Listener-directed tone hyperarticulation:
The effects of noise and hearing loss on Mandarin tone production

by

SHUANG LIU

B.A., Sun Yat-sen University, 2013

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Master of Arts

Department of Linguistics

2015

This thesis entitled:
Listener-directed tone hyperarticulation: The effects of noise and hearing loss on Mandarin tone production
written by Shuang Liu
has been approved for the Department of Linguistics

_____
(Dr. Rebecca Scarborough)

_____
(Dr. Kathryn Arehart)

_____
(Dr. Kevin Cohen)

Date: May 14, 2015

The final copy of this thesis has been examined by the signatories, and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

IRB protocol # 15-0043

Abstract

Liu, Shuang (M.A., Linguistics)

Listener-directed tone hyperarticulation:
The effects of noise and hearing loss on Mandarin tone production

Thesis directed by Dr. Rebecca Scarborough

Speakers tend to accommodate listeners when communicate in difficult situations. Noise and hearing loss can induce similar or different difficulties for listeners due to the nature of noise and hearing loss. Speakers may be aware of the characteristics of difficulties experienced by listeners and make speech modifications accordingly. The current study aims to explore the similarities and differences between the effects of white noise and a simulated hearing loss on listener-directed Mandarin tone production. Mean *f0* and tone space dispersion were measured for four native Mandarin speakers.

Mean *f0* was found to be elevated when addressing a listener in white noise and a listener who was simulated to experience a hearing loss. The modification of tone space dispersion was found to be greater in hearing loss condition than in white noise condition. The results suggested that speakers were aware of the difference between the audibility problem and the clarity problem on the listener and adjusted their speech accordingly. The results can be explained in the model of Lindblom's H&H theory. When speakers are aware that the listener's access to information is blocked by some barriers, speakers will accommodate the listener by producing hyperarticulation in certain dimensions according to the nature of the barrier.

CONTENTS

Tables

Figures

1. INTRODUCTION

In daily communication, how details of speech are presented depends on the environment, physical and emotional state of the talker, and the composition of the audience (Uchanski, 2008). Two parties, speaker and listener, have influences on the form of speech. As modeled in Hyper- and Hypo-speech (H&H) Theory by Lindblom (1990), ideal speech keeps a good balance between listener-oriented output (clarity of the speech) and speaker-oriented output (economy of effort). Speech production is an adaptive process where utterances are modified on a continuum of hyper- and hypo-speech. Such adaptations are made based on the speaker's understanding of difficulties or advantages a listener has upon the access to the source of information and the speaker's judgement on the short-term demands for explicit signal information.

The hyperarticulated end of the continuum in Lindblom's H&H theory always involves a distinct speech style called "clear speech". According to Uchanski (2008), clear speech is adopted by speakers when speaking in a difficult communication situation, such as in a very noisy or reverberant environment or when talking to a hearing-impaired person. Two frequently studied clear speech styles are infant-directed speech (IDS) and speech produced in noise.

Though recently some researchers (Martin et al., 2015; McMurray et al., 2013) argued that IDS was hypo-speech but not hyper-speech, previously IDS was treated as hyper-speech and found to be clearer than adult-directed speech (ADS). Vowels were found to be further apart in acoustic space in IDS, which might contribute to infants' language acquisition (Kuhl

et al., 1997; Andruski and Kuhl, 1996). Prosodic patterns in IDS were found to be more informative than that in ADS and might facilitate infants understand the intent of communication (Fernald, 1989).

Foreigner-directed speech (FDS), a less frequently studied speech style, was recently linked to IDS in some studies. Listener-oriented forces on speech modification in IDS and FDS are similar in the sense that both a child/infant and a foreigner are linguistically inferior to the speaker due to their limited linguistic capacity (Biersack et al., 2005). Similarities and differences between IDS and FDS in terms of acoustic-phonetic adjustments were investigated. Uther et al. (2007) found vowels were equally hyperarticulated in IDS and FDS while pitch was higher in IDS than in FDS. Biersack et al. (2005) found prosodic features for child-directed speech (CDS), a similar speech style to IDS, and FDS were different. The above studies suggested that speakers were able to capture characteristics of the two groups of listeners and address them accordingly.

A similar connection may be drawn between speech to a listener in noise and speech to a hearing-impaired listener. In laboratory settings, a clear speech style is sometimes elicited by the instruction that to speak as clearly as possible as if one communicates in *a noisy environment* or with *a hearing-impaired listener*, as done in the landmark study of Picheny et al. (1985). It seems that speech produced in noise (to a listener), can be linked with speech to the hard of hearing in the sense that perceptual difficulties induced by noise to a normal-hearing listener is similar to perceptual deficit experienced by a person with impaired hearing. For example, in studies of hearing loss, using masking noise with normal listeners is a common

approach to simulate the effect of elevated thresholds associated with hearing impairment (Moore, 2007b). However, a listener with certain type of hearing loss may experience different problems than a listener in noise.

If we compare closely the effect of noise and the effect of a hearing-impaired listener on speech modification by taking into account different types of noise and different types of hearing-impaired listeners, we may notice that difficulties faced by listeners in the two cases may vary, and speakers may be correctly aware of the difference between these difficulties and adjust their speech accordingly to accommodate listeners, just like what they do when addressing a foreigner and an infant.

2. BACKGROUND

2.1 Perception difficulties induced by hearing loss and noise

First let's briefly review some types of hearing loss and see how they cast different perception deficits on a listener. Hearing loss can be divided into conductive hearing loss, sensorineural hearing loss and mixed hearing loss (a combination of the former two types). Without considering more details, the nicely general description given to the public by American Speech-Language-Hearing Association (ASHA) gives us a good idea of the similarities and differences of symptoms of these types of hearing loss.

It is written as:

Conductive hearing loss occurs when sound is not conducted efficiently through the outer

ear canal to the eardrum and the tiny bones (ossicles) of the middle ear. Conductive hearing loss usually involves a reduction in sound level or the ability to hear faint sounds[1].

Sensorineural hearing loss (SNHL) occurs when there is damage to the inner ear (cochlea), or to the nerve pathways from the inner ear to the brain. …SNHL reduces the ability to hear faint sounds. Even when speech is loud enough to hear, it may still be unclear or sound muffled[2].

(ASHA does not give a description of the symptoms of a mixed hearing loss. One may think of the symptom as a combination of the above symptoms.)

Obviously, both conductive hearing loss and SNHL result in an elevation of the threshold, but only SNHL resulted in a clarity problem in addition to the audibility problem. In other words, it is easy for listeners with SNHL to confuse different speech sounds even when the speech is audible.

Similarly, a noise may cause only an audibility problem or both audibility and clarity problems. A white noise, which is a pure energetic masker, may only result in the speech to be inaudible. An energetic masker has some spectral overlap with the speech signal. Brungart (2001) reported that performance in speech perception when the noise was an energetic masker decreased monotonically with decreasing signal-to-noise ratio (SNR).

Noise like a competing talker may also result in the speech to be inaudible because of the spectral overlap. However, even when both the target speech and the competing talker are audible, listeners may be still unable to disentangle the element of the target speech from the

[1] http://www.asha.org/public/hearing/Conductive-Hearing-Loss/

[2] http://www.asha.org/public/hearing/Sensorineural-Hearing-Loss/

competing talker (Brungart, 2001).

Perception difficulties induced by hearing loss and noise are summarized in Table 1.

Table 1 General perception difficulties induced by hearing loss and noise.

		Audibility problem (threshold problem)	Clarity problem (suprathreshold problem)
hearing loss	Conductive hearing loss	√	
	Sensorineural hearing loss	√	√
noise	**white noise**	√	
	competing talker	√	√

In terms of speech perception, different types of hearing loss and different types of noise may induce similar or different perceptual difficulties for the listener. In terms of speech production, similarities or differences in perceptual difficulties on the listener's side may be aware by the speaker and speakers may make similar or different speech modifications accordingly. Lu and Cooke (2008) proposed that while the task of speech production was different from the task of speech perception, production might be influenced by perceptual concerns. Speakers may be able to predict perceptual difficulties in communicative environment at the ears of their interlocutor.

To investigate if the assumption on the speaker's side is true, the current study chose to present white noise and to simulate a hearing loss which causes both an audibility problem and a clarity problem on the listener.

2.2 Cross-linguistic concern

Adjustments were made to both prosodic and phonological properties in clear speech.

According to the excellent review of clear speech made by Smiljanić and Bradlow (2009), previous studies have shown that clear speech involves "a decrease in speaking rate (longer segments as well as longer and more frequent pauses), wider dynamic pitch range and greater sound-pressure levels", etc. English clear speech has been consistently shown to have a feature of vowel space expansion, i.e., distances between vowel categories in the acoustic F1 × F2 space were found to be increased. Formant targets were better achieved, thus making vowel categories more salient and probably less perceptually confusable.

Smiljanić and Bradlow (2009) also pointed out that numerous studies on the production of clear speech have been done on English while studies in other languages have received considerably less attention. However, it is very probable that clear speech modifications at both prosodic and phonological levels are shared by speakers of different languages. While investigating how global prosodic features are adjusted from conversational speech to clear speech across languages, it will be especially interesting to explore how the degree of contrast between language-specific phonological categories is adjusted. A systematic understanding of how acoustic cues maintain and strengthen phonological contrast should be built on cross-linguistic investigation of clear speech patterns.

To date, clear speech studies on tonal language are few. The only two found by the author were the study on Cantonese Lombard speech conducted by Zhao and Jurafsky (2009) and the study on Mandarin infant-directed speech conducted by Liu et al. (2007). For tonal languages, the most distinct phonologically contrastive categories are lexical tone categories. Compared to the effect of clear speech on segmental phenomena, lexical tone has received very little

attention. Thus lexical tone naturally became the focus for Zhao and Jurafsky's and Liu et al.'s studies as well as the current study.

Following Bradlow et al. (1996) 's rationale of using vowel space expansion as a measure of contrast between vowels, Zhao and Jurafsky calculated tone space dispersion in their 2009 study to measure the contrast between Cantonese lexical tones. The current study aimed to follow the rationale of Bradlow et al (1996) and Zhao and Jurafsky (2009) to measure contrast between Mandarin lexical tones.

Conducting experiments on Mandarin tone production not only enables us to explore the phonological feature, tone space, but also the prosodic feature, mean *f0*. By measuring fundamental frequency, mean *f0* and tone space can be explored at the same time. These two features may interact with each other in the process of speech modification.

In summary, the present study aimed to investigate Mandarin tone production when a speaker is speaking to a listener who is subjected to the influence of a white noise and/or a hearing loss. Certain prosodic and phonological speech modifications of lexical tone are expected to be explained by the nature of difficulties experienced by the listener.

3. RESULTS OF PREVIOUS STUDIES AND HYPOTHESES

3.1 Noise induced speech modification

Noise has long been known to affect speech. It was first found by Etienne Lombard in 1911 that vocal effort was increased when talkers spoke in noise. This effect is called after the

scientist's name "Lombard". Lombard suggested that this modification was because speakers could not hearing themselves in noise. Vocal effort was increased in order to monitor their voice through auditory feedback. Most of the classic findings in this type of study can be represented by a relatively recent study conducted by Summers et al. (1988). They found that amplitude, duration and *f0* were all increased while speakers were *talking in noise alone*.

The following paragraphs will review in detail some of the mean *f0* results reported by previous studies. Figure 1 is the results from Summers et al. (1988), comparing the mean *f0* in quiet, 80, 90 and 100 dB SPL for two speakers. Both speakers produced higher *f0* in noisy environments than in quiet. Lombard speech of non-tonal languages is characterized by elevated mean *f0.*

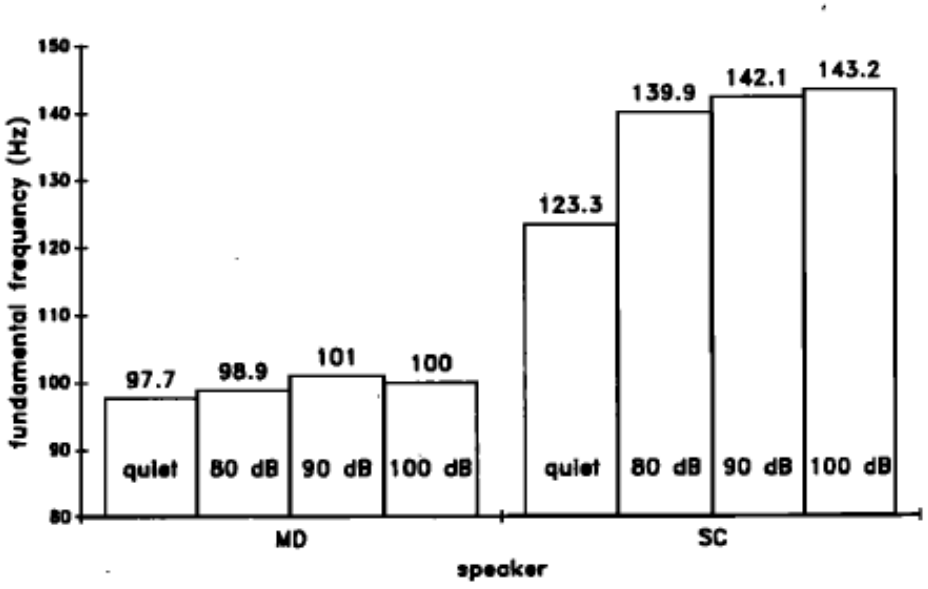


Figure 1 Mean fundamental frequency values for words produced in quiet, 80, 90, and 100 dB SPL of masking noise. After Summers et al. (1988), Fig. 3.

However, it is easy to discover that talker SC produced much higher *f0* when speaking in noise than in quiet, but the *f0* produced by talker MD is just a little bit higher in noise than in

quiet. The two talkers performed differently maybe because there was no communication involved in this study as the author explained. There was no motivation for the speakers to consciously change their speech even with noise presented in the headphones. Lane and Tranel (1971) also suggested that when speaking to themselves, speakers do not need to make hyperarticulation adjustments to let themselves hear better their own voice. Lombard effect may become more obvious in a communicative environment, namely, an environment where a listener is involved and the listener is also subjected to the energetic masking of noise.

Based on this rationale, Cooke and Lu (2010) designed speaking tasks with communication and without communication. They did find that in communicative tasks, *f0* produced in a noisy environment was elevated more than in non-communicative tasks. The results are shown in Figure 2.

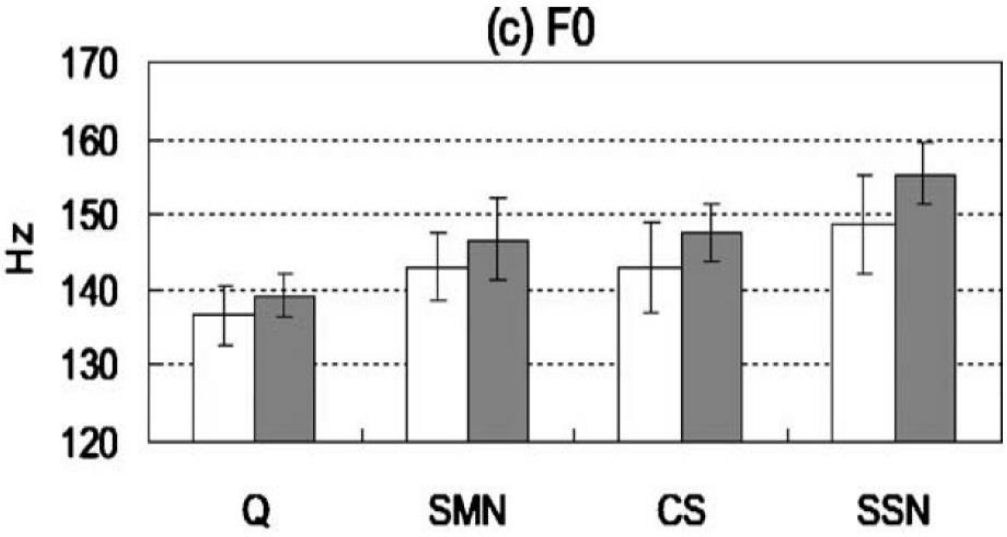


Figure 2 Mean fundamental frequency produced in quiet, masking of speech-modulated noise, competing speech and speech-shaped noise. After Cooke and Lu (2010), Fig.1. White bars represent the results for non-communicative tasks. Gray bars represent the results for

communicative tasks. "Q" is quiet (no noise). "SMN" is speech-modulated noise. "CS" is competing speech. "SSN" is speech-shaped noise.

All types of noise were presented to both speakers and listeners in the communicative task. Obviously, gray bars (communicative) are higher than white bars (non-communicative). We may conclude that speaking alone in noise may result in an increase in *f0*, but the amount of elevation is limited compared to that when speaking to *a* listener in noise. Difference between gray bars and white bars was due to speaker's awareness of difficulties experienced by the listener. The current study is based on the rationale of H&H theory (Lindblom, 1990), where a listener's access of information is a concern for the speaker's speech modification. Thus the non-communicative task is not of interest to the current study.

In addition to the results on mean *f0* of English from Summers et al. (1988) and Cooke and Lu (2010), results on mean *f0* of Cantonese lexical tone from Zhao and Jurafsky (2009) are also related to the interest of current study. They found that when speaking in a white noise, *f0* of all 6 Cantonese tones were raised. Though no listeners were involved in Zhao and Jurafsky's study, their results may further indicate that when the listener is in noise, speakers will produce Mandarin lexical tones with higher *f0*.

As for the phonological feature, tone space, Zhao and Jurafsky (2009) reported that Cantonese lexical tone space dispersion was not increased from quiet to noise. However, in their study speakers spoke in noise alone. The author doubted that a communicative environment may make a difference. For the moment, no strong predictions can be made on the change of Mandarin tone space dispersion when speakers are addressing a listener in noise. Tone space dispersion may or may not be increased in this condition.

Above all, the prosodic feature mean *f0* is expected to be elevated at the level of Mandarin lexical tones when the speaker is aware that the listener is hearing in noise. Tone space dispersion may or may not be increased.

3.2 Hearing-impaired listener induced speech modification

Picheny et al. (1986) found that *f0* was slightly higher when speakers were instructed to speak as clearly as possible as if speaking in a noisy environment or to a hearing impaired person than were instructed to speak conversationally. To date, no study has reported that there was a rising in *f0* when talking to real hearing-impaired listeners. The effect of a real hearing-impaired listener on *f0* may be similar to that of an imagined hearing-impaired listener. It seems that higher *f0* in speech modification could be due to the speaker's feeling that the listener is subject to a threshold problem. The speaker's feeling can result from either a real or imagined hearing loss on the part of the listener. In terms of mean *f0*, noise and hearing loss may have similar effects on speech modification, that is, an elevation in *f0*.

Contrast between phonologically contrastive categories, for example vowels, were found to be increased when addressing a hearing-impaired listener. Vowel space expansion, as introduced in section 2.2, is an indicator of how vowel categories are contrastive from each other. Ferguson and Kewley-Port (2007) found that for some speakers produced a larger vowel space expansion when they were instructed as "*It is important that you speak clearly, so that a hearing-impaired person would be able to understand you*" than when talkers were instructed to read sentences as they would in everyday conversation. In the instruction describing

speaking to a hearing-impaired listener, the phrase “speak clearly” and the verb “understand” indicated that the listener may be experiencing some suprathreshold problems.

Vowels are important for achieving meaning contrast in both tonal and non-tonal languages while lexical tone has a similar role in a tonal language as vowel quality. If a speaker produces larger vowel space dispersion when he is aware that the listener is not able to hear clearly, will a tonal language speaker produce larger tone space dispersion in the same situation?

Following the rationale of using vowel space expansion as an indicator of vowel contrast, Zhao and Jurafsky (2009) found that Cantonese tones of low-frequency words were produced with more contrast than those of high-frequency words, as measured by tone space dispersion. Low-frequency words are less predictable than high-frequency words according to Zhao and Jurafsky (2009). Though no listener was presented in Zhao and Jurafsky’s study, speakers might be aware that low-frequency words were more likely to become less clear for a listener, so they produced more contrast between lexical tones for those words. Zhao and Jurafsky’s results may further indicate that when Mandarin words are not clear for a listener due to hearing loss, speakers will produce more contrast between lexical tones for the listener.

Summary of hypotheses:

A listener in white noise and a listener with simulated hearing loss may show similar effects on a speaker’s Mandarin tone production. The similarity in speech modification is expected to be in the form of an elevation in mean *f0* (a prosodic feature). Speakers may be aware that both white noise and hearing loss cause an audibility problem on the listener.

The presence of white noise and hearing loss to a listener may cause a different effect on a speaker's Mandarin tone production. This difference is expected to be shown in the size of tone space dispersion (a phonological feature). Speakers may be aware that it is more difficult for a listener with simulated hearing loss to hear clearly than a listener in white noise.

4. METHODOLOGY

4.1 The language: Mandarin

Mandarin has four tones in its lexical tone inventory. Tone 1 is a high-level tone. Tone 2 is a mid-rising tone. Tone 3 is a low-dipping tone. Tone 4 is a high-falling tone. Typical *f0* contours of the four tones, produced by a pilot speaker in the current study, are shown in Figure 3.

