonation with the female voice was also found to be significantly better than that with the male voice.

The accuracy rates for identifying sentence or question intonation in Liu's (2009) study were not significantly different between statements and questions. However, a difference between statements and questions in the accuracy rate of identification was indicated in Yuan's (2004) study. Yuan (2004) investigated identification of intonation type by 16 native Mandarin speakers. The identification ratio, which was calculated as the number of correct responses divided by that of total responses, was found to be significantly higher for statements (Mean = 0.98) than for yes-no questions (Mean = 0.8). It appears that statement intonation was more easily identified than question intonation. However, since participants were forced to choose one of the two intonation types as a response, the results do necessarily indicate that statement intonation is easier to identify than question intonation. It may be that statement intonation is the unmarked form and therefore participants were likely to choose this response unless they explicitly heard the marked intonation form.

Tone Type Effect. Both Yuan (2004) and Liu (2009) also investigated the effect of the lexical tone type of the sentence-final syllable on the perception of statement/question intonation. Since both lexical tone and intonation are encoded by F0, it was hypothesised that lexical tone would influence the perception of intonation. Yuan (2004) found that the tone of the last syllable in an utterance affected the perception of question intonation but not the perception of statement intonation. Identification of question intonation was found to improve if the lexical tone of the final syllable was a falling tone (Tone 4). If the lexical tone of the final syllable was a rising tone (Tone 2), perception of question intonation was worse. Liu (2009) also found that an utterance with a falling tone (Tone 4) in the final syllable resulted in a higher accuracy rate in identifying questions and statements. In addition, with the accuracy rates for the identification of statement and question intonation combined, Liu (2009) found the performance to be worse when the final syllable of an utterance was a high level tone (Tone 1) compared to other tones.

Both Yuan's (2004) and Liu's (2009) studies demonstrated a significant effect of the tone type in the sentence-final syllable on the perception of question intonation, with Tone 4 (high-falling) resulting in the best performance. However, the tone effect on the statement identification was only identified in Liu's (2009) study and the findings regarding Tones 1 and 2 were inconsistent between the two studies. These differences may be due to the fact that Liu's (2009) study employed sentences with a focus in different locations whereas Yuan's (2004) study employed sentences with a neutral focus. Also, participants in Liu's (2009) study were required to identify the intonation pattern and the focus position, whereas participants in Yuan's (2004) study were required to identify intonation pattern only. The difference between the two studies regarding the stimuli used and the complexity of the subject's task may have led to the differences in the findings.

Focus Effect. In Mandarin, like many languages, the focus of a sentence is encoded by intonation. The focus may occur in the beginning, middle, or end of a sentence, or the sentence may be without a focus (neutral focus). Yuan (2004) found that identification of statements was not affected by a focus at the beginning or middle of an utterance compared to identification of statements in an utterance with a neutral focus. However, a focus at the end of an utterance made statement intonation more difficult to identify (88% correct) than a neutral focus (98% correct) or a focus at the beginning (98% correct) or middle of an utterance (98% correct). Similarly, Liu (2009) found that identifying statement intonation is significantly more difficult when the focus is in the sentence-final position (77.5% correct) than when the focus is neutral (91.6%) or in the initial (92.7% correct) or medial (89.3% correct) position. As for identification of questions, Yuan (2004) found that accuracy was worse when the focus was neutral (80% correct) or in the initial position (81% correct) compared to medial (90% correct) and final position (90% correct). Liu (2009) also found that identifying question intonation was more difficult when the focus was in the initial position (86.6% correct) and easiest when the focus was in the final position (96.9% correct). The accuracy rate for identifying questions was higher for a sentence with a neutral focus (91.4%) compared to a medial focus (87.1%).

In summary, both Yuan's (2004) and Liu's (2009) studies have shown that native Mandarin speakers are able to use intonation alone to distinguish statements from questions. The accuracy rate in the question/statement identification was also found to be affected by the type of lexical tone in the sentence-final position as well as the placement of sentence focus.

1.2.2.3 Acoustic Markers of Declarative Questions in Mandarin

Findings from studies regarding the acoustic markers of declarative questions in Mandarin have also demonstrated that intonation was used to differentiate between statements and questions. Yuan (2006) carried out an acoustic analysis of 130 sentences produced by 4 female and 4 male native Mandarin speakers (1040 tokens) to determine how the differences between statement and question intonation in Mandarin is achieved through speech intonation. The study identified three elements of intonation to be relevant to the differentiation between questions and statements. The three elements included pitch, loudness, and duration. Based on findings from the acoustic analysis, Yuan (2006) found that compared to statements, questions have a higher overall F0 curve, a higher overall intensity curve, and a higher final tone intensity. In addition to these general differences, the main findings related to particular tones are shown as follows.

1. Tone 3 in the sentence-final position was found to pull down the question F0 curve to the statement curve.

2. Tones 3 and 4 in the sentence-final position have a longer duration in questions than in statements.
3. Tone 2 in the sentence-final position has the largest intensity difference between question and statement and Tone 4 the smallest.
4. Tone 2 in the sentence-final position has a steeper F0 slope in questions than in statements.
5. Tone 4 in the sentence-final position shows no difference in the F0 slope between questions and statements.

Based on these findings, Yuan (2006) proposes three mechanisms to explain the difference in intonation between questions and statements in Mandarin. These mechanisms may be employed to produce the marked question intonation. The first mechanism is to use an overall higher F0 for questions compared to statements. The second is to produce higher strengths for the sentence-final tones for questions compared with statements. Higher strengths are indicated by an increase in F0 and intensity toward the end of the sentence and a longer duration of the final syllable. The third mechanism for marking questions is a tone-dependent mechanism, whereby a falling tone in final position is flattened and a rising slope is steepened. These findings from Yuan's (2006) study were based on utterances with a neutral focus.

It was found in Liu's (2009) study that when the focus was at the beginning of a question, there was a rise in the overall pitch. However, it was also found that when the focus was in the middle of a question, the pitch rise was only seen from the point of focus onwards. Therefore, it appears that the focus can be seen as the point at which the statement and question intonation curves begin their divergence. From this point, the F0 of question intonation rises exponentially when compared to that of statement intonation. As for utterances with a neutral focus, Liu and Xu (2007) found that a neutral tone in the sentence-final position has a falling contour in both statements and questions but the fall was steeper in statements than in questions.

In summary, the findings of Yuan (2006) and Liu (2009) show that the following perceptual cues may be used in Mandarin to distinguish question intonation from statement intonation:

1. Pitch: Question intonation has a higher overall F0 curve, diverging at the point of focus and increasing exponentially so that it is highest at the final syllable. Falling tones in final position are flattened, and rising tones are steepened.
2. Loudness: Question intonation has a higher overall intensity than statement intonation, with intensity increasing toward the end of a sentence so that it is loudest at the final syllable.
3. Duration: Question intonation generally has shorter syllables than statement intonation, except for the final syllable which is generally longer.

1.2.3 Taiwanese Intonation

The Taiwanese lexical tone system differs from Mandarin in both tonemes and tone sandhi. In particular, the tone change in Taiwanese depends mainly on the prosodic context while that in Mandarin on the tonal context (Peng, 2008). As these differences may be important in the perception of question/statement intonation, this section provides information regarding the lexical tone system of Taiwanese and some studies of Taiwanese intonation.

1.2.3.1 Taiwanese Lexical Tones

Taiwanese has seven phonologically distinctive tones, two of which are checked tones. The seven tones are high-level, mid-level, low-rising, low-falling, high falling, low-falling checked, and high-falling checked (Peng, 2008). Taiwanese tones include three tone registers, low, mid, and high, and three contours, level, rising, and falling (see Table 2).

Table 2. The seven Taiwanese tones, with examples of minimally contrastive mono-syllabic words. The numerical scales for the checked tones are underlined and the pitch associated with the glottal stop is indicated as “0”.

Tone type	Description	Numerical scale	Example
1	High-level	55	/si^{55}/ “poem”
2(6)*	High-falling	51	/si^{51}/ “to die”
3	Low-falling	21	/si^{21}/ “four”
4	Low-falling checked	20	/sit$^{\underline{20}}$/ “to lose”
5	Low-rising	24	/si^{24}/ “hour”
7	Mid-level	33	/si^{33}/ “yes”
8	High-falling checked	50	/sit$^{\underline{50}}$/ “solid”

* Tones 2 and 6 are the same.

A major difference between the Mandarin and Taiwanese tone systems is that Mandarin uses F0 contour for tone contrasts but Taiwanese uses both F0 contour and F0 height (Pan, 2008). Chen (2005) examined the F0 and intensity ranges of speakers of Taiwanese and Mandarin speakers. Results showed that Taiwanese speakers displayed a larger intensity range and a smaller lowest intensity when compared to the Mandarin speakers.

Taiwanese also has a much more extensive system of tone sandhi compared to Mandarin (see Appendix 2). In Taiwanese tone sandhi, all tones can undergo changes except when the tone is in isolation or in the end of a phrase or a sentence (Cheng, 1968, 1973). Tone change in Taiwanese generally follows a consistent change pattern (see Appendix 2).

For the unchecked tones in Taiwanese, a regular tone change rule applies, with Tone 5 (low-rising) changing to Tone 7 (mid-level), Tone 7 to Tone 3 (low-falling), Tone 3 to Tone 2 (high-falling), Tone 2 to Tone 1 (high-level), and Tone 1 to Tone 7 (see Appendix 2). Two of these five tone change rules, including the change from Tone 5 (low-rising) to Tone 7 (mid-level) and the change from Tone 2 (high-falling) to Tone 1 (high-level), show a pattern of levelling off the lexical tone of a word when undergoing tone change. The tone changes from Tone 1 (high-level) to Tone 7 (mid-level) and from Tone 7 (mid-level) to Tone 3 (low-falling) show that the high and mid-level tones drop to a lower pitch height when undergoing tone change. These trends, along with the change from Tone 3 (low-falling) to Tone 2 (high-falling), appear to move tones in the direction toward the average pitch level and have an effect of equalising the overall pitch level. Although these tone changes are not dependent on the tone context, the levelling off effect may suggest a coarticulatory effect as it allows for a smoother assimilation in pitch over the course of an utterance.

On the other hand, a close inspection of some of the examples shown in Appendix 2 also revealed that most of the tone-dependent but tone-text-independent rules of tone change in Taiwanese have an effect of increasing the pitch height distance between the ending pitch level of a word and the starting pitch level of the word which immediately follows (e.g., See Appendix 2, from Tone 1 to Tone 7; /t^{h}in^{55}/: "sky", /t^{h}in^{33} tieng^{51}/: "up in the sky") resulting in a more distinctive marking of word junction. In other words, the tone change rules in Taiwanese appear to make the slope of the local between-word F0 contour steeper (i.e., a more abrupt change in F0) to signal word boundaries. A more distinctive F0 contrast between words may provide listeners an advantage in perceiving the word boundaries in a word sequence.

In summary, the tone sandhi of Taiwanese is characterised by a tone change rule that tends to level off tone registers for a global assimilation in pitch but increase the slope of the local pitch shift in the word boundary. As the tone of the last syllable of an utterance undergoes a tone change in one tone language (e.g., Mandarin) but not another (e.g., Taiwanese), an investigation on the effect of the sentence-final lexical tone of these two types of tone languages on the acoustic difference between declarative questions and statements may provide an insight into how speech intonation may be ruled by a universal physiological constraint of speech and voice production or speech perception.

1.2.3.2 Studies of Taiwanese Intonation

Peng and Beckman (2003) describe three constituents important to the Taiwanese prosody, including syllable, tone sandhi group, and intonational phrase. Based on an acoustic analysis of a database of spoken Taiwanese, a global pitch change in Taiwanese was found and considered to be a property of the intonational phrase rather than the syllable (Peng & Beckman, 2003). One of the global pitch patterns identified included a gradual rise on the last tone sandhi group for some question types and another a gradual fall on the final tone sandhi group for declarative statements (Peng & Beckman, 2003). For statements, a distinction between F0 declination, which refers to a gradual F0 decline over the course of an utterance, and final lowering, which is the F0 decline at the end of an utterance, has been noted (Pan, 2008). The focus effect on the syllable length in Taiwanese has been found in unchecked tones to vary by tone type (Wong et al., 2008). As for speech perception, Lin and Repp (1989) have shown, using synthetic speech stimuli, that the perception of tonal distinction in Taiwanese was related to both pitch height and pitch contours. Specifically, it was found that pitch height could be used alone to differentiate between tones with similar pitch contours, namely, Tones 1 ("high-level" and 7 ("mid-level") and Tones 2 ("high-falling" and 3 ("low-falling"). In contrast, pitch contour was found to be used predominantly to differentiate between tones with dissimilar contours such as Tones 2 ("high-falling") and 5 ("low-rising"). These findings suggest that tonal distinction is multifaceted and thus pitch height and pitch contour, along with tonal and prosodic contexts, may all provide the redundant acoustic cues for the differentiation between declarative statements and questions.

1.3 Factors Affecting Question/Statement Perception

Given the differing prosodic systems of English, Mandarin, and Taiwanese, there are several factors that may affect the perception of question/statement intonation. These include factors related to the language itself and factors related to the language experience of the listeners.

1.3.1 Language

Of the three languages involved in the current study, pitch, intensity, and duration, are all used to produce intonation. These features are also used to produce lexical tone in Mandarin and Taiwanese and lexical stress in English. Several differences in intonation between tonal and non-tonal languages have been noted. Firstly, in a study comparing the temporal scope of question/statement intonation in

English and Mandarin, question intonation in English was found to be characterised by an F0 rise which started from the first content word, while the F0 rise in Mandarin questions was found to occur later (Liu, (2009). Secondly, the F0 rise in question intonation was found to have an effect on the pitch targets of all stressed syllables in English but not in Mandarin (Liu, 2009). In Mandarin, the pitch targets of the lexical tones were not found to be affected by question intonation; however, the pitch accent contour of stressed syllables in English depended on the overall intonation contour. Lastly, Chen (2005) compared the F0 and intensity of speech from three languages; Mandarin, Taiwanese, and American English. Results showed that both tonal languages, Mandarin and Taiwanese, displayed significantly larger F0 range and intensity range compared to American English.

There is also evidence that question/statement intonation may differ between different tone languages. Ma, Ciocca, and Whitehall (2011) studied the interaction of lexical tone and intonation in Cantonese and concluded that the use of perceptual cues may differ across languages. This was based on the observation that their findings differed somewhat from a similar study of Mandarin (Yuan, 2004). In contrast to the result found with Cantonese, perception of Mandarin statements was not found to be influenced by the type of lexical tone in the sentence final position. Ma et al. (2011) suggest that the differences found may be due to the differences between in the lexical tones systems of Cantonese and Mandarin. Specifically, Mandarin question intonation is characterised by an overall rise in F0 (Connell, 1983; Yuan, 2004), whereas Cantonese employs a rising boundary tone on the final syllable (Ma, Ciocca, & Whitehill, 2006).

Xu and Mok (2012a) found that Mandarin speakers were less accurate in identifying question/statement intonation in Mandarin when compared to Cantonese speakers identifying question/statement intonation in Cantonese. A further experiment was also conducted using tokens with the final syllable being cut off (Xu & Mok, 2012b). This had the effect of significantly decreasing Cantonese accuracy rates but not Mandarin. The differences found were presented as evidence of Mandarin's use of a more global pitch rise compared to Cantonese's use of a final boundary rise tone, and that these differing mechanisms resulted in the different perceptual accuracy.

1.3.2 Language Experience

There is a strong body of evidence that native speakers of tone languages display differences in the processing of pitch information. For example, Chandrasekarn, Krishnan, and Gandour (2007) compared mismatch negativity measures (MMN) from Mandarin native speakers and American English native speakers. The MMN is a measure derived from an electroencephalographic signal monitoring the mismatch response elicited by a deviant sound interspersed among a series of standard auditory stimuli. In the task of listening to an oddball stimulus which consisted of a repeated Mandarin tone followed by a deviant tone, the Mandarin group showed higher amplitudes and longer MMN latencies in response to the change in stimulus than the English group. This finding suggests that early cortical processing of pitch contours is somewhat dependent on language experience, in other words, MMN responses depend on the relative saliency of the acoustic cue in the listener's native language.

Dichotic perception tests have also been used to examine the effect of language experience on tone processing. Wang, Jongman, and Sereno (2001) found that Mandarin native speakers display a significant right ear advantage during dichotic testing with Mandarin tone stimuli. However, English native speakers with no tone language experience did not display a preference for either ear while doing this task. The left hemisphere superiority, which was demonstrated through the right ear advantage in the Mandarin speaking group, was interpreted as evidence of linguistic processing of the tones.

Even though tone language speakers display perceptual differences in lexical pitch perception, this does not transfer to the perception of intonation pitch. Braun and Johnson (2011) found that when compared to Mandarin native speakers, Dutch native speakers were less likely to be attentive to pitch movements (rise or fall) in nonsense words when the rise was on the first syllable of a two syllable word. Mandarin speakers were more attentive to this pitch movement because it imitated possible lexical information in their native language. However, Dutch speakers were attentive to the pitch movement when it occurred on the final syllable because this imitates question or statement information in Dutch.

1.4 Hearing Loss and Intonation Perception

A review of the literature reveals few studies investigating the impact of hearing loss on the perception of question/statement intonation. However, there are other types of studies which are suggestive of possible difficulties perceiving intonation contours caused by hearing loss. These include studies investigating the perception of intonation encoding emotion and those investigating the production of intonation by persons with hearing loss.

Pereira (1996) studied the perception of intonation encoding anger, sadness, happiness, and neutrality. The study involved two groups, one of 40 normally hearing adults and one of 39 post-lingually deafened users of hearing aids. The hearing aid group had a pure tone average ranging from 13 to 95 dBHL. Results showed that the hearing aid group performed significantly worse on the intonation perception tasks with an overall average correct identification rate of 65%. The normal hearing group had an average correct identification rate of 85%. Most and Aviner (2009) compared four groups of 10 participants between the ages of 10 and 17 years old. These four groups included a normal hearing group, a hearing aid group (severe to profound bilateral hearing loss), an early implantation (before six years old) cochlear implant group, and a late implantation (after six years old) cochlear implant group. Participants were asked to choose between six different emotions on the intonation perception task. All three hearing loss groups performed significantly worse when compared to the normal hearing control group. No significant differences were found between the three groups with hearing loss. Results from these two studies show that, even with amplification, individuals with hearing loss may have a reduced ability to perceive intonation.

Intonation in the speech of hearing impaired individuals has also been found to have decrease prominence. Allen and Arndorfer (2000) analysed the question/statement intonation of six children with severe to profound hearing loss, and six children with normal hearing. They found that although the hearing loss children used pitch, intensity, and duration to produce intonation in a similar way to the normally hearing children, the contrasts between question and statement intonation were significantly less pronounced. In the perceptual experiment, listeners were significantly less likely to correctly identify the question/statement in the speech of the hearing impaired children. The implication of this study is that even with amplification, the hearing impaired children were not fully able to access and utilise intonation cues in their language development.

1.4.1 Hearing Loss and Pitch Perception

Moore and Carlyon (2005) listed several changes in the functioning of the cochlea in people with hearing loss which may contribute to difficulties with pitch perception. These included the bandwidth broadening of auditory filters and the reduction in the phase locking sensitivity caused by damage to the cochlear hair cells. Both of these factors can result in decreased frequency resolution, which in turn can adversely affect the ability to resolve harmonics to perceive pitch accurately. The reduction in the active mechanism of the outer hair cells can also cause abnormalities in the timing of the travelling wave. These abnormalities may disrupt the cross-correlation of different points on the basilar membrane (harmonics) and consequently pitch perception. Finally, Moore and Carlyon (2005) stated that parts of the cochlea may consist of completely non-functioning inner hair cells or nerves. These areas are known as “cochlear dead regions” and can result in neural excitation occurring at areas of the cochlea not normally associated with that frequency. When a pitch falls in a cochlear dead region, subjects sometimes report a noise-like sound or a change in pitch, although a low pitch may sometimes be perceived correctly even when occurring in a cochlear dead region (Moore & Carlyon, 2005).

Studies have shown that there is great variability in the pitch perception abilities of people with hearing impairment (Moore & Carlyon, 2005). A common finding among subjects with cochlear hearing impairment is poorer results on frequency difference limen tests (e.g. Moore & Peters, 1992; Simon & Yund, 1993). This results in greater difficulty in resolving complex tone harmonics which may be required to determine pitch. Although cochlear hearing loss can involve the bandwidth broadening of auditory filters (Oxenham, 2008), a strong correlation between broader auditory filters and larger frequency difference limens has not been found (Moore & Peters, 1992). This suggests that pitch perception may require temporal coding as well as place coding (Oxenham, 2008). Frequency modulation difference limens are also generally adversely affected by cochlear hearing loss (Moore & Carlyon, 2005; Moore & Skrodzka, 2002). It is thought that this deficiency is caused by disruptions to both place and temporal mechanisms (Moore & Skrodzka, 2002).

Cochlear hearing loss may also affect F0 difference limens. Moore and Peters (1992) found that discrimination of fundamental frequency by people with hearing impairment was significantly worse than that by normal hearing participants. They also found that F0 discrimination among older participants with normal hearing was significantly worse than that of younger participants, suggesting a possible age effect. Bernstein and Oxenham (2006) found an association between larger-than-normal F0 difference limens and broader-than-normal auditory filters.

Souza et al. (2010) investigated the link between F0 difference limens and the perception of rising and falling intonation. They used two groups of normal hearing participants, older and younger, and three conditions, unprocessed speech, electroacoustic simulation, and cochlear implant simulation (vocoder). A similar finding to Moore and Peters (1992) was that the older participants had significantly poorer F0 difference limens. Presented with synthetic diphthongs with the F0 rising or falling at 12 different rates, participants were asked to identify the tokens as rises or falls. Although none of the participants were hearing impaired, the older group performed significantly worse than the younger group. Under the unprocessed speech and vocoder conditions, poorer performance on the intonation perception task was associated with poorer F0 difference limens. This suggests that identification of F0 movement relies to some extent on the ability to detect difference in individual F0s.

However, it should be noted that good performance on the F0 difference limens task was not always associated with good performance on the intonation perception task.

1.4.2 Hearing loss and Temporal Fine Structure

Individuals with cochlear hearing loss may have a reduced ability to process temporal fine structure (TFS). TFS refers to the rapid changes in a wave form which occur around a centre frequency and are conveyed through phase locking (Moore, 2008). Although the more slowly varying temporal envelope has been regarded as more important for speech perception, TFS has been found to have several important roles (Moore, 2008). By combining the temporal envelope cues from one utterance and the TFS cues from another, Smith, Delgutte, and Oxenham (2002) found that temporal envelope cues dominated for speech perception while TFS cues were more important for pitch perception and sound localisation. Moore et al. (2006) also found temporal fine structure was important for determining pitch when unresolved harmonics below Harmonic 14 were present. However, a further study found the temporal envelope was more important for determining pitch when only harmonics above Harmonic 14 are present (Moore, Glasberg, Low, Cope, & Cope, 2006).

When TFS cues are deleted from a signal, temporal envelope cues still provide good speech intelligibility (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995). However, this worsens when background noise is present, suggesting that TFS also plays a perceptual role in separating speech from noise (Moore, 2008; Qin & Oxenham, 2003). Drullman’s studies (1995a, 1995b), suggest that the TFS in the troughs of the amplitude envelope of a speech signal are important for this separation of speech from noise. His studies investigated word identification under several different signal-to-noise conditions. Using speech in noise, he found that when the noise was eliminated from the peaks of the speech signal, intelligibility (speech recognition threshold) was not affected. However, when the signal was eliminated from the troughs of the amplitude envelope, a 1.5dB improvement in signal-to-noise ratio was required in order to achieve the same speech recognition threshold score.

Several studies have shown that people with cochlear hearing loss have a reduced ability to make use of TFS cues. Hopkins and Moore (2007) found that when compared to normally hearing participants, participants with moderate cochlear hearing loss were less able to use TFS to discriminate complex tones. Hopkins and Moore (2008) found that adding TFS to a vocoder speech signal would improve speech perception in noise for normally hearing individuals. However, subjects with cochlear hearing loss gained less benefit from the added TFS, suggesting they had deficiencies in processing TFS. Studies have also shown that individuals with cochlear loss are less able to make use of TFS cues in detecting frequency modulation (Moore & Skrodzka, 2002), distinguishing inter-aural phase differences (Lacher-Fougere & Demany, 2005), and discriminating the F0 of complex tones (Moore & Moore, 2003). There is also some evidence that loss of the ability to process TFS cues may cause difficulties in speech recognition (Buss, Hall, & Grose, 2004). Although it is not yet clear exactly why cochlear hearing loss can result in a loss in processing the TFS cues, Moore (2008) suggests that the reduced precision of phase-locking may be an important factor.

Age may also be an important factor in the reduction of the ability in processing TFS. Studies have shown that older people without hearing loss may still have difficulty in the processing of F0 information (Moore & Peters, 1992; Souza et al., 2010) and in distinguishing speech from noise (Pichora-Fuller, Schneider, & Daneman, 1995). It has been speculated that this decline is caused by a loss in neural synchrony which may accompany aging (Gates, Feeney, & Higdon, 2003; Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007).

