Available online at www.sciencedirect.com
ScienceDirect
Journal of Phonetics 37 (2009) 231–247
Journal of Phonetics
www.elsevier.com/locate/phonetics

The effect of lexical frequency and Lombard reflex on tone hyperarticulation

Yuan Zhao*, Dan Jurafsky

Department of Linguistics, Stanford University, Margaret Jacks Hall, Bldg. 460, California 94305, USA

Received 13 April 2007; received in revised form 20 February 2009; accepted 2 March 2009

Abstract

Among the many factors that affect phonetic variation, two of them, low word frequency and speech in noise (Lombard speech), are linked to stronger or fuller realizations of segments. Noise and low usage frequency are both factors that may lead to perceptual difficulty suggesting that the strengthening due to these two factors may be related. To understand how these two factors relate to each other and to see if they play a similar role with respect to suprasegmental phenomena like tone, we investigated the effect of ambient noise and word usage frequency on tone production in Cantonese. We recorded Cantonese monosyllabic words of high and low usage frequency in both normal and Lombard speech styles, controlling for segmental and other factors and measured *f*0. We found that both ambient noise and lexical frequency influence tonal production. All six tones are produced with higher *f*0 in the presence of noise. Usage frequency affects *f*0 realization as well, but only for words with mid-range (mid-level or mid-rising) tones. Low-frequency words with mid-range tones are produced with higher *f*0 than frequent words. Extending earlier work showing greater vowel-space dispersion in less frequent words, we also found that the *f*0 trajectories of less frequent words are more dispersed than that of their more frequent counterparts, especially for mid–low tones. This dispersion does not occur for Lombard speech. Our results suggest that different aspects of speech production may account for the strengthening of speech in noise and in low-frequency words.

1. Introduction

Understanding the causes of variation in the surface form of words is crucial to building models of spoken word production and recognition. One of the earliest-reported correlates of phonetic variation is word frequency, noted by Sibawayhi, the Arabic grammarian of the 8th century (Al-Nassir, 1993; Carter, 2004) and well-known by early modern linguists (Schuchardt, 1885; Zipf, 1929; *inter alia*). A number of more recent studies show that frequent or predictable words are shorter and display various other lenited characteristics such as reduced vowels and simplified or deleted codas compared to less frequent words. Less frequent words are longer and have stronger or fuller forms (Aylett & Turk, 2006; Bell, Brenier, Gregory, Girand, & Jurafsky, 2009; Bybee, 2000; Fidelhotz, 1975; Fosler-Lussier & Morgan, 1998; Hooper, 1976; Munson, 2007; Pluymaekers, Ernestus, & Baayen, 2005; Rhodes, 1992, 1996). Strengthened and longer forms are also associated with another cause of variation: ambient noise. Speech produced in noise is often referred to as Lombard speech (Lombard, 1911). Modifications due to noise vary somewhat depending on the type of noise and the task, but in general Lombard speech has increased duration and energy, higher *f*0, and segmental effects such as changes in spectral tilt and an increase in the energy center of gravity (e.g. Draegert, 1951; Junqua, 1993; Lane & Tranel, 1971; Summers, Pisoni, Bernacki, Pedlow, & Stokes, 1988).

Thus both (low) word frequency and Lombard speech are associated with strengthening such as increased duration and energy and, at least for Lombard speech, higher *f*0. Although these phenomena have not previously been considered together, the same explanation has often been independently offered for both: phonetic strengthening may aid perception, either by the hearer or by the speaker (in monitoring their own speech). Thus for example the Hyperarticulation and Hypoarticulation (H&H) model (Lindblom, 1990) claims that speakers will produce strengthened phonetic forms to counteract

---

*Corresponding author.
E-mail address: yuanzhao@stanford.edu (Y. Zhao).

0095-4470/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2009.03.002

difficulties in comprehension of the sort that might arise from low word predictability or from ambiguity due to noise. Such a model is supported by a wide variety of other evidence (summarized below) showing that less predictable elements are more hyperarticulated than more predictable elements (e.g., Aylett, 2000; Bell et al., 2003; Raymond, Dautricourt & Hume, 2006; Jurafsky, Bell, Gregory, & Raymond, 2001; Liberman, 1963).

Noise and usage frequency are linked in another way. In a Bayesian or "noisy channel" model of comprehension, a comprehender chooses the interpretation which is most probable given the input. For word recognition, a Bayesian recognizer (Norris, 2006; Jurafsky, 2003; Norris & McQueen, 2008) selects the word which has the highest probability given the speech input. Bayesian models integrate two kinds of probability for each candidate word: the *prior probability* of the word (how expected the word is *a priori*) and the *likelihood* of the acoustic evidence given the word (how much the acoustic input matches our expectations of how this word should be acoustically realized). Low word frequency decreases the prior probability of the word; noisy acoustics decreases the likelihood of the acoustics, given the word.

Despite the fact that Lombard speech and word usage frequency can both play a role in models like H&H and Bayesian recognition, we know very little about the similarities and differences of their effects on phonetic realization.

Another important gap in our understanding of these sources of strengthening is related to their impact on suprasegmental phenomena, such as lexical tone. Recent research has suggested that speakers might make use of prosodic prominence as a linguistic means to achieve smooth signal redundancy and therefore improve communication robustness (Aylett, 2000; Aylett & Turk, 2004). We know that both usage frequency and noise have impacts on *f*0; rare words are more likely to have a pitch accent in English and Dutch (Pan & Hirschberg, 2000), and *f*0 is generally higher in Lombard speech (Draegert, 1951; Junqua, 1993; Lombard, 1911). Assuming both word frequency and ambient noise are linked with changes in *f*0 as a consequence of a smooth profile of signal redundancy, we know very little about how these two phenomena, which emphasize different aspects of the predictability of the signal, behave similarly or differently. Similarly, we do not know how strengthening would be manifested on *f*0 phenomena like lexical tone. Lexical tone is contrastive; a change in *f*0 can change the identity of a lexical tone. Does the phonemic nature of lexical tone influence the effect of word frequency and the Lombard reflex on *f*0 in tone languages?

In the next two sections, we summarize some of what is known about usage frequency and Lombard speech, before turning to our new study.

1.1. Word frequency and production

The idea that more frequent words tend to be shorter than less frequent words was actually first noticed in the 8th century by the Arabic grammarians (Al-Nassir, 1993; Carter, 2004) and rediscovered by linguists in the 19th and early 20th century (Schuchardt, 1885; Zipf, 1929, among others). Modern studies first demonstrated the effect of frequency on lenited pronunciation. Fidelhotz (1975), for example, showed that low-frequency words like *forfend* were less likely to have a schwa vowel in the first syllable than high-frequency words like *forget*. Rhodes (1992, 1996) showed that /t/s in low-frequency words were less likely to flap than /t/s in high-frequency words. Bybee (2000) found that deletion was less likely in low-frequency words than in high-frequency words.

Recent studies have found similar effects in natural spoken corpora. Bybee (1996) showed that word-final /t/ and /d/ deletion rates in a corpus of spoken Chicano English were lower in low-frequency words (34.3%) than in high-frequency words (54.5%). Gregory et al. (1999) and Jurafsky et al. (2001) confirmed this lower rate of final /t/ and /d/ deletion for low-frequency words in a corpus of American English telephone conversations by using multiple regression to control for factors such as segmental context, rate of speech, word-length in phones, and neighboring disfluencies. They also found that low-frequency words were longer than high-frequency words. Patterson and Connine (2001) investigated flap production in the telephone speech corpus Switchboard. They found that the probability of occurrence of flaps in a medial /t/ position was a function of the lexical frequency of the carrier word. The probability of flapping was smaller for lower frequency words (76%) than high frequency words (95%). Aylett and Turk (2004) examined syllable duration in the HCRC Map Task Corpus of task-oriented dialogs between Glaswegian English speakers. They found that syllables from low-frequency words were longer than syllables from high-frequency words, after controlling for many other factors.

More recent studies on vowel production, such as Aylett and Turk (2006), found that vowels showed increased centralization with increased predictability. Munson (2007) had subjects read high-frequency and low-frequency CVC words and studied the resulting vowel-space dispersion. He found that low-frequency words had an expanded vowel space compared to high-frequency words.

In general, whether all these results are caused by a strengthening of low-frequency words or a weakening or lenition of high-frequency words is still not well understood. Some models argue that both factors may play a role (see, e.g., Bell et al., 2009) and other models such as Aylett and Turk (2004) propose that pronunciation strength is inversely related to word frequency and therefore no default realization is required.