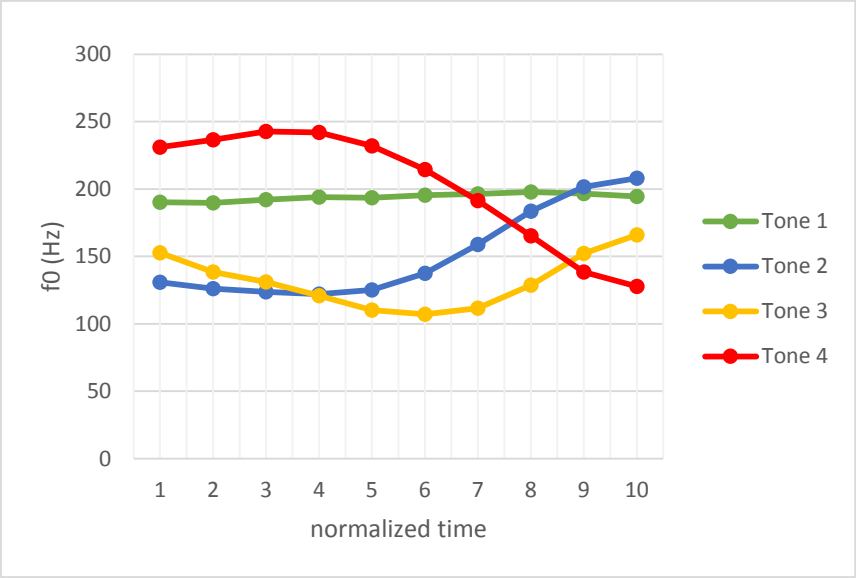


Figure 3 f0 contours of lexical tones in Mandarin produced by a pilot speaker.

The sound of a Chinese character contains one syllable and one tone. For example, the character "马" is read as /ma/ with the low-dipping tone, meaning "horse". A word in Chinese can contain one or more characters. The definition of the term "word" in Chinese is an unclear one. Thus, "word" will not be used in following sections. Instead, "character" is used.

4.2 Stimuli

Since lexical frequency was known to have effects on lexical tone production (Zhao and Jurafsky, 2009), the current study only used high-frequency Mandarin characters as the stimuli. *A List of Commonly Used Characters in Modern Chinese* was published by the *National Language Committee of People's Republic of China* in 1988[3]. This list contains 3,500 characters which are relatively commonly used in media and daily life compared to other Chinese characters[4]. Among these 3,500 characters, 2,500 characters were found to be the mostly commonly used ones. The 20 characters (5 syllables × 4 tones) used in the current study were selected from the list of 2,500 characters[5]. Subjects who received at least primary and secondary education in China should be quite familiar with those characters. The 20 characters were listed in Table 2.

Table 2 Characters used as stimuli

Tone / Syllable	Tone 1 (level)	Tone 2 (rising)	Tone 3 (dipping)	Tone 4 (falling)
di	低	敌	底	地
ge	哥	隔	葛	个
guo	锅	国	裹	过
liu	溜	刘	柳	六
ma	妈	麻	马	骂

[3] 关于发布《现代汉语常用字表》的联合通知: www.china-language.gov.cn/wenziguifan/shanghi/013a.htm

[4] 《现代汉语常用字表》说明: www.china-language.gov.cn/wenziguifan/shanghi/013b.htm

[5] 常用字（2500 字）笔画顺序表: www.china-language.gov.cn/wenziguifan/shanghi/013c.htm

4.3 Participants

Eight native Mandarin speakers, four males and four females, participated in the experiment for $10/h compensation. Subjects' ages ranged from 22 to 27. All of them were born and educated in China before they came to the United States for college or graduate school. All subjects reported no known speech, language or hearing disorders.

4.4 Experiment design and test procedures

To create a communicative environment in which a speaker was not just speaking to himself, either a real listener or an imagined listener could be used. Previous studies conducted research in both ways. Scarborough et al. (2007) reported that additional acoustic-phonetic adjustment might be made to imagined listeners rather than real listeners in listener-directed speech. To avoid the possible additional adjustment, the current study invited a confederate listener to sit in the sound booth with the speaker during recording.

All subjects participated in two sessions. In the first session, speakers were told that the listener had normal hearing. Indeed, the listener had normal hearing from a previous hearing screening. Speakers had a short communication with the listener prior to recording, so they experienced that the listener had normal hearing. In the second session, speakers were told that the listener had impaired hearing. Indeed, the listener wore a pair of bright yellow ear plugs to reinforce the impression of the speaker that the listener is a hearing-impaired one. Importantly, the speaker was told that the listener was subjected to a clarity problems in addition to the

audibility problem. Though the real hearing loss experienced by the listener in the second session was a conductive one, which might only involve a threshold problem, the symptoms were described as a combination of a threshold problem and a suprathreshold problem. All speakers had very limited knowledge of hearing loss, so they were not able to judge if the suprathreshold problem is true or not. The term "sensorineural hearing loss" and "conductive hearing loss" were not included in the instruction to avoid any unnecessary confusion for speakers.

In the second session, the listener wearing earplugs sit next to the speaker. Prior to recording, Mandarin description was read loudly to the speaker when the listener was also in the booth. The Mandarin description was "他不光听不见，还听不清，他不太容易听出来正确的音是哪个" which could be translated as *He cannot hear faint sounds, also he cannot hear clearly. He cannot easily recognize what the sound was.* The confederate listener was told to act as though he suffered from a clarity problem. The listener asked the experimenter, "What did you just say about me? (你刚才说我什么?) I didn't hear it clearly (我没听清)". Speakers saw that the listener had heard something about himself but did not hear clearly the experimenter's description. Some speakers also had very short conversation with him, such as asking him "You can't hear well now? (听不见啦?)". The confederate listener behaved similarly and kept emphasizing that he couldn't hear clearly.

In both sessions, the speaker was told to read the character appeared on the screen to the listener with the carrier sentence "这个字是__" (*This character is__*) as if he was dictating the listener. A similar carrier sentence was used in Scarborough and Zellou (2013). The listener

was sitting in the booth facing the door but not the computer screen, so the speaker knew that the listener could not see what the character was. During recording, speakers, facing the listener, spoke to him but the listener did not respond to the speaker. Speakers performed according to their impression of the listener's symptoms. No response was designed here to avoid additional adjustments that were out of control.

In both sessions, speakers talked in both noisy and quiet environments. The noise was a white noise presented binaurally through a SENNHEISER HD 280 headphone at 70 dB SPL to the speaker and through an ATH-M40sf headphone to the listener. A sound level meter with ear simulator was used to calibrate the level of the noise at the headphone prior to each experiment. The speaker clearly knew that the white noise was presented to the listener as well. The speaker was told to imagine the situation as he was speaking to the listener next to a waterfall.

The fact that the speaker was also wearing headphones could result in inadvertent increase of vocal effort as reported by some previous studies that *f0* was elevated when speaking alone in noise. However, there were also results shown that *f0* was only elevated to a minimum amount when speaking alone in noise. *f0* result from Summers et al. 1988 is a good example. Though we cannot rule out the difference made by the noise at the ear of the speaker, we can say that when the goal is to speak to someone who is listening in the same environment noise, most portion of the speech modification must be deliberate but not inadvertent.

If the speaker did not wear headphones, the speaker would not know how difficult the situation is for the listener, because no response from the listener was designed in this study for

the purpose of experiment control. The task would be more unnatural and the speech modification would result from the speaker's pure imagination of a noise presented at the ear of the listener.

Two types of listeners (normal hearing and hearing-impaired) and two types of environments (quiet and noisy) resulted in four experiment conditions. All speakers spoke to normal hearing listeners in the first session and hearing-impaired listeners in the second session. In each session, all speakers spoke in quiet first and then in noise. A long break, about 15 minutes, was inserted between the two sessions and a short break, about 5 minutes, was put between two environment conditions in each session. The experiment procedure was summarized in Table 3.

Table 3 Experiment procedure

<table>
<tr><th>session</th><th>listener</th><th>environment</th><th>Condition</th></tr>
<tr><td rowspan="3">1</td><td rowspan="3">normal hearing (NH)</td><td>quiet (Q)</td><td>NH-Q</td></tr>
<tr><td colspan="2">short break (5 min)</td></tr>
<tr><td>noise (N)</td><td>NH-N</td></tr>
<tr><td colspan="4">long break (15 min)</td></tr>
<tr><td rowspan="3">2</td><td rowspan="3">hard of hearing (HOH)</td><td>quiet (Q)</td><td>HOH-Q</td></tr>
<tr><td colspan="2">short break (5 min)</td></tr>
<tr><td>noise (N)</td><td>HOH-N</td></tr>
</table>

Under each experiment condition, each speaker produced the 20 characters shown in Table 2 (Data structure is reproduced in Table 10 in section 5.2.2). Before recording, characters were shown to the speaker to make sure that they were indeed familiar with those characters. In each experiment condition, the 20 characters were randomly presented on the

screen for the speaker. Every character is displayed in relatively large size as shown in the following figure.

Figure 4 Example of a character shown on the computer screen.

The experimenter controlled the presentation by monitoring the speaker's sound. After she heard the speaker finished a character, she pressed a button, then the next character appeared. The recording took place in the sound booth of the Phonetics Lab at University of Colorado Boulder. The audio was recorded using an Earthworks M30 measurement microphone. The sampling rate was 44.1 kHz.

4.5 Labeling and measurement

The production of four subjects were discarded because of the following reasons:

a. The first subject (male) experienced an extra-long NH-quiet condition (the first condition) because some technical problems happened to the microphone. He stopped and restarted many times. This somehow made him become nervous, which resulted in unnatural production.

b. There was one character which should be pronounced as tone 1 but one subject (female)

pronounced it as tone 4. This does not mean that this character was less familiar to the subject than other characters but because this character indeed have two pronunciations. This resulted in unequal number of productions of tone 4 and tone 1 in each experiment condition, so the sample size of tone 1 and that of 4 are different for this subject.

c. Two subjects (one male and one female) produced pervasive creaky voice with large amount of irregular pulses in the NH-quiet condition, which made the pitch measurement almost impossible for that condition.

The productions of the other four subjects, two males and two females, were carefully measured and became the data.

The long sound file of each experiment condition was segmented into short sound files of individual characters. Character boundaries were hand-labeled and annotated by the author using Praat. *f0* contours in the vowel were automatically measured for each character by a commonly used Praat script *ProsodyPro* created by Xu (2013).

In order to accurately compare the differences between f0 contours over multiple tokens, the *f0* values of the tone trajectory of each character were measured at 10 evenly spaced time points as done by Zhao and Jurafsky (2009).

For each character, the domain of time normalization and *f0* measurement was the vocalic segment. All preceding consonants were discarded. The boundaries between consonants and vowels were manually labeled by the author.

f0 tracking during creaky voice was manually corrected as shown by the following figures. In Figure 5, pulses were incorrectly tracked which resulted in extremely high *f0* values for

creaky voice. Figure 6 shows manually corrected pulses.

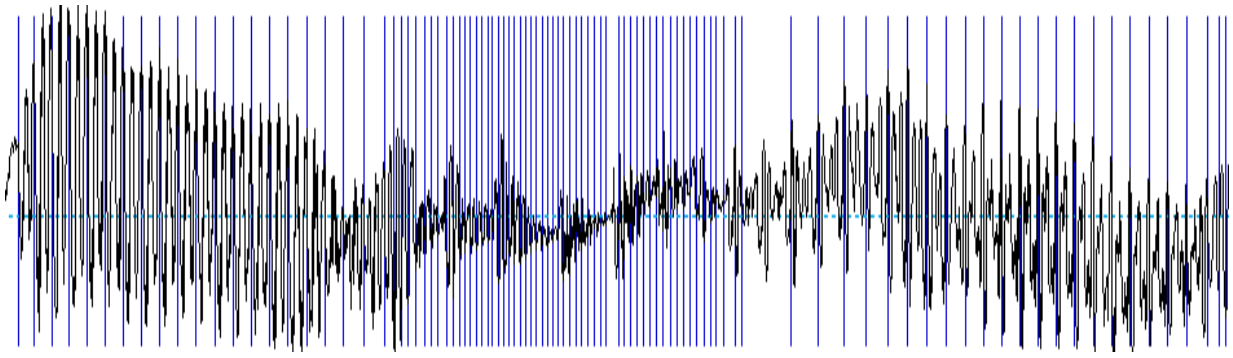


Figure 5 Wrong pulses automatically measured by Praat.

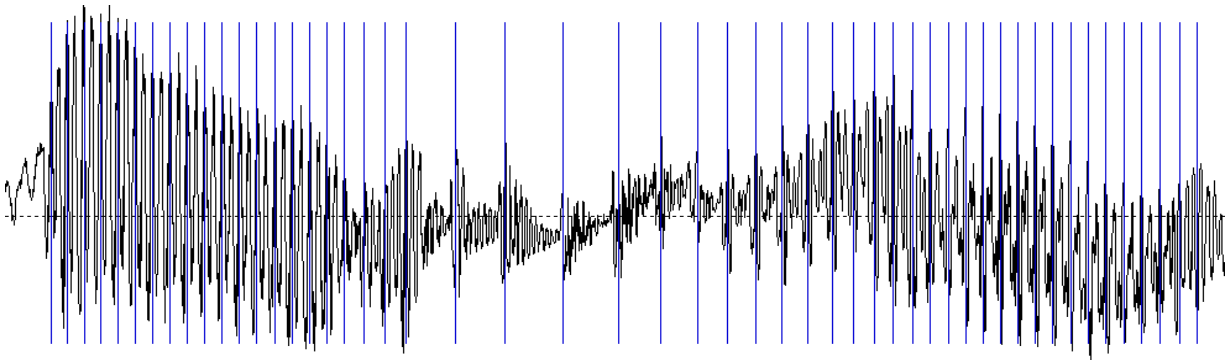


Figure 6 Manually corrected pulses.

5. RESULTS AND DISCUSSION

5.1 Effect of noise and hearing loss on *f0*

The mean *f0* of a single production was calculated by averaging across the ten points. For example, the following figure shows a single production of tone 2. The mean *f0* of this production is the average of the values at the 10 points, as shown by the single dot.

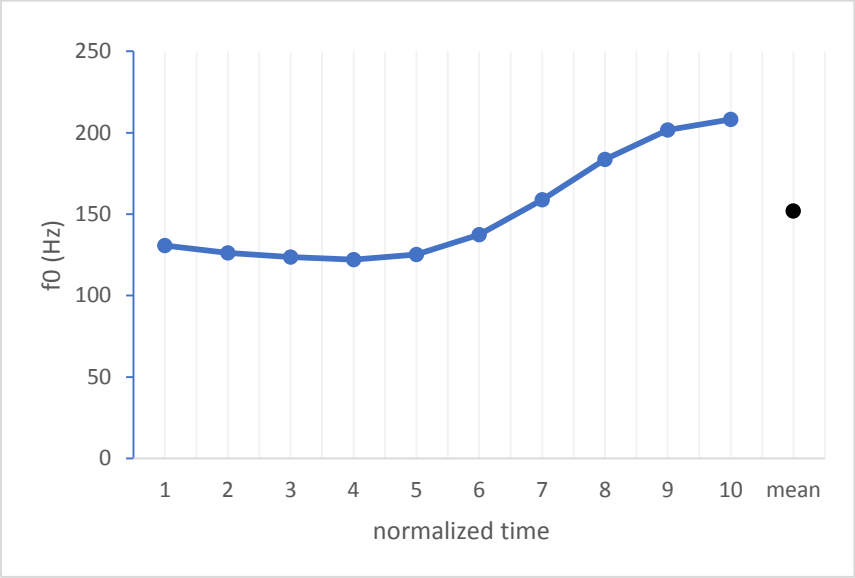


Figure 7 A single production of tone 2 and the mean f0 averaged across time.

A Repeated Measures ANOVA was performed to examine whether environment (quiet vs. noise), listener type (normal hearing vs. simulated hearing impaired), and tone type (tones 1–4) had a significant influence on mean *f0* and also whether interactions happened between these main effects. ANOVA results for main effects and interactions were summarized in Table 4.

Table 4 ANOVA results of main effects and interactions for mean f0.

	df	F statistic	Significance
Environment	(1, 19)	67.019	< .001
Listener	(1, 19)	121.563	< .001
Tone	(3, 57)	183.697	< .001
environment × tone	(3, 57)	5.789	< .01
listener × tone	(3, 57)	15.557	< .001

5.1.1 Main effects

Environment had a significant effect on mean *f0* (F (1, 19)[6] = 67.019, p < 0.001). Speaking in noise resulted in a higher *f0* compared to speaking in quiet (M_noise = 231.791, M_quiet = 199.912). For each tone in Figure 8 below, the orange bar representing mean *f0* produced in noise is higher than the blue bar representing mean *f0* produced in quiet.

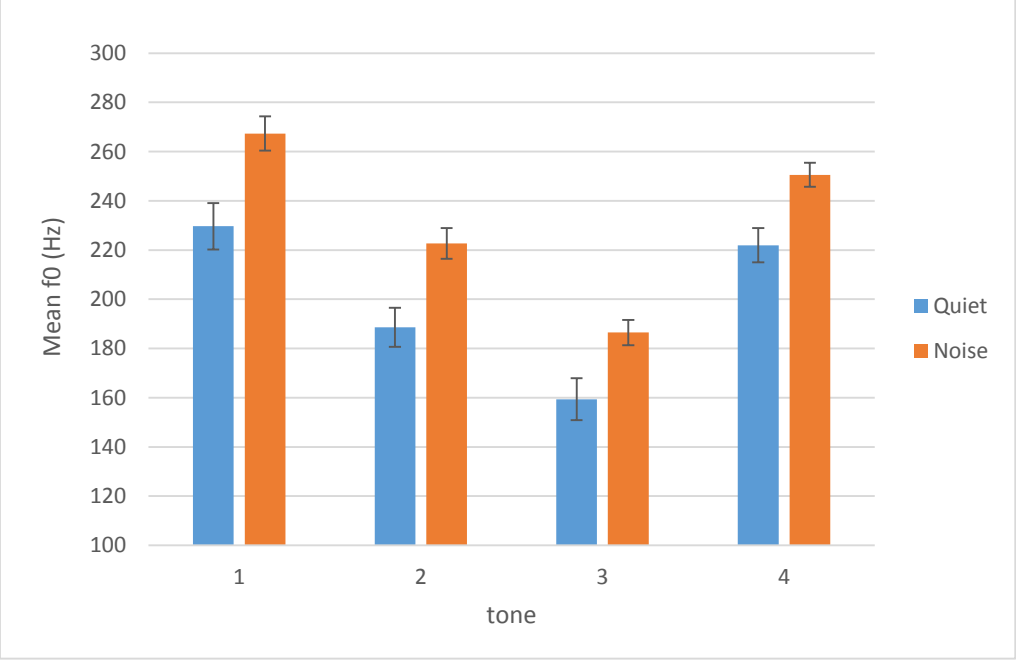


Figure 8 Mean f0 produced in quiet and noise broken up by tones. Standard errors are shown by the error bars.

[6] 20 observations for one tone under each condition resulted from 4 speakers × 5 syllables. Syllable was not a factor for the current study. See Table 2 in section 4.2 and Table 10 in section 5.2.2.

The trends can also be seen from *f0* contours. Figure 9 shows the *f0* contours of four tones produced in quiet and noise conditions averaged across listeners, syllables and speakers. Orange lines represent *f0* contours produced in noise and blue lines represent that produced in quiet. From the beginning to the end of the *f0* contour, the orange dot is always higher than the blue dot.