In summary, cochlear hearing loss and aging have been associated with a reduced ability to process TFS. A loss in TFS processing ability has been found to lead to deficits in skills associated with the processing of F0, which is important for pitch perception, localisation, and detection of speech in noise. Based on these observations, it is possible that the perception of intonation, which also relies on F0 information, may also be adversely affected by the reduction of TFS processing ability.

1.5 Cochlear Implants and the Encoding of Pitch

The perception of intonation by cochlear implant users is limited by physiological and surgical factors as well as factors related to the speech processing strategy used by the cochlear implant. An important surgical factor is the insertion of the implant into the cochlear. Because the implants are not usually fully inserted into the cochlear, the range of available frequencies will be less than that of a normally hearing cochlea (Moore & Carlyon, 2005). The insertion of the electrode array is also not likely to be completely accurate. Ketten et al. (1998) found that, out of 20 cochlear implantees, the insertion of the most apical electrode corresponded to a tonotopical cochlear location ranging from 387 Hz to 2,596 Hz. The most apical electrode generally encodes frequencies below 240 Hz, and therefore in many cases there will be a mismatch between encoded frequency and place. Even when electrodes are correctly positioned, impulses generated cannot always be guaranteed to stimulate the corresponding auditory nerves. This may be due to abnormal flow of current, or to the presence of cochlear dead regions resulting from neural degeneration (Moore & Carlyon, 2005).

In normal acoustic hearing, both temporal phase locking information and place of excitation on the basilar membrane combine to convey pitch information. The pitch conveyed by cochlear implants can also be conveyed by both temporal and place cues. Studies of patients with a unilateral cochlear implant and some residual hearing in the other ear have shown that as more basally located electrodes are stimulated, the perceived pitch increases (Boëx et al., 2006; Dorman et al., 2007). Studies have also found that manipulation of the pulse rate of a single electrode can result in a change of the perceived pitch (Pfingst et al., 1994; Zeng, 2002). This effect has been found to be evident only for the pulse rate up to about 300 Hz, suggesting this is the upper boundary for temporal coding (McKay, McDermott, & Carlyon, 2000; Zeng, 2002). Although both temporal changes and place changes were considered to affect pitch, they have been found to be related to separate perceptual dimensions (McKay et al., 2000; Tong, Blamey, Dowell, & Clark, 1983). It has been suggested that it is the combination of the two types of cue which may be important in decoding pitch (Carlyon & Deeks, 2002).

Speech processing strategies used by cochlear implants have important implications for the perception of intonation. The most common speech processing strategies used in cochlear implants (CIS and SPEAK) do not provide the same correspondence between temporal and place cues as acoustic hearing. Instead, the same pulse train rate is applied to all electrodes and this pulse is frequency modulated by the temporal envelope extracted from the incoming signal (Moore & Carlyon, 2005). Therefore, only temporal envelope information is conveyed by the cochlear implant, with temporal fine structure being discarded (Oxenham, 2008). Due to the importance of temporal fine structure cues in the detection of F0 information (Moore, 2008; Zeng et al., 2004), this loss is a key factor in adversely affecting the pitch and intonation perception of cochlear implant users.

1.5.1 Pitch Contour Perception of Cochlear Implant Users

Many studies have found that cochlear implants were more beneficial than hearing aids to children with severe to profound hearing loss (e.g. Blamey et al., 2001). However, these studies tend to look at the perception of segmental features of speech. When comparison has been conducted on the perception of suprasegmental features of language, such as intonation and syllable stress, this advantage is not evident. For example, studies have found that cochlear implant users may have difficulties with several aspects of language encoded by the fundamental frequency, including intonation (Most & Peled, 2007), tone perception (Ciocca, Francis, Aisha, & Wong, 2002; Luo & Fu, 2006), and word emphasis (Meister, Landwehr, Pyschny, Wagner, & Walger, 2011).

Most and Peled (2007) compared the suprasegmental perception of Hebrew speaking children with cochlear implants to those with hearing aids. All of the children in the cochlear implant group used a Nucleus 24 with the ACE processing strategy. They found that hearing aid users performed significantly better than the cochlear implant group on measures of syllable stress identification and intonation identification. The intonation identification task involved listening to a recording and distinguishing statement and question intonation. The cochlear implant group had a mean accuracy of 42.5%, which was significantly lower than that for the hearing aid group with profound hearing loss (80.82%), and for the hearing aid group with severe hearing loss (98.99%). The results show a clear advantage for the hearing aid users in perceiving intonation. Caution may be required in interpreting these results as eight of the ten cochlear implant users had been implanted at a relatively late age of over 6 years old. Previous research has shown that early implantation results in significantly better speech perception and language measures (Goswami & Johnson, 2010; Harrison, Gordon, & Mount, 2005); therefore, age may also have had an effect in this study.

A study of 26 child cochlear implant users (Peng, Tomblin, & Turner, 2008) also found that participants had difficulties with the production and perception of intonation. The children in this study ranged from 7 to 21 years old and the age at implantation ranged from 1.48 to 6.34 years. The normal hearing control group was age-matched to the cochlear implant group. The cochlear implant group was found to score significantly lower on both production and perception intonation tasks when compared to participants with normal hearing. The intonation perception task was identification of question or statement intonation. The cochlear implant group had an average accuracy of 70.13%, which was significantly lower than that for the control group (97.11%). The production and perception task results were moderately correlated, and there was also considerable intersubject variability. Because of the cross-sectional nature of the participants, factors such as age at implantation and length of cochlear implant experience may have influenced the results. However, the study suggests that in general, cochlear implant users do have difficulty in perceiving question/statement intonation, including users who were implanted at an early age (below 1.5 years). All participants used the Nucleus 22 or 24 cochlear implant with either the SPEAK or ACE processing strategy. Those with SPEAK processing performed better on the identification task than those with ACE processing. However, this was confounded by the fact that the SPEAK users had a significantly longer period of experience with the implant.

1.5.2 Electric and Acoustic Hearing Combined

In order to improve perception of suprasegmental features, two options have been proposed to preserve acoustic hearing in cochlear implant users. One option is the use of a fully inserted cochlear implant in one ear and a hearing aid on the other ear. The other option is the use of an electro-acoustic cochlear implant (EAS), whereby a partial array is inserted so that higher frequencies are conveyed by the cochlear implant and lower frequencies are conveyed by a hearing aid. Both of these options require some residual hearing in the low frequencies in at least one ear (Talbot & Hartley, 2008).

Most et al. (2011) studied the perception of suprasegmental features by adults with a unilateral cochlear implant and a hearing aid on the opposite ear. Participants were required to discriminate between statement and question intonation under two conditions. One condition was to listen through the cochlear implant only and the other was to listen with the cochlear implant and hearing aid together (i.e., bimodal condition). They found the bimodal condition resulted in a 75% accuracy rate, which was significantly higher than that in the cochlear implant only condition (55.07%). Although the overall group average was significant, it should be noted that individual results varied considerably. Some participants benefitted greatly from the hearing aid, others benefitted only somewhat, and some did not receive any benefit. An important reason for this variation may have been the different degrees of residual hearing in the non-implanted ear of the participants. A significant negative correlation between perception of suprasegmental features and the pure tone average in the non-implanted ear was found. Overall, the results suggest a benefit of the bimodal condition in perceiving intonation, although the extent of benefit was dependant on the degree of residual hearing in the non-implanted ear.

1.6 Simulated Hearing Loss

Simulated hearing loss has been used in several studies as a method to isolate specific factors associated with cochlear hearing loss. Baer and Moore (1993, 1994) and Nejime and Moore (1997) advocate the use of simulated hearing loss on the grounds that it enables researchers to investigate one specific aspect of hearing impairment without the confounding effects of another. For example, Baer and Moore (1993) used simulated hearing loss to isolate the effects of reduced frequency selectivity from other factors such as reduced audibility, reduced dynamic range, and reduced temporal resolution. Another advantage of using simulated hearing loss is that participants with normal hearing can be assumed to have a similar perceptual experience. In contrast, when individuals with hearing loss are used in a perceptual study, different past perceptual experiences may result from differences in the exact nature of cochlear impairment and/or processing abilities (Adams & Moore, 2009; Baer & Moore, 1993). There is, however, a limitation to studies of simulated cochlear hearing loss in that not all aspects of cochlear impairment may be simulated. Therefore, simulation of cochlear hearing loss cannot be said to yield findings reflective of true cochlear hearing loss, but rather it provides information about certain aspects of cochlear hearing loss.

1.6.1 Cochlear Hearing Loss

Several methods have been used to simulate hearing loss and cochlear implant processing. Kumar and Yathiraj (2009) used band-pass filters to simulate three different types of hearing loss configurations. This method was effective in providing information about the effect of reduced audibility at certain frequencies on the perception of specific classes of phonemes. Low-pass filtering (with cut-off frequencies of 300-600Hz) has also been used in several studies to preserve prosodic meaning while discarding segmental meaning (Munro, 1995; Nazzi, Bertoncini, &

Mehler, 1998; Xu & Mok, 2012b). Xu and Mok (2012b) used low-pass filters with a cut-off frequency of 150 to 300 Hz (depending on the speaker's fundamental frequency) to examine perception of question/statement intonation in Mandarin and Cantonese. With this cut-off frequency, participants could not understand the sentences; however, intonation identification accuracy rate was still approximately 85% in Cantonese and 67% in Mandarin.

In order to investigate impaired frequency selectivity, Baer and Moore (1993, 1994) used spectral smearing. This method involved manipulating the speech signal to resemble excitation patterns evoked in a cochlear with auditory filters broadened by factors of 3 and 6 when compared with normal auditory filters. This technique allowed the researchers to separate out effects of impaired frequency selectivity from other cochlear deficits. Baer and Moore (1993) found that spectral smearing adversely affected the detection of speech in noise but did not have a significant effect on detection of speech in quiet. In a follow-up study, results were found to be consistent for the detection of speech in the presence of interfering speech (Baer & Moore, 1994). It should be noted that the methods used in these studies for spectral smearing replicated broader auditory filters, but did not replicate reduced time coding in the cochlea.

Simulated hearing loss has also been used to investigate loudness recruitment associated with cochlear impairment (Duchnowski & Zurek, 1995; Moore & Glasberg, 1993). In order to replicate the impaired cochlea's loss in compressive nonlinearity which results in loudness recruitment, Moore and Glasberg (1993) employed filters to separate a speech signal into a number of frequency bands and then applied an expansive non-linearity differently to each band. A high frequency hearing loss was also simulated by including more envelope expansion at the high frequencies. Adams and Moore (2009) used a similar method to examine the effect of noise on speech rate judgment. Speech stimuli were filtered into four bands, and each band was attenuated by an amount corresponding to threshold elevation due to hearing loss. Loudness recruitment simulation was added to this simulation of elevated thresholds.

Nejime and Moore (1997) combined three aspects of cochlear impairment in their hearing loss simulation. Reduced frequency selectivity was simulated according to the spectral smearing method of Baer and Moore (1993, 1994) and threshold elevation and loudness recruitment was then simulated according to the method of Moore and Glasberg (1993). Intelligibility of speech in noise was found to be impaired for conditions simulating both a moderate flat hearing loss and a moderate to severe high frequency sloping loss. Amplification was then applied to the simulated hearing loss signals to replicate hearing aid use. Under this condition it was found that intelligibility of speech in noise was reduced compared to that of a normal control group. The results were interpreted as evidence that hearing aids cannot compensate for all aspects of cochlear impairment.

1.6.2 Temporal Jittering

In a different approach, Pichora-Fuller et al. (2007) used temporal jittering to simulate loss of neural synchrony which may occur in aging auditory systems. In contrast to the spectral smearing technique of Baer and Moore (1993, 1994), the temporal jittering method was designed to disrupt periodicity cues but preserve spectral information. This results in a loss of TFS cues. Their aim was to test the hypothesis that a disruption to periodicity cues, as would occur with a loss of neural synchrony, would result in a decrease in speech perception. Results of the study confirmed that word recognition in noise was significantly reduced when temporal jittering was applied to frequencies below 1.2 kHz. Pichora-Fuller et. al. (2007) applied the jitter only to frequencies below 1.2 kHz in order to replicate the loss in neural synchrony only in the area of the cochlea where phase-locking occurs. A follow-up study found that this was also true for the 1.2-7 kHz range (MacDonald, Pichora-Fuller, & Schneider, 2010). The results were compared to distortion through spectral smearing. It was found that while both temporal jittering and spectral smearing contributed to speech recognition difficulties at higher frequencies, spectral smearing at low frequencies did not result in the same observed difficulties.

Since the ability to process TFS may impact the perception of intonation, the current study used temporal jittering as a way to simulate this aspect of cochlear hearing loss. A loss in neural synchrony and decreased audibility are problems likely to co-occur in individuals with sensorineural hearing loss (Dillon, 2012; Moore & Carlyon, 2005). Therefore, different hearing loss configurations combined with a loss in neural synchrony will be simulated by including tokens with high-pass or low-pass filters combined with temporal jittering.

1.7 Research Questions and Importance

The present study aimed to investigate two main research questions. Firstly, the impact of a loss of temporal fine structure on speech intonation was examined. This was done by using temporally-jittered and vocoder-processed speech signals in an acoustic analysis and a perceptual experiment. Secondly, the universal nature of the impact of the hearing loss simulations on the perception of intonation was examined. This was done by comparing the acoustic and perceptual findings for a non-tonal language (English) and two tonal languages (Mandarin and Taiwanese).

Findings from this study may contribute to several areas of research. Firstly, intonation will be added to a number of other aspects of listening which have been studied in regard to a loss in TFS. Secondly, the results will add to the understanding of how hearing loss may impact the pitch-related aspects of speech perception in typologically different languages. Clinically, this information may be used when considering the amplification needs of patient from differing linguistic backgrounds.

1.8 Hypotheses and Rationales

This study involves the comparison of declarative questions and declarative statements in tonal and non-tonal languages through the use of an acoustic analysis and a perceptual experiment. There are three main hypotheses proposed in this study.

Hypothesis 1: Based on the general findings from previous studies of question/statement intonation, it is hypothesised that declarative statements have a falling pitch and declarative questions have a rising pitch toward the end of the sentence in both tonal and non-tonal languages.

Hypothesis 2: As pitch has been shown in the literature to be the primary feature of speech intonation in both stress and tone languages, and pitch perception has been found to be compromised in individuals with hearing loss, it is hypothesised that listeners of both tonal and non-tonal languages will perform significantly worse on tokens with degraded temporal fine structure (i.e., temporally jittered and vocoder-processed) when compared to non-processed tokens.

Hypothesis 3: Since F0 height and contour play a role in tonal distinction, and this may interact with speech intonation, it is hypothesised that tonal language listeners will perform significantly worse than non-tonal language listeners in the perception of intonation on signals processed with simulated hearing loss.

2 Method

The current study investigated the effect of simulated hearing loss on the acoustics and perception of speech intonation in one non-tonal language (English) and two tonal languages (Mandarin and Taiwanese). The study consisted of two stages. The first stage involved the recording of the declarative questions and declarative statements produced in English, Mandarin, and Taiwanese and an acoustic analysis of these recordings. The second stage involved a signal manipulation process to simulate different types of hearing loss, followed by an acoustic analysis of the processed signals and a perceptual experiment to evaluate the impact of these hearing loss simulations on the listeners' ability to detect declarative questions.

2.1 Participants

Individuals whose native language was New Zealand English, Mandarin, or Taiwanese were recruited from the campus of the University of Canterbury (Christchurch, New Zealand). Subject inclusion criteria were native adult New Zealand English, Mandarin, or Taiwanese speakers without any history of speech or hearing problem. For each target language, two females and two males were included as speakers and five female and five male were included as listeners. Participants consisted of a total of eight speakers, including four native English speakers and four native speakers of Mandarin and Taiwanese, and a total of 20 listeners, including 10 native English speakers and 10 native speakers of Mandarin and Taiwanese. All native speakers of Mandarin and Taiwanese were born in Taiwan. The speakers' age ranged from 23 to 47 years (Mean = 31.9 years, SD = 8.5). Results from a t test revealed that the two speaker groups (English and Mandarin-Taiwanese) showed no significant age difference ($t = 1.259$, $df = 6$, $p = 0.255$). The listeners' age ranged from 19 to 55 years (Mean = 36.4 Years, SD = 11.4). The two listener groups (English and Mandarin-Taiwanese) also showed no significant difference in age (t = -1.637, df = 18, p = 0.119).

For the Mandarin-Taiwanese native speakers serving as speakers in this study, the age of starting to learn English as a second language ranged from six to 12 years old and the duration of stay in New Zealand ranged from two to 18 years. For the Mandarin-Taiwanese native speakers serving as listeners in this study, the age of starting to learn English as a second language ranged from eight to 12 years old. Before the experiment, all listeners had their hearing screened to 30dBHL at four frequencies (0.5, 1, 2, 4 kHz) and were found to have hearing thresholds within this screening limit. All participants were also informed of the purpose and the procedure of the study (Appendices 3 and 4) and signed the consent forms (Appendix 5), which had all been approved by the institutional human subject ethics review board.

2.2 Participants' Tasks

The speakers' task was to read a list of 28 phrases as either a question or a statement, with native English speakers reading in English only and native Mandarin and Taiwanese speakers in English, Mandarin, and Taiwanese in separate sessions. The phrases were semantically equivalent across languages (see Appendix 6). The stress or tone type for the last syllable in each phrase was shown in Appendix 7. Each phrase could potentially be used as a question or as a statement. When reading as questions, speakers were instructed to imagine they were checking information. Results from a one-way Analysis of Variance (ANOVA) revealed that the number of syllables did not significantly vary by language [F(2, 81) = 0.929, p = 0.399], with the number of syllables ranging from two to nine syllables (Mean = 3.54 syllables, SD = 1.27). In the participant groups serving as speakers, each native English speaker was asked to produce a total of 280 utterances (28 phrases X 2 sentence types X 5 trials) and each native speaker of Mandarin and Taiwanese a total of 840 utterances (28 phrases X 2 sentence types X 5 trials X 3 languages). As for the participants serving as listeners in this study, their task was to listen to a total of 112 different utterances (8 phrases X 7 signal types X 2 sentence types) in their own native language and judge one sentence at a time whether it was a question or a statement.

2.3 Instrumentation

The microphone recording system included a digital voice recorder (Sony PCM-M10), a headset condenser microphone (AKG C420), and a mixer (Eurorack MX602A). The sampling rate was set at 96 kHz, with a 24-bit resolution. The acoustic signals recorded through this system were directly digitised and saved as "WAV" files. The digital files were later processed and analysed using a computer. Sentences segmented out from the original recordings were further processed with six types of simulated hearing loss.

For signal manipulation, a locally developed algorithm written in MATLAB 7 (The Mathworks, Inc.) was used to perform high pass (cut-off frequency = 1.2 kHz), loss pass (cut-off frequency = 1.2 kHz), temporal jittering, envelop-vocoder processing (Band-pass filter: 0.05-6 kHz; Number of vocoder bands = 12; Envelope cut-off frequency = 0.05 Hz), and normalisation. For signal analysis, the TF32 acoustic analysis software (Milenkovic, 1987) was used. For signal playback during the perceptual experiment, a pair of headphones (Philips SHN9500) was connected to a laptop. A locally developed algorithm written in Visual C++ was used to present the stimuli in a predetermined random sequence and record the listener's response onto a text file convertible to an Excel spreadsheet. The SPSS statistical software (Version 19) was used for statistical analysis.

2.4 Signal Manipulation

The signals to be processed with various types of hearing loss simulation were selected from the original signals, with one declarative question and one declarative statement chosen for each of the 28 sentences in each of the four speakers. Among the five trials each speaker produced for each of the 28 sentences in each of the two sentence types (i.e., question and statement), the utterance showing the highest Maximum F0 in the last syllable (MaxF0-last) was chosen for questions and that showing the lowest MaxF0-last was chosen for statements. This selection was to maximise the difference between questions and statements in the signals to be manipulated. The rationale of this selection was based on the observation in previous studies that declarative questions tended to show a raised pitch toward the end of a sentence.

Signals were processed using temporal jittering in order to simulate a loss in neural synchrony. As phase-locking becomes less precise as a result of hair cell loss or damage that accompanies hearing loss, the distribution of interspike intervals may become degraded. Pichora-Fuller et al. (2007) simulated such a reduction in phase-locking sensitivity by introducing random normally distributed delays to the signal. In their study, the jittered waveform $y(t)$ was a version of the original waveform $x(t)$ whereby a time delay σ was introduced so that $y(t) = x[t - \sigma(t)]$. The two variables in the jitter were the distribution of the delay over time and the rate the delays change over time. To achieve random normal distribution, Pichora-Fuller et al. (2007) made the time delay proportional to a low-pass noise amplitude and the standard deviation of the delays equal to the RMS amplitude of the noise. The high frequency cut-off of the noise was varied in order to vary the rate of the change in delay over time. The current study followed this method except that in order to achieve random normal distribution, a random number was generated by a computer program. Three levels of temporal jittering were specified, namely, 1.4, 2.7, and 4. These were the numbers used in the MATLAB jitter function representing the standard deviation in milliseconds of the distributions from which the random numbers were drawn. Further measurement found that the actual standard deviations of distributions resulting from the MATLAB function were 0.4, 0.8, and 1.2 milliseconds respectively. The temporal jittering simulation with a jitter standard deviation of 4 was also combined with high and low pass filters (cut off frequency = 1.2 kHz) to simulate different configurations of hearing loss.

In addition, hearing through cochlear implants was simulated through a vocoder processing algorithm to allow for a perceptual comparison between the loss of TFS through temporal jittering and that through cochlear implant processing strategies. The use of a vocoder to simulate the processing of a cochlear implant is a widely used research methodology (e.g. Shannon et al., 1995; Van Tasell, Soli, Kirby, & Widin, 1987; Won et al., 2012). The current study used a commonly used procedure in the design of the vocoder, similar to that as outlined by Stone, Fullgrabe, and Moore (2008). Specifically in this study, the signal was divided into 12 channels ranging from 0.5 kHz to 6 kHz, with equal-octave widths. The envelope of each channel was extracted by half-wave rectification and low-pass filtering. This envelope was used to modulate a band of white noise. The noise was band-pass filtered according to each channel's frequency range before and after modulation. The resulting signals from the 12 channels were summed.

In summary, the six types of hearing loss simulation employed in this study included signals superimposed with a temporal jittering of 1.4 ("Jit140"), 2.7 ("Jit270"), and 4 ("Jit400") respectively, those being high passed at 1.2 kHz and superimposed with a temporal jittering of 4 (“HighPass-Jit400”), those being low passed at 1.2 kHz and superimposed with a temporal jittering of 4 (“LowPass-Jit400”), and those being processed with a simulation of a 12-electrode cochlear implant (“Vocoder”). The seven types of signal conditions, including the original signals and the six types of modified signals, were normalised to the same intensity level for the perceptual test.

2.5 Procedure

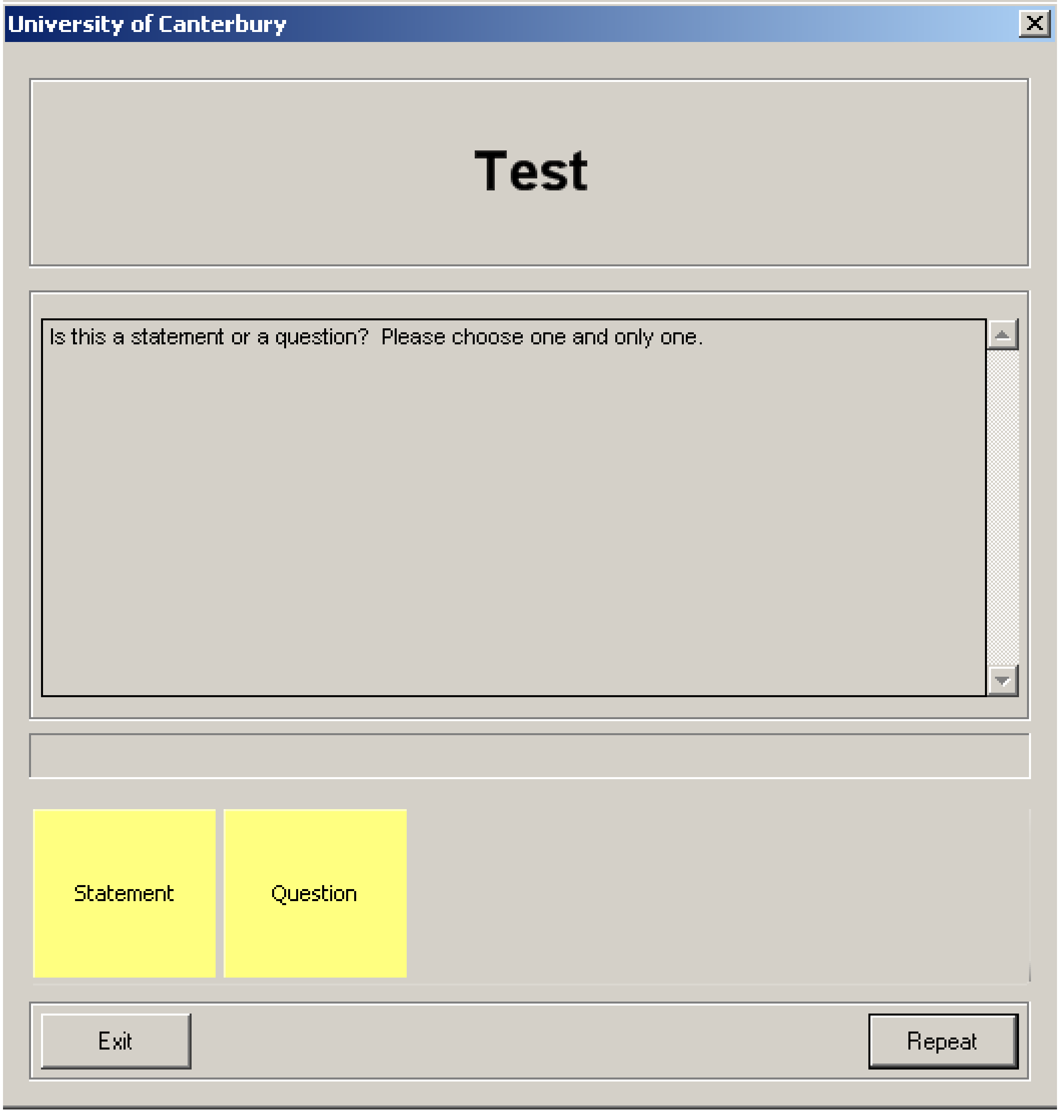
The procedures used in conducting the speech recordings and the perceptual experiment are described separately as follows.