Many experiments have also focused on factors closely related to frequency, including predictability and frequency-weighted neighborhood density. In a classic early study, Liberman (1963) showed that words were less reduced when they occurred in less predictive sentence contexts. Gregory, Raymond, Bell, Fosler-Lussier & Jurafsky (1999), Jurafsky et al. (2001) and Raymond et al. (2006) found that (contextually) less predictable content words were longer and less likely to have /t/ or /d/ deleted. Bell et al. (2003) found that unpredictable function words had less vowel reduction, less coda deletion and were longer. Pluymaekers et al. (2005) showed similar effects of local predictability on reduction in seven Dutch adverbs. Aylett and Turk (2004) found that the predictability and also the given/new status of a word affected its duration. In addition to predictability, the frequency-weighted neighborhood of a word has also been shown to influence phonetic production. Scarborough (2004), for instance, found that French words from dense neighborhoods had higher degree of nasal and vowel-to-vowel coarticulation. Munson and Solomon (2004) reported that vowels were produced with a higher degree of dispersion in words with many neighbors.

In summary, lexical frequency and other related lexical factors have been shown to affect a wide variety of phonetic variables, including vowels (longer, less use of schwas, and expanded vowel space) as well as consonants (longer, less flapping, less palatalization, and less coda deletion). Lexical frequency and predictability have also been shown to have an effect on prosodic phenomena: low frequency or low predictability words are more likely to receive pitch accent (Pan & Hirschberg, 2000). One important class of phonetic phenomena, however, has not been investigated for the influence of word frequency: lexical tone.

Should we expect to see the effect of frequency on tone? It is very likely that there is a frequency effect, given any of the various models that have been proposed to explain the effect of word frequency on phonetic reduction at the segmental or prosodic level. These include the H&H Articulation model of Lindblom, which suggests that speakers use hyperarticulation or hypoarticulation based on the potential needs of the hearer (Lindblom, 1990), or models based on lexical access (Munson, 2007) mediated by an articulatory pipeline (Pluymaekers et al., 2005), models based on smoothing language redundancy (Aylett & Turk, 2004) or models of prosodic strength (Kochanski & Shih, 2003). We do not intend in this paper to compare or contrast these different potential models, and indeed it seems likely that frequency effects reflect more than one such causal factor. We note, however, that all of these models explain segmental or prosodic weakening of frequent forms with general mechanisms, suggesting that the effects should be observed across a variety of surface phonetic manifestations including the use of *f*0 to realize lexical tone.

1.2. Lombard speech

As early as 1911, Etienne Lombard observed that talkers adjusted their speech and increased vocal effort when speaking in noise. He proposed that the acoustic modification was a result of maintaining the ability to self-monitor through the auditory feedback loop. Later studies reported various aspects of this acoustic modification. For instance, Draegert (1951) found, in utterances with the presence of noise, an increase in intensity, a change in the distribution of spectral energy, and modified vowel quality. The observed hyperarticulation effect in Lombard speech extends to suprasegmental production as well: changes in the prosodic characteristics of Lombard speech. Hanley and Steer (1949) showed that in the presence of noise speakers talked much louder and increased the duration of the utterance. Draegert (1951) and Junqua (1993) found that when speaking with masking noise, speakers not only reduced speech rate, but also used higher *f*0. Other studies further show that speakers raise *f*0 to different degrees according to different levels of noise (Summers et al., 1988).

Some other studies argue that the phonetic hyperarticulation in Lombard speech is not just a result of our self-monitoring needs. The acoustic enhancement in noise serves to improve communicative effectiveness in a degraded environment (Letowski, Frank, & Caravella, 1993) and to optimize information transfer to listeners (Lane & Tranel, 1971). One inadequacy of earlier research on Lombard speech is that it focuses on ambient noise as the sole variable, without taking into consideration the tasks and the linguistic content involved in speech production. Only recently have studies started to show that communicative factors and the informational status of the word might also play a role in the execution of this noise-induced hyperarticulation. For instance, Lu and Cooke (2008) examined the acoustic differences in the Lombard realization of utterances from pairs of speakers with or without a collaborative task involved. In one condition, subjects were asked to perform a puzzle-solving task with a partner when speaking in noise, while in the other condition, subjects were asked to speak aloud while solving the puzzle alone without interacting with a partner. They found that both groups produced acoustic modification consistent with the Lombard effect. However, the Lombard effect produced by the group performing a collaborative task with a partner had a stronger effect. Besides communicative factors, recent studies such as Patel and Schell (2008) showed that word meaning also plays a role in the degree of phonetic strengthening in Lombard speech. They found that at moderate noise levels, most word types appeared to be uniformly boosted in *f*0, intensity and duration. However, as the noise level increased, words referring to agents, objects and locations were disproportionately elongated and raised in *f*0 compared with the rest. They argue that linguistic content shapes the magnitude of the Lombard effect, with *f*0 and duration serving as primary cues for marking the saliency of those information-bearing word types.

While we thus know quite a bit about English Lombard speech, no systematic study has looked at the impact of the Lombard reflex on the realization of lexical tones in tonal languages. It is not clear whether the prosodic changes observed in English Lombard speech (such as higher *f*0) would also be found in the production of lexical tones. If Lombard articulation plays a universal role in increasing the rate of vocal fold vibration, we would expect to see the same rise in *f*0 in Lombard realizations of tones as we observe in English Lombard speech (e.g., Draegert, 1951; Junqua, 1993; Summers et al., 1988). Alternatively, if the Lombard reflex plays its role in a more complex way, we might expect to see different effects on lexical *f*0 than on the intonational contour, given lexical *f*0's phonemic nature.

1.3. Our approach

In this study we offer a preliminary investigation of the role of usage frequency and ambient noise on tone production in Cantonese. We examine words of low usage frequency and words of high usage frequency, in noise and no-noise conditions. Our goal is to investigate how the factors of rarity and noise affect the realization of *f*0, and to understand the similarities and differences between the two.

We hypothesize that the tones of low-frequency words and tones produced in the presence of noise will both be hyperarticulated when compared with their high-frequency and no-noise counterparts. We can imagine a few possible scenarios for the realization of this hyperarticulation. If the "function" of hyperarticulation is to make elements louder and/or phonetically stronger in other aspects, whether for our own needs of self-monitoring or for the benefit of listeners, *all* tones should be realized with higher *f*0, since a higher *f*0 increases the spectral energy overall. Higher *f*0 is in general correlated with increased vocal efforts in articulation and with signal intensity (e.g. Junqua, 1993; Picheny, Durlach, & Braida, 1986; Rostolland, 1982). This scenario thus predicts an overall increase in mean *f*0 for all lexical tones, whether they are high or low tone.

Alternatively, if the function of the hyperarticulation is to make elements more *distinct* within the phonemic tone inventory, we might expect to see higher tones getting higher and lower tones get lower, or dynamic tones getting "stretched", so that the high *f*0 point of the dynamic tone gets higher and the low point gets lower, or some other such reconfiguration in the use of the tone space to increase the contrast between tones.

We therefore propose to investigate the details of how noise and usage frequency affect the *f*0 realization of words with each lexical tone, focusing on whether we see the tones behave similarly or differently from each other.

We will also examine the relationship of these two factors with other potential variables including the gender of the speaker.

2. Method

2.1. Test language: Cantonese

Our test language is Cantonese, a Chinese language with a rich inventory of both static and dynamic tones. Cantonese has six contrastive tones, including three level tones that contrast in *f*0 height (high-level: tone 1; mid-level: tone 3; low-level: tone 6), two rising tones (high-rise: tone 2; mid-rise: tone 5), and a low-falling tone (tone 4) (Cheung, 1986).[1] Fig. 1 shows the *f*0 contours of the six Cantonese tones that we will be investigating.