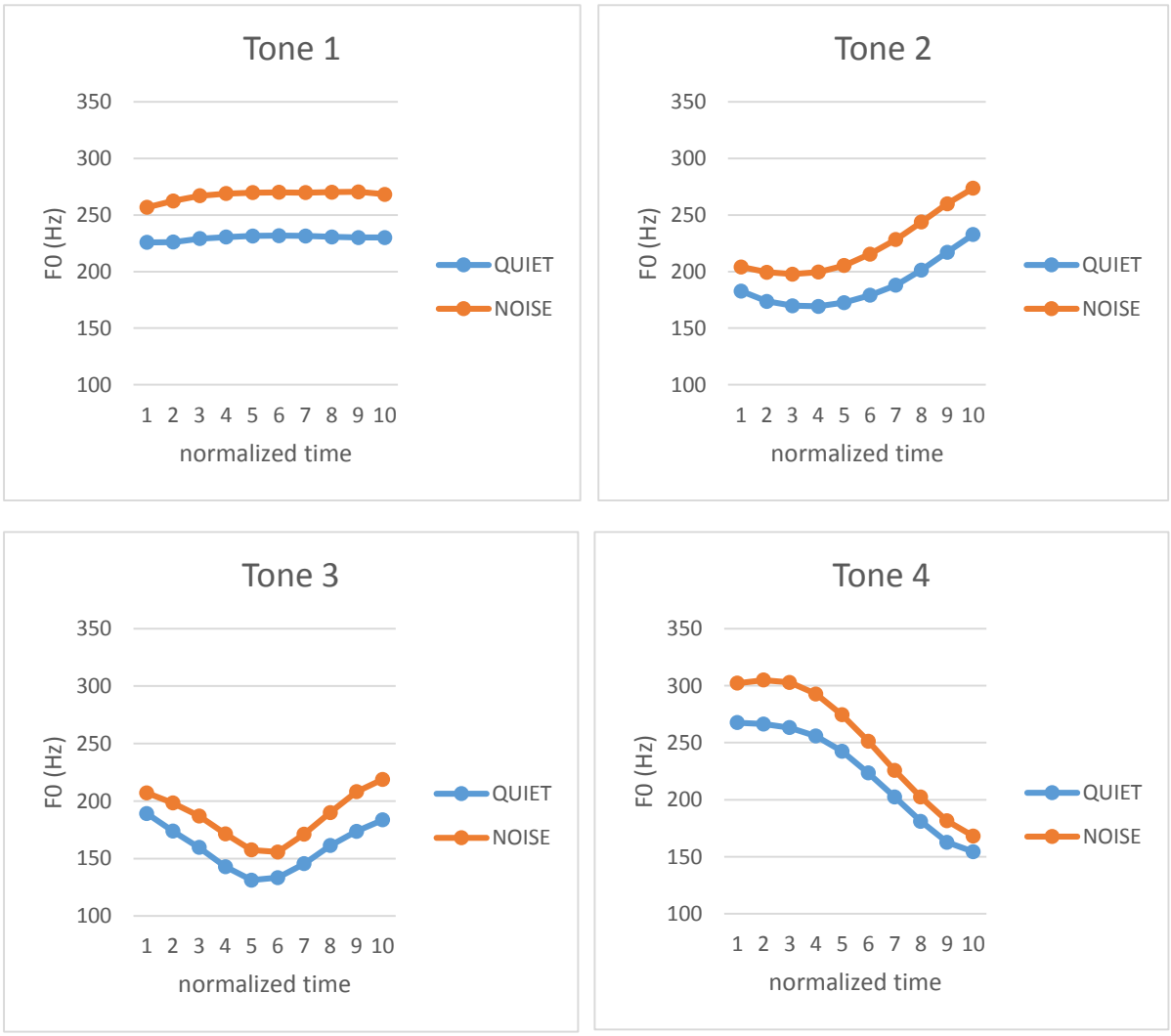


Figure 9 f0 contours of four tones produced in quiet and noise averaged across listeners.

Similar to the environment factor, listener had a significant effect on mean *f0* (F (1, 19) =

121.563, p < 0.001) too. Speaking to the hard of hearing produced higher *f0* than speaking to the normal hearing (M_HOH = 239.051, M_NH = 192.652). For each tone in Figure 10, the yellow bar representing the mean *f0* produced for the hard of hearing is higher than the green bar representing that produced for the normal hearing.

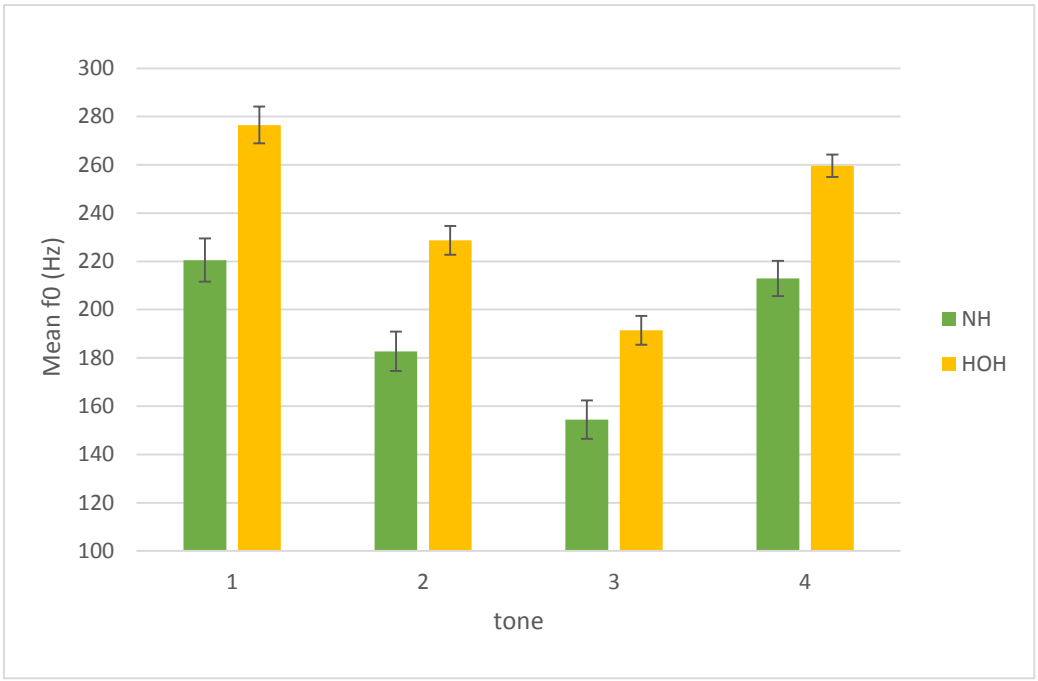


Figure 10 Mean f0 produced for normal hearing listener (NH) and hard-of-hearing listener (HOH) broken up by tones. Standard errors are shown by the error bars.

The trends can also be seen from *f0* contours. Figure 11 plots the *f0* contours of four tones produced for the normal hearing and the hard of hearing averaged across environments, syllables and speakers. In each tone category, the yellow dots are higher than the green dots from the beginning to the end of the contour.

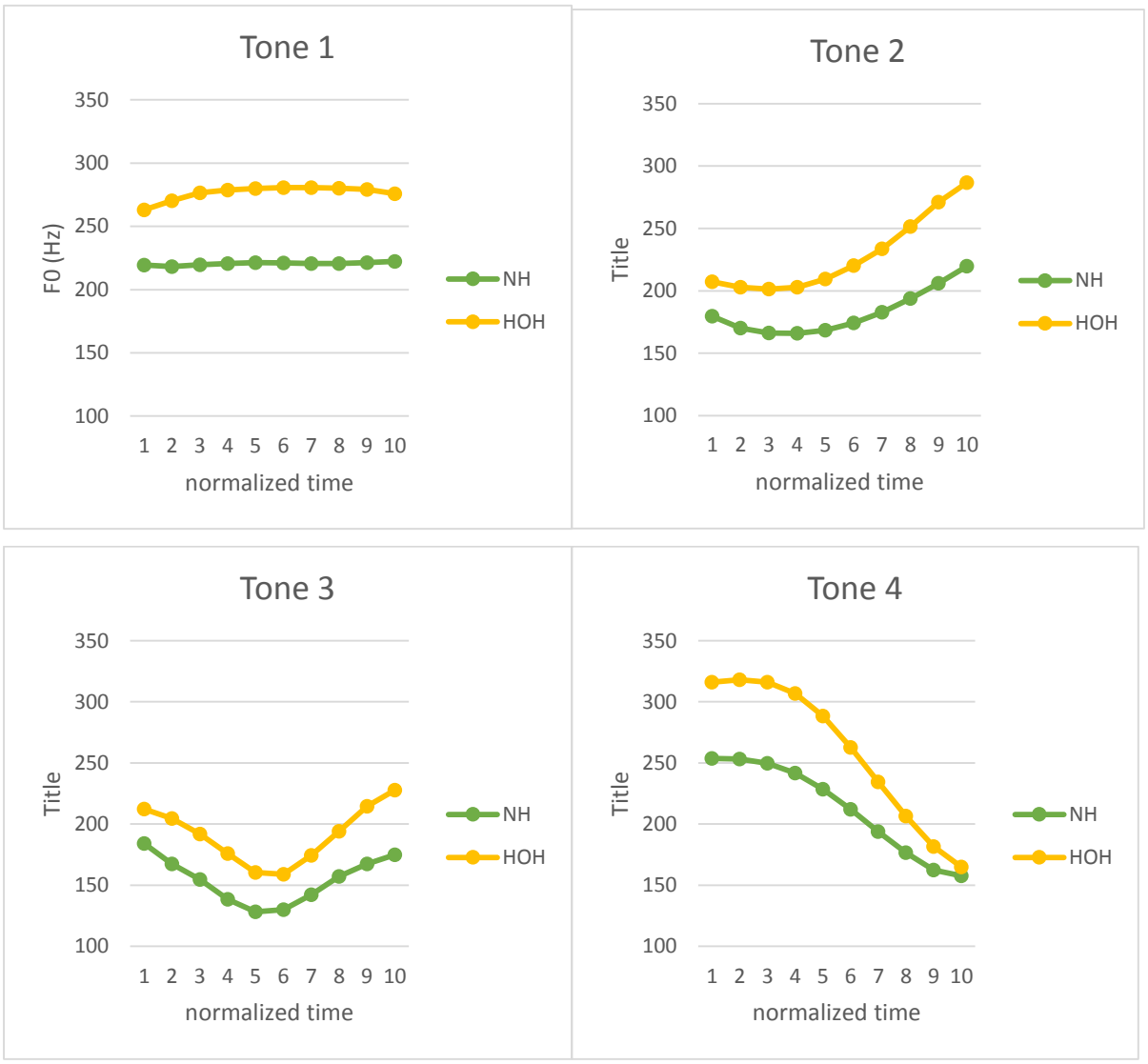


Figure 11 f0 contours of four tones produced for the normal hearing (NH) and the hard of hearing (HOH) averaged across environments, syllables and speakers

Needless to report, tone type had a significant effect on mean *f0* (F (3, 57) = 183.697, p < 0.001). It can be seen from Figure 8 and 10 above that tone 1 always had the highest mean *f0* while tone 4 was always the second highest. Mean *f0* of Tone 2 was lower than that of tone 4 but higher than that of tone 3.

The above results showed that both listener and environment had a significant influence on the mean *f0* of Mandarin lexical tones. The effects were similar in the sense that mean *f0*

went higher when speakers were addressing a listener who was hearing in white noise or who had simulated hearing loss. It seems that speakers were aware that the threshold problem was shared by a listener with hearing loss and a listener hearing in white noise.

5.1.2 Interactions

There was an interaction between environment and tone (F (3, 57) = 5.789, $p < 0.01$) and also an interaction between listener and tone (F (3, 57) = 15.557, $p < 0.001$). It can be seen from Figure 8 and 10 above that for each tone mean *f0* was significantly increased from NH to HOH and from Quiet to Noise. The amount of change for different tones might be different, but the significance of change were the same for all tones. However, the change of mean *f0* from tone to tone were not significant in all environments and all listener conditions.

In noise conditions, mean *f0* of tones in pair were significantly different from each other for all pairs (all p-values were less than 0.005 in post-hoc Bonferroni tests). This can be seen from the right panel of Figure 12 below that all bars in pair were significantly different from each other. In quiet conditions, mean *f0* of tones in pair are significantly different from each other for all pairs except the pair of tone 1 and tone 4. Post hoc Bonferroni tests showed that in quiet mean *f0* of tone 1 and that of tone 4 were not significantly different from each other (M_tone1 = 229.696, M_tone4 = 221.938, p = 0.541). It can be seen from the left panel of Figure 12 that the black bars in pair are not so different from each other in height.

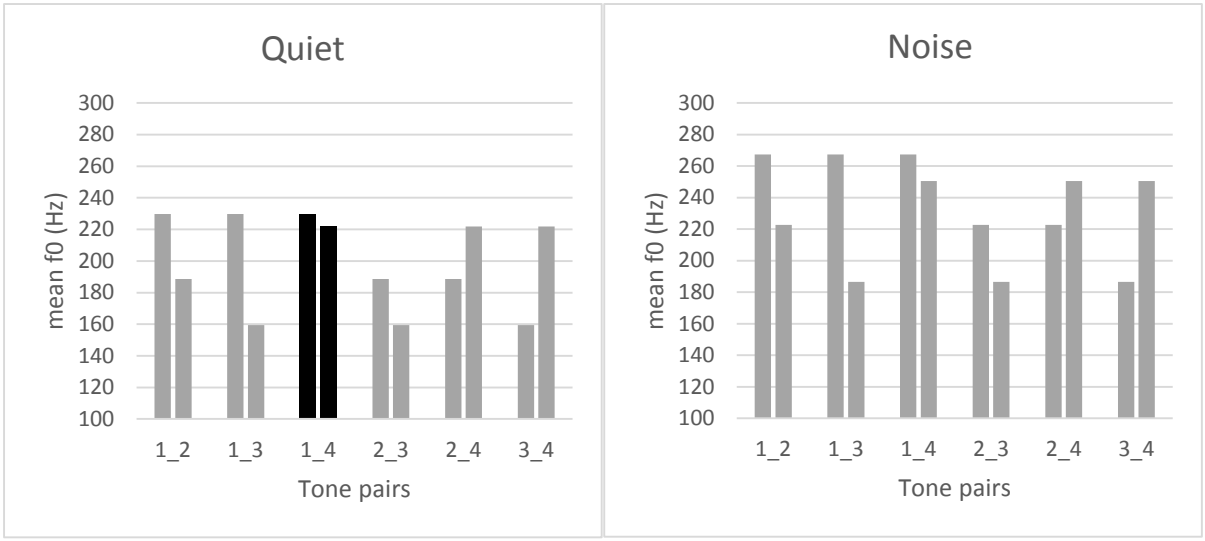


Figure 12 Mean f0 of tone pairs in quiet and noise averaged across listener conditions.

In addition, it can be seen from Figure 12 above and Table 5 below that the *f0* differences between most tones in pair were more obvious in Noise than in Quiet.

Table 5 Mean difference of mean f0 of tones in pair in Quiet and Noise.

tone pairs	MD in Quiet		MD in Noise
1_2	41.058	<	44.643
1_3	70.321	<	80.863
1_4	7.758 (not *)	<	16.779
2_3	29.263	<	36.22
2_4	33.3	>	27.864
3_4	62.563	<	64.084

Similarly, in HOH conditions, all tones in pair were significantly different from each other in mean *f0* (all p-values were less than 0.05) as can be seen from the right panel of Figure 13 below. In NH conditions, all tones in pair were significantly different from each other in mean *f0* except the pair of tone 1 and tone 4, as can be seen from the left panel of Figure 13. Post-hoc Bonferroni tests showed that in the NH conditions tone 1 and tone 4 are not significantly

different from each other (M_tone1 = 220.538, M_tone4 = 212. 935, p = 0.241)

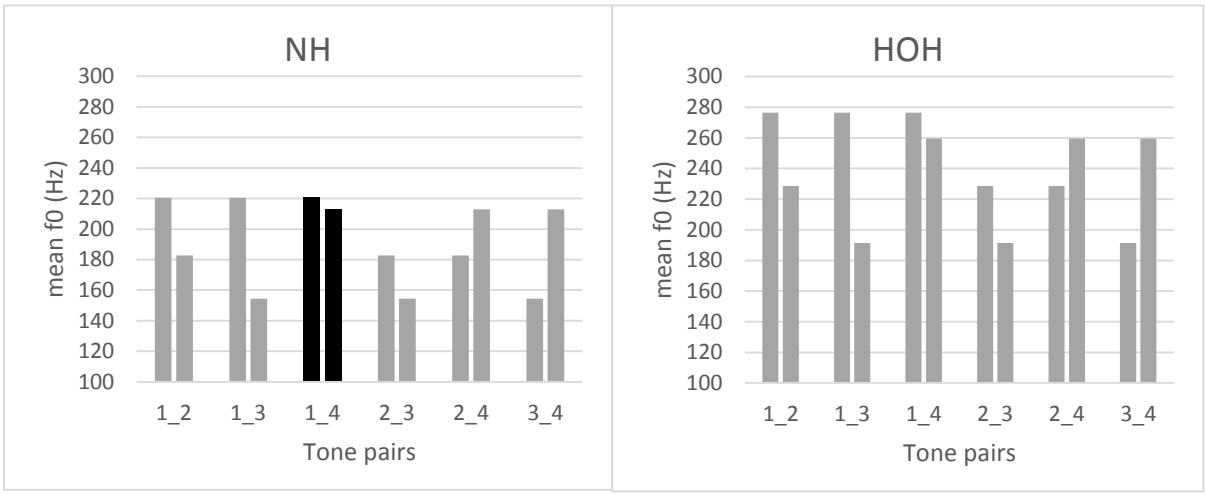


Figure 13 Mean f0 of tone pairs in NH and HOH averaged across environment conditions.

In addition, it can be seen from Figure 13 above and Table 6 below that the *f0* differences between most tones in pair were more obvious in HOH than in NH.

Table 6 Mean difference of mean f0 of tones in pair in NH and HOH.

tone pairs	MD in NH	relationship	MD in HOH
1_2	37.836	<	47.865
1_3	66.106	<	85.078
1_4	7.603 (not *)	<	16.934
2_3	28.27	<	37.213
2_4	30.233	≈	30.931
3_4	58.503	<	68.144

These interactions suggested that in more difficult conditions (Noise and HOH), the difference between mean *f0* of different tones might be bigger than that in easy conditions (Quiet and NH). This is especially a case for the difference between tone 1 and tone 4 in terms of statistical significance.

5.1.3 Interim discussion

In the current study, *f0* in Mandarin tone production became higher when addressing a listener hearing in noise or a listener with simulated hearing loss. This similarity may be due to the fact that both white noise and hearing loss cause an audibility problem for the listener. Speakers were aware this fact and made similar speech modifications.

Does that mean the elevation of *f0* was a deliberate speech modification to overcome audibility problem for the listener? Intuitively the strategy for overcoming the audibility problem is to increase the vocal intensity. In the current study, sound intensity was not reported, but during the experiment the experimenter monitored the waveform of speech through the software Audacity. When the gain of the microphone and the distance between the speaker and the microphone werr maintained, the amplitude of speech waveform was much larger in NH_noise, HOH_quiet and HOH_noise conditions than in the baseline NH_quiet condition. Apparently speakers increased their voice intensity when addressing the listener in noise and/or with a hearing loss. It seems that it is not necessary to raise *f0* to make the sound more audible. Why *f0* was raised?

A coupling effect that the *f0* always rises with intensity is a well-accepted phenomena (Gramming et al., 1988). Some researchers claimed that increasing *f0* in noise was unconscious and a by-product of other speech modifications such as increasing vocal intensity (Cooke and Lu, 2009). If this is true, then the raising in *f0* was only an indicator of the speaker's deliberate effort in overcoming the audibility problem and raising *f0* was not the goal of speech modification. Other researchers claimed that speakers can control *f0* while manipulating vocal

intensity (Bordon et al., 1994), which means speakers can increase vocal intensity to make speech more audible while keeping the *f0* the same. If this is true, then the observed increase in *f0* was the goal of speech modification at least in the HOH_quiet condition, if the direct influence of the noise on the speaker could not be ruled out completely (see section 4.4).

Either the *f0* increase was conscious or unconscious, it seems reasonable to take increase in *f0* as an *indicator* of the speaker's deliberate speech modification. Put another way, an elevation of *f0* indicates that the speaker was tending to accommodate the listener who had an audibility problem, but it does not mean that the goal of speech modification was to raise *f0*.

5.1.4 Summary

In terms of signal degradation in audibility, the influences of white noise and hearing loss on the listener were similar. Speakers elevated the mean *f0* when addressing a listener listening in noise and when addressing a listener who was simulated to experience a hearing loss.

In the next section, we move onto the difference between the influences of white noise and simulated hearing loss. The hearing loss was simulated to cause a suprathreshold/clarity problem for the listener. White noise may not have such influence on the listener (see section 2.1 and 2.2). Speakers may be aware of the difference and make different speech modifications when addressing the two types of listeners.

5.2 Effect of noise and hearing loss on tone space dispersion

5.2.1 Method review

Though duration and turning point (the time point at which the *f0* contour changes from falling to rising) could be the contrastive acoustic/perceptual features for Mandarin tones (Shen and Lin, 1991; Moore and Jongman, 1997), the most salient contrast between tones might be manifested by the difference between *f0* contours of various tones types.

The dispersion of different tones' *f0* contours was chosen as the measurement for the difference/contrast between tones for the current study. The rationale for calculating tone space dispersion followed the intuition of calculating vowel space dispersion developed by Bradlow et al. (1996). Zhao and Jurafsky (2009) established a valuable example of measuring tone space dispersion in Cantonese. The method used in the current study was built on their metric.

In Bradlow et al (1996), mean Euclidian distance of individual vowels from the center of the F1-F2 vowel space was used to represent vowel space dispersion. In Figure 14 below, the location of each vowel token is decided by its F1 and F2 measurement. The point in the center represents the center of gravity of the speaker's vowel space. The length of each line represents a vowel token's distance from the center. A certain talker's vowel space dispersion is calculated as the mean of these distances.