2.5.1 Recording of Sentences

Each participant serving as a speaker was seated in a sound treated booth. The headset microphone connected to the digital recorder was placed off axis approximately 5 cm away from the participant’s lips. After the microphone was secured in place, the participant was asked to perform the speaker’s task while the experimenters activated the recording system. The 28 phrases for each language were listed in five different pre-determined random orders. Each of the five lists was randomly repeated in separate sessions, once with each phrase read as a declarative statement and once as a declarative question. The native speakers of Mandarin and Taiwanese were asked to complete the Mandarin reading sessions, followed in order by the Taiwanese and English reading sessions. All speakers were allowed to take a break and have a sip of water any time after finishing one list. The whole recording session for each speaker was completed within one hour.

2.5.2 Perceptual Experiment

Participants serving as listeners were seated in a quiet room. After the headphones were placed over the listener's ears, the listener was asked to perform the listener's task. For each of the seven signal conditions to be tested in this study, eight statements and eight questions were randomly chosen from the four speakers and the 28 phrases, with each utterance selected only once. Two label boxes ("statement" and "question") were shown on the computer screen for the listener to select by clicking the mouse (see Figure 1). The listeners were allowed to repeat each trial if needed. The listeners were given a short break half way through the experiment.



Is this a statement or a question? Please choose one and only one.

Figure 1. The computer interface for the perceptual experiment.

2.6 Acoustic Analysis

From each of the sentences recorded, six acoustic measures were obtained. From each of the selected vowels embedded in each of the sentences processed with the seven signal conditions, two acoustic measures were obtained.

2.6.1 Original Signals Only

The original recorded sentences were used to extract six acoustic measures, including the maximum F0 of the last syllable (MaxF0-last), the maximum F0 of the whole sentence excluding the last syllable (MaxF0-rest), the maximum RMS value of the last syllable (MaxRMS-last), the maximum RMS value of the whole sentence excluding the last syllable (MaxRMS-rest), the ratio between MaxF0-last and MaxF0-rest (F0ratio), and the ratio between MaxRMS-last and MaxRMS-rest (RMSratio).

To derive these measures, each sentence was first segmented out from a time waveform display of an original sounds file on a computer screen. After selecting the boundaries of the target sentence from the displayed signals, the investigator examined the spectrographic displays of the signals and listened to the auditory playback of the selected section to verify the selection. The selected segment was exported as a separate file. Each sentence was later displayed to allow for cursor selection of the last syllable or the segment excluding the last syllable. As both Mandarin and Taiwanese are both monosyllabic languages, the maximum F0 of the last syllable of a sentence was the same as that of the last word in these two languages. In English, the maximum F0 of the last syllable of a sentence was that of the last syllable in the last word.

The “pitch trace” and “RMS trace” modules of TF32 were used to allow for automatic derivation of the measures of the maximum F0 or RMS values of the selected segment. Specifically, the MaxF0-last and MaxRMS-last values were measured by placing the cursor between the starting and ending point of the last syllable and obtaining the pitch and RMS trace readings. The maximum F0 or RMS of the remaining syllables in the sentence (MaxF0-rest and MaxRMS-rest) were measured by placing the cursor between the starting and ending point of the rest of the sentence's syllables and obtaining the pitch and RMS trace readings. After the readings from the selected segments were copied onto the spreadsheet, measures of F0ratio and RMSratio were calculated by dividing the MaxF0-last (or MaxRMS-last) values by the MaxF0-rest (or MaxRMS-rest) values.

2.6.2 Seven Types of Signals

The original and the six processed ("Jit140", "Jit270", "Jit400", "HighPass-Jit400", "LowPass-Jit400", and "Vocoder") sound files were used to extract two vowel-based acoustic measures: percent jitter (%Jit) and spread of spectral energy around the dominant frequency (MomentCOV). To yield these vowel-based measures, three corner vowels, including /i/, /a/, and /u/, embedded in the end of a sentence were segmented out from three chosen sentences [English: Sentences 3 ("tea": /i/), 12 ("too": /u/), and 25 ("nana": /a/); Mandarin: Sentences 8 ("ba": /a/), 14 ("qi": /i/, and 25 ("zu": /u/); Taiwanese: Sentences 16 ("si": /i/), 22 ("a": /a/), and 24 ("yu": /u/)] expressed in both statement and questions. A 50-millisecond segment from the middle portion of the embedded vowel was cursor selected and exported as a separate file.

The vowel files were submitted to the batch processing function in the "Jitter" module to extract the measure of %Jit. As jitter refers to cycle-to-cycle frequency variation, a higher %Jit would indicate deterioration of voice quality. For derivation of MomentCOV, the Moment One (i.e., mean frequency) and Moment Two (i.e., standard deviation of the frequency) values were obtained through the moment analysis function in the “Spec” module. The measure of MomentCOV was calculated by dividing Moment Two by Moment One. A higher MomentCOV would indicate greater fluctuation of the dominant pitch.

2.7 Statistical Analysis

The six acoustic measures obtained from the original signals, the two vowel-based acoustic measures obtained from seven types of signals (“Original”, Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”), and the perceptual data were submitted to separate statistical analyses. The significant level was set at 0.05.

2.7.1 Acoustic Data

Data obtained from the original signals were submitted to statistical tests to determine how declarative questions and statements differed on six acoustic measures, including MaxF0-last, MaxF0-rest, MaxRMS-last, MaxRMS-rest, F0ratio, and RMSratio. Data obtained from the original and the processed signals were submitted to statistical tests to evaluate the impact of different types of hearing loss simulation on two vowel-based measures, including %Jit and MomentCOV.

2.7.1.1 Original Signals Only

The average values for each of the six measures obtained from each of the two sentence types (i.e., question and statement) were calculated for each speaker in each of the four language groups. The four language groups included the native English, Mandarin, and Taiwanese productions and the non-native English productions obtained from the native speakers of Mandarin and Taiwanese. To identify which acoustic measure may differentiate between questions and statements, a two-way (2 sentence types X 4 language groups) Multivariate Mixed Model ANOVA (MANOVA) was conducted, with sentence type (i.e., question and statement) treated as a within-group factor and language (i.e., English, Mandarin, Taiwanese, and non-native English) as a between-group factor. Follow-up univariate ANOVAs were conducted if the MANOVA results showed any significant effect.

2.7.1.2 Seven Types of Signals

Two acoustic measures, %Jit and MomentCOV, were obtained from a total of 504 vowel tokens (7 signal conditions X 3 languages X 2 sentence types X 3 vowel types X 4 speakers). These vowels were segmented from sentences produced by the native speakers for each of the three languages under study, namely, English, Mandarin, and Taiwanese. To evaluate the effect of simulated hearing loss on these measures, which were associated with pitch distortion and thus had a potential impact on the question/statement identification, a four-way (7 signal conditions X 3 languages X 2 sentence types X 3 vowels) Mixed Model MANOVA was conducted, with vowel (/i/, /a/, and /u/), sentence type (question and statement), and signal condition (“Original”, Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) treated as within-group factors and language (English, Mandarin, and Taiwanese) as a between-group factor.

2.7.2 Perceptual Data

The percentage of correct responses in identifying a sentence as a declarative question or statement was calculated for each listener and defined as the identification accuracy rate. A four-way (2 sentence types X 7 signal conditions X 3 languages X 2 listener genders) Mixed Model ANOVA, with sentence type (question and statement) and signal condition (“Original”, Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) treated as within-group factors and language (English, Mandarin, and Taiwanese) and listener gender (female and male) as between-group factors, was conducted mainly to determine how different types of hearing loss simulation affected the accuracy rate in identifying a question or a statement and whether the effect might vary by language, gender, or sentence type.

As mentioned in the literature review, previous studies of question/statement intonation have consistently found that identification of statements is more accurate than identification of questions. This includes studies of normally hearing individuals (Ma et al., 2011; Xu & Mok, 2012a; Yuan, 2006) and studies of individuals with hearing loss and cochlear implants (Most et al., 2011). Ma et al. (2011) argue that that this is evidence of a bias towards perception of statements, meaning that the default choice is a statement, and participants will not identify question intonation unless they clearly perceive the acoustic markers. To eliminate such potential bias introduced by the different strategies listeners used in performing the perceptual task, a “proportion of the area” (PA) measure for the dichotomous identification task was derived (Green & Swets, 1966). In this study, the PA measure is defined as an average of hit rate (i.e., the number of correct statement detection divided by the total number of statements presented) and correct rejection rate (i.e., the number of correct question detection divided by the number of questions presented).

A three-way (3 languages X 2 genders X 7 signal conditions) Mixed Model ANOVA, with language (English, Mandarin, and Taiwanese) and listener gender as the between-groups factors and signal condition as the within-groups factor, was conducted on PA mainly to determine whether signals processed with simulated hearing loss would result in a change of the performance in the question/statement detection task.

2.8 Reliability

As all measurements were extracted automatically through the software, a 100% measure-remeaure reliability was ensured. However, measurement variability may arise from variations in segmentation, especially in signals that have been highly distorted and thus susceptible to errors in computerized pitch extraction. To gauge the stability of the measurement of %Jit and MomentCOV, a 500-Hz pure tone was generated, through a computer synthesizer, with a time duration of 500 ms. The synthesized pure tone was further processed with the “Vocoder” or manipulated with 12 different levels of temporal jittering, including jitter standard deviations set at the levels of 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4 (“Jit140”), 1.64, 2.7 (“Jit270”), 3.4, 4 (“Jit400”), and 7.2 respectively. The original sine waves, the pure tones that have been processed with “Vocoder”, and those superimposed with different levels of temporal jittering were each measured six times, each time with the same predetermined scheme of time segmentation. All of the 14 signals were segmented in the same way, with each individual segment starting at the point that is 0, 50, 100, 150, 200, or 250 milliseconds after the onset of the signal and the corresponding segment having a duration of 500, 400, 300, 200, 100, and 50 milliseconds.

Figure 2 shows the mean %Jit measures of the calibrating signal with different levels of temporal jittering respectively. The mean %Jit values for the pure tones undergoing 12 different levels of temporal jittering ranged from 1.01 to 12.83. The pure tone processed with “Vocoder” yielded a mean %Jit value of 9.73. The coefficient of variation (COV) of the %Jit values measured was also calculated, for each of the 14 signals, by dividing the standard error of the %Jit values by the mean

%Jit value. The COV of the %Jit measure for signals at the 12 temporal jittering levels was found to range from 2.09% to 4.48% (Mean = 4.15%, SD = 2.9%). As for the pure tones processed with the "Vocoder", the measurement variability for %Jit was found to be relatively low, with a COV value of 2.2.

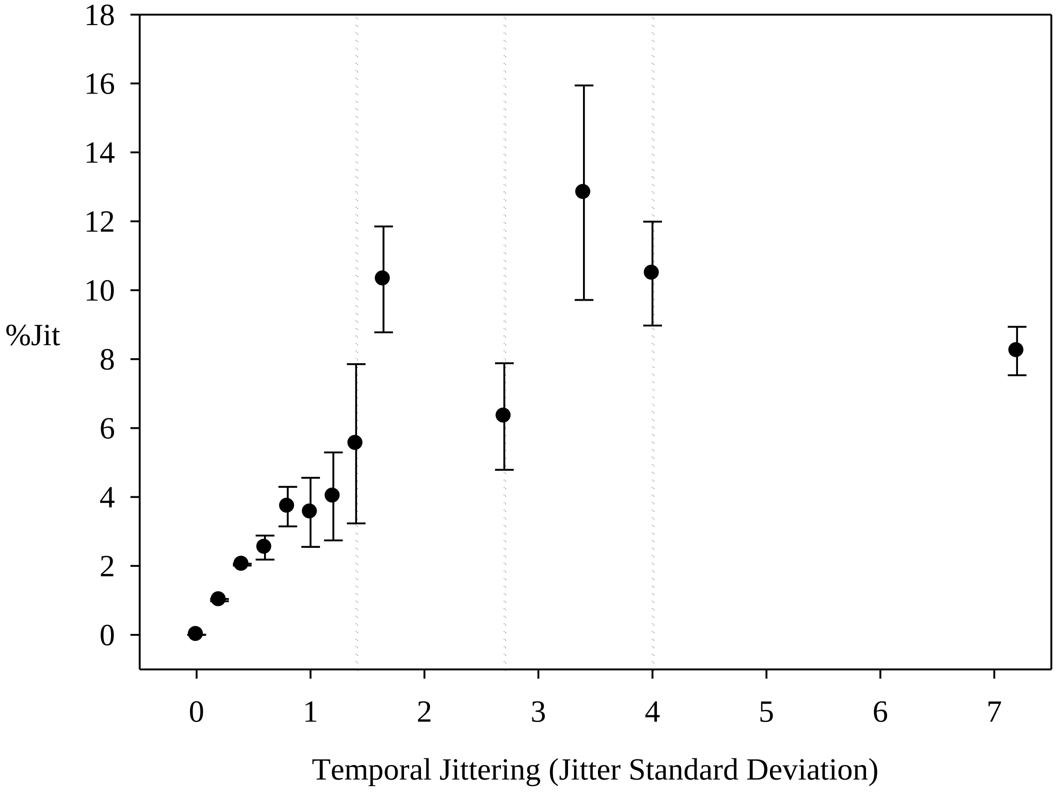


Figure 2. The means and standard errors (in error bars) of the measures of percent jitter (%Jit) for a 500 Hz pure tone processed with different levels of temporal jittering. The dotted lines indicate the level of temporal jittering used in this study for simulating hearing loss, including the level with a jitter standard deviation value of 1.4 ("Jit140"), 2.7 ("Jit270"), and 4 ("Jit400").

Figure 3 shows the MomentCOV measures of the calibrating signal processed through different levels of temporal jittering. The mean MomentCOV values for the pure tones undergoing 12 different levels of temporal jittering ranged from 0.06 to 0.59 (see Figure 3). The pure tone processed with “Vocoder” yielded a mean MomentCOV value of 1.08. The MomentCOV value appears to increase linearly with the increase of temporal jittering. However, when the temporal jittering value exceeds 1, the average MomentCOV value starts to level off and the variation of the measures greatly increases (see Figure 3). The coefficient of variation (COV) of the MomentCOV values measured was calculated, for each of the 14 signals, by dividing the standard error of the MomentCOV values by the mean MomentCOV value. The COV of the MomentCOV measure for signals at the 12 temporal jittering levels was found to range from 2.09% to 4.48% (Mean = 4.15%, SD = 2.9%). As shown in Figure 3, the measurement variation of MomentCOV starts to increase greatly with temporal jittering value exceeding 1. As for the pure tones processed with the “Vocoder”, the measurement variability for MomentCOV was found to be 10.9%, which was much higher than signals superimposed with temporal jittering.

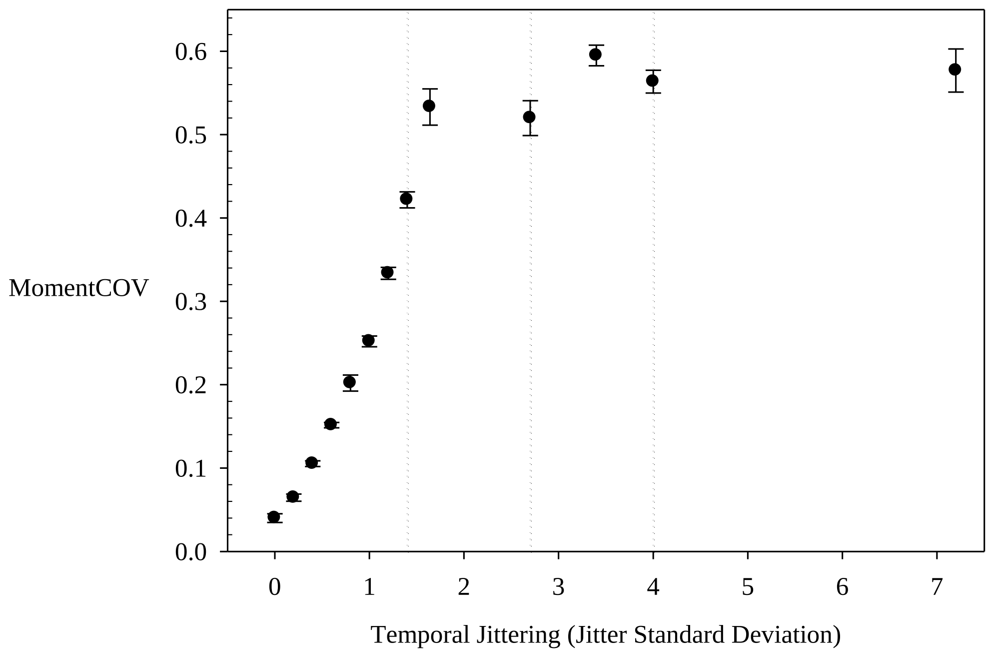


Figure 3. The means and standard errors (in error bars) of the measures of MomentCOV for a 500 Hz pure tone processed with different levels of temporal jittering. The dotted lines indicate the level of temporal jittering used in this study for simulating hearing loss, including the level with a jitter standard deviation value of 1.4 ("Jit140"), 2.7 ("Jit270"), and 4 ("Jit400").

In summary, the COV of both %Jit and MomentCOV was below 50%, indicating that the measurement variability due to variation in signal segmentation could be considered adequate. A linear relationship between temporal jittering and %Jit and MomentCOV was observed for pure tones with temporal jittering below 1. However, the measurement of %Jit was more variable than that of MomentCOV, especially at the higher level of temporal jittering. In addition, the pure tones processed with the "Vocoder" modification yielded a higher mean %Jit than those superimposed with temporal jittering below 1.64 and a higher mean MomentCOV value than all of the pure tones processed with temporal jittering at the levels tested (i.e., 1, 2.7, and 4) in this study.

3 Results

Statistical results are organised in this chapter to address the two main objectives of the current investigation, that is, (1) to identify the acoustic differences between declarative questions and declarative statements across three languages, and (2) to determine the impact of simulated hearing loss on the acoustic differences between declarative questions and statements and on the auditory-perceptual detection of declarative questions across three languages.

3.1 Acoustic Markers of Declarative Questions

Results from the two-way (2 sentence types X 4 language groups) Mixed Model MANOVA conducted on the six acoustic measures obtained from the original signals showed a significant sentence type effect [Pillai's Trace = 0.973, $F(6, 7) = 42.36$, $p < 0.001$, $\eta_p^2 = 0.97$] but no significant language group effect [Pillai's Trace = 1.278, $F(18, 27) = 1.113$, $p = 0.391$] or sentence type by language group interaction effect [Pillai's Trace = 1.249, $F(18, 27) = 1.07$, $p = 0.427$]. Follow-up univariate ANOVAs revealed a significant sentence type effect on RMSratio [$F(1, 12) = 95.05$, $p < 0.001$, $\eta_p^2 = 0.89$], MaxF0-last [$F(1, 12) = 67.005$, $p < 0.001$, $\eta_p^2 = 0.85$], F0ratio [$F(1, 12) = 26.786$, $p < 0.001$, $\eta_p^2 = 0.69$], MaxRMS-last [$F(1, 12) = 21.001$, $p = 0.001$, $\eta_p^2 = 0.64$], and MaxF0-rest [$F(1, 12) = 17.637$, $p = 0.001$, $\eta_p^2 = 0.60$] but no significant sentence type effect on MaxRMS-rest [$F(1, 12) = 1.17$, $p = 0.301$, $\eta_p^2 = 0.089$]. As shown in Table 3, the mean values of all of the acoustic measures were higher in declarative questions than in statements. The mean F0ratio was higher than one in declarative questions but lower than one in statements, indicating that F0 was generally elevated toward the end of a sentence above the level of the previous segment in declarative questions.

Table 3. Descriptive statistics for the six acoustic measures with data from native English, Mandarin, and Taiwanese and non-native English productions combined.

	Statement			Question		
Measure	n	Mean	SD	n	Mean	SD
MaxF0-last (in Hz)	16	175.92	15.87	16	263.70	24.11
MaxF0-rest (in Hz)	16	192.16	15.31	16	225.07	20.23
MaxRMS-last (in V)	16	1.10	0.08	16	1.38	0.10
MaxRMS-rest (in V)	16	1.52	0.13	16	1.58	0.12
F0ratio	16	0.92	0.01	16	1.21	0.06
RMSratio	16	0.78	0.03	16	0.92	0.02

With all four language groups combined, the average MaxF0-last for male speakers was 124 Hz (SD = 9.58; Ranged from 117.01 to 144.96 Hz) for statements and 190 Hz (SD = 36.74; Ranged from 146.14 to 237.42 Hz) for questions. The average MaxF0-last for female speakers was 228 Hz (SD = 29.44; Ranged from 188.93 to 264.19 Hz) for statements and 338 Hz (SD = 69.97; Ranged from 266.12 to 434.73 Hz) for questions.

To further investigate whether the F0ratio and RMSratio of a statement can be used to predict the F0ratio of its corresponding declarative question in the four types of language productions, including Taiwanese, Mandarin, English, and the non-native English productions obtained from the native Taiwanese-and-Mandarin speakers, a stepwise multiple regression, using the F0ratio and RMSratio of a statement (SF0ratio and SRMSratio) as the independent variables and the F0ratio of the corresponding declarative question (QF0ratio), was conducted on each of the four language data sets. Results from this multiple regression analysis for each of the six language data sets and detailed discussions are included in Appendix 8.

3.2 Effects of Simulated Hearing Loss

The effects of signal condition (“Original”, “Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) on the acoustic measures and the perceptual scores are shown separately as follows.

3.2.1 Acoustic Effects

Results from the four-way MANOVA conducted on the two vowel-based measures, including %Jit and MomentCOV, showed significant signal condition effect, sentence type effect, vowel by signal condition interaction effect, and vowel by signal condition by language interaction effect (see Table 4).

Table 4. Results from the four-way (7 signal conditions X 3 languages X 2 sentence types X 3 vowels) Mixed Model MANOVA conducted on the two vowel-based measures (%Jit and MomentCOV).

Factor	Pillai's Trace	F	Hypothesis df	Error df	p	η_p^2
Signal Condition (C)	1.539	30.008	12	108	< 0.001*	0.769
Sentence Type (T)	0.741	11.423	2	8	0.005*	0.741
Language (L)	0.246	0.630	4	18	0.647	0.123
Vowel (V)	0.268	1.390	4	36	0.257	0.134
C*T	0.244	1.248	12	108	0.260	0.122
C*L	0.336	0.909	24	108	0.590	0.168
C*V	0.559	3.488	24	216	< 0.001*	0.279
T*L	0.289	0.760	4	18	0.565	0.144
T*V	0.308	1.639	4	36	0.186	0.154
L*V	0.631	2.074	8	36	0.065	0.315
C*T*L	0.196	0.489	24	108	0.977	0.098
C*T*V	0.195	0.974	24	216	0.501	0.098
C*L*V	0.498	1.49	48	216	0.030*	0.249
T*L*V	0.364	1.001	8	36	0.452	0.182
C*T*L*V	0.382	1.063	48	216	0.375	0.191

*Significant at 0.05 level.

3.2.1.1 Percent Jitter

Follow-up univariate ANOVAs revealed a significant effect on %Jit only for signal condition [$F(6, 54) = 14.059$, $p < 0.001$, $\eta_p^2 = 0.61$]. Figure 4 shows the average %Jit values for each of the seven signal conditions with all languages, sentence types, and vowels combined. As shown in Figure 4, %Jit was highest in the “Vocoder”, “LowPass-Jit400”, and “Jit400” conditions, which were not significantly different from one another. The increase of temporal jittering resulted in an increase of %Jit, with the average %Jit being significantly higher in the “Jit400” condition than in the “Jit270” condition and significantly higher in the “Jit270” condition than in the “Jit140” condition. The average %Jit obtained from signals in the “Original” condition was not significantly different from that in either “Jit140” or “HighPass-Jit400” condition.