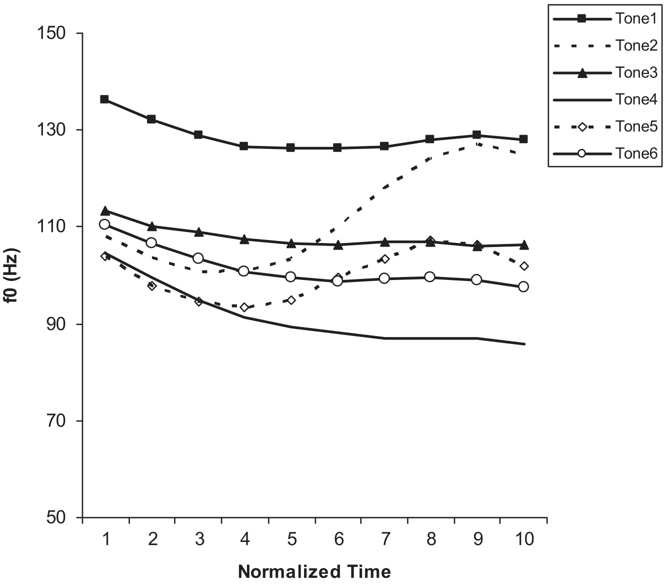

Fig. 1. *f*0 contours of Cantonese tones.

2.2. Subjects

Eight native Hong Kong Cantonese speakers, four males and four females, participated in the experiment for $10 compensation for each session. Subjects' age ranged from 20 to 52. All of them were born and raised in Hong Kong and had lived in the United States for less than 3 years. All speakers were at least college-educated and were literate. The speakers reported no known speech, language or hearing disorders.

2.3. Materials

The stimuli consisted of 92 words paired between high-frequency usage and low-frequency usage, and balanced across tones (words are listed in the appendix). A Chinese talking syllabary of the Cantonese dialect (Huang, 1997) was used as the reference for selecting words of each standard citation tone. Word token frequency was calculated from Academia Sinica's Balanced Corpus of Modern Chinese (Hsu & Huang, 1995; Huang & Chen, 1992), which has about five million words of written Chinese sampled from different genres and styles.

---

[1] There are three additional "checked tones", i.e. syllables ending with unreleased stops, which mainly differ with the other tones in duration. Checked syllables are systematically shorter in duration than unchecked ones but do not differ from the unchecked tones in pitch height, and were not included in the study.

The raw token frequency was transformed into log frequency. High-frequency words were chosen from the top quartile of the frequency range and low-frequency words were chosen from the bottom quartile of the frequency range. The mean log frequency of high-frequency words was 3.655 and that of low-frequency words was 0.969. Ideally one would calculate the word frequency from a psycholinguistically designed corpus of Cantonese. Chinese, however, is diglossic; Cantonese is a spoken language and it is common in reading experiments like ours to use written frequencies. The Chinese written by Cantonese speakers is the same as standard written Chinese except for dialect-specific lexical items. Due to the unavailability of a balanced Cantonese corpus, we therefore based our counts on the Sinica corpus of written Chinese. We did check our frequencies against a corpus of Chinese presumably written by and/or for Cantonese speakers: the corpus of Hong Kong Newspaper text released by the Linguistic Data Consortium (LDC, 2000), containing 18,147 news articles (about 10 million Chinese characters total). All our high-frequency and low-frequency words were still in the same frequency bin according to this LDC written Cantonese corpus.

All high/low frequency word pairs were monosyllabic with CV or CVC structure. The syllable structure within each pair of words was kept constant. Syllables were chosen with the initial consonant, vowel and coda matched within each pair as much as possible. Voicing and the manner of the initial consonants were matched as much as possible. Contrastive vowel length was also consistent within each pair. Only words with nasal codas were used in CVC word pairs and in each pair the same nasal coda was selected. We tried to avoid using homophone pairs to minimize confounding factors (see Caramazza, Costa, Miozzo, & Bi, 2001; Jescheniak & Levelt, 1994). χ^2 tests showed that the high-frequency word list and the low-frequency list did not differ significantly in the manner of the initial consonant (χ^2 [2] = .913, $p < .633$), the length of the vowel (χ^2 [1] = .289, $p < .591$) or the presence/absence of the coda (χ^2 [1] = .013, $p < .908$). Due to our attempt to control for phonological and frequency constraints, we ended up with a few words in our high-frequency stimuli that could be considered "closed class" according to some classifications (such as Zhang, 2003). We will come back to this limitation at the end.

After a set of potential stimuli was chosen, it was given to two native Cantonese speakers to rate the familiarity of the words on a scale from 1 to 10. The final 92 stimuli used in the study were all rated above 6 on the familiarity scale of 10. The mean rating of high-frequency words' familiarity was 7.2 and that of low-frequency words was 6.9.

2.4. *Procedure*

The recording took place in the sound booth of the Phonetics Lab at Stanford University. Subjects participated in two recording sessions. In the first session they recorded the word list under normal speaking condition. Subjects were presented with words in traditional Chinese orthography, in a random order one at a time on a computer screen. They produced the words in isolation self-paced without intervention and the audio was recorded through a Shure microphone (SM10A). After two weeks, subjects came to the lab again and recorded the same set of materials while listening to a signal containing 75 dB white noise. The noise was calibrated by a sound-level meter prior to the experiment and it was presented binaurally through a Sennheiser headphone (HD 280). Both sessions were recorded using a Panasonic Professional DAT recorder at a sampling frequency of 44.1 K. The data were later analyzed with Praat (Boersma & Weenink, 2000).

2.5. *Measurements*

The sound files for both conditions were segmented into words, using the inter-word silence as the criterion. Word boundaries were hand-labeled and annotated by the first author using Praat. After the boundaries were hand-delineated, *f*0 was automatically measured for each word by a commonly used Praat script created by Xu (2006).

In order to capture as much tonal information as possible, the *f*0 values of the tone trajectory of each word were measured at ten equidistant points. Since the *f*0 at the onset of the vowel may be perturbed by the preceding consonant (e.g. Hombert, Ohala, & Ewan, 1979) the *f*0 of the initial vocalic segment was not taken into consideration and hence *f*0 was measured starting 10% of the way through the vocalic segment from the voicing onset (10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%).

3. Analysis and results

3.1. *Effect of noise and lexical frequency on f0*

In order to examine the effect of noise and usage frequency on *f*0, we calculated the mean *f*0 for each time point by averaging across all tokens of the same tone type at the same time point produced by all speakers. Fig. 2 shows the *f*0 trajectories of six tones produced in normal and Lombard conditions; Fig. 3 plots the *f*0 trajectories of six tones separately for high-frequency and low-frequency words.

A Repeated Measures ANOVA was used to examine whether speaking condition (normal vs. Lombard), lexical frequency category of the word (high vs. low) and tone types (tones 1–6), as well as speakers' gender had a significant influence on the production of words' overall mean *f*0 calculated by averaging across the ten points of all tones of all speakers.

Lexical frequency and speaking condition both had significant main effects on mean *f*0 ($F_{\text{lexical\_frequency}}(1,5) = 18.929, p < .007; F_{\text{speaking\_condition}}(1,5) = 6.503, p < .043$). Low-frequency words were produced with a higher *f*0 and speaking in noise also resulted in a higher *f*0. In addition,