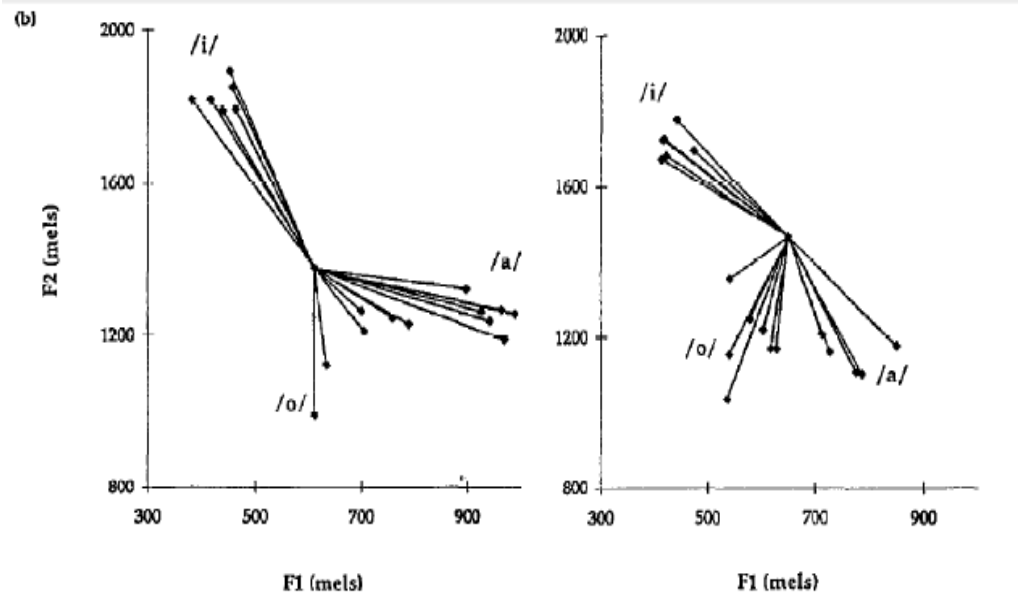


Figure 14 Illustration of vowel space dispersion. After Bradlow et al. (1996).

Since this metric was found to be correlated with word frequency (Munson and Solomon, 2004), Zhao and Jurafsky (2009) adopted the intuition when developing the measurement for tone-space dispersion. Therefore, tone-space dispersion was calculated as the mean Euclidean distance of individual tones from the center of the speaker's fundamental frequency space.

Two important issues are needed to be discussed regarding the measurement of tone-space dispersion. The first issue, mentioned by Zhao and Jurafsky (2009), is about the center of the tone space, i.e., the centroid. The two options are to use a single global value for all time points or a specific value for each time point. Zhao and Jurafsky (2009) claimed that changes in tone space might not happen homogenously across the tone trajectory. The current study agrees with this claim. As can be seen from Figure 15 below, each tone has a specific fundamental frequency value at a certain time point. The dispersion of the observed values at a certain time point changes as the time point moves (the dispersion changes as a function of time). Thus to

use a single global value for all time points is inappropriate in this theoretical account. For example, to measure the dispersion at the fifth time point and the eighth time point in Figure 15, two different centroids are calculated by the observed values at the two time points.

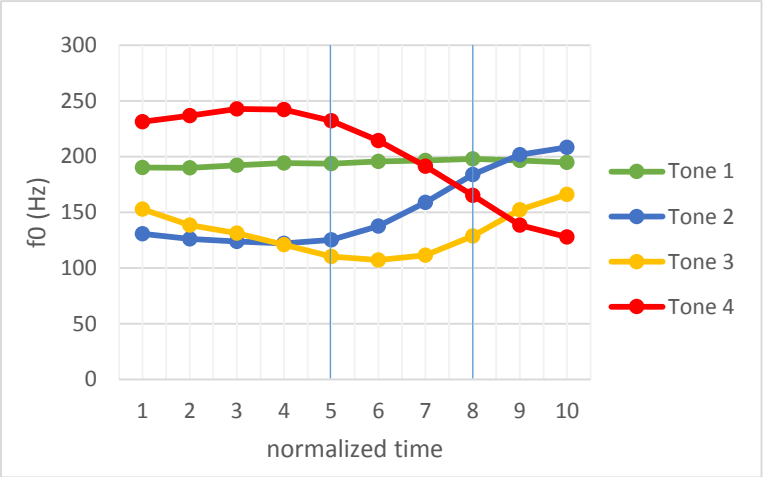


Figure 15 Example of observed values and dispersion at two time points.

Another important issue is on what scale the dispersion should be measured. One could choose either the acoustic scale (Hertz) or the perceptual scale (Semitone, Mel or Bark, etc.). Previous studies measured vowel space dispersion on both scales according to the particular research interests. For example, Bradlow (1995) compared the vowel categories of English and Spanish on Hertz scale to reveal the principles that determine the acoustic realizations of the vowel of these two languages. In her other study (Bradlow, 1996), the F1 and F2 measurements of English and Spanish vowels were plotted on the perceptually motivated mel scale for the purposes of the comparison between the acoustic and perceptual vowel categories. Moreover, when speech intelligibility is involved in the research interest, vowel space was always measured on a perceptual unit scale such as mel or bark (e.g., Bradlow et al., 1996; Wright, 2004).

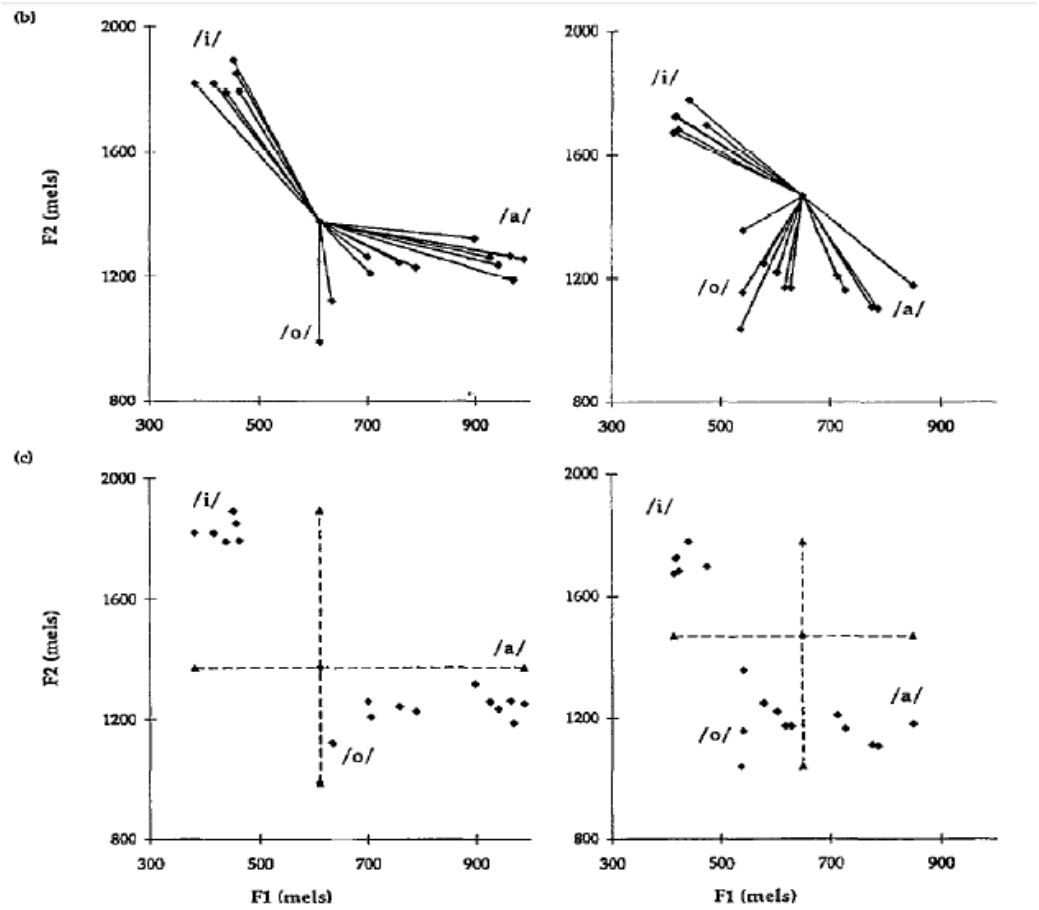


Figure 16 Panels on the left show the vowel space of a high-intelligibility speaker while panels on the right show that of a low-intelligibility speaker. After Bradlow et al. (1996).

A related issue is on which scale the center of the tone space should be found. It is reasonable to find the center on the acoustic scale if the goal is to measure dispersion on the acoustic scale. Similarly, if one is going to measure dispersion on a perceptual scale, the center should be found on the perceptual scale. For example, Bradlow et al. (1996) found the center on mel scale when exploring the correlation between speech intelligibility and vowel space. In Figure 16 above, all measurement of vowel formant frequency values in Hertz were converted to mel. The center of the vowel space, i.e. the central point in panel (b), was found by the point of intersection of the dashed line segments in panel (c). Dashed line segments represent the range of the F1/F2 observed values in mel.

If understood correctly, Zhao and Jurafsky (2009) found the center of the tone space on the Hertz scale while their goal was to measure tone dispersion on the semitone scale. This method might be problematic. Since the calculation of tone-space dispersion in the current study was built on but not same as their metric, in the following sections first we review their metric to see if it is indeed problematic and if corrections are needed to be made.

First we take a hypothetical example to show how Zhao and Jurafsky's measurement was performed. Let 100 Hz, 150 Hz, 250 Hz and 350 Hz be the observed values in Hertz. Using Zhao and Jurafsky's method, one should first get the mean of these four values. In this case, the mean is 212.5 Hz. This mean is called by Zhao and Jurafsky the *central f0*. In Figure 17 below, observed values in Hertz are shown by the blue dots on the horizontal axis. The mean, 212.5 Hz, is shown by the black triangle dot on the horizontal axis.

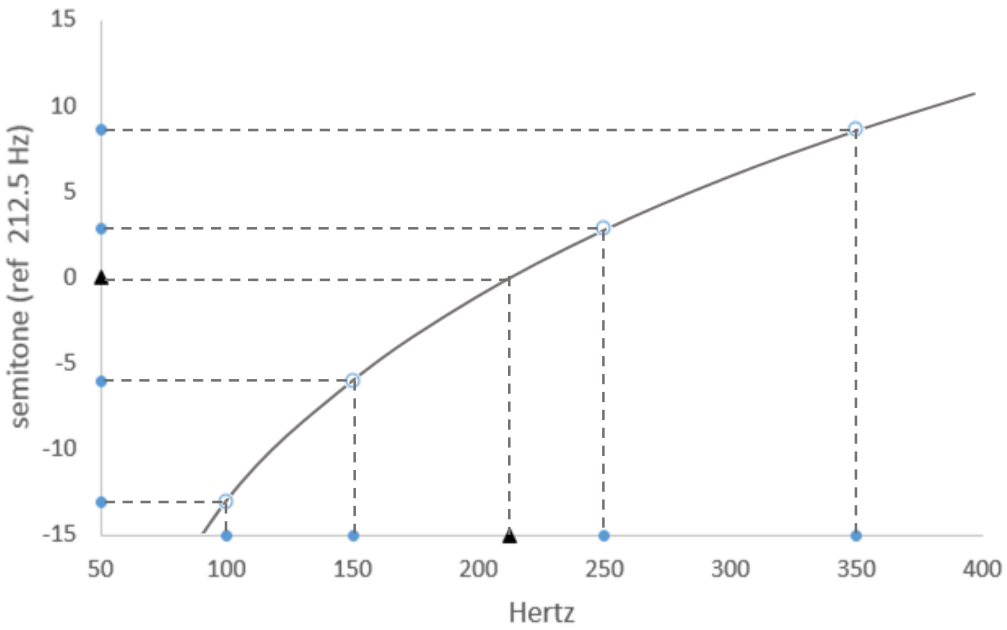


Figure 17 Illustration of Zhao and Jurafsky's metric of tone space dispersion.

Their next step is to find the differences in semitone (st) between 100Hz and 212.5Hz,

150Hz and 212.5 Hz, 250 Hz and 212.5Hz and 350Hz and 212.5Hz. Finding the difference in semitone between the observed Hertz value and the Hertz mean is equivalent to assigning a semitone value for the observed Hertz value using the Hertz mean as the reference.

The formula is:

$$\mathrm{y} = 12 * \log_2 \frac{x}{212.5}$$

where *x* is the observed Hertz value and y is the assigned semitone value.

The function is shown by the curve in Figure 17 above. The assigned semitone values for the Hertz values are -13.0 st, -6.0 st, 2.8 st and 8.6 st, which are shown by the blue dots on the vertical axis. The black triangle dot on vertical axis represents the semitone value for 212.5 Hz, which is 0 st, because 212.5 Hz is the reference. The final step is to calculate the mean of the absolute values of those semitone results, that is, 7.6 st.

Zhao and Jurafsky's method actually found out the mean of distances between blue dots and the black dot on the vertical axis. However, the black dot is not the center for those blue dots on the semitone scale. If the goal is to find the dispersion of the observed values on semitone scale, probably the center should be found directly on the semitone scale. One should not use a semitone value converted from a "center" Hertz value as the center of the observed values on the semitone scale.

If one uses mean of the observed values on the semitone scale as the center of gravity on the semitone scale, the value is -1.9 st, shown by the red dot on the vertical axis in Figure 18 below. The dispersion is calculated as 7.6 st, which is same as the result given by Zhao and

Jurafsky's method. Does that mean it does not matter to use the either the black dot or the red dot to measure the dispersion of blue dots on the vertical axis?

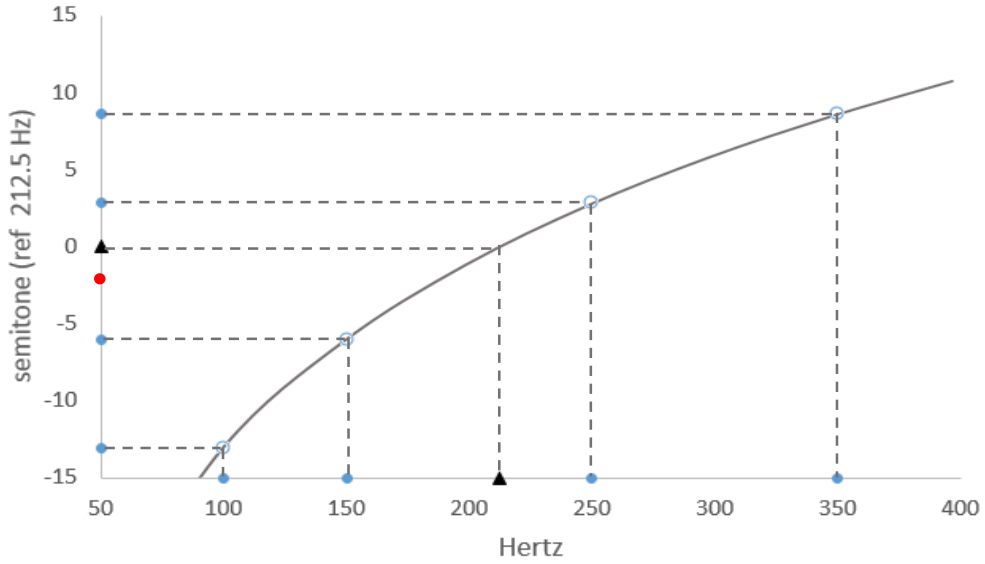


Figure 18 The red dot is a more appropriate center for blue dots on vertical axis than the black dot.

The following example illustrates that the answer is no.

Why in the above case did the red dot and the black dot give us the same result? This is because the black dot and the red dot are in the same interval (They are all located between the middle two blue dots). As shown by the following illustration, as long as dot O is located between dot B and dot C, the mean distance from A, B, C and D to O is always equal to $\frac{BC+AD}{4}$.

However, when dot O is not between dot B and dot C, the mean distance is equal to $\frac{AD+CO+BO}{4} = \frac{AD+BC+2CO}{4}$ which is bigger than $\frac{BC+AD}{4}$.

Therefore, on the vertical axis in Figure 18, if the red dot and the black dot are located in different intervals, the dispersion given by using the red dot and the black dot as the center will be different. For another set of hypothesized data, 50 Hz, 100 Hz, 150 Hz and 350 Hz (blue dots on horizontal axis in Figure 19 below), the red dot and the black dot do not locate in the same interval on the vertical axis. This time the black dot gives the measurement of dispersion as 10.87 st which is higher than the result, 10.18 st, given by the red dot.

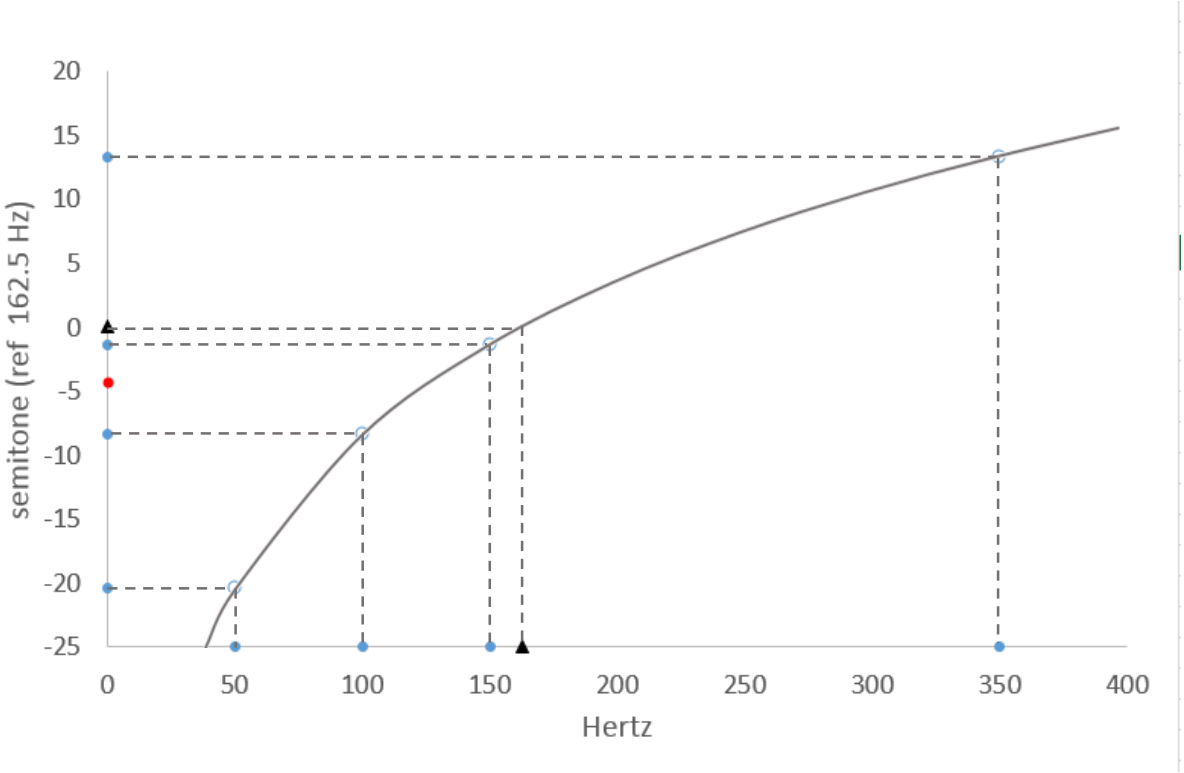


Figure 19 Another set of hypothetical data. The mean of those Hertz values is 162.5 Hz shown by the black triangle dot on the horizontal axis. Converting the observed values using 162.5 Hz as the reference gives -20.41, -8.41, -1.39, 13.28 on the semitone scale, shown by blue dots on vertical axis. 162.5 Hz corresponds to 0 st shown by black dot on vertical axis. The mean of converted semitone values is -4.23 st, shown by red dot on the vertical axis.

Zhao and Jurafsky's method calculated neither the mean deviation of the observed values

on the Hertz scale nor the mean deviation of the observed values on the semitone scale. The current study applied a correction to their method. The current study found a center of the observed values on the semitone scale directly with the goal of measuring dispersion on semitone scale. Mean deviation was used as the measurement of dispersion.

The formula for calculating mean deviation is

$$\text{Mean Deviation} = \frac{\sum|x - \bar{x}|}{n}$$

where *n* is the number of observed values, *x* is a certain observed value, $\bar{x}$ is the mean of all observed values.

5.2.2 Detailed description of method

In this study, dispersion was measured on the semitone scale because it was assumed that the speech modification was listener directed. For future intelligibility research, results on semitone scale is more meaningful. The trends in tone space dispersion measured on Hertz scale were pretty similar to that on semitone scale. Section 5.2.3 will report results on semitone scale. Main results on Hertz scale were attached at the end in appendix.

Figure 20 below shows an example of the raw data on the Hertz scale. It is the productions of a pilot speaker when speaking in HOH_quiet. Five curves of the same color represent the production of five characters under the same tone category (5 syllables × 1 tone).