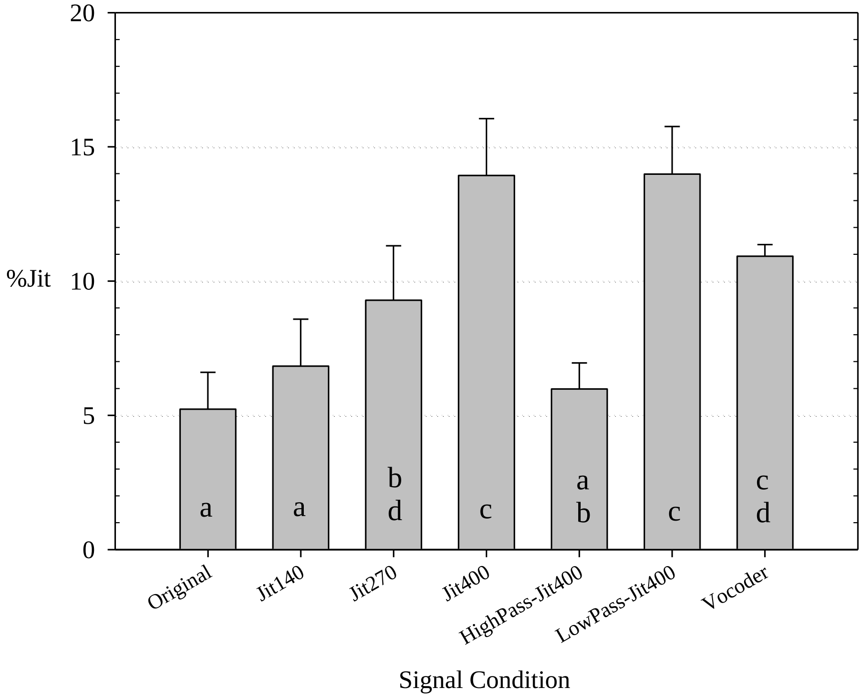


Figure 4. The average %Jit values for each of the seven signal conditions with all languages, sentence types, and vowels combined. Significantly different conditions were marked with different letters.

3.2.1.2 MomentCOV

As for MomentCOV, there was a significant signal condition effect [F(6, 54) = 102.962, $p < 0.001$, $\eta_p^2 = 0.92$], sentence type effect [F(1, 9) = 20.549, $p = 0.001$, $\eta_p^2 = 0.7$], vowel by signal condition interaction effect [F(12, 108) = 5.685, $p < 0.001$, $\eta_p^2 = 0.39$], vowel by signal condition by language interaction effect [F(24, 108) = 2.118, $p = 0.005$, $\eta_p^2 = 0.32$], and vowel by signal condition by language by sentence type interaction effect [F(24, 108) = 1.794, $p = 0.023$, $\eta_p^2 = 0.29$].

English. For English, a significant effect on MomentCOV was found only for the signal condition effect [F(6, 18) = 75.252, $p < 0.001$, $\eta_p^2 = 0.96$] and the vowel effect [F(2, 6) = 11.981, $p = 0.008$, $\eta_p^2 = 0.8$]. As shown in Figure 5, the “Vocoder” condition resulted in the highest MomentCOV. The “Original”, “HighPass-Jit400”, and “LowPass-Jit400” were not significantly different on the MomentCOV measure. The “Jit270” condition resulted in a significantly higher MomentCOV than “Jit140”, suggesting that a higher level of temporal jittering was associated with a higher MomentCOV. However, the “Jit400” condition was not significantly different from the “Original”, “Jit140”, and “Jit270” conditions. With all sentence types and signal conditions combined, the English vowel /a/ (Mean = 1.914, SD = 0.108) showed a significantly higher MomentCOV than the English /i/ (Mean = 1.852, SD = 0.163) and /u/ (Mean = 1.555, SD = 0.135).

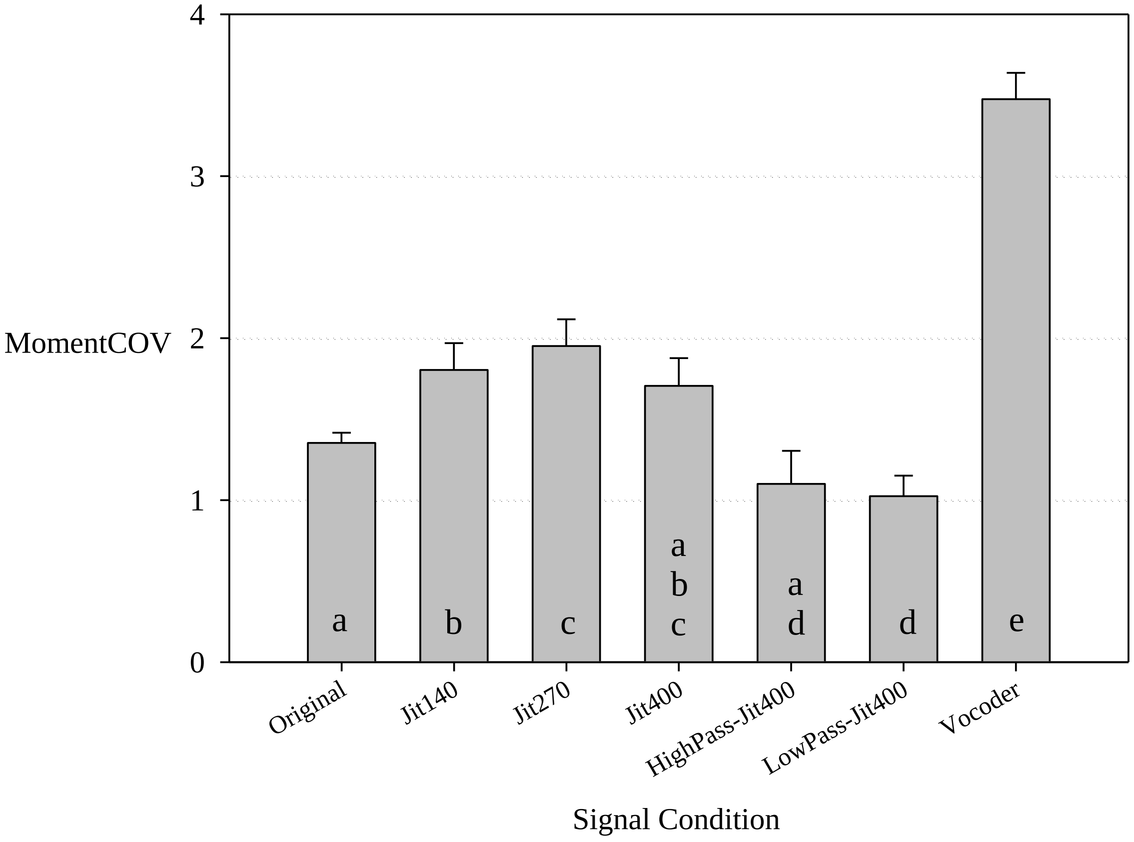


Figure 5. The average MomentCOV values obtained from native English productions in each of the seven signal conditions (“Original”, “Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) with all sentence types and vowels combined. Significantly different conditions were marked with different letters.

Mandarin. For Mandarin, a significant effect on MomentCOV was found only for the signal condition effect [$F(6, 18) = 17.077$, $p < 0.001$, $\eta_p^2 = 0.85$] and the vowel by signal condition interaction effect [$F(12, 36) = 2.277$, $p = 0.028$, $\eta_p^2 = 0.43$]. For both /a/ and /u/ in Mandarin, pairwise comparisons using the Sidak test revealed no significant difference on MomentCOV between any two signal conditions. For the Mandarin /i/, a significant difference was found between "Vocoder" and "LowPass-Jit400", with the former having a significantly higher average MomentCOV than the latter (see Figure 6).

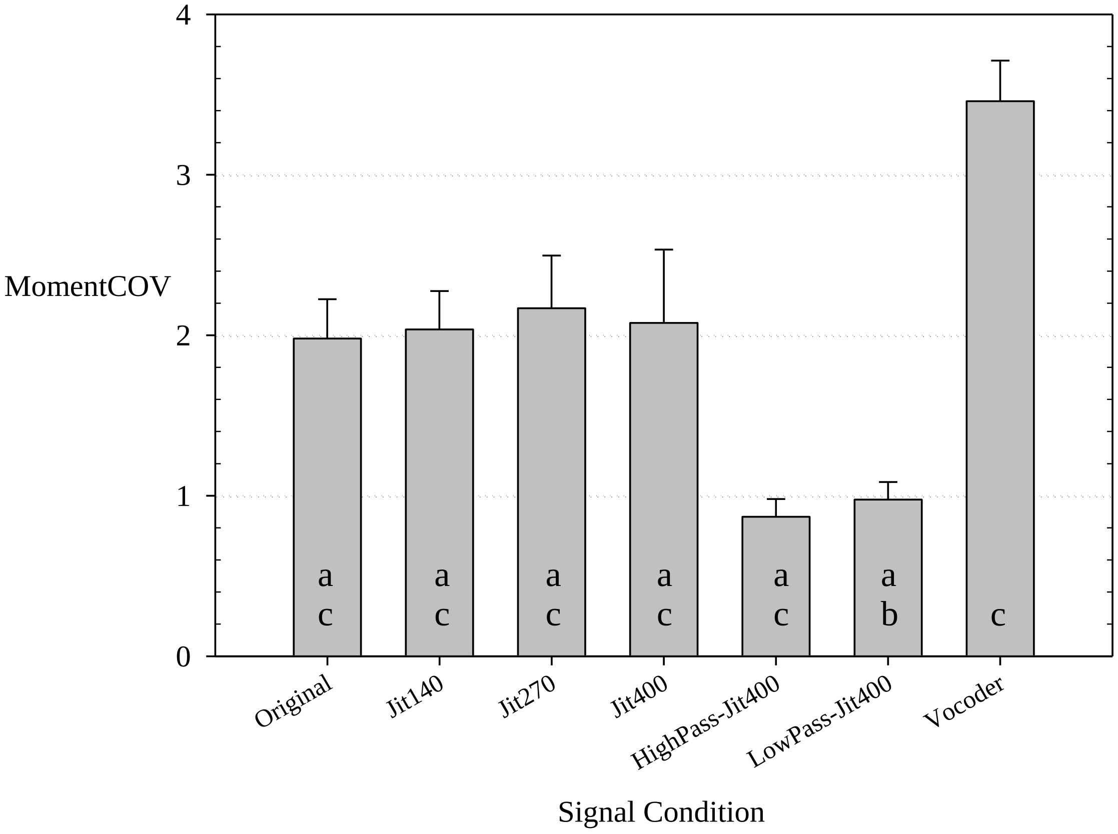


Figure 6. The average MomentCOV values obtained from native Mandarin /i/ productions in each of the seven signal conditions ("Original", "Jit140", "Jit270", "Jit400", "HighPass-Jit400", "LowPass-Jit400", and "Vocoder") with all sentence types and vowels combined. Significantly different conditions were marked with different letters.

Taiwanese. For Taiwanese, a significant effect on MomentCOV was found only for the signal condition effect [F(6, 18) = 46.697, $p < 0.001$, $\eta_p^2 = 0.94$], sentence type effect [F(1, 3) = 11.21, $p = 0.044$, $\eta_p^2 = 0.79$], vowel by signal condition interaction effect [F(12, 36) = 6.063, $p < 0.001$, $\eta_p^2 = 0.67$], and vowel by signal condition by sentence type interaction effect [F(12, 36) = 3.081, $p = 0.004$, $\eta_p^2 = 0.51$].

For the Taiwanese /i/, significant sentence type effect [F(1, 3) = 54.902, $p = 0.005$, $\eta_p^2 = 0.95$] and signal condition effect [F(6, 18) = 8.907, $p < 0.001$, $\eta_p^2 = 0.75$] on MomentCOV were found. On average, the Taiwanese /i/ had a significantly higher MomentCOV when embedded in the declarative statements (Mean = 2.133, SD = 0.166) than in the declarative questions (Mean = 1.63, SD = 0.146). For the Taiwanese /i/, a significant difference between signal conditions on MomentCOV was found only between "Vocoder" and "LowPass-Jit400" and between "Vocoder" and "HighPass-Jit400", with the "Vocoder" condition having a significantly higher average MomentCOV than both "LowPass-Jit400" and "HighPass-Jit400" (see Figure 7).

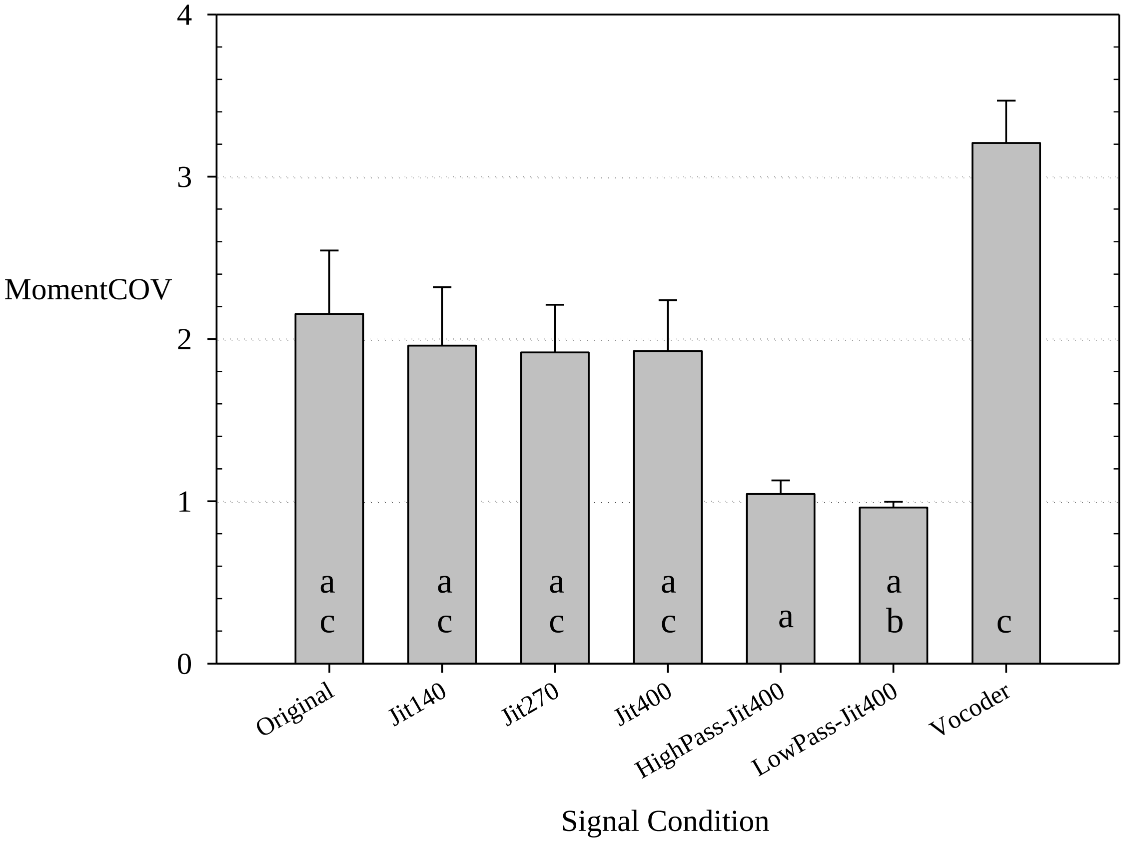


Figure 7. The average MomentCOV values obtained from native Taiwanese /i/ productions in each of the seven signal conditions (“Original”, “Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) with all sentence types and vowels combined. Significantly different conditions were marked with different letters.

For the Taiwanese /a/, only a significant signal condition effect [F(6, 18) = 75.806, $p < 0.001$, $\eta_p^2 = 0.96$] on MomentCOV was found. For the Taiwanese /a/, the “Vocoder” condition had a significantly higher MomentCOV than all the other signal conditions (see Figure 8).

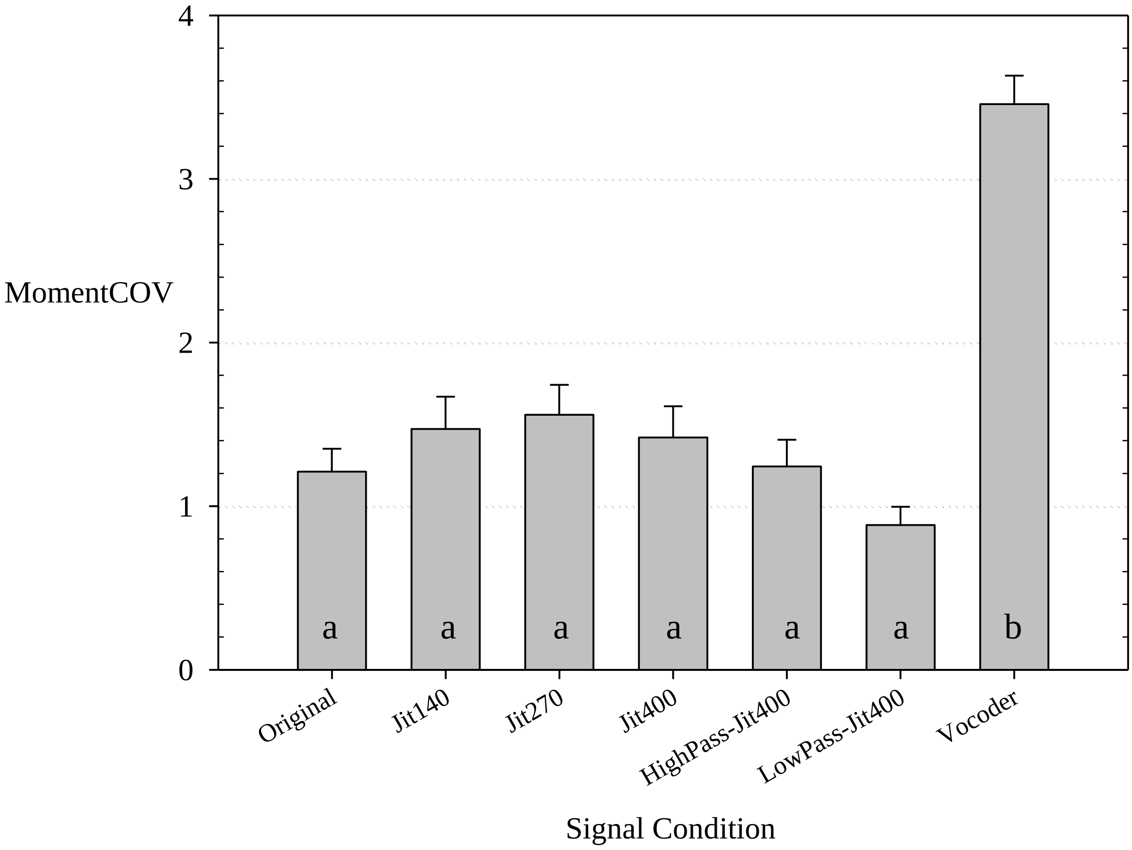


Figure 8. The average MomentCOV values obtained from native Taiwanese /a/ productions in each of the seven signal conditions (“Original”, “Jit140”, “Jit270”, “Jit400”, “HighPass-Jit400”, “LowPass-Jit400”, and “Vocoder”) with all sentence types and vowels combined. Significantly different conditions were marked with different letters.

For the Taiwanese /u/, only a significant signal condition effect [$F(6, 18) = 42.963$, $p < 0.001$, $\eta_p^2 = 0.94$] on MomentCOV was found. For Taiwanese /u/, the "Vocoder" condition showed a significantly higher MomentCOV than both "HighPass-Jit400" and "LowPass-Jit400" conditions and the "HighPass-Jit400" condition showed a significantly higher MomentCOV than the "LowPass-Jit400" condition (see Figure 9).

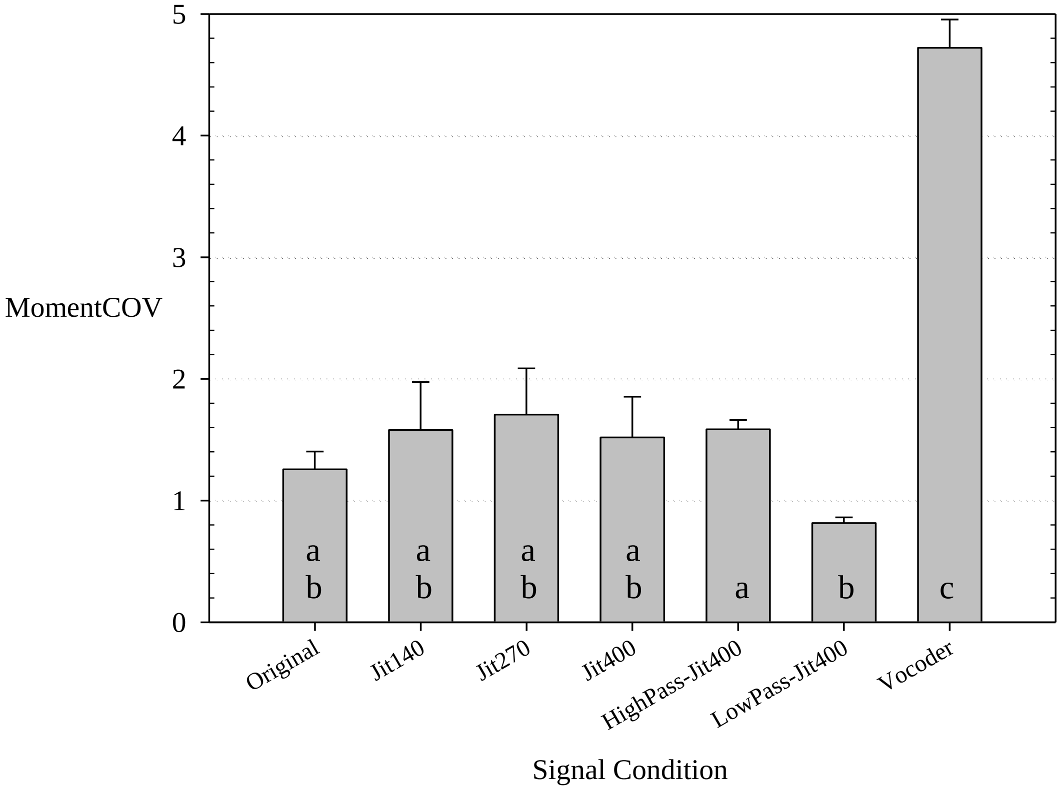


Figure 9. The average MomentCOV values obtained from native Taiwanese /u/ productions in each of the seven signal conditions ("Original", "Jit140", "Jit270", "Jit400", "HighPass-Jit400", "LowPass-Jit400", and "Vocoder") with all sentence types and vowels combined. Significantly different conditions were marked with different letters.

3.2.2 Perceptual Effects

Results from the perceptual data included the general comparisons on the measures of Percent Correct and PA for question/statement detection in different signal conditions and an error analysis for each of the three language groups.

3.2.2.1 Percent Correct

Results from the four-way (2 sentence types X 7 signal conditions X 3 languages X 2 genders) Mixed Model ANOVA conducted on the Percent Correct scores for question/statement identification revealed significant sentence type effect, signal condition effect, language effect, and signal condition by sentence type interaction effect (see Table 5).

Table 5. Results from the four-way (2 sentence types X 7 signal conditions X 3 languages X 2 genders) Mixed Model ANOVA conducted on the percent correct scores.

Factor	F	Hypothesis df	Error df	p	η_p^2
Signal Condition (C)	46.832	6	144	< 0.001*	0.661
Sentence Type (T)	28.596	1	24	< 0.001*	0.544
Language (L)	5.627	2	24	0.010*	0.319
Gender (G)	0.027	1	24	0.872	0.001
C*T	7.203	6	144	< 0.001*	0.231
C*L	1.660	12	144	0.082	0.122
C*G	1.526	6	144	0.174	0.060
T*L	1.952	2	24	0.164	0.140
T*G	0.361	1	24	0.554	0.015
L*G	0.251	2	24	0.251	0.021
C*T*L	0.666	12	144	0.782	0.053
C*T*G	0.353	6	144	0.907	0.014
C*L*G	1.277	12	144	0.238	0.096
T*L*G	0.073	2	24	0.929	0.006
C*T*L*G	0.666	12	144	0.782	0.053

*significant at the 0.05 level.

Post-hoc pairwise comparison tests using the Bonferonni method showed that the average Percent Correct score for the English stimuli (Mean = 86.5%) was significantly higher than for the Taiwanese stimuli (Mean = 73.8%) while that for the Mandarin stimuli (Mean = 77%) was not significantly different from that for either English or Taiwanese stimuli. Across all of the seven signal conditions, the average Percent Correct score was significantly higher for the statement stimuli than for the question stimuli (see Figure 10). For the question stimuli, the “HighPass-Jit400” and “Vocoder” conditions resulted in significantly lower Percent Correct scores than all the other signal conditions (see Figure 10). For the statement stimuli, the “LowPass-Jit400” and “Vocoder” conditions yielded significantly lower average Percent Correct scores than the “Original”, “Jit140”, and “Jit270” signal conditions (see Figure 10).