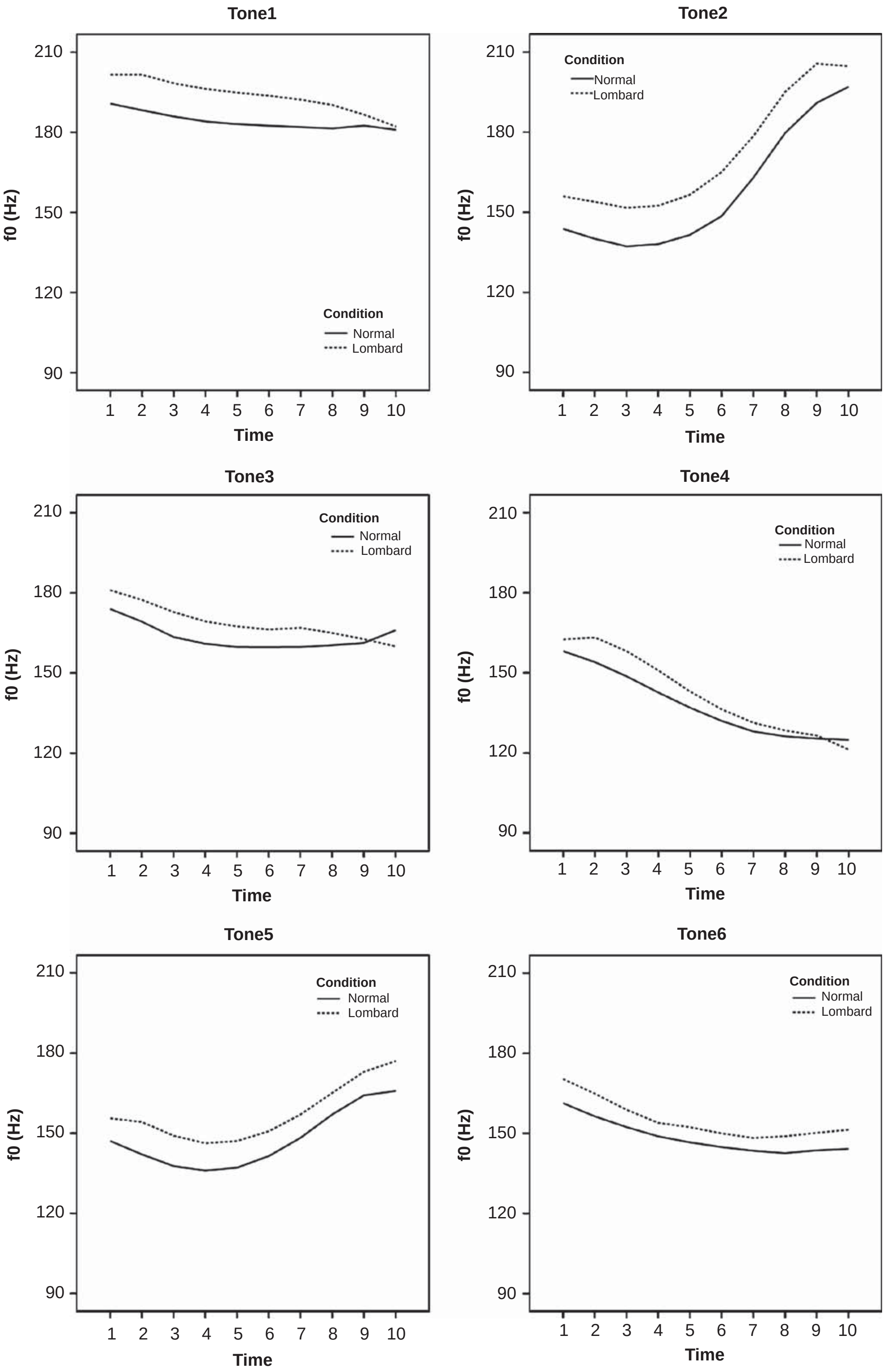

Fig. 2. $f0$ trajectories of six tones produced in normal and Lombard conditions averaged across speakers.

tone type and gender, not surprisingly, had significant main effects on the mean $f0$ ($F_{\text{tone\_type}}(5,25) = 49.648$, $p < .000$; $F_{\text{gender}}(1,5) = 66.756$, $p < .000$).

There was also a significant interaction between tone type and lexical frequency ($F(5,25) = 4.786$, $p < .003$). The interaction, visible in Fig. 3, suggested that the change in $f0$ between high-frequency and low-frequency words differed by the tone type and not every tone was equally subject to the influence of lexical frequency. The two tones that were influenced by usage frequency were mid-level tone 3 and low-rising tone 5, both of which were mid-range tones in Cantonese. *Post-hoc* Bonferroni tests show that lexical

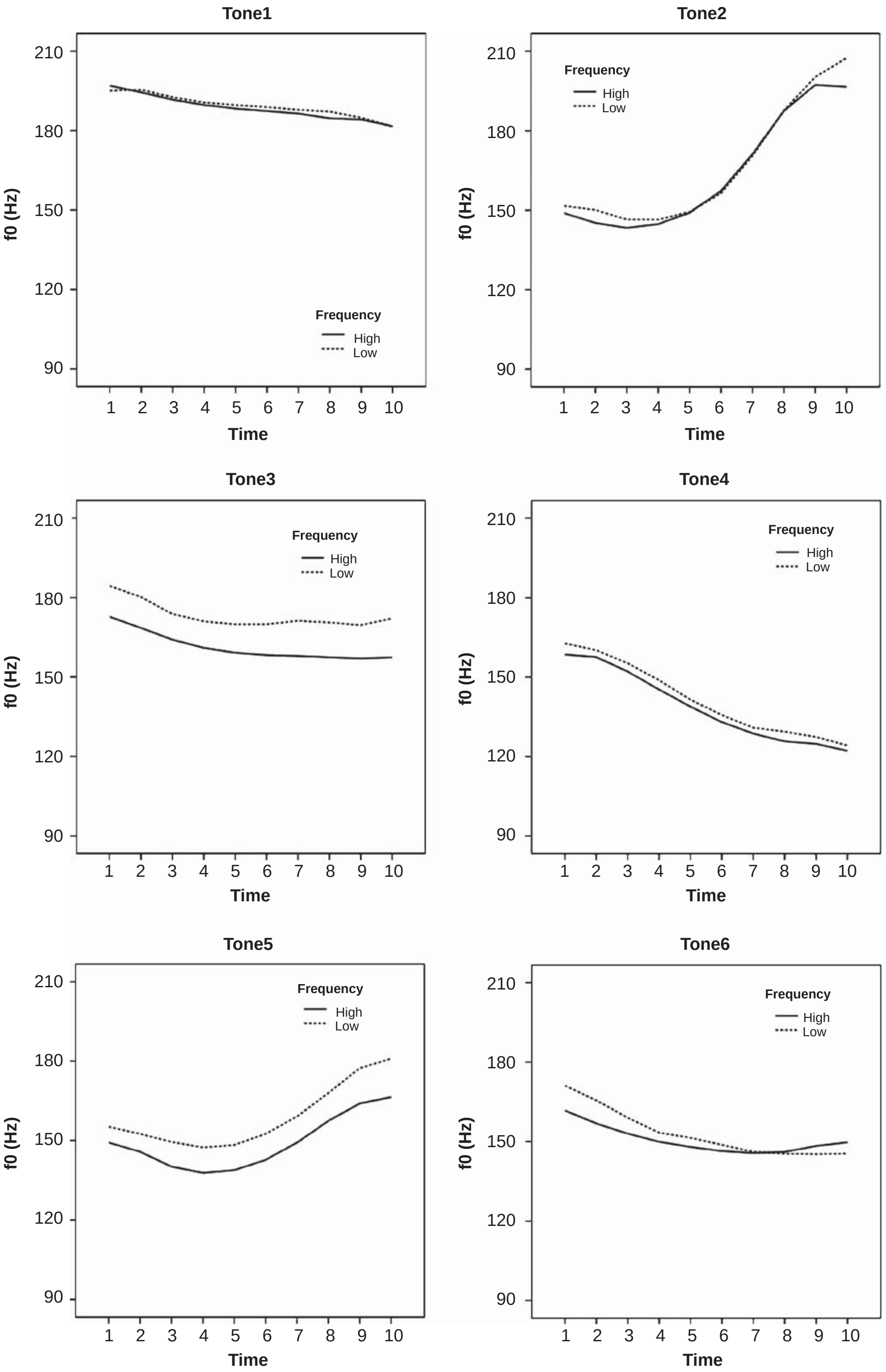

Fig. 3. *f*0 trajectories of six tones for high-frequency and low-frequency words averaged across speakers.

frequency had a significant effect on tone 3 ($F(1,7) = 13.147$, $p < .008$) and tone 5 ($F(1,7) = 8.151$, $0 < .025$). Low-frequency words of mid-level tone 3 were produced higher in *f*0 than high-frequency words ($M_{\text{low-frequency}} = 173.096$, $M_{\text{high-frequency}} = 162.182$). The *f*0 of low-frequency low-rising tone 5 words was also higher than that of high-frequency tones 5 words ($M_{\text{low-frequency}} = 155.502$, $M_{\text{high-frequency}} = 151.200$).

In contrast with the frequency effect, there was no interaction between ambient noise and tone type. As suggested by Fig. 2, the Lombard effect on lexical tones was a universal raising in *f*0 across all tone types.

3.2. Interim discussion

Our results show that speech style and lexical frequency both influence tone production. Consistent with previous findings in non-tonal languages like English (Junqua, 1993), Lombard production of tones in Cantonese is characterized by raised *f*0. The lack of interaction with tone type suggests that this rise occurs across all tone types. In contrast, usage frequency does not affect all tones equally. Low-frequency words of mid-level (tone 3) and low-rising (tone 5) tones are more likely to be raised in *f*0.