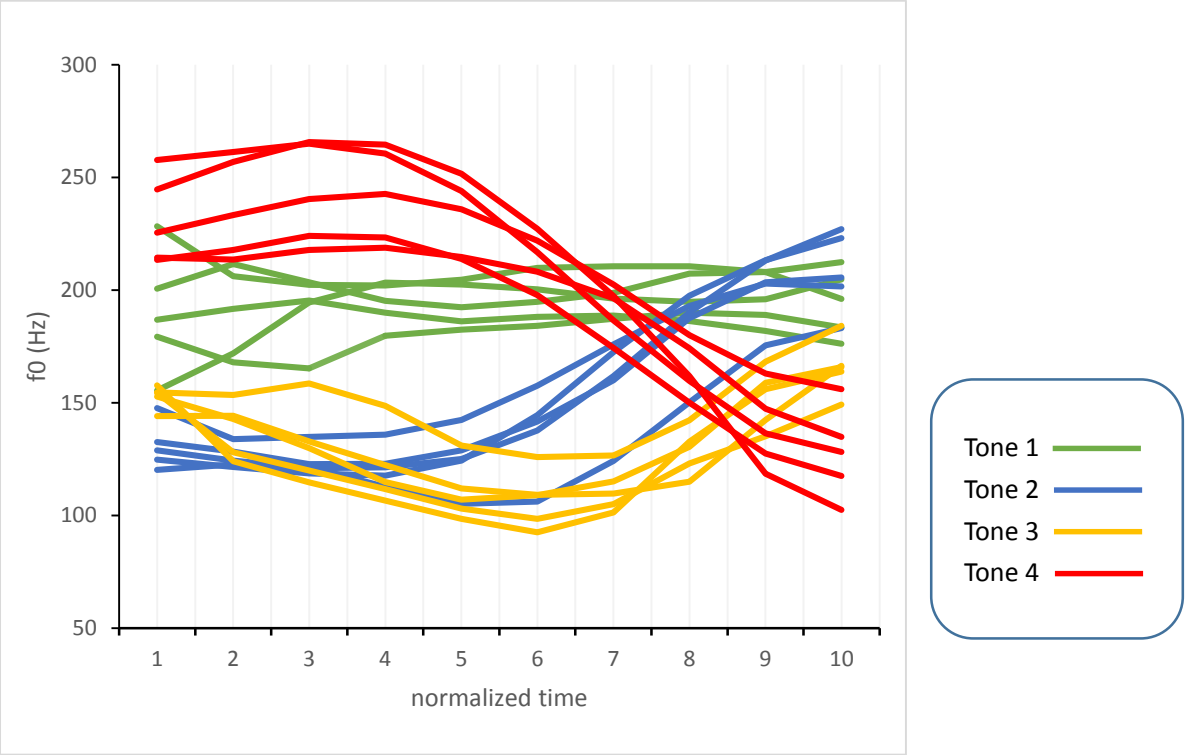


Figure 20 Observed values on the Hertz scale for a pilot speaker in HOH_quiet..

To measure dispersion on the semitone scale, all observed Hz values were converted to semitone values with 50 Hz as the reference frequency. (It does not matter which frequency is used as the reference, because the ultimate goal is to calculate the distance in semitones. The distance in semitones will only depend on the Hz values of the two frequencies that are being compared.)

The formula is

$$y = 12 * \log_2 \frac{x}{50}$$

where *x* is the Hertz value and y is the semitone value.

Figure 21 below shows the converted values on the semitone scale.

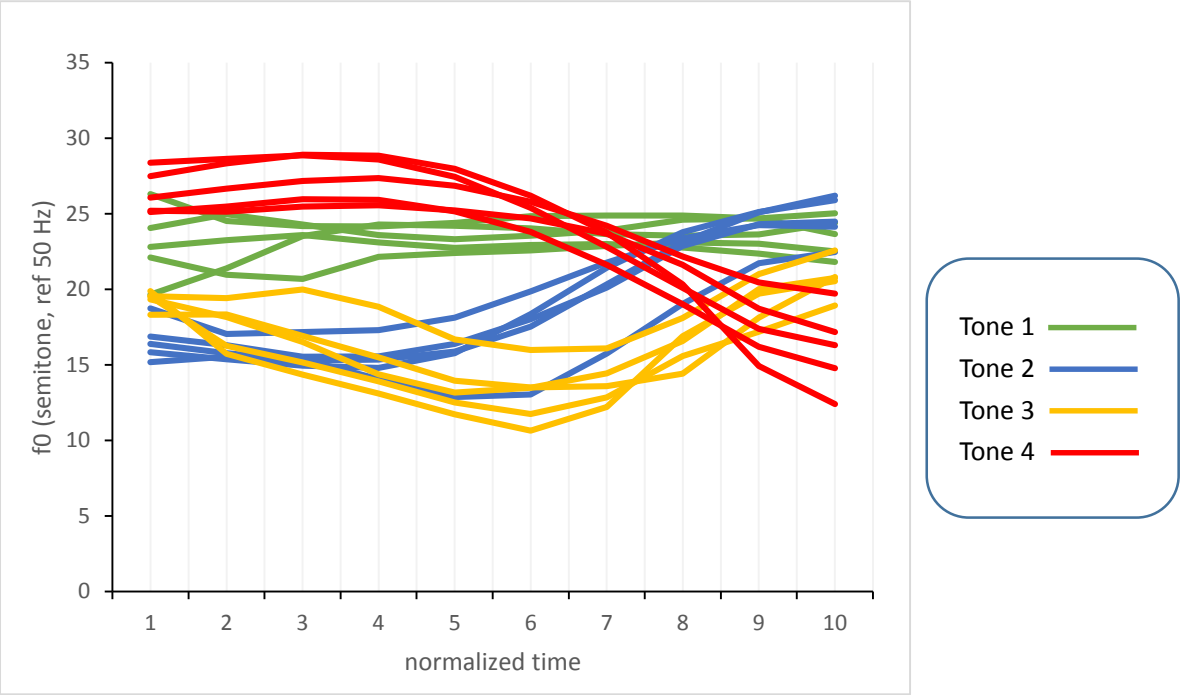


Figure 21 Production of a pilot speaker in experiment condition HOH_quiet on semitone scale (reference: 50Hz).

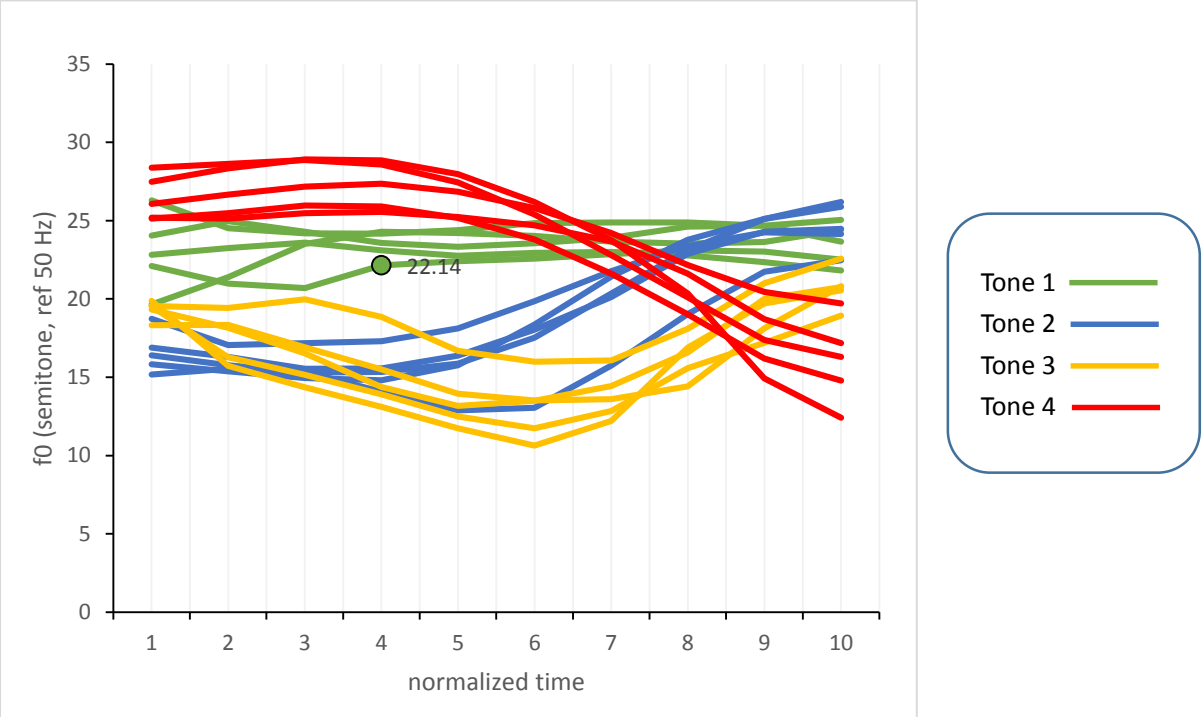


Figure 22 Observed value of contour n *at time point* i*, represented by* $O_n(t_i)$.

$O_n(t_i)$ represents the observed value of contour *n* at time point *i* (*n* ranges from 1 to 20; *i* ranges from 1 to 10). The green dot in Figure 22 above shows the observed value of a green contour at time point 4, which is 22.14 semitone (reference: 50 Hz).

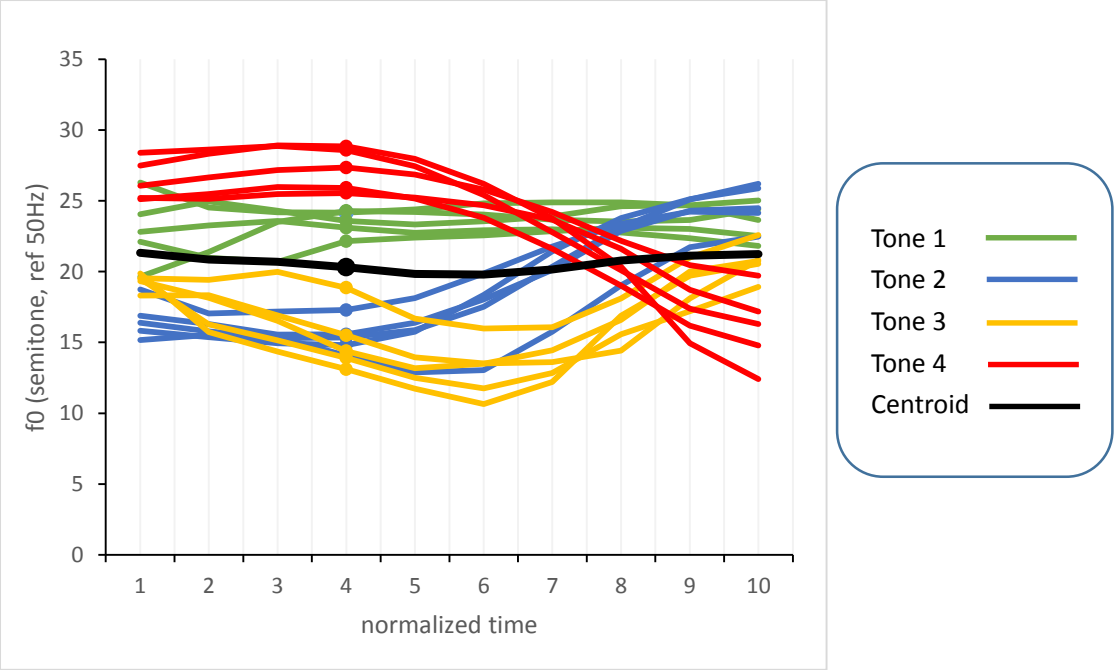


Figure 23 Centroid at time point i, *represented by* $C\ (t_i)$

At time point *i*, the centroid is the average of the observed values of 20 contours at time point *i*.

$$\mathrm{C}\ (t_i) = \frac{1}{20}\sum_{n=1}^{20} O_n(t_i)$$

In Figure 23 above, the black dot is the centroid at time point 4. The centroid changes as a function of time, which is shown by the black curve.

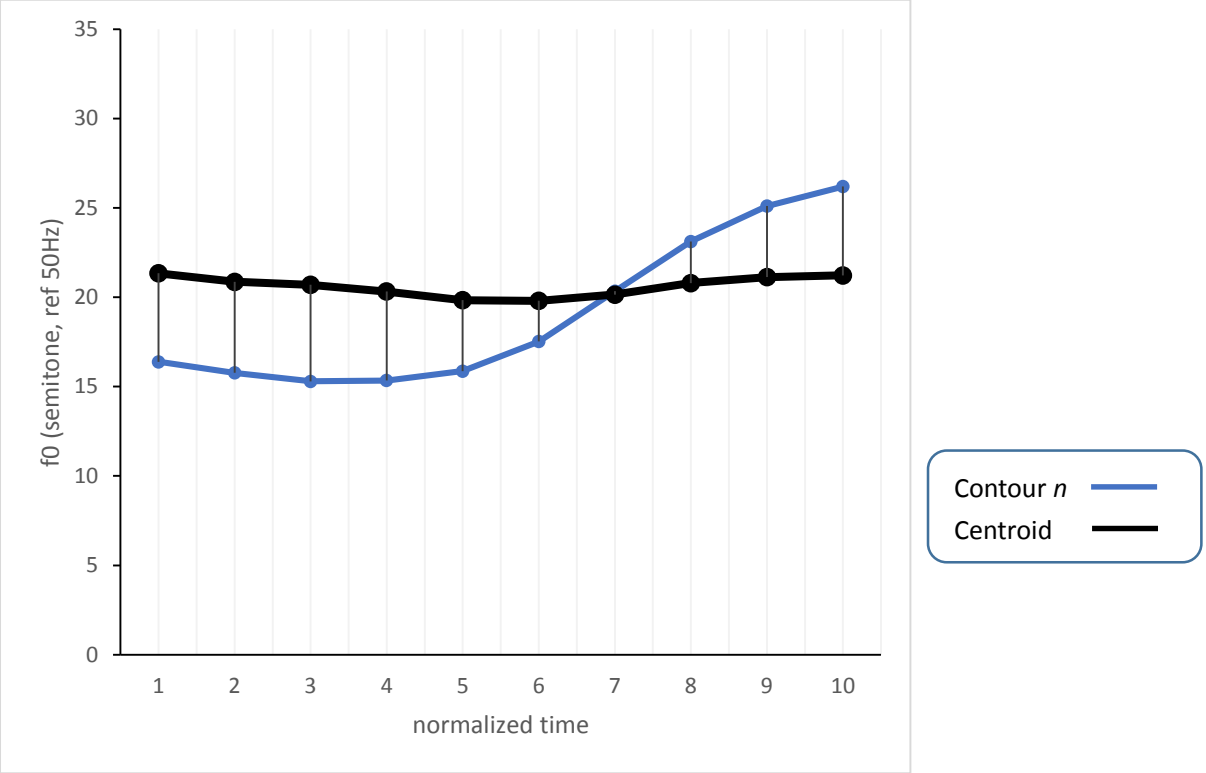


Figure 24 Distance of contour n from the centroid, represented by $D\ (R_n)$

For a single contour, its distance from the centroid was calculated as the mean of its distances from the centroid at ten time points.

$$\mathrm{D}\ (R_n) = \frac{1}{10}\sum_{i=1}^{10} |\ O_n(t_i) - C\ (t_i)\ |$$

In Figure 24, the average of the lengths of the 10 vertical line segments is the distance of the blue curve from the black curve.

A Centroid was calculated for each speaker under each experiment condition. Each speaker produced the *f0* contours of 20 characters (5 syllables × 4 tones) under each experiment condition. Data structure is summarized in Table 7.

Table 7 Data structure

CONDITION	NH_quiet				NH_noise				HOH_quiet				HOH_noise			
Tone	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
speaker 1																
speaker 2																
speaker 3																
speaker 4																

Cells shaded by diagonal lines (equivalent to Table 2) share a centroid.

5.2.3 Results and discussion

Table 8 ANOVA results of main effects and interactions for tone space dispersion.

	df	F statistic	Significance
Listener	1, 19	8.38	< .01
Environment	1, 19	0.516	0.481
Tone	3, 57	20.727	< .001
listener × environment	1, 19	16.367	< .002
listener × tone	3, 57	6.585	< .002

5.2.3.1 Main effects

A Repeated Measures ANOVA was performed to examine whether listener type (normal hearing vs. hearing impaired) and environment (quiet vs. noise) as well as tone types (tones 1–4) had a significant influence on the degree of dispersion in tone space.

Listener type has a significant effect on tone space dispersion ($F(1, 19) = 8.38$, $p < 0.01$). Speakers used a more dispersed tone space when speaking to the hard of hearing (Mean Difference = 0.29 st).

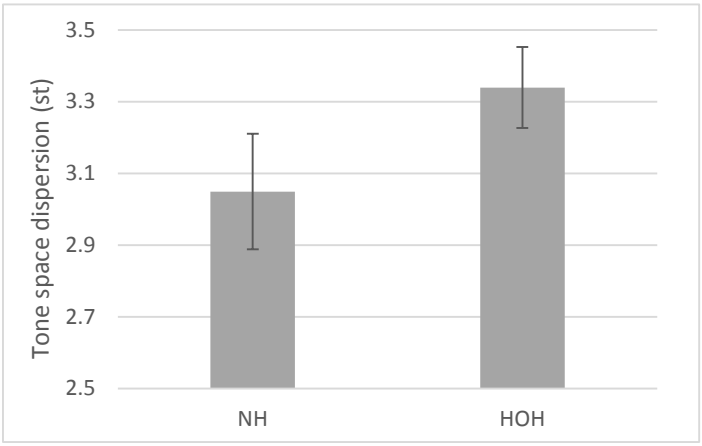


Figure 25 Tone space dispersion for normal-hearing listener and hard-of-hearing listener averaged across environments and tones with standard errors.

Table 9 Mean and standard error values for Figure 25.

	Mean	Standard error	MD	p
NH	3.049 st	0.171	0.290 st	< 0.01
HOH	3.339 st	0.108		

Unlike listener type, environment was found to have no significant effect on the adjustment of tone space ($F(1, 19) = 0.516$, $p = 0.481$), which indicated that speakers did not use a more dispersed tone space when speaking in the presence of noise.

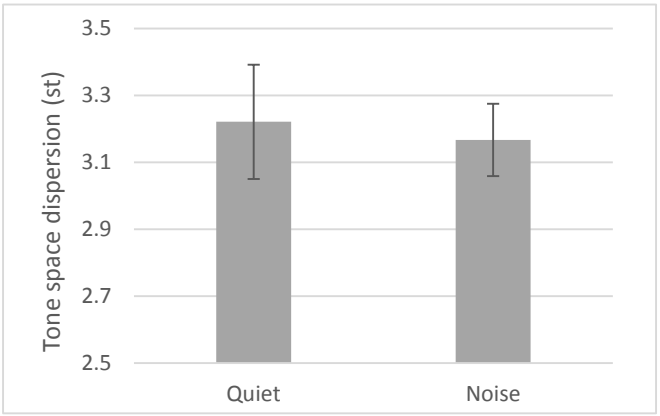


Figure 26 Tone space dispersion for quiet environment and noise environment averaged across listeners and tones with standard errors.

Table 10 Mean and standard error values for Figure 26.

	Mean	Standard error	MD	p
Quiet	3.221	0.161	0.054 st	0.481
Noise	3.167	0.113		

The above results supported the main hypothesis that speakers addressed the listener with simulated hearing loss and the listener hearing in noise differently in terms of tone space dispersion. The simulated hearing loss had an effect on tone space dispersion while the noise did not. Speaker produced more contrast between lexical tones to accommodate the listener who had a suprathreshold/clarity problem.

Tone type also had a significant effect on tone space dispersion (F (3, 57) = 20.727, p < .001). Tone 3 and tone 4 were further away from the centroid than Tone 1 and Tone 2 (all p-values in pairwise comparisons were less than 0.005). Difference between tone 1 and tone 2 as well as the difference between tone 3 and tone 4 were not statistically significant. Figure 27 below shows distances from centroid for each tone averaged across experiment conditions.

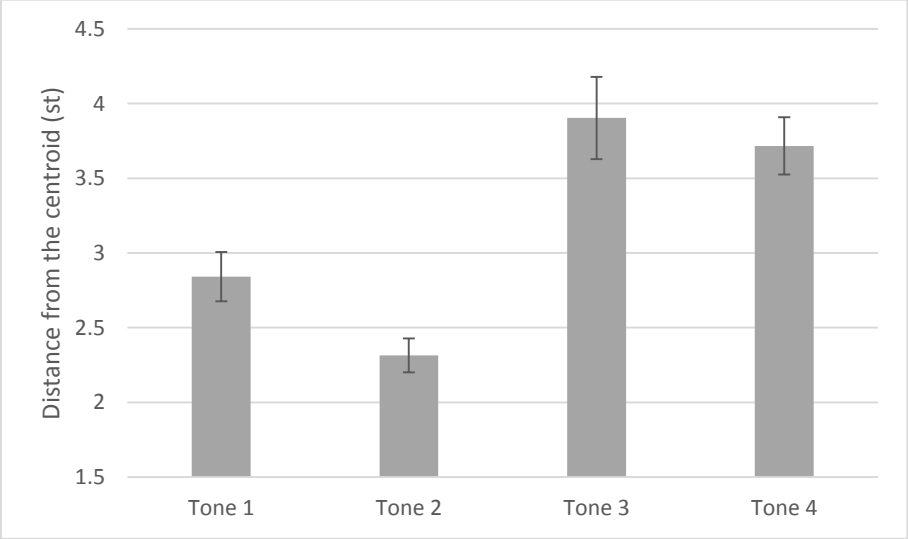


Figure 27 Distance from the centroid for each tone averaged across experiment conditions.