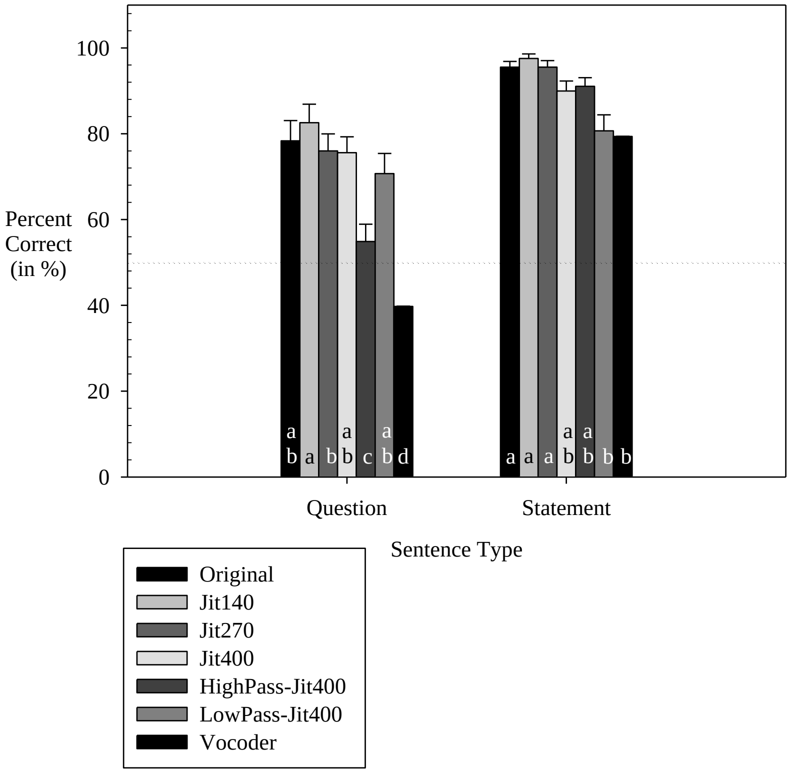


Figure 10. The average Percent Correct scores obtained from the perceptual task for question and statement stimuli presented in each of the seven signal conditions. Significantly different conditions were marked with different letters.

3.2.2.2 Proportion of the Area (PA)

Results from the three-way (3 languages X 2 genders X 7 signal conditions) Mixed Model ANOVA conducted on the measure of PA revealed significant language effect [F(2, 24) = 4.649, p = 0.02, η_p^2 = 0.279], signal condition effect [F(6, 144) = 44.223, p < 0.001, η_p^2 = 0.648], and language by signal condition interaction effect [F(12, 144) = 1.883, p = 0.041, η_p^2 = 0.136] but no significant gender effect [F(1, 24) = 0.152, p = 0.701, η_p^2 = 0.006], gender by signal condition interaction effect [F(6, 144) = 2.033, p = 0.065, η_p^2 = 0.078], gender by language interaction effect [F(2, 4) = 0.276, p = 0.761, η_p^2 = 0.022], or language by gender by signal condition interaction effect [F(12, 144) = 1.386, p = 0.179, η_p^2 = 0.104]. The PA measures was found to be significantly higher for the English group than the Taiwanese group in two signal conditions, including "HighPass-Jit400" (English: 82.4%; Taiwanese: 65.5%) and "LowPass-Jit400" conditions (English: 86.4%; Taiwanese: 64.7%).

Figures 11 to 13 show the mean PA values across the seven signal conditions for English, Mandarin, and Taiwanese respectively.

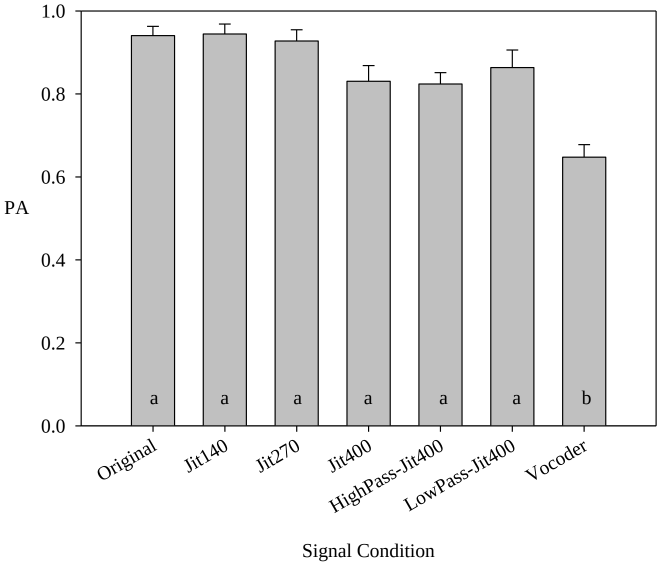


Figure 11. Mean and standard error of the PA measure across seven signal conditions for the English stimuli.

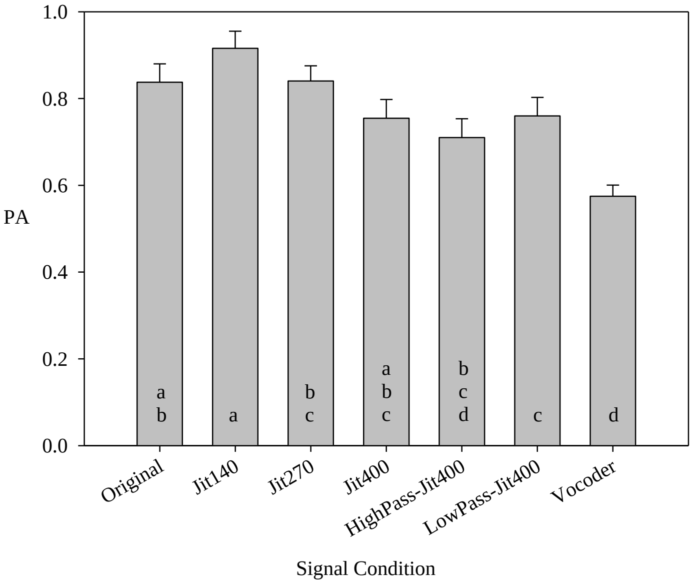


Figure 12. Mean and standard error of the PA measure across seven signal conditions for the Mandarin stimuli.

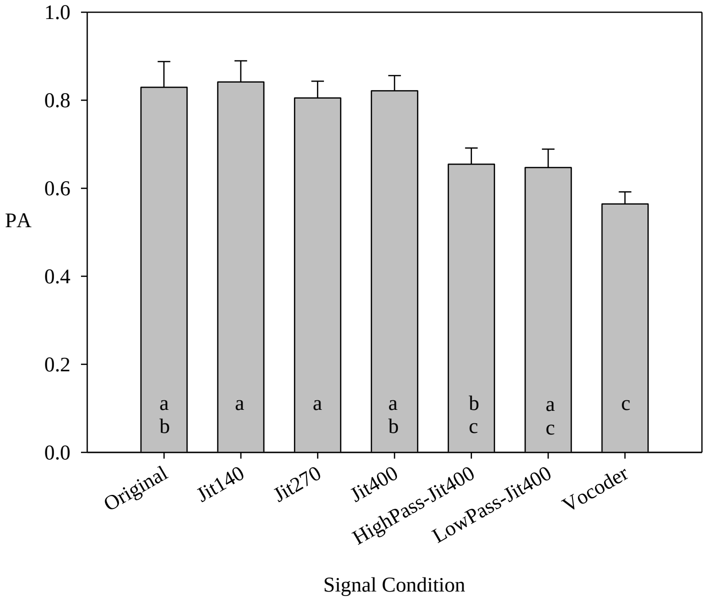


Figure 13. Mean and standard error of the PA measure across seven signal conditions for the Taiwanese stimuli.

3.2.2.3 Error Analysis

The distribution of errors was compared across different categories of lexical stress and lexical tone and between different speaker genders.

3.2.2.3.1 Stress and Lexical Tone

To determine whether the error rate in detecting question/statement varied by the lexical tone of the last syllable, an error analysis was conducted for each of the three language groups. The error rate for each type of lexical tone was calculated as the error count for the lexical tone type (stress vs. unstressed in English and different tone types for Mandarin and Taiwanese) divided by the total count of errors and then multiplied by 100. The stress effect was investigated in English while the tone effect was investigated in Mandarin and Taiwanese.

English. For English, each of the 28 sentences was categorised into “stressed” or “unstressed” based on whether the last syllable in the sentence was in a stressed or unstressed syllable. Of the 28 English phrases, 12 had stressed final syllables and 16 unstressed (see Appendix 7). Result of a paired t test conducted on the error rate revealed no significant stress effect ($t = -0.471$, $df = 9$, $p = 0.649$).

Mandarin. For Mandarin, each of the 28 sentences was categorised into Tone 0, 1, 2, 3(2), or 4 based on which lexical tone of the last syllable in the sentence was (see Appendix 7). It is noteworthy that Tone 3 in Mandarin often changes into Tone 2 at the end of an utterance, either a phrase or a sentence. A one-way RM ANOVA was conducted on the error rate, with tone type treated as the within-groups factor. Results showed a significant tone effect [$F(4, 36) = 6.899$, $p < 0.001$, $\eta_p^2 = 0.434$]. As shown in Figure 14, Tones 0 (neutral) had a significantly higher error rate than all other tones except for Tone 3 (low-dip-rise).

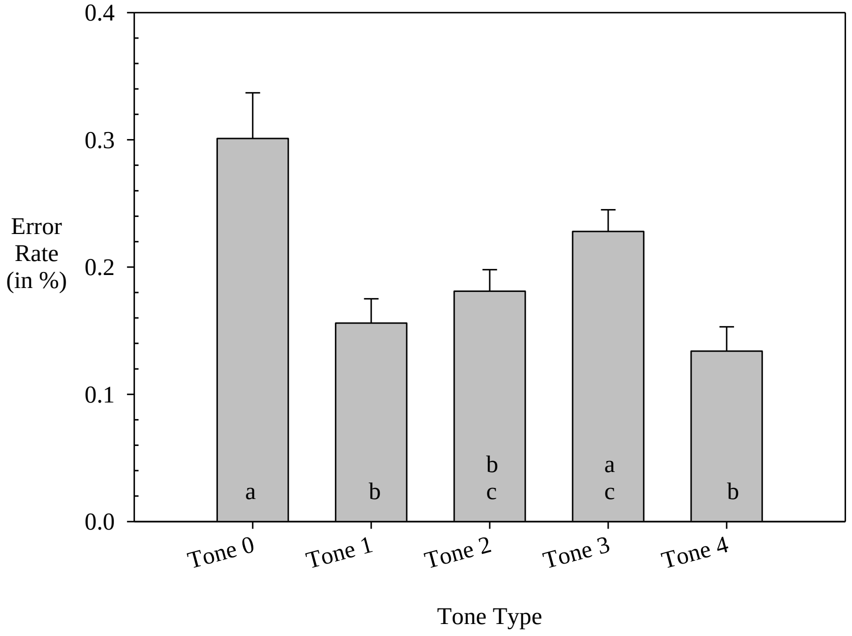


Figure 14. The mean error rate for each of the seven tone types in Mandarin.

Taiwanese. For Taiwanese, each of the 28 sentences was categorised into Tone 1, 2, 3, 4, 5, 7, or 8 based on which lexical tone of the last syllable in the sentence was (see Appendix 7). A one-way RM ANOVA was conducted on the error rate, with tone type treated as the within-groups factor. Results showed a significant tone effect [$F(6, 54) = 11.129$, $p < 0.001$, $\eta_p^2 = 0.553$]. As shown in Figure 15, pairwise comparisons using the Bonferroni test showed that Tones 5 (low-rising), 1(high-level), and 2 (high-falling) had a significantly higher error rate than Tones 3 (low-falling), 4 (low-falling checked), 7 (mid-level), and 8 (high-falling checked).

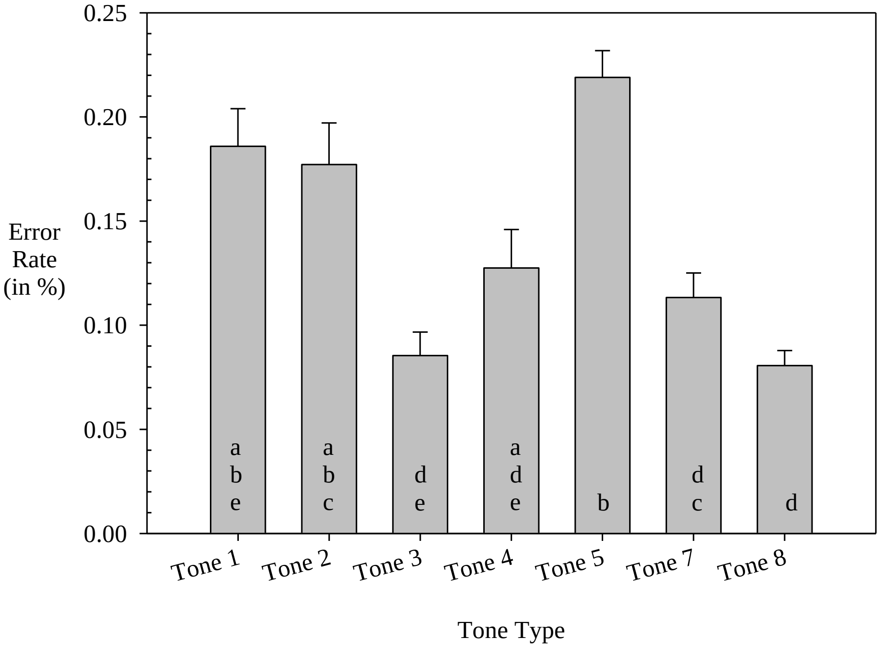


Figure 15. The mean error rate for each of the seven tone types in Taiwanese.

3.2.2.3.2 Speaker Gender

To determine whether the error rate in detecting question/statement intonation varied by the speaker gender, the error rate for each speaker gender was calculated as the error count for each speaker gender divided by the total count of errors and then multiplied by 100. A four-way (2 speaker genders X 2 signal conditions X 3 languages X 2 listener genders) Mixed Model ANOVA was conducted, with speaker gender and signal condition treated as the within-group factors and language and listener gender as the between-group. Consequently, significant signal condition effect, language effect, and signal condition by speaker gender interaction effect were found (see Table 6). Post-hoc pairwise comparisons using the Bonferroni method revealed that the Taiwanese groups showed a significantly higher mean error rate (Mean = 26.3%) than both Mandarin (Mean = 22.9%) and English groups (Mean = 13.4%), which were not significantly different from each other on the mean error rate. This finding was consistent with the finding for the percent correct measure. Figure 16 shows the mean error rates for females and males in each of the seven signal conditions. As shown in Figure 16, stimuli with male voices resulted in a significantly higher error rate than those with female voices in the “Jit140”, “Jit400”, and “HighPass-Jit400” conditions. In contrast, stimuli with female voices resulted in a significantly higher mean error rate than those with male voices in the “LowPass-Jit400” condition (see Figure 16).

Table 6. Results from the four-way (2 speaker genders X 2 signal conditions X 3 languages X 2 listener genders) Mixed Model ANOVA conducted on the error rates.

Factor	F	Hypothesis df	Error df	p	η_p^2
Signal Condition (C)	31.561	6	144	< 0.001*	0.568
Speaker Gender (SG)	2.306	1	24	0.142	0.088
Language (L)	5.548	2	24	0.010*	0.316
Listener Gender (LG)	0.027	1	24	0.872	0.001
C*SG	5.785	6	144	< 0.001*	0.194
C*L	1.079	12	144	0.382	0.082
C*LG	1.414	6	144	0.213	0.056
SG*L	0.762	2	24	0.478	0.060
SG*LG	0.039	1	24	0.845	0.002
L*LG	0.227	2	24	0.799	0.019
C*SG*L	0.703	12	144	0.747	0.055
C*SG*LG	0.561	6	144	0.760	0.023
C*L*LG	0.976	12	144	0.474	0.075
SG*L*LG	0.046	2	24	0.955	0.004
C*SG*L*LG	0.520	12	144	0.899	0.042

*significant at the 0.05 level.

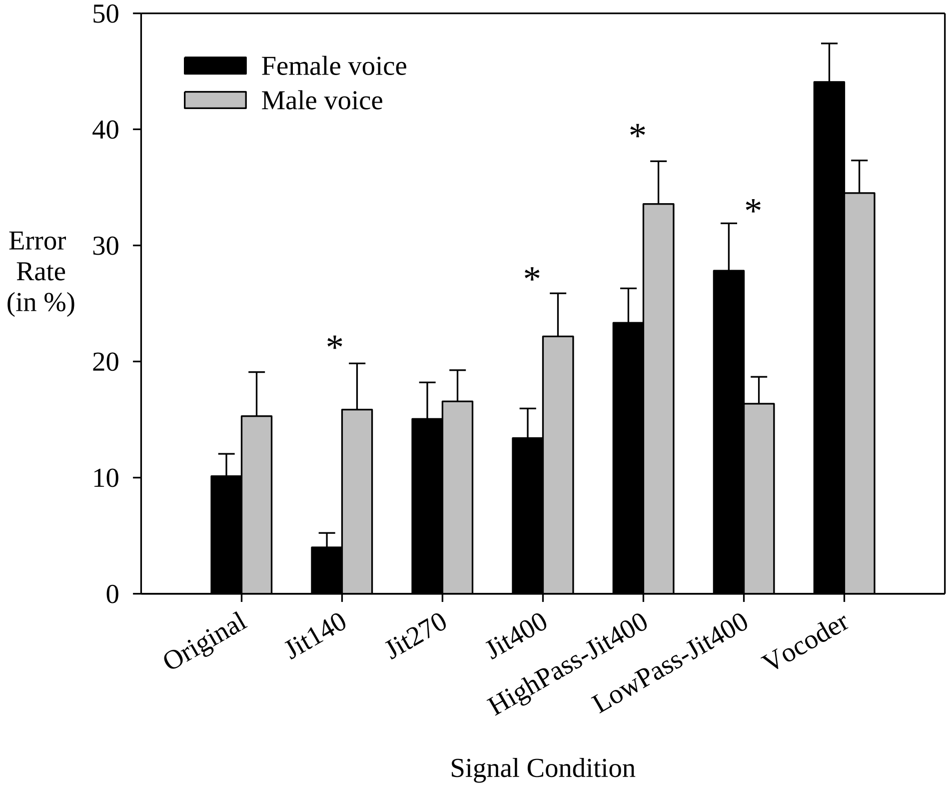


Figure 16. The means and standard errors of the error rate measure for female and male voices across seven signal conditions. Pairs with significant between-gender differences were marked with an asterisk ("*").

4 Discussion

This chapter discusses the results of the study in relation to the research hypotheses and other relevant studies. Clinical implications of the study will then be discussed followed by a discussion on limitations of the study and possible directions for future research.

4.1 Study Findings In Relation to Research Hypotheses

This study was designed to examine three hypotheses. The first hypothesis was that declarative statements would have a falling pitch and declarative questions would have a rising pitch at the end of the phrase in both tonal and non-tonal languages. The second hypothesis was that listeners of both tonal and non-tonal languages would perform significantly worse on tokens with degraded temporal fine structure (temporally-jittered and vocoder-processed) compared to non-processed tokens. The third hypothesis was that tonal language listeners would perform significantly worse than non-tonal language listeners in the perception of intonation on all signals processed with simulated hearing loss.

In relation to Hypothesis 1, declarative questions in all three languages were found to have a significantly higher F0 on the whole and higher F0 on the final syllable compared to statements. The F0ratio, which is the ratio between the maximum F0 of the last syllable and that of the rest of the sentence, was also significantly higher in questions than in statements. These findings indicate that rising pitch is indeed a feature of question intonation in all three languages. In addition to F0, the RMS ratio and RMS of the final syllable were also found to be significantly higher in questions than in statements. This indicates that the intensity of the final syllable in a sentence is also an acoustic feature that differs between statements and question in all three languages. No significant difference was found in the acoustic measures of intonation when the English native speaker's recordings were compared to the second language English speakers' recordings. Both the pitch and intensity markers of intonation can therefore be considered to be common to native and non-native English speakers.

Hypothesis 2 was partly supported in that listeners from all three languages did perform significantly worse, as shown in both percent correct and PA measures, on the vocoder-processed condition when compared to the non-processed (original) condition for both question and statement identification tasks. Listeners performed worse for vocoder-processed tokens when compared to all other processing conditions. The difference between the vocoder-processing condition and all other processing conditions was statistically significant in all cases (using the PA measure) with the exception of the comparison with the "HighPass-Jit400" condition in Mandarin, and the "LowPass-Jit400" and "HighPass-Jit400" conditions in Taiwanese. Therefore, it can be concluded that vocoder-processing, with the parameters used in this study, is disruptive to identification of question/statement intonation in English, Mandarin, and Taiwanese.

Temporally jittered signals without band pass filtering, including signals processed with the highest level of temporal jitter in this study ("Jit400"), did not result in significantly worse intonation identification results (both percent correct and PA measures) than the original signals. This indicates that temporal jittering alone, at the levels used in this study, is not sufficient to disrupt perception of question/statement intonation.

When compared to the original signals using the percent correct scores, temporally jittered signals with additional high-pass filtering ("HighPass-Jit400") resulted in significantly worse question intonation identification and those with additional low-pass filtering (“LowPass-Jit400”) worse statement intonation identification. Using the PA measure, the “HighPass-Jit400” or “LowPass-Jit400” signals showed lower average PA measures than the original signals but the differences between these signals and the original signals were not significant except for the “LowPass-Jit400” in Mandarin. The PA measure can be considered more robust than the percent correct score as it was designed to eliminate any potential bias resulting from the two-alternative forced-choice perceptual task. In addition, the acoustic findings from the Mandarin data revealed no significant difference between the original signals and the temporally jittered signals with additional filtering on both %Jit and MomentCOV measures except that temporally jittered signals with additional low-pass filtering showed a significantly higher %Jit than the original signals. Therefore, the acoustic and perceptual findings in this study provide some evidence that temporally jittered signals with additional low-pass filtering may compromise the perception of question/statement intonation.

In relation to Hypothesis 3, the Taiwanese group performed significantly worse on the perceptual task when compared to the English group using the percent correct score. No significant differences were found between Mandarin group and the English or Taiwanese group. When examining the PA measure, the difference between the English and Taiwanese groups was found to be significant only for the two filtered conditions (“HighPass-Jit400” and “LowPass-Jit400”). These findings suggest that perception of question/statement intonation under simulated hearing loss conditions is to some extent dependant on language. However, this effect is not based only on a tonal/non-tonal language distinction, rather, probably on more specific details of the intonation and tone or stress in each language. Further discussions on this language effect are included in Sections 4.2.4 to 4.26 and Appendix 8.

4.2 Study Findings In Relation to Previous Research

The following section will discuss how findings from the current study relate to previous research on acoustic measures of intonation, perception of intonation by cochlear implant users, and speech perception through temporal jittering and other types of hearing loss. The effect of language, tone type, stress, and speaker gender on intonation perception will also be discussed.

4.2.1 Acoustic Measures of Intonation

All three languages displayed a higher overall pitch in questions as compared to statements. This applies to the maximum F0 of the final syllable in an utterance, the maximum F0 for the rest of the utterance, and the ratio between these two F0 measures. The finding of pitch rising for declarative questions was common to all three languages and in agreement with previous studies of English (Hedberg et al., 2004; Kochanski et al., 2004), Mandarin (Fang Liu, 2009; Yuan, 2006), and Taiwanese (Peng & Beckman, 2003).

In a follow-up stepwise multiple regression analysis of the acoustic data (see Appendix 8), the relationship of the statement F0ratio (SF0ratio) and statement RMSratio (SRMSratio) with the corresponding question F0ratio (QF0ratio) was examined. This relationship was found to vary between languages, with the SF0ratio being a strong predictor of the corresponding QF0ratio in Taiwanese and Mandarin, but not in English. The increase in F0ratio from statements to questions was found to be greater in Taiwanese when compared to Mandarin. In addition, the SRMSratio was also predictive of QF0ratio in Mandarin. These results suggest that although all three languages employ rising intonation for declarative questions, the rise in tonal languages is more closely related to the statement F0 contour than in a non-tonal language. This may be explained by the use of F0 in the formation of lexical tones in Taiwanese and Mandarin. The F0 in tonal languages encodes both intonation and lexical pitch, and therefore the intonation contour is to some extent influenced by the lexical tones. However, because the overall English F0 contour is not constrained by lexical tones, more variability is possible, as seen in the results of the multiple regression analysis. The greater increase in F0ratio in Taiwanese as compared to Mandarin may be due to the fact that Taiwanese has more tone types. Taiwanese also uses F0 height contrasts in differentiating tones whereas Mandarin does not. These differences may mean that a greater increase in F0 ratio is required in order to make the statement and question intonation contours distinctive.

4.2.2 Perception of Vocoder-Processed Signals

Several studies have shown that cochlear implant users have a reduced ability to correctly identify question/statement intonation. The current study used vocoder-processed speech to simulate cochlear implants and found an average accuracy rate in the identification of question/statement intonation of 61% (English 67%, Mandarin 60%, Taiwanese 55%). These results are comparable to a group of Hebrew speaking cochlear implant users who had question/statement identification accuracy (uncorrected scored) of 71% (Most & Peled, 2007) and a group of English speaking cochlear implant users who had an accuracy rate of 70.13% (Peng et al., 2008).