The fact that Lombard speech leads to higher *f*0 for all tones, while usage frequency only influences part of the tonal inventory, suggests that the two kinds of hyperarticulation may play different functional roles. The universality of higher *f*0 in noise suggests a *strengthening*-oriented hyperarticulation, functioning to phonetically strengthen some acoustic dimensions, such as *f*0 or intensity, whether for self-monitoring for auditory feedback (Lombard, 1911) or in a listener-directed effect similar to the many kinds of clear speech such as speech to the hard-of-hearing (Picheny et al., 1986).

Since low word frequency, by contrast, only affects a subset of tones, it seems unlikely to be functioning to increase the phonetic strength; there is no obvious reason to limit phonetic strengthening to only a subset of the tones. By contrast, a *distinction* model, which is sensitive to the phonemic structure of the tonal inventory, and whose function was to make tones more perceptually distinct from each other, might predict changes in only a subset of the tones, perhaps to help separate some specific neighboring tones of high confusability.

In order to tease apart these two models, then, it is important to come up with a metric for tone distinctiveness, and measure the effects of lexical usage frequency and Lombard speech on this metric. In the next section we introduce such a measure: **tone space**, investigating whether the phonemic structure of the tonal inventory may help in understanding these differential roles of lexical frequency and ambient noise.

4. Hyperarticulation and tone-space adjustment

4.1. Overview

In the previous section we suggested that different functional aims might explain the fact that ambient noise leads to higher *f*0 for all tones, while lexical frequency does not influence all tones uniformly. Perhaps where noise is to *phonetic strength* of individual tones, lexical frequency plays its functional role in making tones more *phonemically contrastive* or *distinct* from one another.

This idea that one function of hyperarticulation could be to make phonemes more distinct was first suggested for segmental domain by Wright (2003), Munson and Solomon (2004), and Munson (2007), who showed that (point) vowels of low-frequency words were more dispersed in the vowel space than those of their high-frequency counterparts. If suprasegmental production is influenced by lexical factors in similar ways to segmental production, we might expect some similar adjustments in the acoustic *tone space* of a tonal language when low-frequency words are produced.

By contrast, there is no evidence from segmental research that Lombard speech leads to more dispersed or contrastive vowels. Lombard or shouted speech does show changes in the vowel space, but the resulting space is not necessarily expanded, and it also shows huge individual differences in the direction and the degree of the shifts (Summers et al., 1988). Indeed, in some cases, due to the shifting of the formant frequencies, loud speech is less intelligible and harder to understand (Rostolland, 1982). This suggests that in the Lombard articulation of lexical tones, we would not expect an expansion of *tone space*, despite the overall increase in *f*0.

If there are indeed expansions in tone space for low-frequency words, we might also expect to see differences between male and female speakers. Cross-linguistically, females make use of a larger acoustic space than male speakers, including pitch range as well as vowel space (e.g. Diehl, Lindblom, Hoemeke, & Fahey, 1996; Henton, 1995). We will therefore also look at possible gender differences in tone space.

In this section we investigate these issues by proposing a measure of **tone space** and its dispersion, and investigating the consequences of raised *f*0 on phonemic contrast within the tone inventory.

4.2. Method

A natural way of defining tone space is to draw on previous work on vowel-space. Models of vowel-space dispersion are based on the mean Euclidean distance from the center of the participants' $F1/F2$ space (see. e.g., Bradlow, Torretta, & Pisoni, 1996; Munson, 2007; Wright, 2003). This metric for vowel dispersion is correlated with word frequency: Munson and Solomon (2004) found that vowels in high-frequency words were less expanded than those in low-frequency words ($M = 2.822$ Bark for high-frequency words; $M = 2.940$ Bark for low-frequency words).

This same intuition seems applicable to tone as well. The closer tone productions are to the center of the acoustic *f*0 space, the less salient the acoustic difference between them, and the less distinctive each tone becomes. Thus if lexical frequency affects production in the tone domain as well, we would predict more tone-space dispersion of low-frequency words than of the high-frequency words.

We propose to measure tone-space dispersion as the mean Euclidean distance of individual tones from the center of the speakers' *f*0 space (acoustic tone space).

How should we measure the center of the speakers' *f*0 space, given that we have multiple estimates of *f*0 across the trajectory of a word? One option is to compare the *f*0 value at each of the 10 points with a single global *f*0 mean. But changes in the tone space may not happen homogeneously across the tonal trajectory of a word. For this reason, rather than keeping a single global value for the center of the speaker's tone space, we will compare tone values at each of the 10 points along the trajectory with the speakers' mean computed just for that point.

We thus define the **central *f*0 at each time point *k*, $CF0_k$** as the mean *f*0 a speaker uses for the time point *k* along the *f*0 trajectory, averaged across words. That is, for each of the ten equidistant points along the trajectory for every word, we compute $CF0_k$, the central *f*0 at the time point *k*, as the average over all words *j* spoken by the speaker of $f0_k^i$, the *f*0 value of word *i* at *k*:

$$CF0k = \frac{1}{j}\sum_{i=1}^{j} f0_k^i \tag{1}$$

Given this mean value, we compute the tonal distance in semitones (st) between an *f*0 value of a particular word *i* at time point *k* and the central *f*0 at *k* as follows:

$$\text{Tonal Distance to Central } f0(TDCF0) = 12\left|\log_2 \frac{f0_k^i}{CF0_k}\right| \tag{2}$$

In this study, we measured *f*0 at ten time points over a number of items. Therefore to obtain the averaged tonal distance from a tonal trajectory to the speaker's central *f*0, we need to get the mean of *TDCF*0 across all ten time points over all items, as shown by the following:

$$\text{Tonal Dispersion} = \frac{1}{10j}\sum_{k=1}^{10}\sum_{i=1}^{j} TDCF0_k^i \tag{3}$$

We then used the measure of the tonal dispersion to investigate whether hyperarticulation in tones incurred an adjustment in the degrees of tonal dispersion as well.

4.3. Results

4.3.1. Tone space of Cantonese

In order to examine quantitatively whether or not Cantonese speakers make use of a reconfigured tone space when producing hyperarticulated tones, we first calculated the tonal centroids used in the production of high-frequency and low-frequency words respectively, and in Lombard and normal conditions.

Fig. 4 shows the *f*0 trajectories of six tones in the normal and Lombard conditions, each shown with the position of central *f*0. We can see once again that all tones have higher *f*0 when produced under noise. However, the distance between each individual tone and the tonal centroid does not change in any obvious way. Thus while noise leads to higher *f*0, it does not seem to lead to more dispersion or contrast among tones.

Fig. 5 shows the tone trajectories of both high-frequency and low-frequency words together with the central *f*0 of the tone space. Unlike the lack of effect we saw with noise, with usage frequency we do see a change in tonal distance: the distance of tones 3, 4, and 6, respective to the tonal centroid is greater in low-frequency words than high-frequency words. The relative positioning of all tones in the tone space also undergoes visible changes when the tones are realized on low-frequency words, as shown in Fig. 6. Level and falling tones are more spread out from each other and the relative distance among tones 1, 3, 4, and 6 increases; the rising tones' *f*0 maxima are also higher in low-frequency words.

Tone-space dispersion was calculated according to Eq. (3) above. A Repeated Measures ANOVA was used to check whether speaking condition (normal vs. Lombard), lexical frequency category of the word (high vs. low), tone type and gender had a significant influence on the degree of dispersion in tone space. The result showed that tone type and the usage frequency of a word had significant main effects on tone-space dispersion ($F_{\text{tone\_type}}(5,25) = 20.877$, $p < .000$; $F_{\text{lexical\_frequency}}(1,5) = 13.376$, $p < .015$). Speakers used a more dispersed tone space when producing low-frequency words than high-frequency ones ($MD = .417$ st). Unlike lexical frequency, speaking condition was found to have no significant effect on the adjustment of tone space $F_{\text{speaking\_condition}}(1,5) = .278$, $p = .620$), which indicated that speakers did not use a more dispersed tone space when speaking in the presence of noise.