5.2.3.2 Interactions and further discussion

For the three main factors, there were two major interactions between them. One was the interaction between listener and environment. The other was the interaction between listener and tones.

The interaction between listener and tone type (F (3, 57) = 6.585, p < .002) indicated that change in the distance of a tone contour from the centroid from NH to HOH differ by tone. Tone 2 and tone 4 were further away from the centroid in HOH conditions than in NH conditions. The distance of Tone 1 from the centroid showed no significant difference between NH conditions and HOH conditions. This was true for tone 3 as well.

Table 11 Comparison of tone space dispersion between NH and HOH for each tone.

	tone	Mean difference	standard error	Significance
NH vs HOH	1	-0.21	0.14	0.137
	2	-0.44*	0.06	< .001
	3	0.11	0.21	0.609
	4	-0.62*	0.14	< .001

Figure 28 below shows the distances from centroid for all tones in NH and HOH averaged across environments. It can be seen that distances from the centroid for tone 2 and tone 4 in HOH are much higher than that in NH. However, for tone 1 and tone 3, the distances were not so significantly different between NH and HOH.

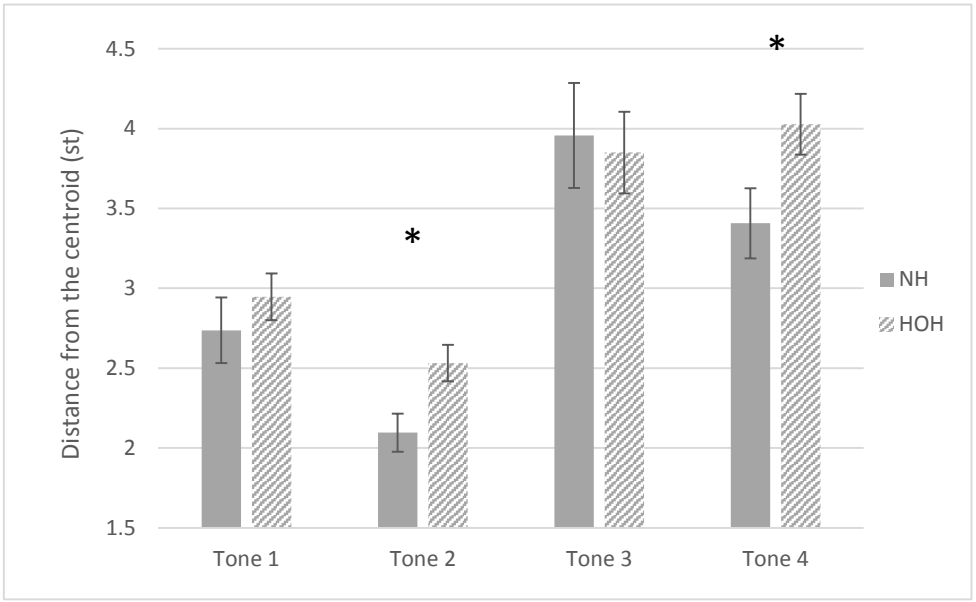
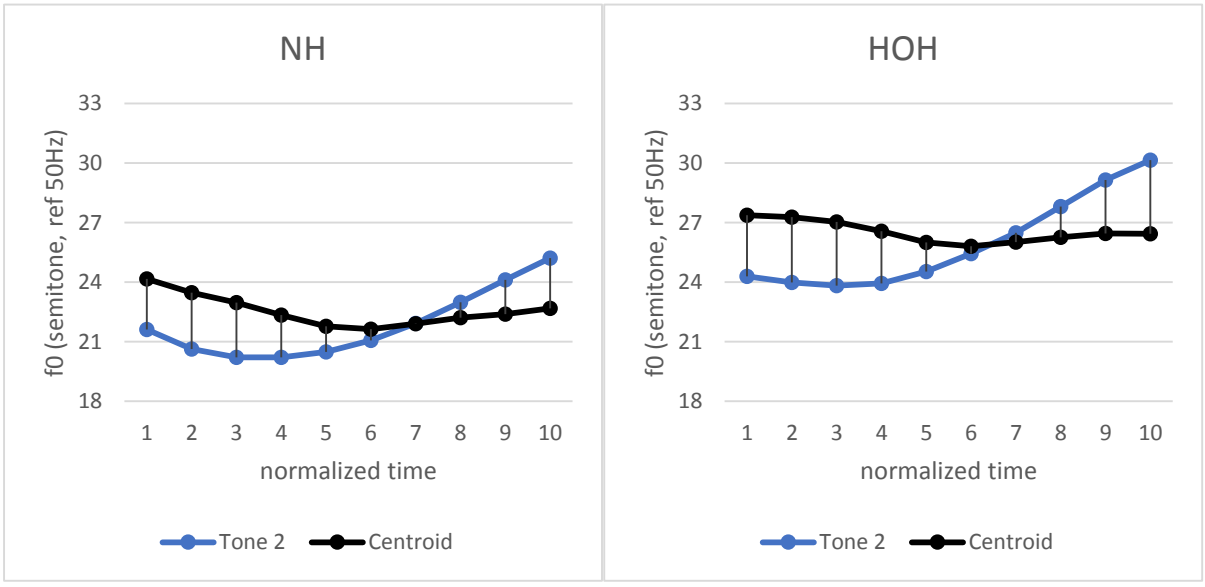


Figure 28 Distance from the centroid for each tone broken up by listener type.

Figure 29 below shows the distance from the contour to the centroid for NH and HOH averaged across environments for tone 2 and tone 4. Comparing the panels on the left and the panels on the right, one may discover that *f0* contours of tone 2 and tone 4 are slightly further away from the centroid in HOH than in NH. The area between the contour and the centroid was increased from NH to HOH.



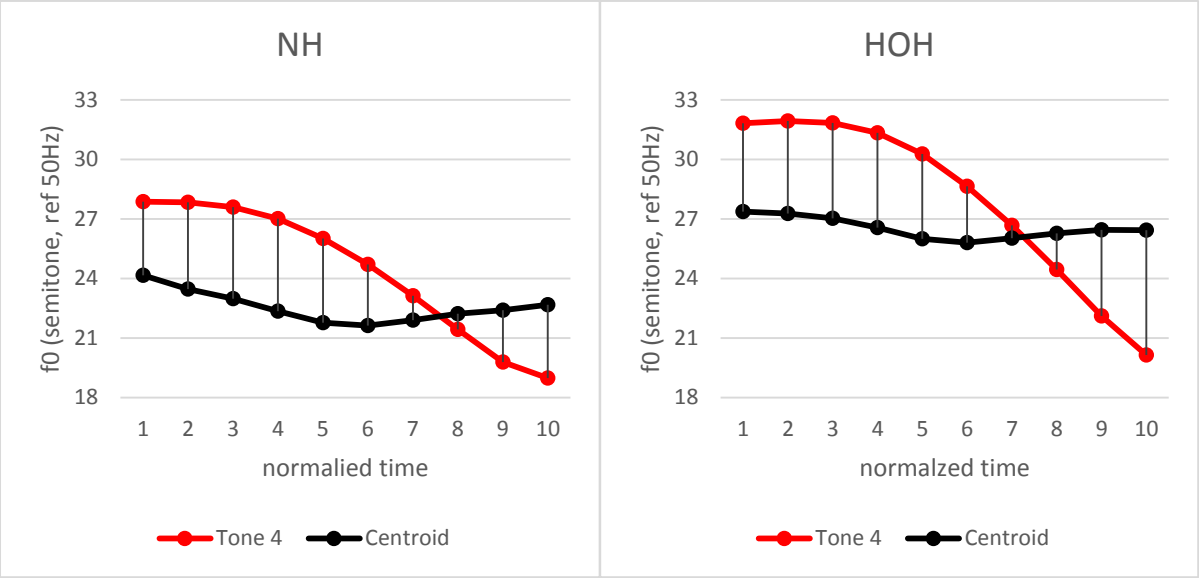


Figure 29 Distance from the centroid for tone 2 and tone 4 for NH listener and HOH listener averaged across environments.

Why do tone 2 and tone 4 show this pattern? Figure 30 below puts *f0* contours of the same tone in one figure to compare how the contour was elevated from NH to HOH for tone 2 and tone 4.

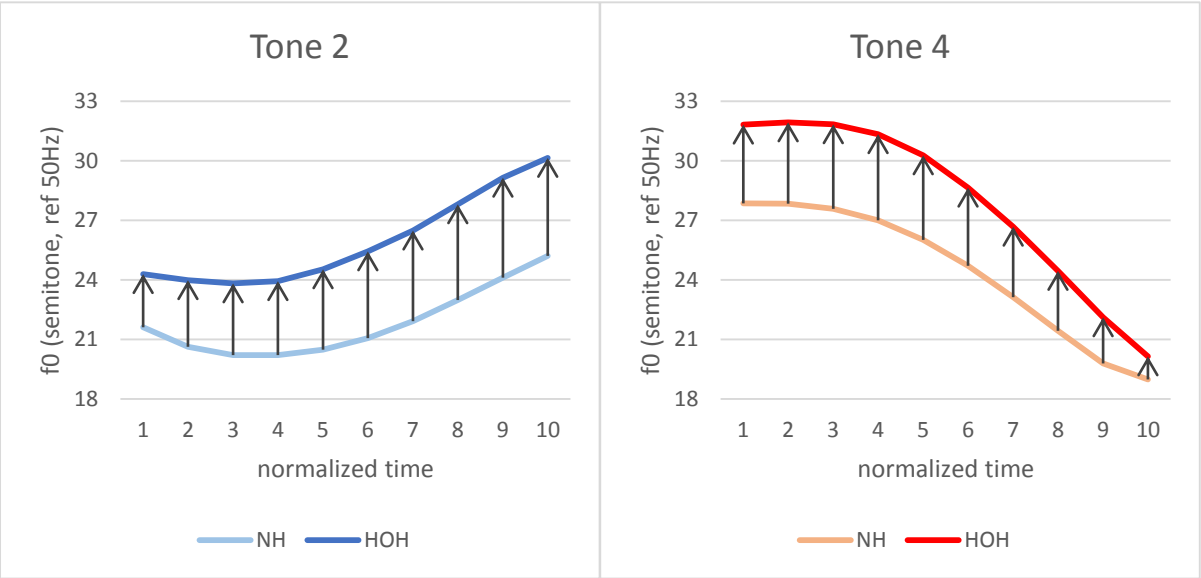


Figure 30 f0 contours of tone 2 and tone 4 in NH and HOH averaged across environments on semitone scale.

It is obvious that the amount of elevation at different time points were different for these tones. The range of *f0* was increased from NH to HOH for tone 2 and tone 4. If tone 2 is characterized by rising at the end and tone 4 is characterized by the high beginning, contours produced in HOH were more typical contours for tone 2 and tone 4 than contours produced in NH.

For tone 2 and tone 4, the contour in NH and that in HOH were not parallel with each other. This caused an increase in the distance from the contour to the centroid from NH to HOH. The pattern could be simplified and illustrated by Figure 31 below.

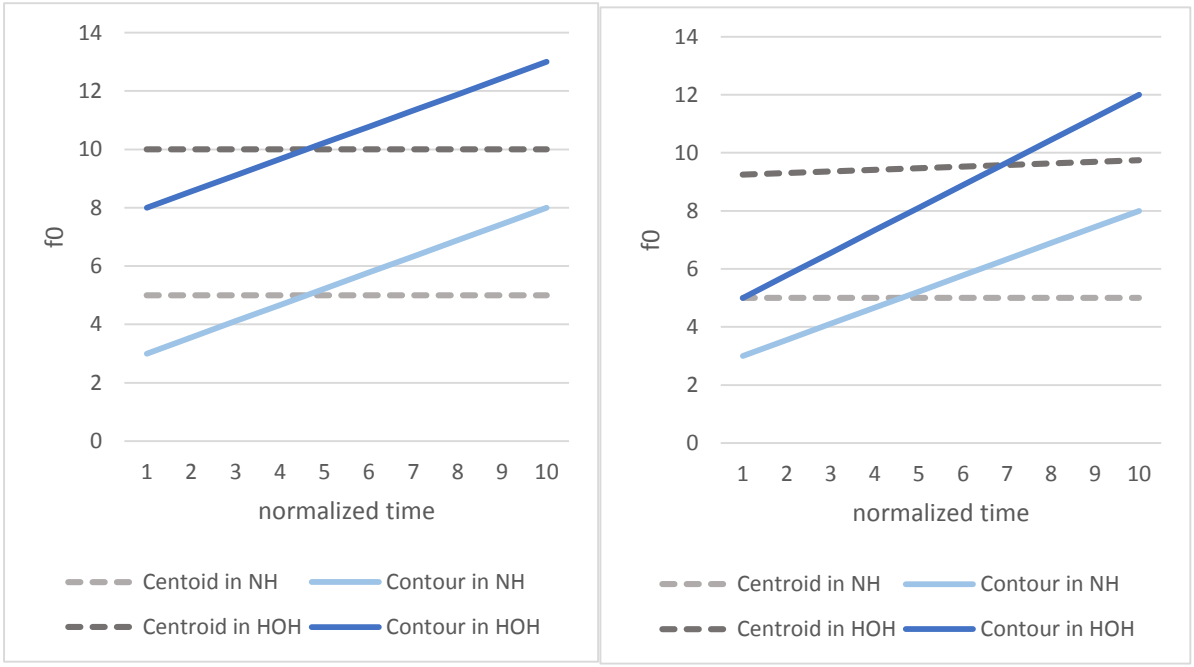


Figure 31 Hypothetical data that illustrates the increase of the distance between contour and centroid from NH to HOH.

In the left panel, the lower solid line (the contour in NH) and the upper solid line (the contour in HOH) are parallel. From the lower solid line to the upper solid line, the amount of elevation is 5 units at each time point. Suppose that other tone contours are all elevated by 5

units at each time point too, then the centroid from NH to HOH is elevated by 5 units at each time point (shown by the elevation from the lower dashed line to the upper dashed line). In this case, the area between the lower solid line and the lower dashed line is equal to the area between the upper solid line and the upper dashed line. The distance from the contour to the centroid did not change while the contour and the centroid were elevated.

In the right panel, the two solid lines are not parallel. The amount of elevation from the lower solid line to the upper solid line changes as a function of time. There is a difference in slope between these two solid lines. Suppose again that other tones are all elevated by 5 units at each time point, then the slope of the elevated centroid (upper dashed line) is close to but bigger than 0. A difference in slope can be found between the two dashed lines. Because the increase in slope of the dashed line was not as big as that of the solid lines, the area between the upper solid line and dashed line is bigger than the area between the lower solid line and dashed line. The distance from the contour to the centroid is increased from NH to HOH in this case. The above is not a rigorous mathematical proof but just a schematic approximation of the situation for tone 2 and tone 4.

Tone 1 and tone 3 show a different pattern from tone 2 and tone 4 because their amount of elevation of the contour from NH to HOH did not change as a function of time. The two contours are almost parallel, as shown in Figure 32 below.

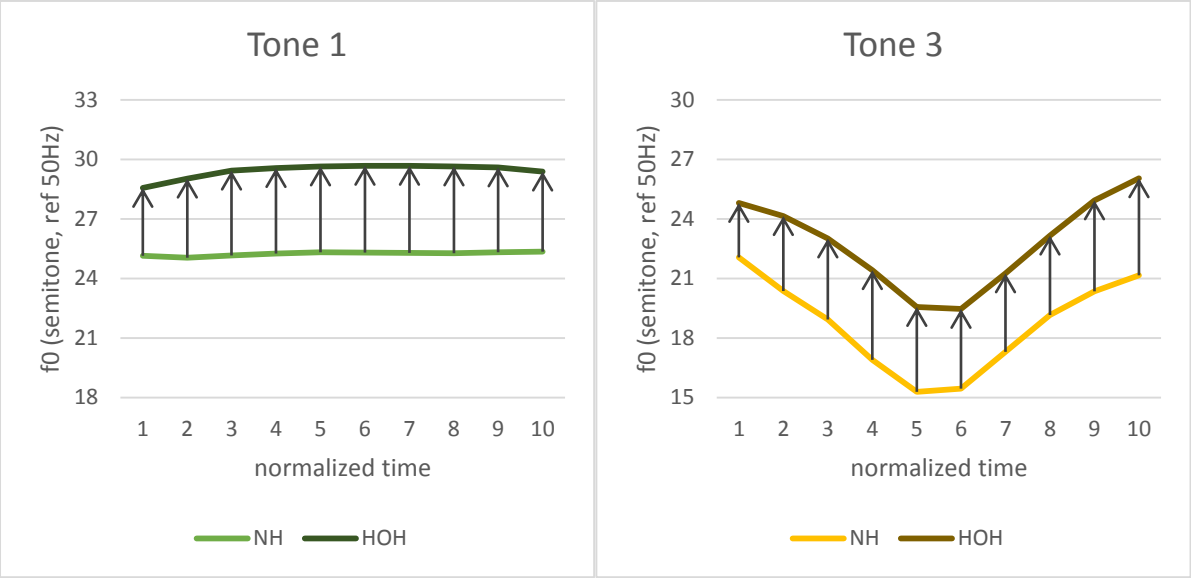


Figure 32 Contours in NH and HOH for tone 1 and tone 3.

Pattern shown in Figure 32 is pretty similar to the pattern shown in the left panel of Figure 31 above. Therefore, it is not surprising that the distance from the contour to the centroid was not significantly changed from NH to HOH for tone 1 and tone 3.

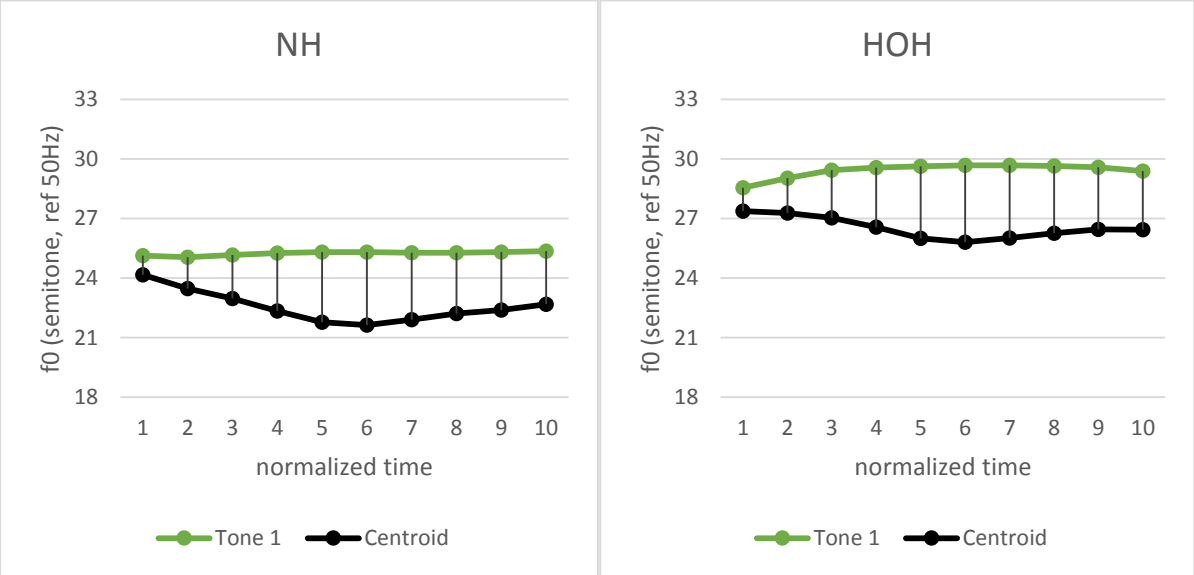


Figure 33 Distance from the contour to centroid for tone 1 in NH and HOH averaged across environments on semitone scale.