Participants in the Peng et al. (2008) and Most and Peled (2007) studies were users of either Nucleus 22 or Nucleus 24 devices and are thus likely to have had more active channels than the vocoder used in the present study. It should also be noted that participants in the present study received bilateral input through headphones, whereas participants in the two above cochlear implant studies were unilateral users and received stimulation through a sound field. Despite these differences, the vocoder perception results from the current study support the finding from previous studies that perception of intonation is adversely affected by cochlear implant processing. The current study would also suggest these perceptual difficulties are likely to be common across different languages types.

4.2.3 Perception of Temporally-Jittered and Band-pass Filtered Signals

The present study did not find evidence that temporal jittering alone (up to the highest level used in this study) reduces perception of question/statement intonation. Previous research has found that temporal jittering of a speech signal results in decreased performance in word identification in noise tasks (MacDonald et al., 2010; Pichora-Fuller et al., 2007). There are several possibilities which may explain the difference in the word perception results of these studies and the intonation perception results of the current study. Firstly, there may have been a difference in the levels of temporal jitter used. Because the signals were processed differently in each study, it is not possible to directly compare these levels. Secondly, the difference in results may be due to the fact that MacDonald et al. (2010) and Pichora-Fuller et al. (2007) used stimuli with noise whereas the current study did not. This is an important factor as Pichora-Fuller et al. (2007) found that word identification of temporally jittered tokens improved as the signal-to-noise ratio improved. Thirdly, the results may vary due to the differing impact of temporal jitter on intonation perception as compared to speech in noise perception. In other words, although the TFS of a speech signal is important for pitch information and for separating speech from noise, a greater degree of temporal jitter may be required to disrupt the perception of pitch contour when compared to the degree required to disrupt the perception of speech in noise.

When temporal jittering was combined with high-pass filtering ("HighPass-Jit400"), the accuracy rate for question intonation identification, as measured by percent correct score, was significantly reduced in all three languages. With the PA measure, the "HighPass-Jit400" signals resulted in lower performance than the original signals in all three language groups but the differences were not statistically significant. It should be noted that temporal jittering was only applied to frequencies below 1.2 kHz. Therefore, it is likely that the lower identification accuracy under the "HighPass-Jit400" condition is due to the high-pass filtering rather than the temporal jittering. This would not be unexpected given the dominance region important for pitch perception (see Section 4.2.7).

The addition of low-pass filtering also decreased identification accuracy rates in all three languages. Although the "LowPass-Jit400" signals showed a significantly lower percent score than the original signals for all three languages in statement intonation identification, only the "LowPass-Jit400" signals in Mandarin showed a significant lower PA measure than the original signals. Since the band pass filter conditions were only used in combination with temporal jittering, it is not clear whether the reduced intonation identification accuracy would result from either the high-pass or low-pass filters alone. Without temporal jittering, it could be expected that intonation identification would be less disrupted by low-pass filtering than by high-pass filtering due to the importance of the lower frequencies in conveying pitch. However, when the only harmonics reserved after filtering were perturbed as in the "LowPass-Jit400" condition, the potential adverse effect of temporal jittering on intonation perception appeared to be aggravated to the point where a significant reduction of accuracy rate could be observed in the Mandarin group.

A previous study has shown that with low pass filtering with cut-off frequencies of 150 to 300 Hz, listeners are still able to identify question/statement intonation at 85% in Cantonese and 67% in Mandarin (Xu & Mok, 2012b). These results are similar to the 77% identification accuracy rate (overall percent correct score for all languages) of the low-pass filtered and temporally jittered condition in the current study (71% in Mandarin). However, the cut-off frequency in the current study was much higher (1.2 kHz), suggesting that the temporal jittering may have been a contributing factor to the performance of intonation identification in the current study.

4.2.4 The Effect of Language

Although few studies have directly compared question/statement intonation perception accuracy in different languages, some differences have been found. Perception accuracy rates have been found to be lower in Cantonese when compared to Mandarin, and this was explained as being due to the use of a final syllable pitch contour in Cantonese as compared to a more global pitch contour in Mandarin (Ma et al., 2011; Xu & Mok, 2012a, 2012b). In the current study, Taiwanese intonation was found to be worse than English (overall percent correct scores). No language effect was found on F0 or RMS measures for the final syllable or for the rest of the sentence. Therefore, the perceptual difference among the three languages cannot be explained in terms of global rise as compared to a boundary tone rise. It is possible that the difference found is due to the influence of lexical tone on intonation in Taiwanese which is not a factor with English (see Section 4.2.5).

It is also possible that the relationship between the statement F0 and the question F0 was a factor in the poorer perceptual scores of Taiwanese when compared to English. The stepwise multiple regression analysis (Appendix 8) found a strong relationship between SF0ratio and QF0ratio in the two tonal languages but not in English. The greater variation between the SF0ratio and QF0ratio in English when compared to Taiwanese may therefore mean that a distinction between the statement and question intonation contours is easier to perceive in English.

The current study also provides preliminary evidence that intonation perception in different languages may be affected differently by hearing loss. Statistical significance was found between the Taiwanese and English accuracy rates for the two band-pass conditions using the PA measure. The low-pass filter and high-pass filter simulate high frequency hearing loss and low frequency hearing loss respectively. Therefore, it would seem that either of these hearing loss configurations would result in more difficulty perceiving Taiwanese intonation than English intonation.

4.2.5 The Effect of Tone

When analysing the effect of the final tone on identification accuracy, it was found that Mandarin speakers performed worst with the neutral tone (i.e., Tone 0) in sentence-final position. This was statistically significant when compared to all tones except for tone 3. This result may be explained by two observations from previous studies. Firstly, the neutral tone is shorter in duration (Chao, 1965), and therefore there is less range over which pitch contour can be detected on the final syllable. Secondly, Liu and Xu (2007) found that neutral tones in sentence final position in both statements and questions have a falling contour, with the statement contour being steeper. The perceptual judgement from the final syllable cue must therefore be made from the steepness of the contour, and this is likely to be a more difficult distinction to make than distinguishing tone contour direction.

The results from the current study differ from previous studies of the interaction between Mandarin tone and question/statement intonation. Yuan (2004) reported that questions were easier to identify with a final syllable at Tone 4 (high-falling), and more difficult to identify at Tone 2 (high-rising). Liu (2009) found that a high tone (Tone 1) adversely affected both question and statement identification. The present study did find Tone 4 in the final syllable to result in the lowest error rate for intonation identification but did not find a significant different between Tone 4 and Tone 1 or 2. As Yuan (2004) and Liu (2009) did not include tokens with the neutral tone on the final syllable, these results were hard to compare.

For Taiwanese, Tone 5 (low-rising) was found to result in the highest error rate in question/statement detection amongst all tone types, and this was statistically significant when compared to Tones 3, 4, 7, and 8. Tone 1 (high) and Tone 2 (high-falling) had the next worst error rates. Tone 5 is the only rising tone in Taiwanese, and therefore the poorer intonation identification results may be due to potential confusion of the rising tone contour and the rising question contour. Studies of other tonal languages have also found that a rising tone on the final syllable can adversely affect question/statement intonation perception. This includes one study of Mandarin (Yuan, 2004) and two studies of Cantonese (Ma et al., 2011; Xu & Mok, 2012b).

4.2.6 The Effect of Stress

The current study found that the presence of stress on the final syllable did not affect the error rate in English question/statement intonation detection. This result is consistent with the finding of Liu (2009), that high or falling pitch accents in statements change to rising when a rising question contour is employed. In other words, unlike lexical tones contours, pitch accent contours do not affect the global pitch contour of the question/statement intonation. Therefore, it would not be expected that question/statement perception be affected by the stress type of the final syllable in an utterance for a stress language such as English.

4.2.7 The Effect of Speaker Gender

Perceptual error rates were significantly worse on tokens with female speakers under the “LowPass-Jit400” condition when compared to male speaker tokens. For male speaker tokens, error rates were significantly worse compared to female speakers under the “Jit140”, “Jit400”, and “HigPass-Jit400” conditions. These results can be explained theoretically with reference to the important harmonic dominance region for determining pitch. For the final syllable measurement, female speakers in this study had an average MaxF0-last of 228 Hz for statements and 338 Hz for questions. Male speakers had an average MaxF0-last of 124 Hz for statements and 190 Hz for questions. Given that harmonics below the sixth harmonic are generally thought to be important for pitch perception (Plack & Oxenham, 2005), the effect of filtering would depend on the upper limit (i.e., the sixth harmonic) of the dominant frequency. For female speakers, the upper limit of the first six harmonics would be 1,368 Hz for statements and 2,028 Hz for questions. For male speakers, the upper limit of the first six harmonics would be 744 Hz for statements and 1,140 Hz for questions. Therefore, the important pitch information for the male statements and questions is missing in the “HighPass-Jit400” condition as the upper limits for male voices (744 and 1,140 Hz) were both below the 1.2 kHz cut-off frequency. This susceptibility of male voices to high-pass filtering for intonation perception is reflected by the poorer perceptual results for the male voices compared to the female voices. In the “LowPass-Jit400” condition, important male pitch information is preserved as both upper limits for male voices were below the 1.2 kHz cut-off frequency for the low-pass filter. For female voices, however, important pitch information for questions will be above the 1.2 kHz cut-off frequency. This susceptibility of female voices to low-pass filtering for intonation identification is reflected by the poorer perceptual results, in the “LowPass-Jit400” condition, for the female speaker tokens when compared to the male speaker tokens.

Furthermore, temporal jitter was only applied to frequencies below 1.2 kHz as this is the region where phase-locking occurs in the cochlea. Therefore, generally speaking, for question intonation, the dominant harmonics for the male voices would have been more affected by the temporal jitter than the female voices. This may explain the significantly poorer results on male voices as compared to female voices under the “Jit140” and “Jit400” conditions. However, caution is needed with this interpretation as without any hearing loss or signal processing, Liu’s (2009) study found that perception of female Mandarin intonation was significantly better than perception of male Mandarin intonation. It should be noted that the low-pass filter condition (“LowPass-Jit400”) was also temporally-jittered, and in this case, perception of male intonation was better than female intonation. Therefore, it appears that the filtering process might have had a more adverse effect on perception of intonation than the temporal jittering process did.

4.3 Clinical Implications

The results of the current study give further support to the need to consider prosodic aspects of speech when providing amplification to patients with hearing loss. Results from perception of the vocoder-processed tokens indicate that cochlear implantees are likely to have difficulties in the perception of intonation, and that this difficulty is common across the three languages studied. Clinically, the results would support the use of low-frequency acoustic amplification when residual hearing is present in cochlear implantees, either through electro-acoustic stimulation (EAS) or through the use of a hearing aid on the ear contralateral to the cochlear implant.

Results also add further support to the need to differentiate between tonal and non-tonal languages when considering hearing aid and cochlear implant prescriptions. The new NAL-NL2 prescription formula for hearing aids provides an option to choose a tonal or non-tonal language prescription, with the tonal language option providing more low frequency gain (Keidser, Dillon, Flax, Ching, & Brewer, 2011). This distinction was based on the importance of low frequencies for lexical tone perception. However, the additional low frequency gain is also likely to benefit perception of suprasegmental prosody.

Finally, the results may provide useful information when counselling patients with hearing loss. Specifically, patients may be counselled about the possible interaction of lexical tone and intonation, and about possible differences in the perception of male and female intonation.

4.4 Limitations and Future Directions

Several limitations to the current study can be observed. Firstly, simulated cochlear hearing loss and cochlear implant simulation was used rather than participants with true hearing loss and cochlear implants. Although using simulations gave the advantage of being able to manipulate specific aspects of the speech signal, further perceptual studies using individuals with cochlear hearing loss and cochlear implants as subjects would be beneficial in order to confirm the results. Secondly, only one aspect of intonation was investigated in this study, namely, the grammatical role of distinguishing statements and declarative questions. Other aspects of intonation, such as the encoding of emotion, may be more complex in terms of tone patterns when compared to question/statement intonation. Therefore, perception of the more subtle changes in emotion intonation may be different from the present findings, which were derived from investigations on question/statement intonation alone. Further research could be undertaken on the effect of hearing loss simulations on these other aspects of speech intonation.

The study was also limited by the signal manipulations used. Seven conditions were chosen to investigate the hypotheses, the number being limited mainly by the concern about listener fatigue and the time allowed for this experiment. However, several additional conditions would have been useful for comparison with the current results, and these may provide direction for further research. These include conditions of high-pass and low-pass filters without any temporal jitter added. This would enable a clearer differentiation between the role of the filter and the role of temporal jitter in the perceptual results. Another useful condition would be the addition of noise to the simulations used in the study. This would enable an analysis of the effects of disruption to temporal fine structure in noisy and calm environments. Furthermore, a measure of word identification under the seven conditions would be useful. This would allow a comparison of the relative impact of a loss in temporal fine structure on intonation as compared to speech perception.

With regard to the MomentCOV measurements, the accuracy was limited by the sample vowels used. For Taiwanese, vowel by signal interaction was found, however given that the lexical tone on these vowels differed, it is possible that this interaction was due to the lexical tone.

A further limitation of the study was that the tokens were not recorded from a natural spoken context. Therefore, in order to ensure the token used had a discernible intonation contour the token with the highest or lowest final syllable pitch was chosen from each of the five trials. Since this selection method may have caused the intonation contours to be more prominent than would be found in natural speech, it would have made the question/statement task easier to identify and thus may restrict the potential generalisability of the present findings to real life situations.

4.5 Conclusion

This study has investigated the effects of several simulated hearing loss conditions on question/statement intonation in English, Mandarin, and Taiwanese. The acoustic findings demonstrate that all three languages employ an overall higher pitch and a rise on the final syllable pitch in declarative questions when compared to statements. Results from the perceptual experiment show that under certain simulated hearing loss conditions, perception of question/statement intonation can vary by language, the lexical tone type of the final syllable, and the gender of the speaker.

The language effect on intonation perception was found between Taiwanese and English under the two band-pass filtered conditions, with Taiwanese listeners performing significantly worse. The lexical tone type effect was found in both Taiwanese and Mandarin, with Taiwanese listeners performing significantly worse on the rising tone and Mandarin listeners performing significantly worse on the neutral tone. The effect of speaker gender was found to vary by signal conditions, with the performance in the question/statement identification task being poorer for male voices than for female voices on the high-pass filtered condition and poorer for female voices than for male voices on the low-pass filtered condition. Finally, the overall results showed that intonation perception was significantly affected by vocoder-processing but not by temporal jittering alone.

5. References

Adams, E. M., & Moore, R. E. (2009). Effects of speech rate, background noise, and simulated hearing loss on speech rate judgment and speech intelligibility in young listeners. *Journal of the American Academy of Audiology, 20*(1), 28-39. doi: 10.3766/jaaa.20.1.3

Allen, G. D., & Arndorfer, P. M. (2000). Production of sentence-final intonation contours by hearing-impaired children. *Journal of Speech, Language, and Hearing Research, 43*(2), 441-455.

Baer, T., & Moore, B. (1993). Effects of spectral smearing on the intelligibility of sentences in noise. *The Journal of the Acoustical Society of America, 94*(3), 1229-1241.

Baer, T., & Moore, B. (1994). Effects of spectral smearing on the intelligibility of sentences in the presence of interfering speech. *The Journal of the Acoustical Society of America, 95*(4), 2277-2280.

Bernstein, J. G. W., & Oxenham, A. J. (2006). The relationship between frequency selectivity and pitch discrimination: Sensorineural hearing loss. *The Journal of the Acoustical Society of America, 120*(6), 3929-3945.

Blamey, P. J., Sarant, J. Z., Paatsch, L. E., Barry, J. G., Bow, C. P., Wales, R. J., Psarros, C., Rattingan, K., Tooher, R. (2001). Relationships among speech perception, production, language, hearing loss, and age in children with impaired hearing. *Journal of Speech, Language, and Hearing Research, 44*(2), 264-285.

Boëx, C., Baud, L., Cosendai, G., Sigrist, A., Kós, M.-I., & Pelizzone, M. (2006). Acoustic to electric pitch comparisons in cochlear implant subjects with residual hearing. *Journal of the Association for Research in Otolaryngology, 7*(2), 110-124. doi: 10.1007/s10162-005-0027-2

Bolinger, D. L. (1958). A Theory of Pitch Accent in English. *Word, 14*(2), 109-149.

Bolinger, D. L. (1978). Intonation Across Languages. In J. Greenberg (Ed.), *Universals of Human Language* (Vol. 2, pp. 471-524). Stanford: Stanford University Press.

Braun, B., & Johnson, E. K. (2011). Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics, 39*(4), 585-594. doi: 10.1016/j.wocn.2011.06.002

Britain, D. (1992). Linguistic change in intonation: The use of high rising terminals in New Zealand English. *Language Variation and Change, 4*(1), 77.

Buss, E., Hall, J. W., & Grose, J. H. (2004). Temporal fine-structure cues to speech and pure tone modulation in observers with sensorineural hearing loss. *Ear and Hearing, 25*(3), 242-250.

Carlyon, R. P., & Deeks, J. M. (2002). Limitations on rate discrimination. *The Journal of the Acoustical Society of America, 112*(3), 1009-1025.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*, 148-156. doi: 10.1016/j.brainres.2006.10.064

Chao, Y. R. (1965). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Chen, S. H. (2005). The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. *The Journal of the Acoustical Society of America, 117*(5), 3225-3230.

Cheng R, L. (1968). Tone sandhi in Taiwanese. *Linguistics, 6*(41), 19-42. doi: 10.1515/ling.1968.6.41.19

Cheng R, L. (1973). Some notes on tone sandhi in Taiwanese. *Linguistics, 11*(100), 5-25. doi: 10.1515/ling.1973.11.100.5

Ciocca, V., Francis, A. L., Aisha, R., & Wong, L. (2002). The perception of Cantonese lexical tones by early-deafened cochlear implantees. *The Journal of the Acoustical Society of America, 111*(5), 2250-2256. doi: 10.1121/1.1471897

Connell, B. A. (1983). Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics, 11*(4), 337.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics* (6th ed). Malden: Blackwell Publishing.

Dai, H. (2000). On the relative influence of individual harmonics on pitch judgment. *The Journal of the Acoustical Society of America, 107*(2), 953-959. doi: 10.1121/1.428276

Dillon, H. (2012). *Hearing Aids*. New York: Thieme.

Dorman, M., Spahr, T., Gifford, R., Loiselle, L., McKarns, S., Holden, T., Skinner, M., Finley, C. (2007). An electric frequency-to-place map for a cochlear implant patient with hearing in the nonimplanted ear. *Journal of the Association for Research in Otolaryngology, 8*(2), 234-240. doi: 10.1007/s10162-007-0071-1

Drullman, R. (1995a). Speech intelligibility in noise: Relative contribution of speech elements above and below the noise level. *The Journal of the Acoustical Society of America, 98*(3), 1796-1798.

Drullman, R. (1995b). Temporal envelope and fine structure cues for speech intelligibility. *The Journal of the Acoustical Society of America, 97*(1), 585-592.

Duanmu, S. (2007). *The Phonology of Standard Chinese.* New York: Oxford University Press.

Duchnowski, P., & Zurek, P. M. (1995). Villchur revisited: Another look at automatic gain control simulation of recruiting hearing loss. *The Journal of the Acoustical Society of America, 98*(6), 3170-3181.

Fletcher, J., Grabe, E., & Warren, P. (2007). Intonational Variation in Four Diaclects of English: The High Rising Tune. In S. A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 390-409). New York: Oxford University Press, USA.

Gao, M. C. F. (2000). *Mandarin Chinese: An Introduction*. Melbourne: Oxford University Press.

Gates, G. A., Feeney, M. P., & Higdon, R. J. (2003). Word recognition and the articulation index in older listeners with probable age-related auditory neuropathy. *Journal of the American Academy of Audiology, 14*(10), 574-581.

Goswami, U., & Johnson, C. (2010). Phonological awareness, vocabulary, and reading in deaf children with cochlear implants. *Journal of Speech, Language, and Hearing Research, 53*(2), 237-261.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Harrison, R. V., Gordon, K. A., & Mount, R. J. (2005). Is there a critical period for cochlear implantation in congenitally deaf children? Analyses of hearing and speech perception performance after implantation. *Developmental Psychobiology, 46*(3), 252-261. doi: 10.1002/dev.20052

Hartmann, W. M. (1996). Pitch, periodicity, and auditory organization. *The Journal of the Acoustical Society of America, 100*(6), 3491-3502.

Hedberg, N., Sosa, J. M., & Fadden, L. (2004). *Meanings and configurations of questions in English.* Paper presented at Speech Prosody, Nara, Japan.

Hopkins, K., & Moore, B. (2007). Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information. *The Journal of the Acoustical Society of America, 122*(2), 1055-1068.

Hopkins, K., Moore, B., & Stone, M. (2008). Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *The Journal of the Acoustical Society of America, 123*(2), 1140-1153.

Jun, S.-A. (2005). *Prosodic typology: the phonology of intonation and phrasing*. New York: Oxford University Press.

Keidser, G., Dillon, H. R., Flax, M., Ching, T., & Brewer, S. (2011). The NAL-NL2 prescription procedure. *Audiology Research, 1*(1), 88-90. doi: 10.4081/audiores.2011.e2

Ketten, D. R., Skinner, M. W., Wang, G., & Vanner, M. W. (1998). In vivo measures of cochlear length and insertions depth of nucleus cochlear implant electrode arrays. *The Annals of Otology, Rhinology & Laryngology, 107*(11), 1-16.

Kochanski, G., Grabe, E., & Coleman, J. (2004). The difference between a question and a statement: A cross-dialect survey. *The Journal of the Acoustical Society of America, 115*(5), 2398-2398.

Kumar, P., & Yathiraj, A. (2009). Perception of speech simulating different configurations of hearing loss in normal hearing individuals. *Clinical Linguistics & Phonetics, 23*(9), 680-687. doi: 10.1080/02699200903072062

Lacher-Fougere, S., & Demany, L. (2005). Consequences of cochlear damage for the detection of interaural phase differences. *The Journal of the Acoustical Society of America, 118*(4), 2519-2526.

Ladd, D. R. (2008). *Intonational Phonology*. Cambridge: Cambridge University Press.

Lin, H.-B., & Repp, B. H. (1989). Cues to the Perception of Taiwanese Tones. *Language and Speech, 32*(1), 25-44. doi: 10.1177/002383098903200102

Liu, F. (2009). *Intonation systems of Mandarin and English: A functional approach.* Ph.D. 3350885, The University of Chicago, United States -- Illinois. Retrieved from http://search.proquest.com.ezproxy.canterbury.ac.nz/ ProQuest Dissertations & Theses A&I database.

Liu, F., & Xu, Y. (2007). *The neutral tone in question intonation in Mandarin.* Paper presented at Interspeech, Antwerp.

Luo, X., & Fu, Q. J. (2006). Contribution of low-frequency acoustic information to Chinese speech recognition in cochlear implant simulations. *The Journal of the Acoustical Society of America, 120*(4), 2260-2266.

Ma, J. K., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America, 120*(6), 3978-3987.

Ma, J. K., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *The Journal of the Acoustical Society of America, 129*(2), 1012-1023. doi: 10.1121/1.3531840

MacDonald, E. N., Pichora-Fuller, M. K., & Schneider, B. A. (2010). Effects on speech intelligibility of temporal jittering and spectral smearing of the high-frequency components of speech. *Hearing Research, 261*(1–2), 63-66. doi: 10.1016/j.heares.2010.01.005

McKay, C. M., McDermott, H. J., & Carlyon, R. P. (2000). Place and temporal cues in pitch perception: are they truly independent? *Acoustics Research Letters Online, 1*(1), 25-30.

Meister, H., Landwehr, M., Pyschny, V., Wagner, P., & Walger, M. (2011). The perception of sentence stress in cochlear implant recipients. *Ear and Hearing, 32*(4), 459-467. doi: 10.1097/AUD.0b013e3182064882

Moore, B. (1973a). Frequency difference limens for narrow bands of noise. *The Journal of the Acoustical Society of America, 54*(4), 888-896.

Moore, B. (1973b). Frequency difference limens for short-duration tones. *The Journal of the Acoustical Society of America, 54*(3), 610-619.

Moore, B. (2003). *An Introduction to the Psychology of Hearing*. Boston: Academic Press.

Moore, B. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology, 9*(4), 399-406. doi: 10.1007/s10162-008-0143-x

Moore, B., & Carlyon, R. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In C. Plack, R. Fay, A. Oxenham & A. Popper (Eds.), (Vol. 24, pp. 234-277). New York: Springer.

Moore, B., & Glasberg, B. R. (1993). Simulation of the effects of loudness recruitment and threshold elevation on the intelligibility of speech in quiet and in a background of speech. *The Journal of the Acoustical Society of America, 94*(4), 2050-2062.

Moore, B., Glasberg, B. R., Flanagan, H. J., & Adams, J. (2006). Frequency discrimination of complex tones; assessing the role of component resolvability and temporal fine structure. *The Journal of the Acoustical Society of America, 119*(1), 480-490.

Moore, B., Glasberg, B. R., Low, K. E., Cope, T., & Cope, W. (2006). Effects of level and frequency on the audibility of partials in inharmonic complex tones. *The Journal of the Acoustical Society of America, 120*(2), 934-944.

Moore, B., Glasberg, B. R., & Peters, R. W. (1985). Relative dominance of individual partials in determining the pitch of complex tones. *The Journal of the Acoustical Society of America, 77*(5), 1853-1860.