There was a significant interaction between tone type and lexical frequency ($F(5,25) = 5.273$, $p < .002$), visible in Fig. 5. *Post-hoc* Bonforroni tests indicated that lexical frequency had a significant effect on the dispersion of (mid-level) tone 3 ($F(1,7) = 24.319$, $p < .002$), tone 4 ($F(1,7) = 23.296$, $p < .002$) and tone 6 ($F(1,7) = 9.304$, $p < .019$). Low-frequency tone 3's *f*0 trajectories ($M = 2.073$) were more dispersed from the tonal centroid than those of high-frequency tone 3 words ($M = 0.926$). The *f*0 trajectories of low-frequency words with (low-falling) tone 4 *f*0 trajectories ($M = 3.123$) were more dispersed than those of high-frequency tone 4 words ($M = 2.73$); and low-frequency tone 6 (low-level) words ($M = 2.24$) were more dispersed than high-frequency tone 6 words ($M = 1.53$) from the tonal centroid. The direction of the dispersion was relative to the position of the centroid as shown in Fig. 5: (mid-level) tone 3 moved up and away from the center, (low-falling) tone 4 and (low-level) tone 6 moved down and away from the center. As a result, the relative tonal distance among tones 3, 4, and 6 increased in low-frequency words, as can be seen in Fig. 7. Therefore, the change in the degree of the dispersion from the tonal centroid seems to focus on the mid–low range of the tone space and varies by tone type.

Fig. 8 shows the tone-space dispersion for each speaker when producing high-frequency and low-frequency words. This suggests that all individual speakers showed some degree of dispersion.

Though all speakers displayed tone-space dispersion, gender had a significant effect on the degree of the dispersion ($F_{\text{gender}}(1,5) = 6.129$, $p < .056$). Females tended

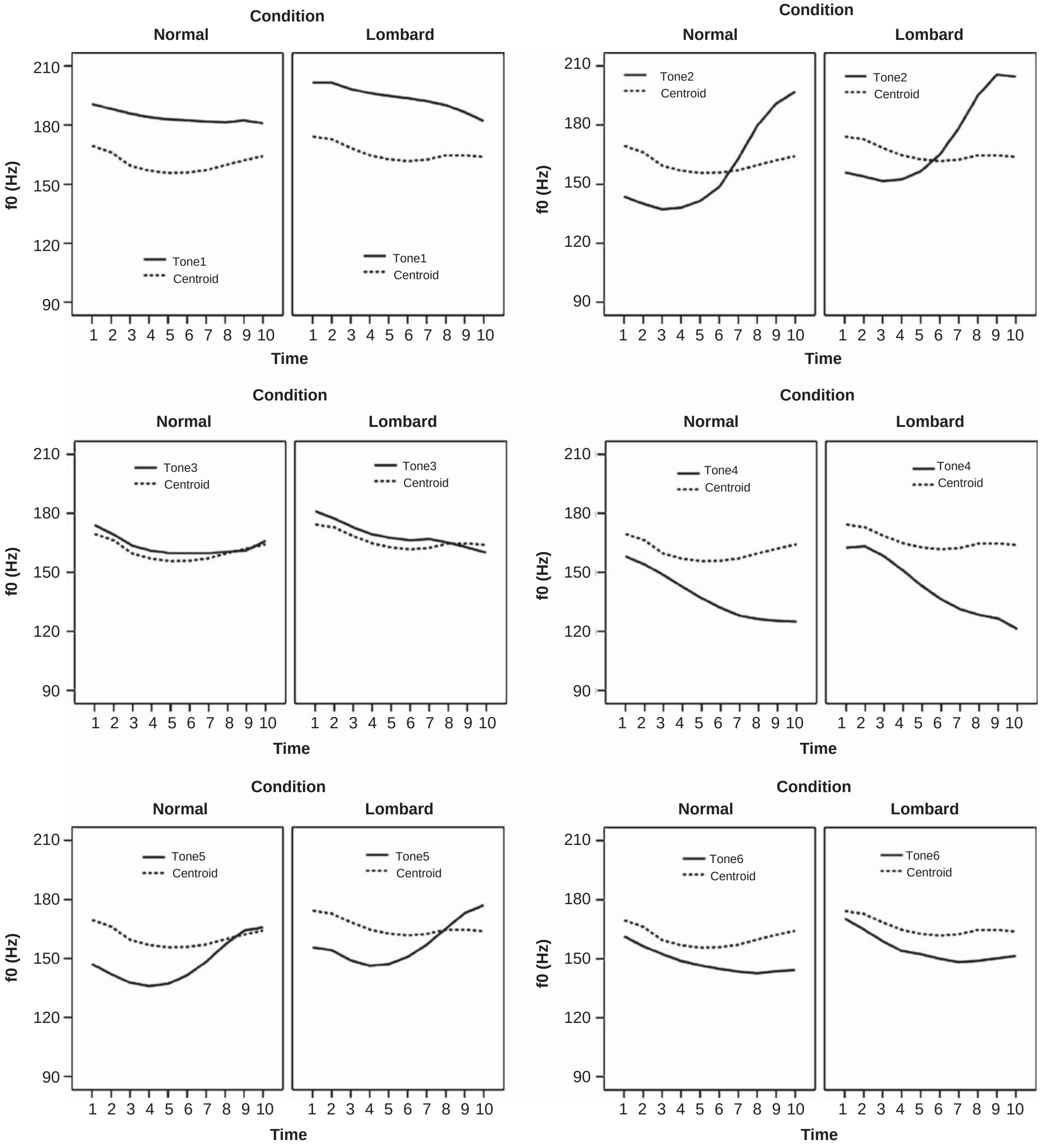

Fig. 4. *f*0 trajectories of six tones relative to tonal centroids in normal and Lombard conditions averaged across speakers.

to use a more dispersed tone space than male speaker for both high-frequency and low-frequency words across speaking conditions ($MD = .411$ st), as shown in Fig. 9.

4.3.2. An alternative measure of dispersion for dynamic tones

The previous section showed that only three tones showed evidence for dispersion: (mid-level) tone 3, (low-falling) tone 4, and (low-level) tone 6. No dispersion was found for the two rising tones, (high-rising) tone 2 and (low-rising) tone 5. We considered the possibility that dispersion in dynamic tones like these rising tones might be realized through an increase in slope, i.e., by lowering the *f*0 min and raising the *f*0 max. This kind of dispersion might not have been captured by our previous metric, since stretching the endpoints might not change the average distance to the tonal centroid. In order to rule out this possibility, we calculated the *f*0 range (the *f*0 differences between highest *f*0 point and lowest *f*0 point along the tone trajectory) for (high-rising) tone 2, and (low-rising) 5, as well as (low-falling) tone 4, to see whether any dispersion appears by this metric.