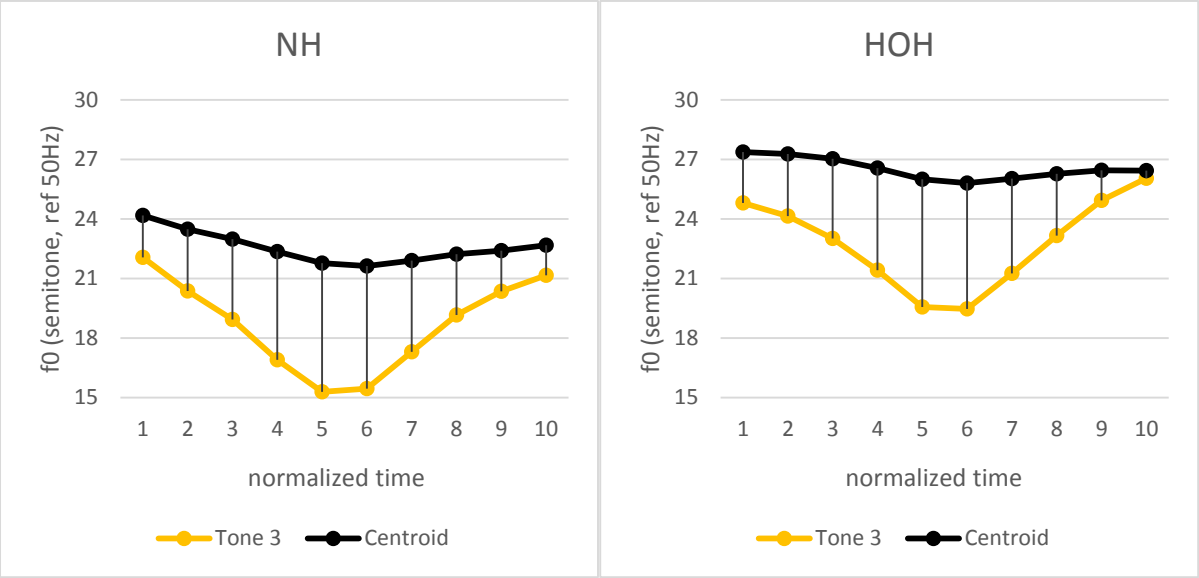


Figure 34 Distance from the contour to centroid for tone 3 in NH and HOH averaged across environments on semitone scale.

Why did the contours of tone 1 and 3 show a parallel pattern from NH to HOH? Tone 1 is characterized by no change in *f0* across time. Thus the two contours in NH and HOH should be parallel otherwise the production is not a typical tone 1. Tone 3 is characterized by the dipping in the middle part. Perhaps because the goal of elevating the overall height of tone 3 and the goal of lowering the middle part contradict with each other to some extent, so the speaker did not produce a more typical tone 3 in HOH. Another possibility is that tone 3 is characterized by the creakiness in the middle. The need of producing creaky voice in the middle part and the need of elevating *f0* could not be satisfied at the same time.

The other interaction between the main factors was the interaction between listener and environment ($F(1, 19) = 16.367$, $p < .002$). This interaction suggested that change in tone space dispersion between the normal-hearing listener condition and the hard-of-hearing listener

condition differed by environments, which can be seen in Figure 35 below.

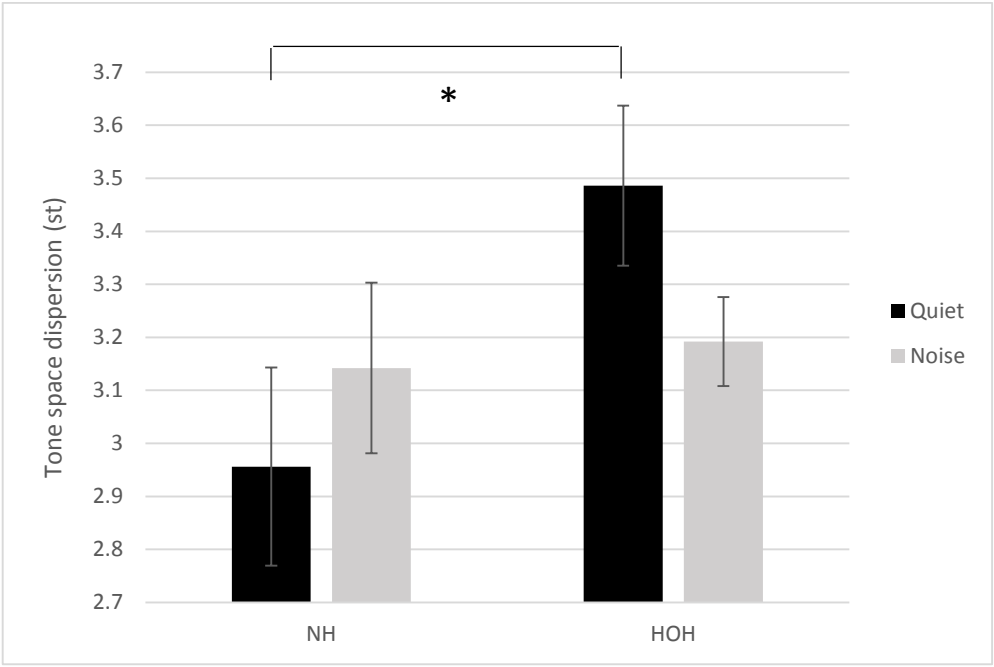


Figure 35 Tone space dispersion of two types of listeners broken up by environment with standard errors.

Table 12 Mean and standard error values for Figure 35.

Condition	Mean	Standard error
NH_quiet	2.956	0.187
NH_noise	3.142	0.161
HOH_quiet	3.486	0.151
HOH_noise	3.192	0.084

Table 13 Post-hoc pairwise comparisons between NH and HOH under different environments.

	Mean Difference	standard error	significance
NH_quiet vs HOH_quiet	-.530	.110	< .001
NH_noise vs HOH_noise	-.050	.123	.688

Post-hoc Bonferroni tests show that in quiet environment the difference between tone space dispersion of NH and HOH was significant ($p < .001$) while in noise the difference was

not significant (p = .688). It can be seen from Figure 35 above that the black bar on the right is much higher than the black bar on the left while the grey bar on the right is just slightly higher than the grey bar on the left. Figure 36 below shows tone contours and the centroid for the four conditions.

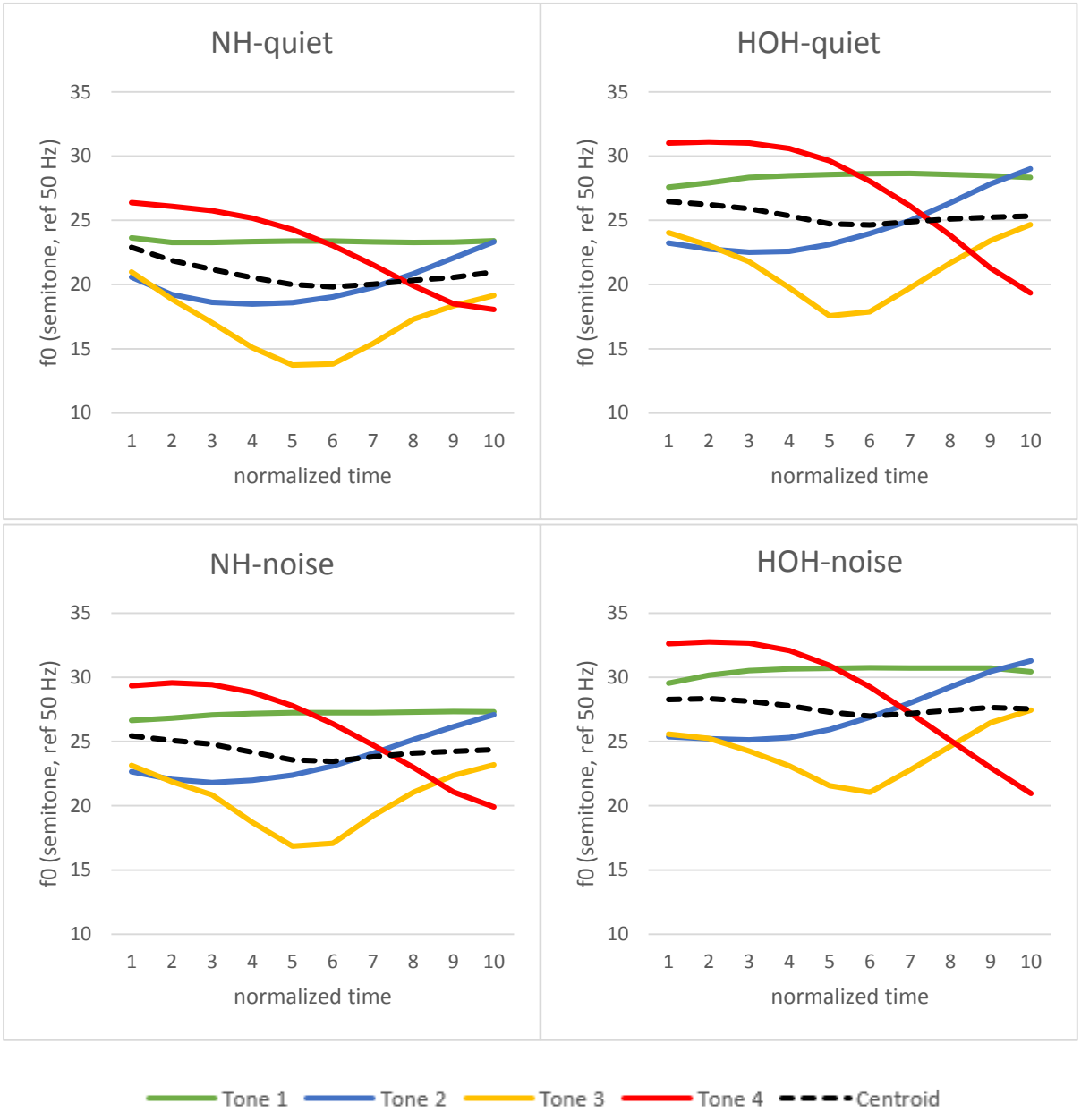


Figure 36 Productions of four tones under each experiment condition averaged across speakers.

From NH_quiet to HOH_quiet (upper two panels), tones went further away from each other and the centroid. From NH_noise to HOH_noise, tones were still close to each other and the centroid.

In Figure 37 below, mean *f0* and tone space dispersion for the four experiment conditions are shown together. One can see that while the *f0* was elevated all the way from the baseline to the fourth condition, tone space dispersion was increased from the baseline to the third condition and then decreased in the fourth condition.

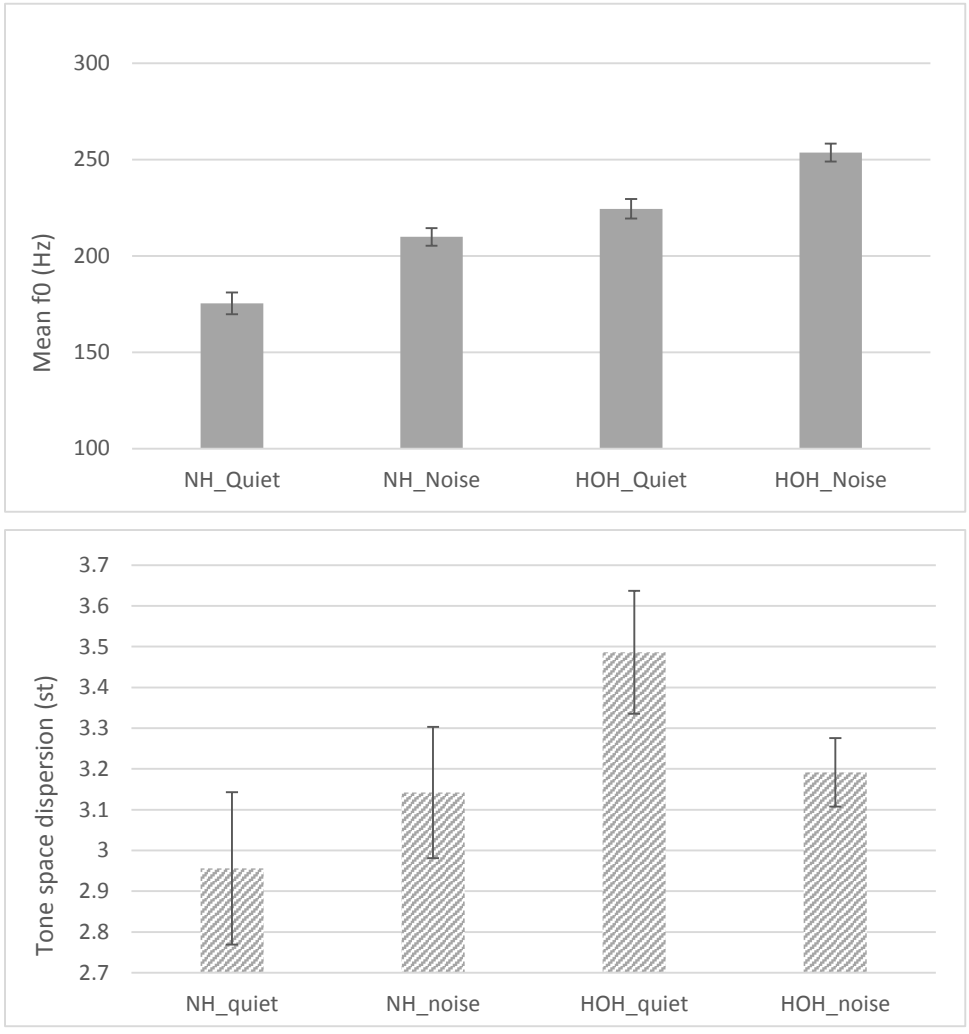


Figure 37 Mean f0 and tone space dispersion for the four experiment conditions.

From baseline to NH_noise, tone space dispersion was increased (MD = 0.19 st). The possible explanation is that the speaker had considered that the listener's ability to discriminate different speech sounds might be hurt by the noise to some extent. Though environment did not have a significant main effect on tone space dispersion, we cannot conclude that white noise did not have any influence on tone space dispersion.

The hearing-impaired listener had a bigger influence on tone space dispersion than the listener in noise. From baseline to HOH_quiet, tone space dispersion was largely increased (MD= 0.53 st) while from baseline to NH_noise the difference was only 0.19 st. Speakers might considered that listener who had both a clarity problem and an audibility problem needed more contrast between lexical tones to hear them clearly than listener who was listening in white noise.

Table 14 Change in f0 and tone space dispersion for two pairs of experiment conditions.

	change in *f0*	change in tone space dispersion
from baseline to NH_noise	34.54 Hz	0.19 st (0.19 × 100 = 19)
from baseline to HOH_quiet	49.06 Hz	0.53 st (0.53 × 100 = 53)

In HOH_noise, speakers produced higher *f0* than in HOH_quiet (MD = 29.22Hz). Speakers might consider that listener in the fourth conditions had a more serious audibility problem than in the third condition, because in the fourth condition the listener was simulated to be experiencing a hearing loss and he was also influenced by the white noise.

In terms of clarity of speech, the fourth condition should be the worst among the four conditions in speakers' minds. Speakers should produce the most contrast between lexical tones

in this condition than in other conditions. However, tone space dispersion was decreased from the third condition to the fourth condition (- 0.29 st). It seems that a decrease in contrast between tones was not a deliberate speech modification because of a physiological constraint.

Table 15 Change in f0 and tone space dispersion from HOH_quiet to HOH_noise.

	change in *f0*	change in tone space dispersion
from HOH_quiet to HOH_noise	29.22 Hz	-0.29 st

Because it is impossible for a speaker to raise *f0* infinitely, there might be a tradeoff between raising *f0* and increasing tone space dispersion. There was an upper limit in a speaker's fundamental frequency profile, it is impossible for a speaker to elevate *f0* and enlarge tone space dispersion simultaneously when he speaks near the upper limit. When both the need of elevating *f0* and the need of keeping good contrast between tones existed, speakers had to prioritize one. From the third condition to the fourth condition, speakers might consider that both audibility and clarity problems became more serious, but they had to solve the audibility problem first, so they produced higher *f0* but smaller tone space dispersion.

In the current study, the white noise was presented at a relatively high level, 70 dB SPL, similar to the level of noise in a noisy restaurant. In the fourth condition, perhaps the combination of 70 dB SPL white noise and the hearing loss on the listener caused speakers put too much effort into speech modification, thus they spoke near the upper limit of speaking fundamental frequency. It would be interesting to look at combination of hearing loss and other levels of white noise, such as 25 dB SPL and 50 dB SPL.

6. GENERAL DISCUSSION

6.1 Summary of results

This study suggests that the effects of noise and a hearing-impaired listener on Mandarin tone production can be similar or different. This was probably because speakers considered both the simulated hearing loss and the white noise involved a degradation in sound level of speech for the listener but only the simulated hearing loss involved a degradation in clarity of speech for the listener.

When addressing a listener who was experiencing white noise and simulated hearing loss, speakers made similar modifications on the *f0* dimension, that is, an elevation of mean *f0*. Both white noise and simulated hearing loss caused degradation of sound level of speech signal for the listener. The elevation of *f0* indicated that speakers had been aware of the similarity of the perceptual deficits induced by white noise and a simulated hearing loss and had tried to accommodate listeners who had an audibility problem.

On the contrary, on a different dimension, tone space dispersion, which is sensitive to the phonemic structure of the tonal inventory, the presence of a hearing loss resulted in hyperarticulation while the presence of a noise did not. Notably, a degradation in clarity of speech was simulated to be experienced by the hearing-impaired listener in this study. Mandarin tones were hyperarticulated in the sense that the contrast between different tone categories was increased, which may have resulted in tones being more perceptually distinct from each other for the listener who couldn’t easily recognize what the speech sound was even

the speech sound was audible.

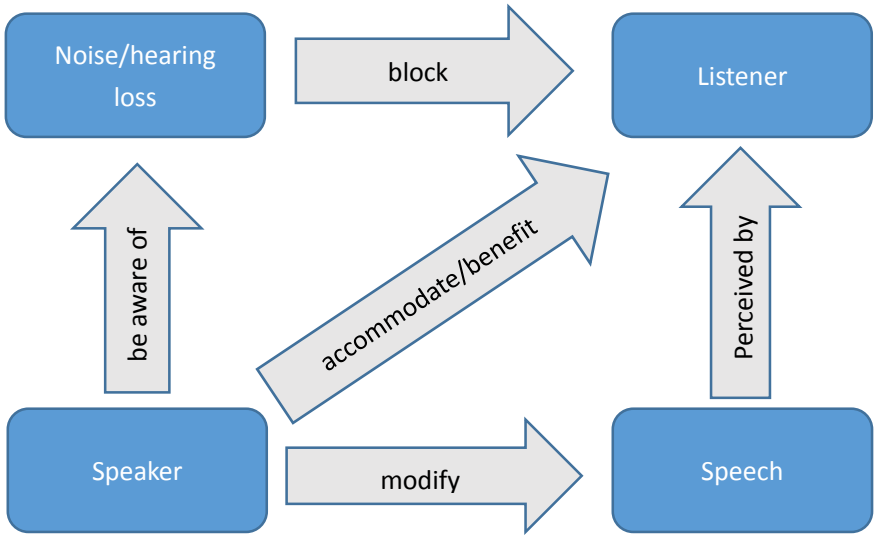


Figure 38 Speakers accommodate listeners subjected to noise or hearing loss.

At the prosodic level, elevation in *f0* indicated that speakers had attempted to overcome audibility problem for the listener. At the phonological level, increase in tone space dispersion indicated that speakers had attempted to overcome clarity problem for the listener. The results fit in the model of Lindblom's H&H theory. When the speaker is aware that the listener's access to information is blocked by some barriers, speakers will accommodate the listener by producing hyperarticulation in certain dimensions according to the nature of the barrier.

6.2 Listener-directed speech modification

Phonetic lab speech is a special speech style. Sometimes it is a reading task in which no listeners are involved while sometimes an imagined listener is given in the instruction. Both tasks require the speaker to read/speak as if they are speaking normally in daily conditions. Intuitively, when one speaks (aloud) in daily condition, a real/imagined listener is always

involved (imagined listener example: when one records a voicemail on the phone). If no real/imagined listener is presented at all, a person does not have to speak (aloud). In a reading task, though, no listener is described in the instruction, but the speaker still speaks (aloud). The author doubts that a "background listener" still exists in a reading task.

Speakers in Zhao and Jurafsky (2009)'s study performed such a reading task. Actual instructions given to speakers were not described. It is not clear if a "background listener" was mentioned, but the motivation of involving noise and low-frequency words in Zhao and Jurafsky (2009)'s study was the **perceptual difficulties** induced by those two factors. Probably speakers linked the perceptual difficulties to a potential listener in that study. Therefore the author doubts that speech modifications in Zhao and Jurafsky's findings were still kind of listener directed. Thus the results of their study and that of the current study may be comparable.

Zhao and Jurafsky (2009) found that *f0* of all 6 Cantonese tones was raised in the presence of white noise. Results of the current study are consistent with their results. In terms of tone space dispersion, though the current study used a slightly different metric than that used by Zhao and Jurafsky (2009), results of the two studies may still be comparable. Actual values given by the two metrics may be different but the abilities of the two metrics to display trends in the data may be not so different.

Zhao and Jurafsky (2009) found that white noise did not have a significant effect on tone space dispersion of Cantonese while word frequency did. The current study found that white noise did not have a significant effect on tone space dispersion of Mandarin while a simulated hearing loss on the listener did. In Zhao and Jurafsky (2009), low-frequency words were linked

with low predictability. In the current study, a simulated hearing loss reduced the listener's ability to discriminate different speech sounds. Perception deficits either due to a simulated hearing loss or word frequency could be considered as suprathreshold/clarity problems. It is not so surprising that the two studies found similar patterns in tone space dispersion.