Moore, B., & Moore, G. A. (2003). Discrimination of the fundamental frequency of complex tones with fixed and shifting spectral envelopes by normally hearing and hearing-impaired subjects. *Hearing Research, 182*(1–2), 153-163. doi: 10.1016/s0378-5955(03)00191-6

Moore, B., & Peters, R. W. (1992). Pitch discrimination and phase sensitivity in young and elderly subjects and its relationship to frequency selectivity. *The Journal of the Acoustical Society of America, 91*(5), 2881-2893.

Moore, B., & Skrodzka, E. (2002). Detection of frequency modulation by hearing-impaired listeners: Effects of carrier frequency, modulation rate, and added amplitude modulation. *The Journal of the Acoustical Society of America, 111*(1), 327-335.

Most, T., & Aviner, C. (2009). Auditory, visual, and auditory-visual perception of emotions by individuals with cochlear implants, hearing aids, and normal hearing. *Journal of Deaf Studies and Deaf Education, 14*(4), 449-464. doi: 10.1093/deafed/enp007

Most, T., Harel, T., Shpak, T., & Luntz, M. (2011). Perception of suprasegmental speech features via bimodal stimulation: Cochlear implant on one ear and hearing aid on the other. *Journal of Speech, Language, and Hearing Research, 54*(2), 668-678. doi: 10.1044/1092-4388(2010/10-0071)

Most, T., & Peled, M. (2007). Perception of suprasegmental features of speech by children with cochlear implants and children with hearing aids. *Journal of Deaf Studies and Deaf Education, 12*(3), 350-361. doi: 10.1093/deafed/enm012

Munro, M. J. (1995). Nonsegmental factors in foreign accent. *Studies in Second Language Acquisition, 17*(01), 17-34. doi: doi:10.1017/S0272263100013735

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756-766. doi: 10.1037/0096-1523.24.3.756

Nejime, Y., & Moore, B. (1997). Simulation of the effect of threshold elevation and loudness recruitment combined with reduced frequency selectivity on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America, 102*(1), 603-615.

Oller, D. K. (1972). On the "Constant Total Amplitude" theory of final syllable lengthening. *The Journal of the Acoustical Society of America, 51*(1A), 102.

Oxenham, A. J. (2008). Pitch perception and auditory stream segregation: implications for hearing loss and cochlear implants. *Trends in Amplification, 12*(4), 316-331. doi: 10.1177/1084713808325881

Pan, H.-H. (2008). Focus and Taiwanese unchecked tones. In C. Lee & M. Gordon (Eds.), *Topic and Focus: Cross-Linguistic Perspectives on Meaning and Intonation*. Dordrecht: Springer.

Peng, S.-C., Tomblin, J. B., Spencer, L. J., & Hurtig, R. R. (2007). Imitative production of rising speech intonation in pediatric cochlear implant recipients. *Journal of Speech, Language, and Hearing Research, 50*(5), 1210-1227. doi: 10.1044/1092-4388(2007/085)

Peng, S.-C., Tomblin, J. B., & Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear and Hearing, 29*(3), 336-351. doi: 10.1097/AUD.0b013e318168d94d

Peng, S., & Beckman, M. E. (2003). *Annotation conventions and corpus design in the investigation of spontaneous speech prosody in Taiwanese.* Paper presented at Spontaneous Speech Processing and Recognition, Tokyo.

Pereira, C. (1996). *"Angry, happy, sad or plain neutral? The identification of vocal afect by hearing aid users".* Paper presented at the Sixth Australian International Conference on Speech Science and Technology, Adelaide.

Pfingst, B. E., Holloway, L. A., Poopat, N., Subramanya, A. R., Warren, M. F., & Zwolan, T. A. (1994). Effects of stimulus level on nonspectral frequency discrimination by human subjects. *Hearing Research, 78*(2), 197-209. doi: 10.1016/0378-5955(94)90026-4

Pichora-Fuller, M. K., Schneider, B. A., & Daneman, M. (1995). How young and old adults listen to and remember speech in noise. *The Journal of the Acoustical Society of America, 97*(1), 593-608.

Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H. E., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hearing Research, 223*, 114-121. doi: 10.1016/j.heares.2006.10.009

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation.* Ph.D. 0356886, Massachusetts Institute of Technology, United States -- Massachusetts. Retrieved from http://search.proquest.com.ezproxy.canterbury.ac.nz/ ProQuest Dissertations & Theses A&I database.

Plack, C., & Oxenham, A. (2005). The Psychophysics of Pitch. In C. Plack, R. Fay, A. Oxenham & A. Popper (Eds.), *Pitch* (Vol. 24, pp. 7-55): Springer New York.

Plomp, R. (1967). Pitch of complex tones. *The Journal of the Acoustical Society of America, 41*(6), 1526-1533.

Qin, M. K., & Oxenham, A. J. (2003). Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. *The Journal of the Acoustical Society of America, 114*(1), 446-454.

Sek, A., & Moore, B. C. J. (1995). Frequency discrimination as a function of frequency, measured in several ways. *The Journal of the Acoustical Society of America, 97*(4), 2479-2486.

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech Recognition with Primarily Temporal Cues. *Science, 270*(5234), 303-304.

Simon, H. J., & Yund, E. W. (1993). Frequency Discrimination in Listeners with Sensorineural Hearing loss. *Ear and Hearing, 14*(3), 190-201.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416*(6876), 87-90.

Snow, D. (2007). Polysyllabic units in the vocalizations of children from 0;6 to 1;11: Intonation-groups, tones and rhythms. *Journal of Child Language, 34*(4), 765-797. doi: 10.1017/s030500090700815x

Souza, P., Arehart, K., Miller, C. W., & Muralimanohar, R. K. (2010). Effects of Age on F0 Discrimination and Intonation Perception in Simulated Electric and Electroacoustic Hearing. *Ear and Hearing, 32*(1), 75-83. doi: 10.1097/AUD.0b013e3181eccfe9

Stone, M. A., Fullgrabe, C., & Moore, B. (2008). Benefit of high-rate envelope cues in vocoder processing: Effect of number of channels and spectral region. *The Journal of the Acoustical Society of America, 124*(4), 2272-2282.

Talbot, K. N., & Hartley, D. E. H. (2008). Combined electro-acoustic stimulation: a beneficial union? *Clinical Otolaryngology, 33*(6), 536-545. doi: 10.1111/j.1749-4486.2008.01822.x

Tong, Y. C., Blamey, P. J., Dowell, R. C., & Clark, G. M. (1983). Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant. *The Journal of the Acoustical Society of America, 74*(1), 73-80.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *The Journal of the Acoustical Society of America, 82*(4), 1152-1161.

Wang, Y., Jongman, A., & Sereno, J. A. (2001). Dichotic Perception of Mandarin Tones by Chinese and American Listeners. *Brain and Language, 78*(3), 332-348. doi: 10.1006/brln.2001.2474

Warren, P. (2005). Patterns of late rising in New Zealand English: Intonational variation or intonational change? *Language Variation and Change, 17*(2), 209-230.

Warren, P., & Britain, D. (2000). Intonation and Prosody in New Zealand English. In A. Bell & K. Kuiper (Eds.), *New Zealand English*. Wellington: Victoria University Press.

Weber, E. G. (1993). *Varieties of Questions in English Conversation*. Philadelphia: J. Benjamins.

Wells, J. C. (2006). *English Intonation: An Introduction*. New York: Cambridge University Press.

Won, J. H., Lorenzi, C., Nie, K., Li, X., Jameyson, E. M., Drennan, W. R., & Rubinstein, J. T. (2012). The ability of cochlear implant users to use temporal envelope cues recovered from speech frequency modulation. *The Journal of the Acoustical Society of America, 132*(2), 1113-1119.

Wong, L. L. N., Vandali, A. E., Ciocca, V., Luk, B., Ip, V. W. K., Murray, B., Yu, H. C., Chung, I. (2008). New cochlear implant coding strategy for tonal language speakers. *International Journal of Audiology, 47*(6), 337-347. doi: 10.1080/14992020802070788

Xu, B. R., & Mok, P. K. (2012a). *Cross-linguistic perception of intonation by Mandarin and Cantonese listeners.* Paper presented at Speech Prosody, Shanghai, China.

Xu, B. R., & Mok, P. K. (2012b, May 26-29). *Intonation Perception of Low-Pass Filtered Speech in Mandarin and Cantonese.* Paper presented at Tonal Aspects of Languages, Nanjing.

Yuan, J. (2004). *Perception of Mandarin Intonation.* Paper presented at the International Symposium on Chinese Spoken Language Processing, Hong Kong.

Yuan, J. (2006). Mechanisms of Question Intonation in Mandarin. In Q. Huo, B. Ma, E.-S. Chng & H. Li (Eds.), *Chinese Spoken Language Processing* (Vol. 4274, pp. 19-30). Berlin: Springer.

Zeng, F.-G. (2002). Temporal pitch in electric hearing. *Hearing Research, 174*, 101-106. doi: 10.1016/s0378-5955(02)00644-5

Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Rio, E. D., Kong, Y.-Y., & Chen, H. (2004). On the dichotomy in auditory perception between temporal envelope and fine structure cues. *The Journal of the Acoustical Society of America, 116*(3), 1351-1354.

Appendix 1. Rules of Tone Sandhi in Mandarin

	Rule	**Description**	**Example**
Tone 3			/hai^{214}/ "sea" 海 /pao^{214}/ "treasure, baby" 寶
	Tone 3 + Tone 2 → **low-dip** + Tone 2	When preceding Tone 2, Tone 3 (low-dip-rise) becomes low-dip, i.e., 214 → 21.	/hai^{21} yang^{35}/ "ocean" 海洋
	Tone 3 + Tone 3 → **Tone 2** + Tone 3 or **Tone 3** + Tone 3 → **Tone 2** + **Tone 2**	When preceding Tone 3, Tone 3 becomes Tone 2.	/hai^{35} ko^{214}/ "seal" 海狗
	Tone 3 + **Tone 3** + Tone 3 → (1) **low-dip** + **Tone 2** + Tone 3 (2) **Tone 2** + **Tone 2** + Tone 3	In a three-word phrase where all words are Tone 3, one of the two following changes applies: (1) the first Tone 3 becomes low-dip and the second Tone 3 becomes Tone 2; (2) the first two Tone-3 words become Tone 2.	(1) /hai^{21} pao^{35} pao^{214}/ "sea baby" 海寶寶 (2) /hai^{35} ti^{35} kuan^{214}/ "underwater museum" 海底館
/i^{55}/ "one" and /bu^{51}/ "no"			/i^{55}/ "one" 一 /pu^{51}/ "no" 不
	Tone 1 /i^{55}/ ("one") or **Tone 4 /bu^{51}/ ("no")** + Tone 4 → **Tone 2** + Tone 4	When preceding Tone 4, the word "one" and "no" become Tone 2.	/i^{35} pian^{51}/ "one piece" 一片 /pu^{35} yong^{51}/ "no need" 不用
	Tone 1 /i^{55}/ ("one") + Tone 1/2/3 → **Tone 4** + Tone	When preceding any tone except for Tone 4,	/i^{51} tian^{55}/ "one day" 一天

	1/2/3	the word “one” becomes Tone 4.	/i^{51} ren^{35}/ “one man” 一人 / i^{51} lao^{214}/ “one elderly” 一老

Appendix 2. Rules of Tone Sandhi in Taiwanese

	Rule	**Description**	**Example**
Regular tone change sequence:* **5 → 7 → 3 → 2 → 1 → 7**			
	Tone 5 + Any Tone → **Tone 7** + Any Tone	When preceding any tone, Tone 5 becomes Tone 7.	/si^{24}/ “hour” 時 /si^{33} kan^{55}/ “time” 時間
	Tone 7 + Any Tone → **Tone 3** + Any Tone	When preceding any tone, Tone 7 becomes Tone 3.	/p^{h}in^{33}/ “nose” 鼻 /p^{h}in^{21} k^{h}ang^{55}/ “nostril” 鼻孔
	Tone 3 + Any Tone → **Tone 2** + Any Tone	When preceding any tone, Tone 3 becomes Tone 2.	/sieng^{21}/ “sacred” 聖 /sieng^{51} jin^{24}/ “saint” 聖人
	Tone 2 + Any Tone → **Tone 1** + Any Tone	When preceding any tone, Tone 2 becomes Tone 1.	/hi^{51}/ “joyful” 喜 /hi^{55} su^{33}/ “happy event” 喜事
	Tone 1 + Any Tone → **Tone 7** + Any Tone	When preceding any tone, Tone 1 becomes Tone 7. ***Note:** Commonly used function words (/t^{h}e/, /beh/, /koh/, /kah/, /k^{h}ah/, /leh/) remain at Tone 1.	/t^{h}in^{55}/ “sky” 天 /t^{h}in^{33} tieng^{51}/ “up in the sky” 天頂
Checked tones	**-/p, t, k/**	When preceding any tone, Tone 4 (-/p, t, k/) becomes	/pid^{20}/ “pen” 筆

(Tones 4 and 8)	**Tone 4** + Any Tone → **Tone 8** + Any Tone	Tone 8.	/pid$^{\underline{50}}$ ki^{21}/ “note” 筆記
	-/h/ **Tone 4** + Any Tone → **Tone 2** + Any Tone	When preceding any tone, Tone 4 (-/h/) becomes Tone 2.	/tsioh$^{\underline{20}}$/ “borrow” 借 /tsioh^{51} chin^{24}/ “borrow money” 借錢
	-/p, t, k/ **Tone 8** + Any Tone → **Tone 4** + Any Tone	When preceding any tone, Tone 8 (-/p, t, k/) becomes Tone 4.	/sit$^{\underline{50}}$/ “solid” 實 /sit$^{\underline{20}}$ k^{h}uang^{24}/ “real power” 實權
	-/h/ **Tone 8** + Any Tone → **Tone 3** + Any Tone	When preceding any tone, Tone 8 (-/h/) becomes Tone 3.	/chioh$^{\underline{50}}$/ “stone” 石 /chioh^{21} t^{h}ao^{24}/ “stone” 石頭
Before /a^{51}/, a diminutive suffix**	**Tone 3** + /a^{51}/ → **Tone 1** + /a^{51}/	When preceding /a^{51}/, Tone 3 becomes Tone 1.	/ki^{21}/ “saw” 鋸 /ki^{55} a^{51}/ “saw” 鋸仔
	Tone 4 (-/p, t, k/) + /a^{51}/ → **Tone 8** + /a^{51}/	When preceding /a^{51}/, Tone 4 (-/p, t, k/) becomes Tone 8.	/tiok$^{\underline{20}}$/ “bamboo” 竹 /tiok$^{\underline{50}}$ a^{51}/ “bamboo” 竹仔
	Tone 4 (/-h/) + /a^{51}/ → **Tone 1** + /a^{51}/	When preceding /a^{51}/, Tone 4 (-/h/) becomes Tone 1.	/toh$^{\underline{20}}$/ “table” 桌 /toh^{55} a^{51}/ “table” 桌仔
	Tone 8 (-/p, t, k/) + /a^{51}/ → **Tone 4** + /a^{51}/	When preceding /a^{51}/, Tone 8 (-/p, t, k/) becomes Tone 4.	/giok$^{\underline{50}}$/ “jade” 玉 /giok$^{\underline{20}}$ a^{51}/ “jade” 玉仔

	Tone 8 (-/h/) + /a^{51}/ → **Tone 7** + /a^{51}/	When preceding /a^{51}/, Tone 8(-/h/) becomes Tone 7.	/ioh$^{\underline{50}}$/ “medicine” 藥 /ioh^{33} a^{51}/ “medicine” 藥仔
Trisyllabic monomorphemic compound **(For Tones 2, 3, 4)****	**Tone 2 + Tone 2** + Tone 2 → **Tone 1 + Tone 1** + Tone 2	The first two tones follow the regular tone change sequence (5 → 7 → 3 → 2 → 1 → 7). For the first two tones, Tone 2 becomes Tone 1.	/san^{51}/ “thin” 瘦 /san^{55} san^{55} san^{51}/ “very thin” 瘦瘦瘦
	Tone 3 + Tone 3 + Tone 3 → **Tone 2 + Tone 2** + Tone 3	For the first two tones, Tone 3 becomes Tone 2.	/t^{h}iann^{21}/ “painful” 痛 /t^{h}iann^{51} t^{h}iann^{51} t^{h}iann^{21}/ “very painful” 痛痛痛
	(-/p, t, k/) **Tone 4 + Tone 4** + Tone 4 → **Tone 8 + Tone 8** + Tone 4	For the first two tones, Tone 4 (-/p, t, k/) becomes Tone 8.	/tsiok$^{\underline{20}}$/ “enough” 足 /tsiok$^{\underline{50}}$ tsiok$^{\underline{50}}$ tsiok$^{\underline{20}}$/ “more than enough” 足足足
	(-/h/) **Tone 4 + Tone 4** + Tone 4 → **Tone 2 + Tone 2** + Tone 4	For the first two tones, Tone 4 (-/h/) becomes Tone 2.	/ba$^{\underline{20}}$/ “solid” 肉 /***ba***^{51} ba^{51} ba$^{\underline{20}}$/ “very solid” 肉肉肉
Trisyllabic monomorphemic compound **(For Tones 1, 5, 7, 8)**	**Tone 1 + Tone 1** + Tone 1 → **Tone 1 (raised) + Tone 7** + Tone 1	The first tone is raised in pitch.	/sio^{55}/ “hot” 燒 /***sio***^{55} sio^{33} sio^{55}/ “very hot” 燒

	Tone 5 + Tone 5 + Tone 5 → **Tone 5 (raised) + Tone 7** + Tone 5	The first tone is raised in pitch.	/teng^{24}/ “long” 長 /***teng***$^{\mathbf{24}}$ teng^{33} teng$^{\mathbf{24}}$/ “very long” 長長長
	Tone 7 + Tone 7 + Tone 7 → **Tone 7 (raised) + Tone 3** + Tone 7	The first tone is raised in pitch.	/ong^{33}/ “prosperous” 旺 /**ong**$^{\mathbf{33}}$ ong^{21} ong^{33}/ “very prosperous” 旺旺旺
	Tone 8 + Tone 8 + Tone 8 → **Tone 8 (raised) + Tone 4** + Tone 8	The first tone is raised in pitch.	/siok$^{\underline{50}}$/ “cheap” 俗 /***siok***$^{\underline{\mathbf{50}}}$ siok$^{\underline{20}}$ siok$^{\underline{50}}$/ “very cheap” 俗俗俗
Adjective /e/ (Neutral tone)	Any tone + **/e/** → Any tone + **/e/-same as the preceding tone**	Follow the previous tone and reduce intensity.	/sio^{55} e^{55}/ “hot” 燒的 /ling^{21} e^{21}/ “cold” 冷的 /ang^{33} e^{33}/ “red” 紅的
De-emphasis	Any tone + **deemphasized word with any tone** → Any tone + **neutral tone**	The tone of a superfluous or deemphasized word is changed to a neutral tone and reduced in intensity while the tone of the emphasized word does not change.	**/lim**$^{\mathbf{24}}$ *sen*^{21} *sin*^{21}/ “Mr. Lin” 林先生 **/kiang**$^{\mathbf{24}}$ *kue*^{21} *k*^{h}*i*^{21}/ “to pass by” 走過去 **/t**$^{\mathbf{h}}$**iao**$^{\mathbf{21}}$ *lok*^{21} *lai*^{21}/ “to jump over” 跳落去 **/p**$^{\mathbf{h}}$**a**$^{\mathbf{21}}$ *si*^{21}/ “to beat to death” 拍死 **/kiang**$^{\mathbf{55}}$ *tioh*^{21}/ “frightened” 驚著 **/bei**$^{\mathbf{51}}$ teng^{21} lai^{21}/ “to buy back”

			買轉來 /**k^{h}uang^{21}** jit^{21} lei^{21}/ “to take a look” 看一下 /**boh^{24}** *k^{h}i^{21}*/ “lost” 無去 (c.f. /boh^{33} khi^{21}/ “didn’t go” 沒有去) /**tsoh^{21}** *lang^{21}*/ “to give in marriage” 做人 (c.f. /tsoh^{51} lang^{24}/ “to carry oneself” 做人) *The emphasized word is bolded and the de-emphasized word is italicized.

Appendix 3. Instructions to Speakers

College of Science

Dr. Emily Lin
Department of Communication Disorders
Email: emily.lin@canterbury.ac.nz
Tel: 64 (3) 364-2987 ext 7080
Email: emily.lin@canterbury.ac.nz

PROJECT INFORMATION

You are invited to participate as a subject in the research project "*Perception of intonation with different configurations of simulated hearing loss in normal hearing adults*."

This study aims to examine the impact of different hearing loss configurations on the perception of speech intonation. Information yielded from this investigation will help identify the limitations of the current speech processing techniques employed in hearing aids and cochlear implants.

Instruction to Speakers

You will be required to attend one voice recording session, which will last less than half an hour.

You will be seated in a sound-treated room and a headset microphone will be placed over your head. You will then be asked to say ten sentences in random orders. You will follow the verbal instruction from the researcher and say these sentences three times as a statement and three times as a question.

This speaking task is perfectly safe and will in no way cause you any discomfort or harm. You may end the task at any time and are free to discontinue participation in this study, including withdrawal of any information you have provided.

The project is being carried out as a requirement for a Master of Audiology thesis conducted by Paul Daniell, a student at the University of Canterbury. The completed thesis will be a public document accessible via the University of Canterbury database. Data will be stored securely at the University of Canterbury for 5 years and then deleted. Results from the project may be used by the research student or supervisor in academic publications or conferences. The project is under the supervision of Dr. Emily Lin, who can be contacted on the telephone number at the top of the page. She will be pleased to discuss any concerns you may have about participation in the project.

Thank you for choosing to take part in this study. Your participation is greatly appreciated.

This proposal has been reviewed and approved by the University of Canterbury Human Ethics Committee.

Appendix 4. Instruction to Listeners

College of Science

Dr. Emily Lin
Department of Communication Disorders
Email: emily.lin@canterbury.ac.nz
Tel: 64 (3) 364-2987 ext 7080
Email: emily.lin@canterbury.ac.nz

PROJECT INFORMATION

*You are invited to participate as a subject in the research project “*Perception of intonation with different configurations of simulated hearing loss in normal hearing adults*.”*

This study aims to examine the impact of different hearing loss configurations on the perception of speech intonation.   Information yielded from this investigation will help identify the limitations of the current speech processing techniques employed in hearing aids and cochlear implants.

Instruction to Listeners

You will be required to attend one testing session, which will last no more than one hour. At the beginning of the session, your hearing will be assessed using standard audiometric testing procedure that will take about 10 minutes. If the hearing screening test indicates a hearing loss, you will be excluded from the experiment and referred to the UC Speech and Hearing clinic.

The task that you will be required to perform will involve listening to a number of presentations of a sentence, and you will be asked to indicate whether it is a question or a statement.

These tests are perfectly safe and will in no way cause you any discomfort or harm. You may end the tests at any time and are free to discontinue participation in this study, including withdrawal of any information you have provided.

The project is being carried out as a requirement for a Master of Audiology thesis conducted by Paul Daniell, a student at the University of Canterbury. The completed thesis will be a public document accessible via the University of Canterbury database. Data will be stored securely at the University of Canterbury for 5 years and then deleted. Results from the project may be used by the research student or supervisor in academic publications or conferences. The project is under the supervision of Dr. Emily Lin, who can be contacted on the telephone number at the top of the page. She will be pleased to discuss any concerns you may have about participation in the project.

Thank you for choosing to take part in this study. Your participation is greatly appreciated.

This proposal has been reviewed and approved by the University of Canterbury Human Ethics Committee.

Appendix 5. Consent Form

College of Science

Dr. Emily Lin
Department of Communication Disorders
Email: emily.lin@canterbury.ac.nz
Tel: 64 (3) 364-2987 ext 7080
Email: emily.lin@canterbury.ac.nz

CONSENT FORM

Project: Perception of intonation with different configurations of simulated hearing loss in normal hearing adults

This research is for a Master of Audiology thesis conducted by Paul Daniell, a student at the University of Canterbury. The completed thesis will be a public document accessible via the University of Canterbury database. Data will be stored securely at the University of Canterbury for 5 years and then deleted. Results from the project may be used by the research student or supervisor in academic publications or conferences.

I have read and understood the description of the above-named project. On this basis I agree to participate as a subject in the project and give consent to publication of the results of the project with the understanding that anonymity will be preserved.

I understand also that I may at any time withdraw from the project, including withdrawal of any information I have provided.

I note that the project has been reviewed ***and approved*** by the University of Canterbury Human Ethics Committee.

NAME (please print): ……………………..…………………………

Signature: ……………………………………………………………………..

Date: ……………………………………………………………………………….