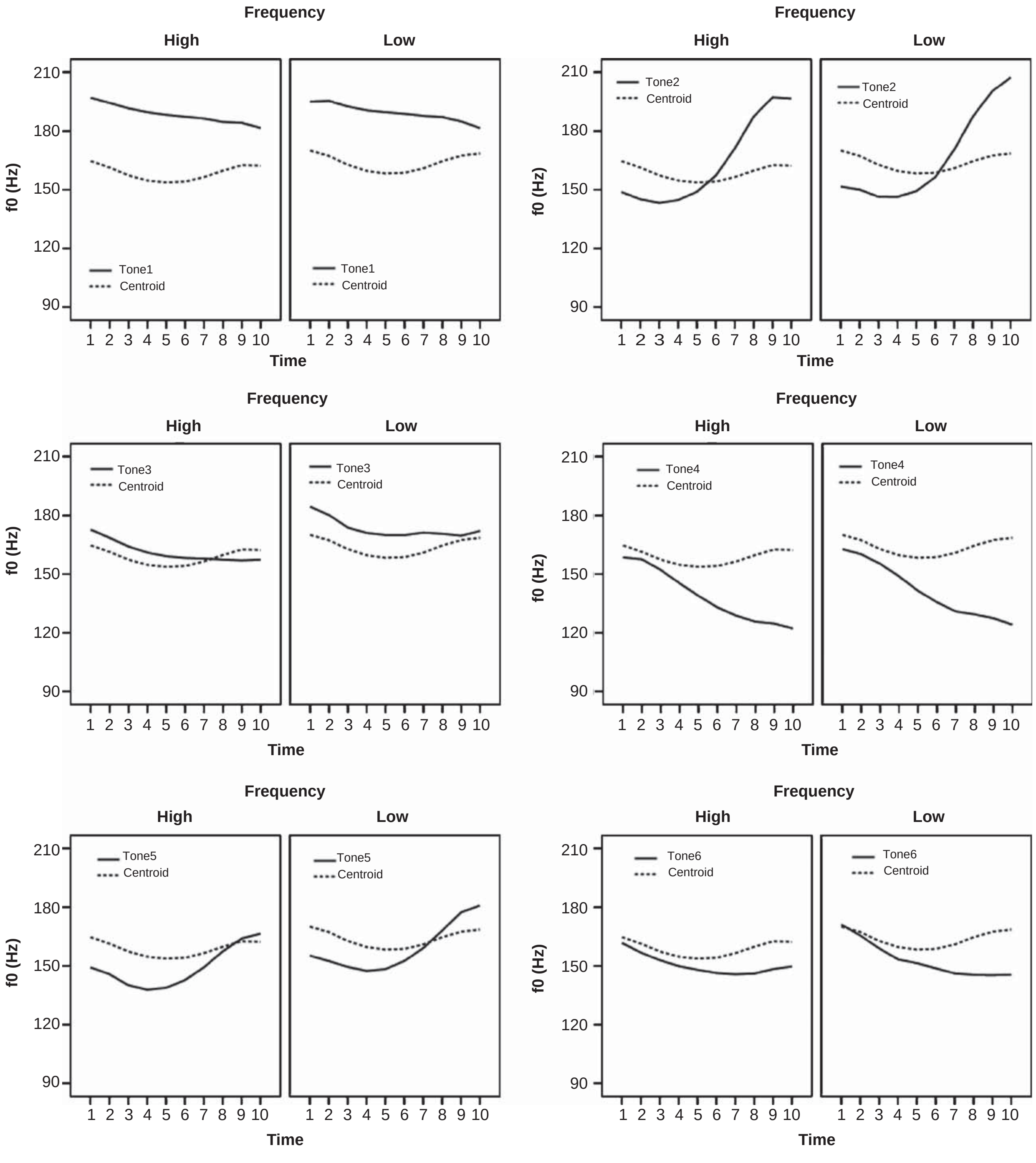

Fig. 5. *f*0 trajectories of six tones relative to tonal centroids for high-frequency and low-frequency words averaged across speakers.

A Repeated Measures ANOVA was used to check whether speech style (normal vs. Lombard), lexical frequency category of the word (high vs. low), and tone type had a significant influence on the dispersion measured by the *f*0 range. The result showed no significant main effects for either speech style ($p = .615$), nor usage frequency ($p = .617$) or interaction between those factors. Tonal dispersion therefore did not target the ‘‘range’’ of the dynamic tones.

4.4. *Interim discussion*

The results on tone-space dispersion suggest that when producing low-frequency words, speakers do tend to use a somewhat adjusted tone space. Dispersion indices show that the reconfigured tone space is more dispersed in the mid–low range when low-frequency words are produced.

On the other hand, noise does not seem to have the same effect on the speakers’ tone space. In Lombard speech, *f*0 is

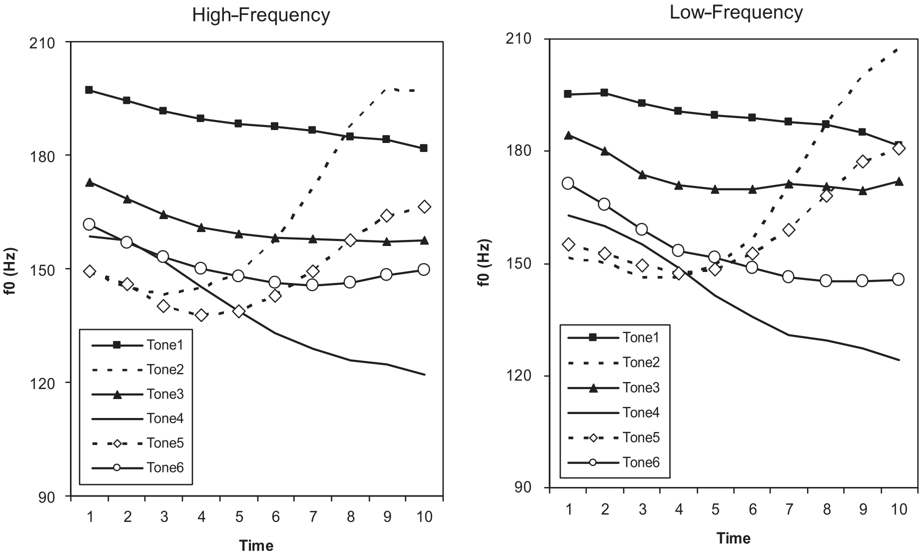

Fig. 6. *f*0 contours of all six tones for high-frequency words and low-frequency words

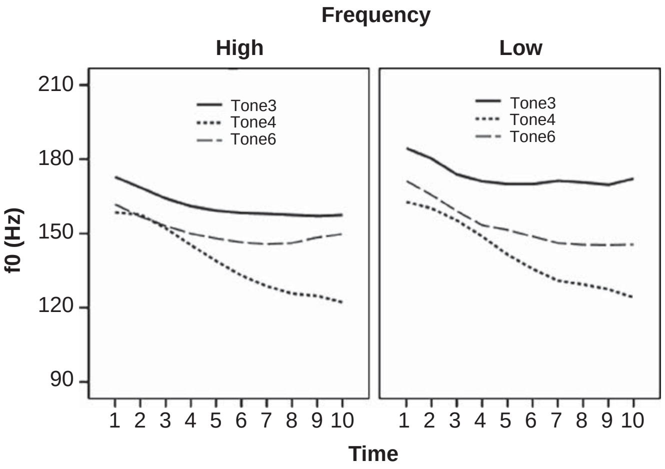

Fig. 7. *f*0 contours of mid–low tones for high-frequency words and low-frequency words

raised universally for all six tones, but this rise does not cause a reconfiguration of the space. Of course the fact that the tone space is more dispersed for low-frequency words does not prove that dispersion is the causal factor in the change of a tone's phonetic realization in low-frequency words. We have merely shown that dispersion is a plausible candidate.

We also found that female speakers make use of a larger acoustic tone space. This is consistent with previous findings on females using a larger pitch range and vowel space than male speakers (Diehl et al., 1996).

5. General discussion

5.1. Result summary

Both lexical frequency and ambient noise influence the phonetic realizations of lexical tones in Cantonese. Ambient noise leads to higher *f*0 for all tone types, compared to the same tone produced without noise. We also found an effect of usage frequency: some (mid-ranged) tones have higher *f*0 in low-frequency words than in high-frequency words. Lexical frequency and Lombard effect therefore join the wide range of other factors that are known to influence tonal variation, such as the assimilatory or dissimilatory contextual effects of neighboring tones (Xu, 1997, 2001). Since in our study tones were elicited in isolation, those effects are above and beyond any such contextual effects.

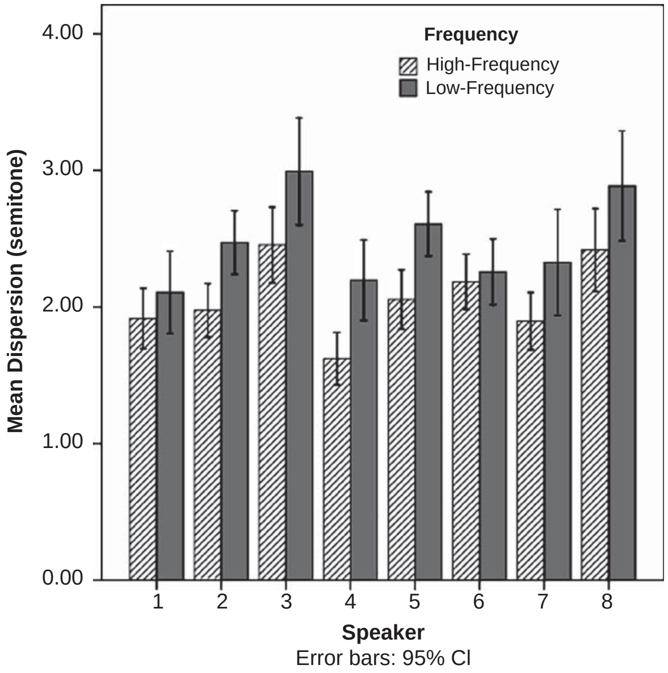

Fig. 8. Mean tone-space dispersion of individual speakers

In addition, we found that while the Lombard reflex affects all tones equally, usage frequency did not. Two mid–low tones (mid-level tone 3 and low-rising tone 5)