Zhao and Jurafsky (2009) found that low-frequency words were produced with larger tone space dispersion than high-frequency words (Mean difference = 0.417 st). The current study found that listeners who had a clarity problem in addition to an audibility problem caused speakers to produce larger tone space dispersion than listeners who only had an audibility problem (Mean difference = 0.290 st). Mean differences in the two studies were all found to be within 0.5 st. The difference between 0.417 st and 0.290 st may be due to many factors. If we reanalyze Zhao and Jurafsky's data with the metric of the current study, mean difference may be found to be slightly smaller than 0.417 st. In addition, lexical tone systems of Cantonese and Mandarin are different, where Cantonese has 6 tones and Mandarin only has 4 tones.

In summary, whether a real/imagined listener is present or not in an experiment, speakers tend to take into account difficulties on the perceptual side when making speech modifications. The specific form of phonetic variation may be in line with the nature of the difficulty.

6.3 Predictions for future intelligibility studies

Many researchers have investigated acoustic-phonetic characteristics of speech with high intelligibility. Predictions of acoustic correlates of intelligibility of Mandarin tone/speech should be based on those studies.

Picheny et al. (1985) recorded speech production when talkers were instructed to talk conversationally or clearly. Then they conducted intelligibility tests in quiet using listeners with real sensorineural hearing loss. They found that the average intelligibility difference between clear speech and conversational speech was 17 percentage points.

Following this study, Picheny et al. (1986) investigated the acoustic characteristics of clear and conversational speech produced in the previous study. They found that vowels were less reduced in clear speech than in conversational speech. They interpreted this difference in terms of change in formant frequencies of vowels. Figure 39 below shows the vowel plot on F1-F2 space in Picheny et al (1986). Though they did not mention the concept of vowel space and claim that vowel space was positively correlated with speech intelligibility, one can see from their plot that vowels were further apart from each other in speech with higher intelligibility (clear speech in their case).

Ferguson (2004) recorded clear and conversational speech in a similar manner as that of Picheny et al. (1985). But her intelligibility test was conducted in a different manner. She used young normal hearing listeners and presented the stimuli in a 12-talker babble background noise. She found that vowel intelligibility in clear and conversational speech for normal-hearing listeners in noise varied widely among 41 talkers.

Following this study, Ferguson and Kewley-Port (2007) selected the speech of talkers who produced big clear speech benefit (BB talkers) and that of talkers who produced no clear speech benefit (NB talkers) to investigate the acoustic characteristics of vowels. They found a larger increase of vowel space expansion from conversational speech to clear speech for BB talkers

than for NB talkers. They concluded that increase in vowel space expansion improved vowel intelligibility. Figure 40 shows their vowel plot in F1-F2 space for BB talkers.

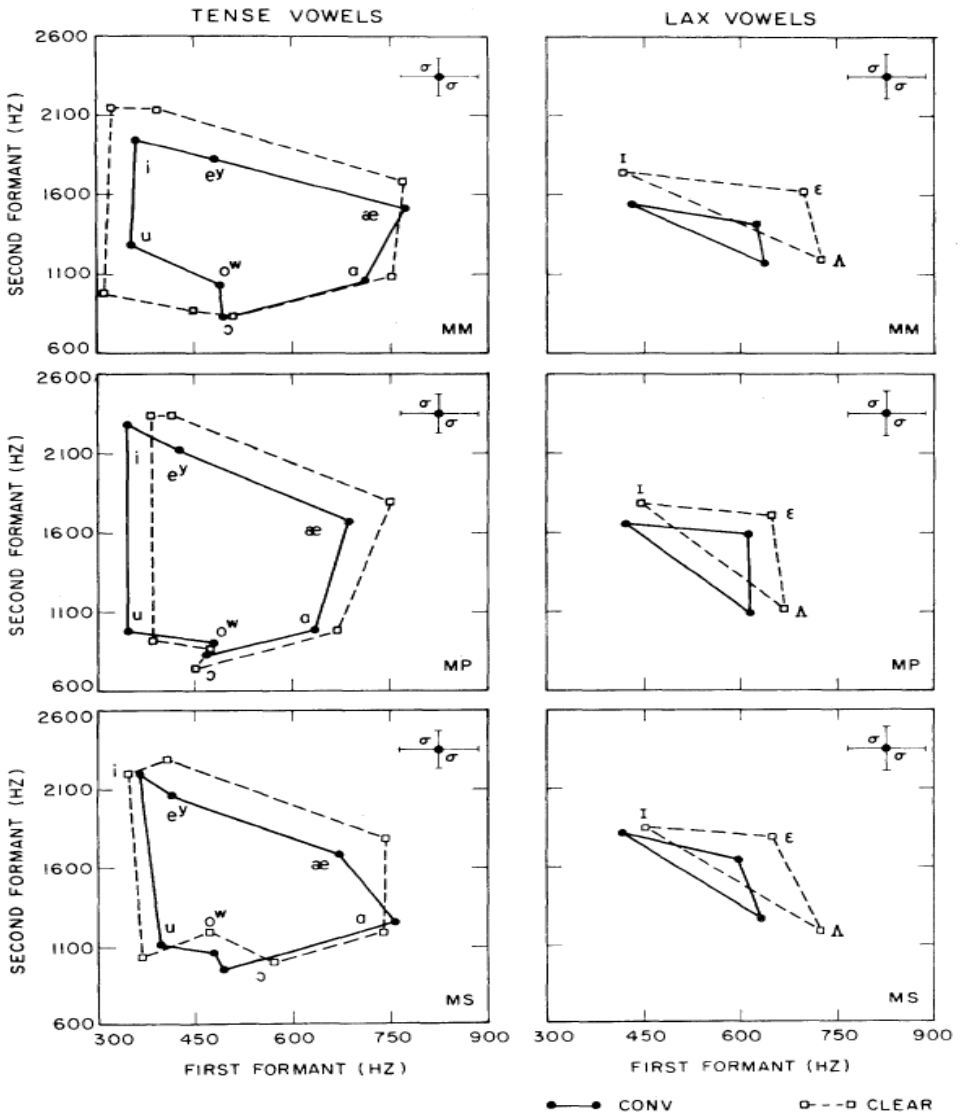


Figure 39 Formant frequency data. Left column shows results for tense vowels and right column for lax vowels; different rows correspond to different speakers; standard deviations are shown in upper-right corner of each graph. Data for conversational speech are indicated by filled circles and solid lines; for clear speech by open squares and dashed lines. After Picheny et al. (1986).

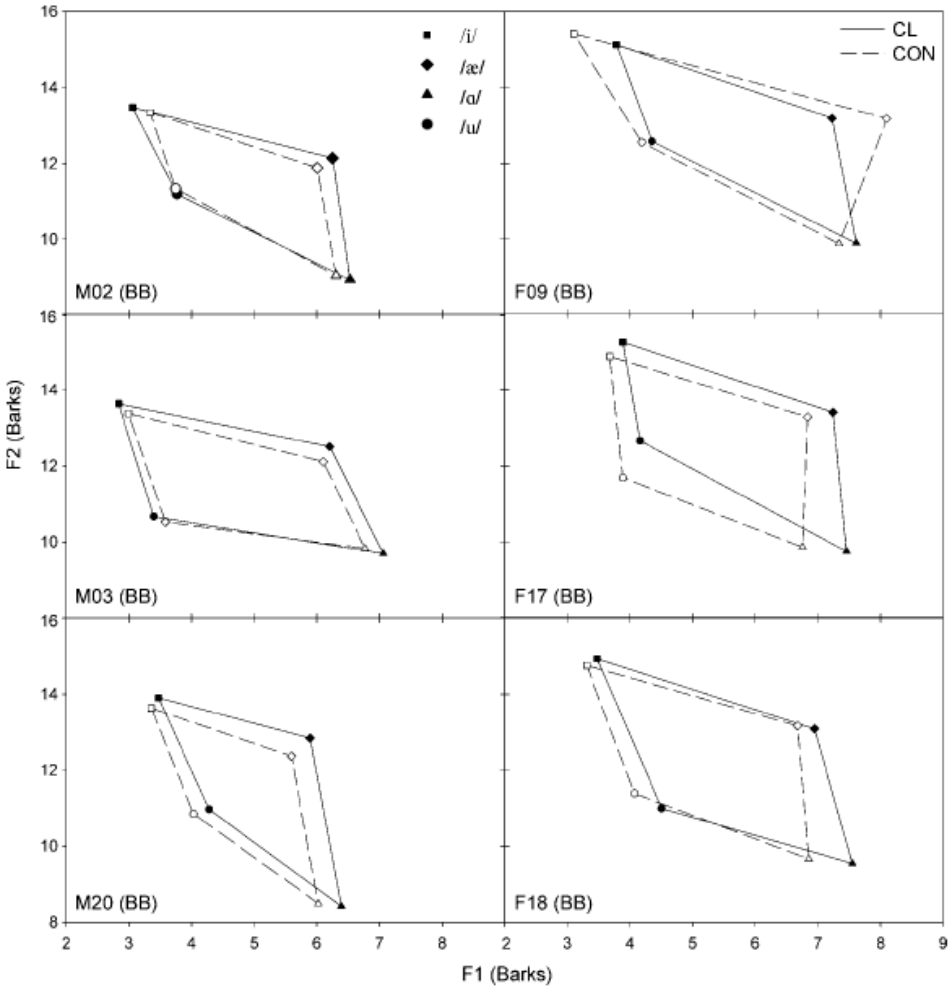


Figure 40 Four-point vowel spaces for talkers who produced big clear speech benefits. CL = clear; CON = conversational. After Ferguson and Kewley-Port (2007).

Besides the results from Picheny et al. (1985) and Ferguson and Kewley-Port (2007), Bradlow et al. (1996) used a speech database with intelligibility scores and they found that larger *f0* range and larger vowel space were generally positively correlated with English speech intelligibility in quiet for normal-hearing listeners. Vowel space expansion is shown by Figure 16 in section 5.2.1.

The above results suggested that a larger vowel space improves speech/vowel intelligibility for listeners with sensorineural hearing loss in quiet and normal-hearing listener in noise and in quiet. Like vowels, lexical tones are important for meaning contrast in tonal

languages. If an intelligibility test is conducted in the future, it is expected that productions with larger tone space dispersion will result in higher intelligibility score than productions with reduced tone space dispersion for a tonal language.

Perhaps a larger lexical tone space expansion is especially important for tonal language users who have sensorineural hearing loss (SNHL). In addition to the general clarity problem described in section 2.1, listeners with SNHL have poorer ability to perceive pitch of complex tones than normal-hearing listeners (Moore, 2007a). They have elevated thresholds for discrimination and reduced ability to take the advantage of *f0* to separate competing sounds as well as difficulty in tracking rapid changes in the pitch (Moore and Carlyon, 2005). This deficit might be more harmful to tonal language users than non-tonal language users. A larger tone space expansion, i.e., more contrast between tone categories, may benefit tonal language users with SNHL in tone discrimination task and speech intelligibility tests.

On the contrary, increase in *f0* might not benefit the listener. Picheny et al. (1985) reported a wider range of *f0,* with a slight bias toward higher *f0,* in high intelligibility speech, however, the higher *f0* difference was not dramatic. Bradlow et al. (1996) reported "overall intelligibility was not correlated with mean fundamental frequency independently of the gender-based difference in overall intelligibility." The gender-based difference was due to the difference in fundamental frequency range. Lu and Cooke (2009) reported that an increase in *f0* did not contribute to intelligibility gain when speech-shaped noise was presented to the listener. Overall, no predictions can be made about the intelligibility of speech with an increase in mean *f0* for English.

This seems to be true for Mandarin as well. For Mandarin tones, an increase in *f0* might result in bigger tone space or smaller tone space. For example, in the current study, both mean *f0* and tone space dispersion was increased from NH_quiet (baseline) to HOH_noise. But from HOH_quiet to HOH_noise, mean *f0* was increased while tone space dispersion was decreased. The intelligibility of Mandarin tones and probably Mandarin speech seems to depend on tone space dispersion but not height of *f0*.

6.4 Limitations

Though experiment instructions and interactions between speakers and the listener were used to make the communication situation as real as possible and maximize conscious speech modification, the difference made by the noise at the ear of the speaker could not be completely ruled out. In noisy conditions, maybe at least a small part of the elevation of *f0* was due to speaker's inadvertent response to the noise but not deliberate speech modification to accommodate the listener. Future experiments should aim at better separating inadvertent and deliberate speech modifications.

In addition, prosodic and phonological speech modifications may be made not only on the continuum of *f0* for Mandarin but also the continuum of voice quality (creaky voice, modal voice and breathy voice). Voice quality seems to play a role in the distinctiveness of tones, especially for tone 3. The current study did not investigate this into detail. The two subjects who produced pervasive creaky voice in the baseline condition may be good subjects whom further experiments could be conducted on.

7. CONCLUSION

The current study explored listener-directed Mandarin tone hyperarticulation when the listener was exposed to white noise or a simulated hearing loss. Results suggested that when speakers realized that the listener's access to information was blocked by some barriers, speakers would accommodate the listener by producing hyperarticulation in certain acoustic dimensions according to the nature of the barrier.

In terms of degradation of audibility of speech, the influences of white noise and hearing loss on the listener were similar. Speech modifications happened when addressing a listener hearing in noise were indeed found to be similar to that when addressing a listener who was experiencing hearing loss. Speakers were aware of the similarity between difficulties induced by white noise and hearing loss and produced a similar phonetic modification, elevation of *f0,* to accommodate the listener.

In terms of degradation of clarity of speech, speakers might have considered that the simulated hearing loss caused a bigger problem for the listener than white noise. Speech modification was indeed found to be different at the phonological level. Speakers produced larger tone space dispersion when addressing the listener with a simulated hearing loss than when addressing the listener experiencing white noise. Interactions between environment and listener suggested a tradeoff between raising *f0* and increasing tone space dispersion when one speaking near his upper *f0* limit. Interactions between tone and listener suggested that different types of tones might have different hyperarticulation mechanisms.

The whole study suggested that listeners' influence on speech modification was obvious. Adjustments observed in various so called listener-directed speech and non-listener-directed speech (e.g., Lombard speech) may all originate from perceptual considerations on the part of the speaker.

REFERENCES

Andruski, J. E., & Kuhl, P. K. (1996). The acoustic structure of vowels in mothers' speech to infants and children. In *Proceedings of the Fourth International Conference on Spoken Language Processing* (pp. 1545–1548). Philadelphia, PA.

Biersack, S., Kempe, V., & Knapton, L. (2005). Fine-tuning speech registers: a comparison of the prosodic features of child-directed and foreigner-directed speech. In *INTERSPEECH* (pp. 2401-2404).

Bordon, G. J., Harris, K. S., and Raphael, L. J. (1994). Speech production: The raw materials-neurology, respiration, and phonation. In *Speech Science Primer: Physiology. Acoustic, and Perception of Speech* (Lippincott Williams and Wilkins, Baltimore, MD), Chap. 4.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America*, 97(3), 1916-1924.

Bradlow, A. R. (1996). A Perceptual Comparison of the /i/–/e/and/u/–/o/Contrasts in English and in Spanish: Universal and Language-Specific Aspects. *Phonetica*, 53(1-2), 55-85.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech communication*, 20(3), 255-272.

Cooke, M., & Lu, Y. (2010). Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America*, 128 (4), 2059-2069.

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: Vowel intelligibility for normal-hearing listeners. *The Journal of the Acoustical Society of America*, 116 (4), 2365-2373.

Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language, and Hearing Research*, 50(5), 1241-1255.

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message?. *Child development*, 1497-1510.

Gramming, P., Sundberg, J., Ternström, S., Leanderson, R., & Perkins, W. H. (1988). Relationship between changes in voice pitch and loudness. *Journal of Voice*, 2(2), 118-126.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., ... & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, 277(5326), 684-686.

Lane, H., & Tranel, B. (1971). The Lombard sign and the role of hearing in speech. *Journal of Speech, Language, and Hearing Research*, 14(4), 677-709.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In *Speech production and speech modelling* (pp. 403-439). Springer Netherlands.

Liu, H. M., Tsao, F. M., & Kuhl, P. K. (2007). Acoustic analysis of lexical tone in Mandarin infant-directed speech. *Developmental Psychology*, 43(4), 912.

Lombard, E. (1911). Le signe de l'elevation de la voix. *Ann. Maladiers Oreille, Larynx, Nez,*

Pharynx, 37, 101–119.

Lu, Y., & Cooke, M. (2008). Speech production modifications produced by competing talkers, babble, and stationary noise. *The Journal of the Acoustical Society of America*, 124(5), 3261-3275.

Lu, Y., & Cooke, M. (2009). The contribution of changes in F0 and spectral tilt to increased intelligibility of speech produced in noise. *Speech Communication*, 51(12), 1253-1262.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., & Cristia, A. (2015). Mothers Speak Less Clearly to Infants Than to Adults A Comprehensive Test of the Hyperarticulation Hypothesis. *Psychological science*, *26*(3), 341-347.

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence?. *Cognition*, *129*(2), 362-378.

Moore, B. C., & Carlyon, R. P. (2005). Perception of pitch by people with cochlear hearing loss and by cochlear implant users. In *Pitch* (pp. 234-277). Springer New York.

Moore, B. C.J. (2007a). Pitch Perception and Frequency Discrimination, in *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues, Second Edition*, John Wiley & Sons, Ltd, West Sussex, England. doi: 10.1002/9780470987889.ch6

Moore, B. C.J. (2007b). Speech Perception, in *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues, Second Edition*, John Wiley & Sons, Ltd, West Sussex, England. doi: 10.1002/9780470987889.ch8

Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin

Chinese tones. *The Journal of the Acoustical Society of America*, 102(3), 1864-1877.

Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of speech, language, and hearing research*, *47*(5), 1048-1058.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1985). Speaking Clearly for the Hard of Hearing I: Intelligibility Differences between Clear and Conversational Speech. *Journal of Speech, Language, and Hearing Research*, 28, 96-103.

Picheny, M. A., Durlach, N. I., & Braida, L. D. (1986). Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech. *Journal of Speech, Language, and Hearing Research*, 29(4), 434-446.

Scarborough, R., & Zellou, G. (2013). Clarity in communication: "Clear" speech authenticity and lexical neighborhood density effects in speech production and perception. *The Journal of the Acoustical Society of America*, 134(5), 3793-3807.

Scarborough, R., Dmitrieva, O., Hall-Lew, L., Zhao, Y., & Brenier, J. (2007). An acoustic study of real and imagined foreigner-directed speech. *Journal of the Acoustical Society of America*, 121(5), 3044

Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language and speech*, 34(2), 145-156.

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and linguistics compass*, 3(1), 236-264.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., & Stokes, M. A. (1988). Effects of noise on speech production: Acoustic and perceptual analyses. *The Journal of the*

Acoustical Society of America, 84(3), 917-928.

Uchanski, R. M. (2008). Clear speech. In *The Handbook of Speech Perception*, edited by D. Pisoni and R. Remez (Blackwell, Malden, MA), pp. 207–235.

Uther, M., Knoll, M. A., & Burnham, D. (2007). Do you speak E-NG-LI-SH? A comparison of foreigner-and infant-directed speech. *Speech Communication*, 49(1), 2-7.

Wright, R. (2004). Factors of lexical competition in vowel articulation. *Papers in laboratory phonology VI*, 75-87.

Xu, Y. (2013). ProsodyPro - A Tool for Large-scale Systematic Prosody Analysis. In *Tools and Resources for the Analysis of Speech Prosody* (pp. 7 - 10). Laboratoire Parole et Langage, France: Aix-en-Provence, France.

Zhao, Y., & Jurafsky, D. (2009). The effect of lexical frequency and Lombard reflex on tone hyperarticulation. *Journal of Phonetics*, 37(2), 231-247.

APPENDIX I

Main results of tone space dispersion measured on Hertz scale

1. Listener

Estimates

			95% CI	
listener	mean	standard error	lower limit	upper limit
NH	31.524	1.312	28.777	34.271
HOH	43.644	.785	42.001	45.286

pairwise comparisons

(I) lis	(J) lis	Mean difference (I-J)	standard error	Sig.
NH	HOH	-12.119	.968	.000

Significance level: 0.05。

adjustment： Bonferroni

2. Environment

Estimates

			95% CI	
env	mean	standard error	lower limit	upper limit
quiet	34.978	1.104	32.667	37.288
noise	40.190	.956	38.188	42.192

Pairwise comparisons

(I) env	(J) env	Mean difference (I-J)	standard error	Sig.
quiet	noise	-5.212	.726	.000

significance level: 0.05

Adjustment: Bonferroni

3. Tone

Estimates

			95% CI	
tone	mean	standard error	lower limit	upper limit
1	33.505	2.257	28.782	38.229
2	28.120	.886	26.266	29.975
3	43.821	2.446	38.700	48.941
4	44.889	1.522	41.702	48.076