Appendix 6. List of 28 Phrases in English, Mandarin, and Taiwanese

Sentence no.	**English**	**Mandarin**	**Taiwanese**
1	Two cups of tea	liang pei cha	neng pue tei
2	One hundred and thirty three	i pai san she san	jit pa san zap san
3	A cup of tea	i pei cha	jit pue tei
4	Everything	i qie	it tsei
5	Everyone	ta jia	ta kei lang
6	More sugar	keng tuo t^hang	k^ha zue t^heng
7	My mom	wuo ma ma	wun ma ma
8	My dad	wuo pa pa	wun pa pa
9	Our nana	wuo meng te zu mu	nan e a ma
10	The tall one	kao te na ke	kuan e hit lei
11	Over there	zai na li	ti hia
12	Me too	wuo ye she	wa ma si
13	Okay	hao le	hoh a
14	Next week	xia xing qi	ao lei pai
15	Two people	liang ke ren	neng e lang
16	In the morning	zai zao shang	t^hao za si
17	Really	zheng te	si jin e
18	At ten thirty	she tien san she te she ho	ti zap tiam san zap hun e si zung
19	In the bag	zai tai ze li	ti tei a lai
20	Just this one	zhe you zhoh ke	kan lang jit lei
21	A cup of coffee	i pei k^ha fei	jit pue ka pi
22	A rusty nail	sheng shio te ting ze	sen senn e ting a
23	Everybody	mei ke ren	mui jit e lang
24	My ex boyfriend	wuo qian ren nan p^heng yo	wa i jing e lang ping yu
25	My nana	wuo te zu mu	wa e a ma
26	The green pencil	lyu se qian pi	tsen sioh e enn pit
27	The red apple	hong p^hing kuo	ang sioh e p^hong koh
28	The world cup	she jie pei	sei kai pue

Appendix 7. Categorization of the last syllable for the 28 Phrases in English, Mandarin, and Taiwanese

Sentence no.	**English (Stress pattern)**	**Mandarin (Tone type)**	**Taiwanese (Tone type)**
1	Stressed	2	5
2	Stressed	1	1
3	Stressed	2	5
4	Unstressed	4	3
5	Unstressed	1	5
6	Unstressed	2	5
7	Stressed	1	4
8	Stressed	0	4
9	Unstressed	2(3)	2
10	Unstressed	0	1
11	Stressed	2(3)	1
12	Stressed	4	7
13	Unstressed	0	4
14	Stressed	2	3
15	Unstressed	2	5
16	Unstressed	4	5
17	Unstressed	0	1
18	Unstressed	4	7
19	Stressed	2(3)	7
20	Stressed	0	1
21	Unstressed	1	1
22	Stressed	0	2
23	Unstressed	2	5
24	Unstressed	2(3)	2
25	Unstressed	2(3)	2
26	Unstressed	2(3)	8
27	Unstressed	2(3)	8
28	Stressed	1	1

Appendix 8. Predicting the F0ratio of a declarative question from the F0ratio or RMSratio of a statement

Results from the multiple regression analysis for each of the six language data sets were shown as follows.

Taiwanese

For Taiwanese utterances, SF0ratio was found to be useful for predicting QF0ratio independently, with a significant Beta of 0.59 ($t = 7.659$, $p < 0.001$). The SRMSratio was removed from the regression model as it showed no significant correlation with QF0ratio (partial correlation = -0.128, $p = 0.179$). The regression model for the Taiwanese data can be written as:

$$\text{QF0ratio} = 0.008 + 1.082 * \text{SF0ratio}$$

This regression model explains 34.2% (Adjusted R-square = 0.342) of the variance of the QF0ratio in Taiwanese. Figure A1 shows the scatter plot of SF0ratio on the x-axis and QF0ratio on the y-axis, with data points marked by the tone type (Tones 1, 2, 3, 4, 5, 7, and 8) of the last word of each Taiwanese sentence. As shown in Figure A1, the trend that a Taiwanese statement showing a higher F0ratio would show a higher F0ratio in its corresponding declarative question applies to all tone types except for Tones 2 ("High-falling"), 3 ("Low-falling"), and 7 ("Mid-level"), where QF0ratio does not appear to be predictable from SF0ratio.

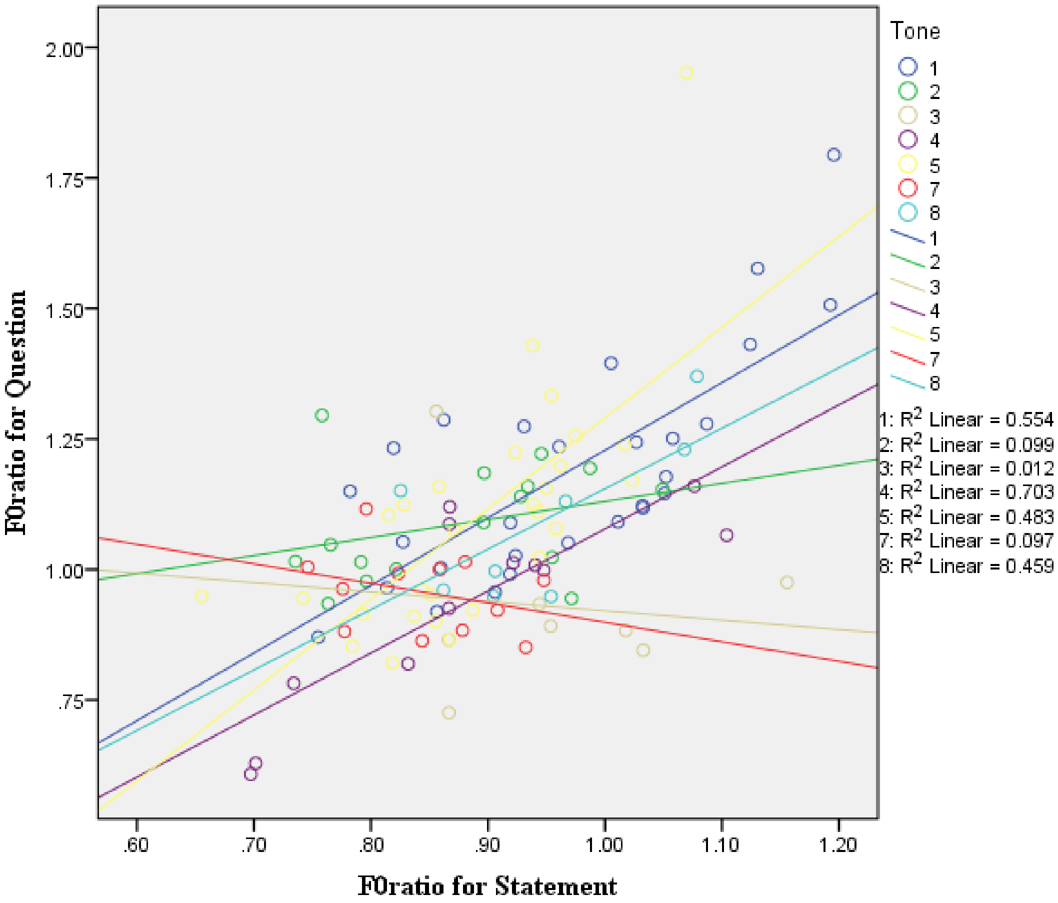


Figure A1. A scatter plot showing the relationship between the F0ratio for a statement and the F0ratio of the corresponding declarative question in Taiwanese. The data points are marked by the tone type (Tones 1, 2, 3, 4, 5, 7, and 8) of the last word of each Taiwanese sentence.

Mandarin

For Mandarin utterances, SF0ratio was also found to be useful for predicting QF0ratio independently, with a significant Beta of 0.664 ($t = 9.307$, $p < 0.001$). The predictive power of SF0ratio decreased slightly to 0.531 ($t = 5.684$, $p < 0.001$) after removing the effect of SRMSratio. The Beta for SRMSratio is 0.202, indicating that the higher the RMSratio of a statement, the higher the RMSratio of its corresponding declarative question. The regression model for the Mandarin data can be written as:

$$\text{QF0ratio} = 0.239 + 0.758*\text{SF0ratio} + 0.138*\text{SRMSratio}$$

This regression model explains 45.4% (Adjusted R-square = 0.454) of the variance of the F0ratio of questions in Mandarin. Figures A2 and A3 show the relationship between QF0ratio and the two predictors (SF0ratio and SRMSratio) respectively, with data points marked by the tone type (Tones 0, 1, 2, 3, and 4) of the last word of each Mandarin sentence. As shown in both Figures A2 and A3, the higher the F0ratio or RMSratio of a statement, the higher the F0ratio of the corresponding declarative question regardless of the tone type of the last word of each Mandarin sentence.

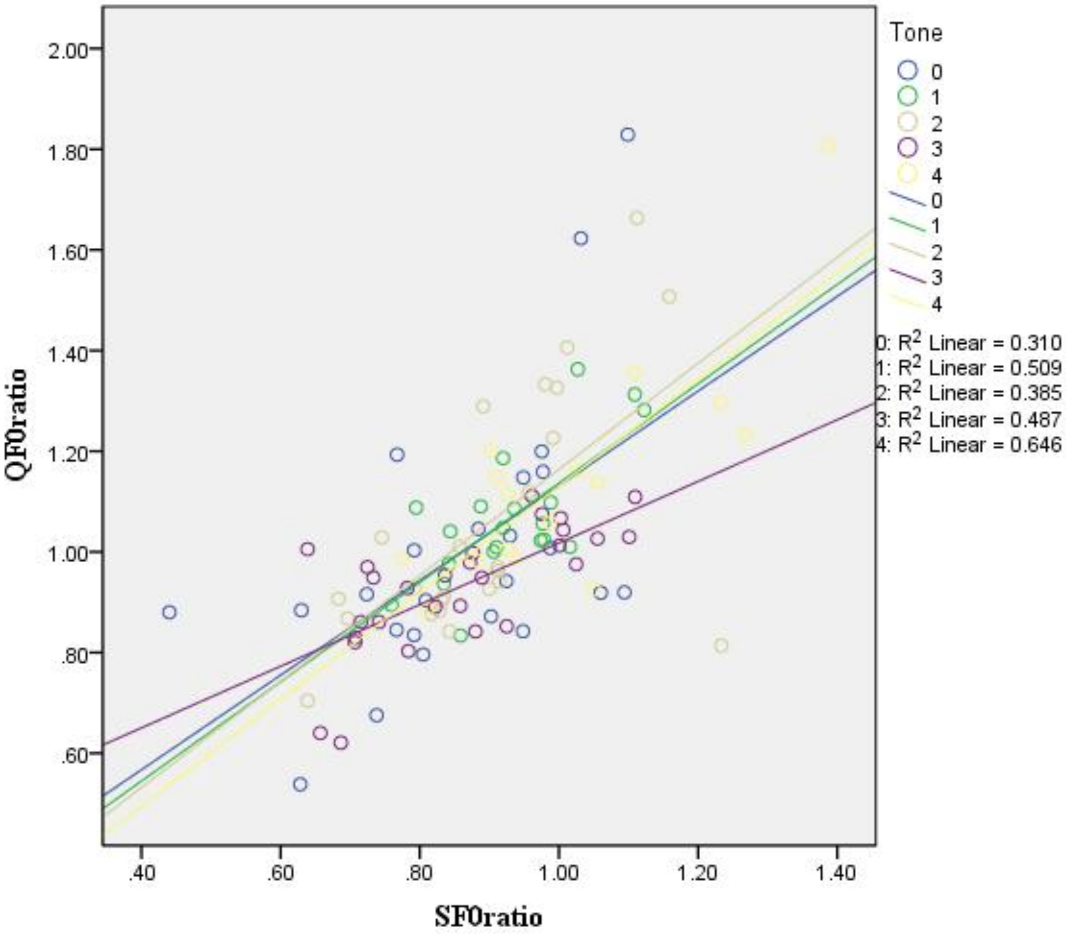


Figure A2. A scatter plot showing the relationship between the F0ratio for a statement (SF0ratio) and the F0ratio of the corresponding declarative question (QF0ratio) in Mandarin. The data points are marked by the tone type (Tones 0, 1, 2, 3, and 4) of the last word of each Mandarin sentence.

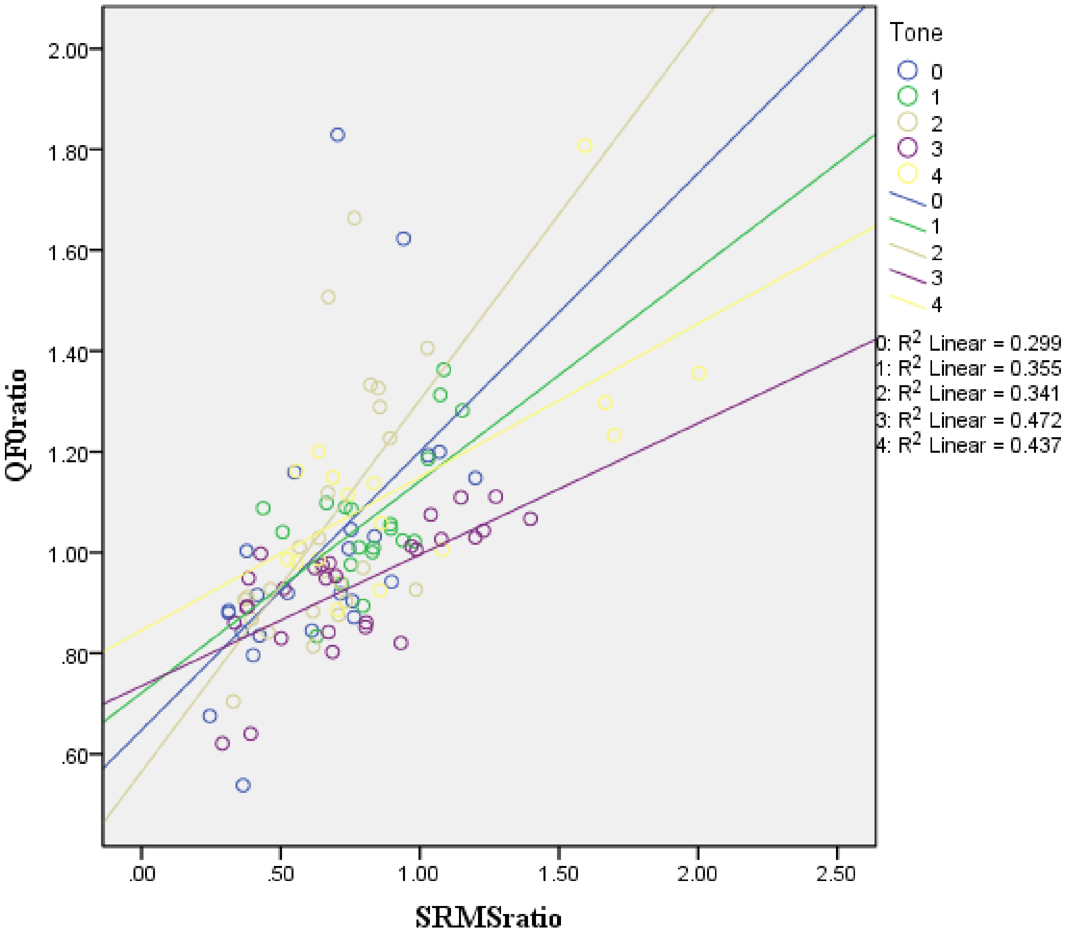


Figure A3. A scatter plot showing the relationship between the RMSratio for a statement (SRMSratio) and the F0ratio of the corresponding declarative question (QF0ratio) in Mandarin. The data points are marked by the tone type (Tones 0, 1, 2, 3, and 4) of the last word of each Mandarin sentence.

English

For English utterances produced by the native English speakers, neither SF0ratio (Beta = 0.163, t = 1.728, p = 0.087) nor SRMSratio (Beta = -0.004, t = -0.044, p = 0.965) was found to be useful for predicting QF0ratio. Figure A4 shows the scatter plot of SF0ratio on the x-axis and QF0ratio on the y-axis, with data points marked by the stress type (stressed and unstressed) of the last syllable of each English sentence. As shown in Figure A4, there was no relationship between QF0ratio and SF0ratio regardless of the stress type of the last syllable of each English sentence.

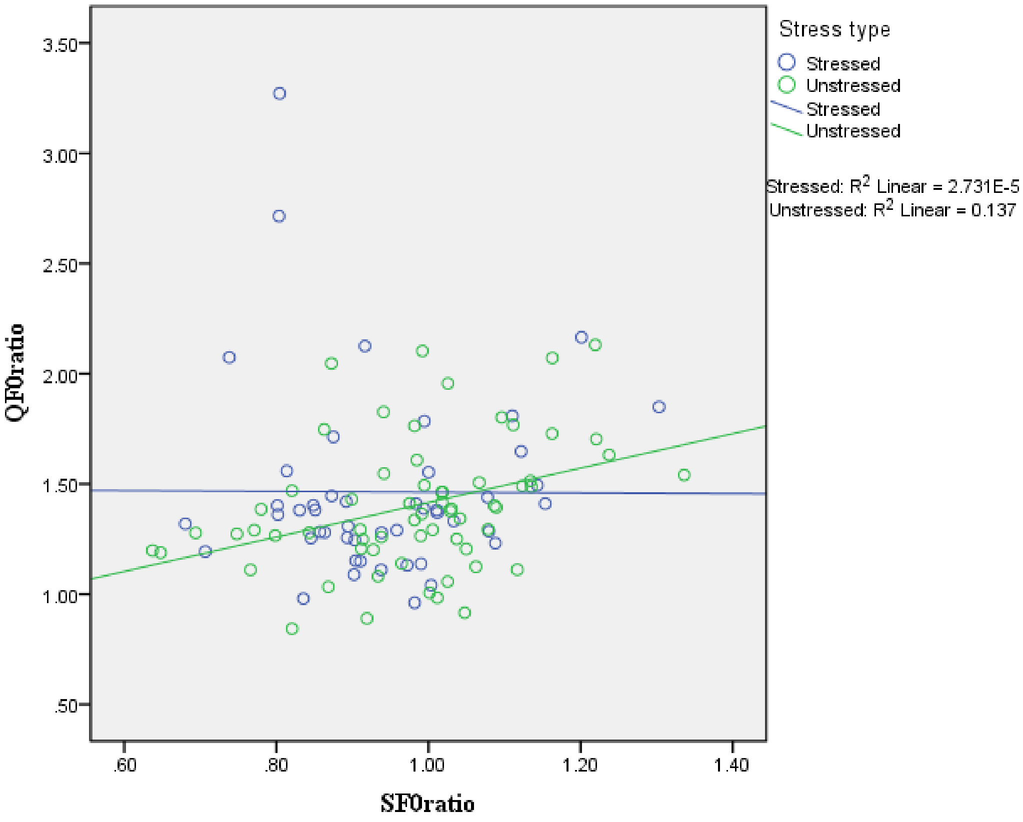


Figure A4. A scatter plot showing the relationship between the F0ratio for a statement (SF0ratio) and the F0ratio of the corresponding declarative question (QF0ratio) in English. The data points are marked by the stress type (stressed and unstressed) of the last syllable of each English sentence.

Native Taiwanese-and-Mandarin Speaker's English

For the English utterances produced by the native Taiwanese-and-Mandarin speakers, SRMSratio was found to be useful for predicting QF0ratio independently, with a significant Beta of 0.218 ($t = 2.345$, $p = 0.021$). The SF0ratio was removed from the regression model as it showed no significant correlation with QF0ratio (partial correlation = 0.057, $p = 0.551$). The regression model for the English data obtained from the native Taiwanese-and-Mandarin speakers can be written as:

$$\text{QF0ratio} = 1.266 + 0.248 * \text{SRMSratio}$$

This regression model, however, explains only 3.9% (Adjusted R-square = 0.039) of the variance of the QF0ratio in the English data obtained from the native Taiwanese-and-Mandarin speakers. As shown in Figure A5, the positive relationship between QF0ratio and SRMSratio found in the native Taiwanese-and-Mandarin speaker's English was only evident in the English sentences where the last syllable was stressed.

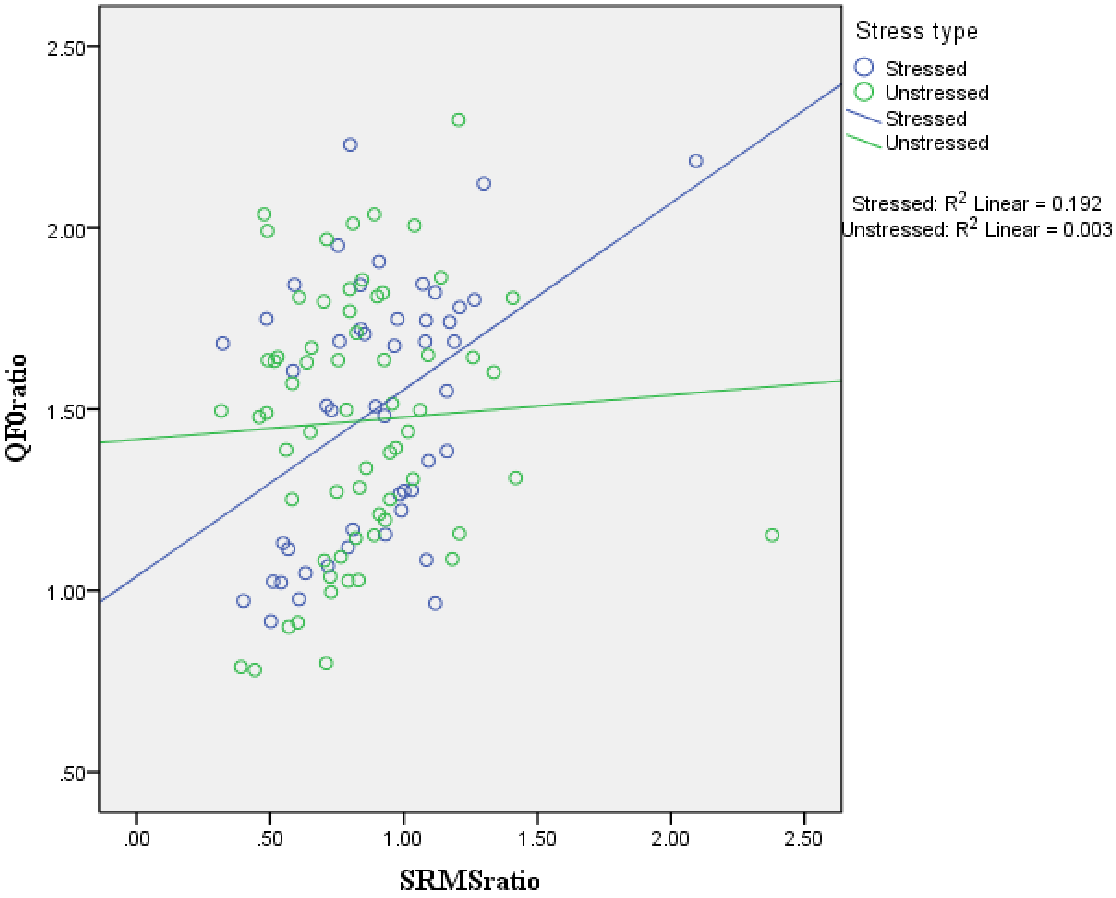


Figure A5. A scatter plot showing the relationship between the RMSratio for a statement (SRMSratio) and the F0ratio of the corresponding declarative question (QF0ratio) in the English produced by the native Taiwanese-and-Mandarin speakers. The data points are marked by the stress type (stressed and unstressed) of the last syllable of each English sentence.

Discussion

For all four types of language productions, the F0 and RMS of the last syllable, as well as the overall F0, of a sentence was found to be significantly higher in a declarative question than in a statement. The finding that F0ratio and RMSratio were both significantly higher in declarative questions than in statements suggests that both F0 and RMS were increased toward the end of a sentence to signal a declarative question. In particular, since the average F0ratio was higher than one for declarative questions, it can be concluded that a declarative question is characterised mainly by a higher F0 toward the end than in the rest of a sentence.

Although no significant language or language by sentence type interaction effect was found for the six acoustic measures investigated in this study, the relationship between the F0ratio of a declarative question and the F0ratio and RMSratio of a statement was found to vary between languages and between native and non-native speakers of English. For both tonal languages, namely, Taiwanese and Mandarin, the F0ratio of a statement was a strong predictor of the F0ratio of its corresponding declarative question. Moreover, Taiwanese, which has more tone types than Mandarin, was found to be associated with a greater increase of F0ratio (slope = 1.082) than Mandarin (slope = 0.758) for declarative questions compared to statements. In addition, the RMSratio of a statement was also found to be a good predictor for the F0ratio of its corresponding declarative question in Mandarin. In contrast, for English, the relationship between F0ratio of a question and that of a statement was absent. It appears that to maintain a sufficient contrast between lexical tones in a tonal language, F0ratio is consistently raised in declarative questions. In non-tonal languages, however, the F0ratio of a question is not raised consistently and is thus less predictable from that of a statement.

The lack of a relationship between the F0ratio of a statement and that of a question in the English utterances produced by the native English speakers was also found in the English utterances produced by the native Taiwanese-and-Mandarin speakers. However, the non-native English speakers seemed to raise the F0ratio of a declarative question in accordance to the RMSratio of its corresponding statement while the native English speakers did not. Although the effect of SRMSratio on QF0ratio appeared to be trivial as indicated by the low predictive power of SRMSratio on QF0ratio (i.e., only 3.9% of variances explained), the relationship shown between SRMSratio and QF0ratio in the non-native English productions suggests that non-native English speakers may be sensitive to the change of RMSratio in a sentence when adjusting the F0ratio in producing a difference between questions and statements in English. In other words, if the RMSratio is high in an English statement, a non-native English speaker is more likely to increase the F0ratio to a greater extent in its corresponding declarative question.

.