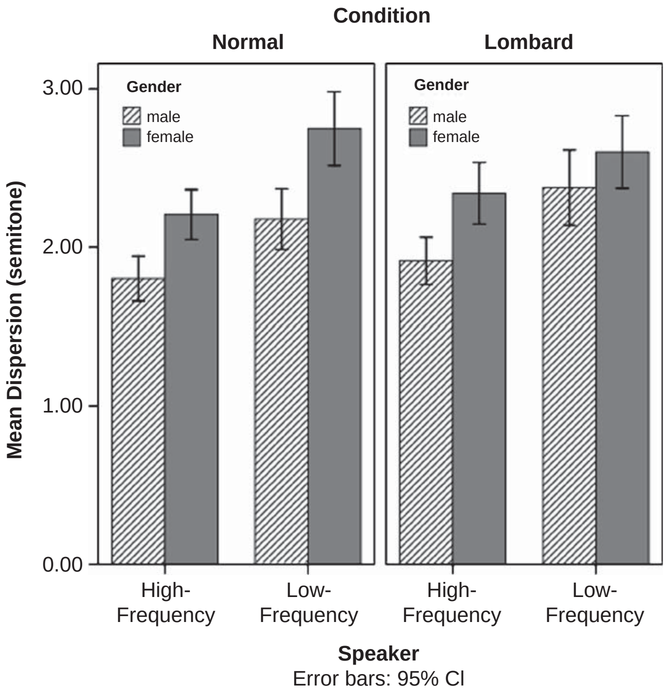

Fig. 9. Mean tone-space dispersion of female and male speakers in normal and Lombard conditions

were significantly affected by usage frequency, while others were not.

One possible explanation for this difference lies in two distinct functional roles of *f*0; to increase *phonetic strength* and to increase *phonemic distinction* or *contrast*. To operationalize this idea of distinction, we introduced a new functional factor called **tone-space dispersion**. We proposed that a speaker's tonal productions lie in a dynamic *tone space*, and that each production has a particular **tonal dispersion** to the tonal centroid. We showed that low-frequency and high-frequency words behaved differently with respect to tone space: low-frequency words were produced with a generally more dispersed tone space than high-frequency words. In particular, mid–low tones (mid-level tone 3, low-falling tone 4, low-level tone 6) were more dispersed from the tonal centroid than their high-frequency counterparts. The fact that the Lombard effect did not result in an increase in tone-space dispersion, instead raising all six tones across-the-board, suggests that for noise, increasing the phonetic strength of individual tones may be a more relevant functional factor than increasing the phonemic contrast among the tones.

5.2. Discussion

This study suggests that different sources of tonal hyperarticulation lead to different consequences for phonetic forms. Hyperarticulation due to ambient noise is phonetic in nature, causing universal raising of *f*0 in all tones, while hyperarticulation due to lexical factors such as usage frequency targets phonemic contrasts and only affects a subset of the tone inventory. For low-frequency words, two tones are higher (mid-level tone 3 and mid-rising tone 5), which moves up the tonal centroid of the tone space, causing mid–low tones such as low-frequency mid-level tone 3, low-falling tone 4 and low-level tone 6 to be more dispersed from the center of the tone space.

Why did dispersion only affect these three tones and not all six? Partial influences are not unheard of: a partial influence of lexical frequency has also been found for vowel production by Wright (2003), who found that frequency or neighborhood density did not affect every vowel. In his study it was the point vowels that were affected by frequency, stretched out from the vowel space center as a way of maximizing perceptual distinctiveness within the vowel space. While it is thus plausible that only part of the tonemic inventory in Cantonese would be subject to lexical influence, it is important to understand why the effect is limited to the mid-tones in particular. In the next few paragraphs, we sketch a potential explanation; deciding if this is indeed a causal factor would of course require further investigation.

One potential functional explanation from the perceptual perspective is that the Cantonese mid–low tones are more perceptually confusable with each other than are other Cantonese tones. We know that, in natural production, the variations of the mid–low level tones overlap greatly with each other. Previous work on Cantonese tone production suggests that in fact the average *f*0 of (mid-level) tone 3 does not significantly differ from the average *f*0 of (low-level) tone 6 (Whitehill, Ciocca, & Chow, 2000). The proximity of mid–low tones in the production *f*0 space results in a high confusion rate in perception. Vance (1977) reported that all stimuli with low average *f*0 (99–120 Hz) were perceived as tone 4. Khouw and Ciocca (2007) showed through a perceptual study that subjects had great difficulty in telling apart the mid–low tones. For instance, (mid-level) tone 3 is often misperceived as (low-level) tone 6, and (low-falling) tone 4 as either tone 3 or tone 6. Our results show that lexical frequency affects the configuration of those tones of high confusability in perception. Increasing dispersion in the mid–low tones may be a way of alleviating this perceptual confusability by increasing phonemic contrasts. By raising tones 3 and 5, the tonal centroid is raised, and tones 3, 4, and 6 become more dispersed and further apart from one another. The fact that tone 5 rises but does not itself become more dispersed may be because it itself is not confusable with the rest of the mid–low tones (tones 3, 4, 6) due to its unique rising contour.

Under this hypothesis, high tones do not expand because they are already less confusable. Some sort of ceiling effect may also be playing a role. Again, this possibility that the perceptual confusability of the mid–low tones is the causal factor is only a hypothesis, and understanding whether it is really implicated in the effect of frequency on mid-range tones needs to be empirically verified.

Finally, one limitation of our study is the possible confound of function words with high-frequency words.

Due to a controversy in the classification of function words in Chinese (e.g., Ma, 2004), it is hard to strictly determine whether a word is a function word or not. But according to some classifications (e.g. Zhang, 2003), a few words in our high-frequency stimuli can be considered "closed class", such as *dou1* and *bei2*. Since these words are spread out across all tone categories, we think their presence is unlikely to have been the causal factor in the rise we found in the mid-tones. In any case, since high frequency is so strongly correlated with function word status cross-linguistically, it remains for future work to explore these two factors factorially (see Bell et al., 2009, for such a study in English).

5.3. *Implications for models of speech production*

The fact that we see hyperarticulation in the production of lexical tone, whether related to lexical frequency or to Lombard speech, needs to be accounted for by any model of lexical production.

One model of speech production that is generally consistent with our results is the Hyperarticulation and Hypoarticulation (H&H) model (Lindblom, 1990). The H&H model assumes that speakers simultaneously balance the needs of the speaker to minimize production effort and the needs of the listener for clear articulation. Words that might be harder for the hearer to hear, because of noise, or to predict, because they are rarer, less contextually predictable, or more confusable, result in more cues for the hearer, such as longer durations, expanded vowel space, more canonical pronunciations, and so on. More recently, some research has offered further details of its functional motivations. For instance, Van Son, Koopmans-van Beinum, and Pols (1998), Jurafsky et al. (2001), Van Son and Pols (2003), Aylett (2000), and Aylett and Turk (2004, 2006) have proposed that hyperarticulation is related specifically to the information value (entropy) or probability of an utterance; more probable or less informative parts of an utterance were somehow phonetically reduced. Pan and Hirschberg (2000) showed that this effect of information value also played its role at the prosodic level; words that were unpredictable or had lower frequency were more likely to be pitch accented by the speaker.

Our results could be interpreted under these H&H-style models, which would indeed predict hyperarticulation for low-frequency words or words produced in noise. However, truly testing an H&H-style explanation would require further experiments to see whether hyperarticulated tones result in better intelligibility for the hearer. In addition, the H&H models do not currently give a way to explain the *strengthening* function of noise (raising *f*0 for all 6 tones), a kind of 'phonetic hyperarticulation', and the *contrast* function of usage frequency (increasing dispersion for mid–low tones), a kind of 'phonemic hyperarticulation'.

Other theorists have proposed possible production mechanisms by which hyper- or hypoarticulation may be implemented. With regards to the execution of prosody in particular, Kochanski and Shih (2003) proposed a model of prosodic strength to accommodate various functional needs in the implementation of prosody. Aylett (2000) and Aylett and Turk (2004) also proposed that predictability-based adjustments in articulation are mediated by prosodic prominence structure. The information value of parts of an utterance is used to determine prosodic prominence, which then determines syllable and segment duration. The models of Aylett (2000), Aylett and Turk (2004), and Kochanski and Shih (2003) all suggest that the prosodic realization of low-frequency words or tones produced in degraded environments would be phonetically strengthened due to their low *a priori* predictability, high information load or the need of enhanced intelligibility in noisy transmission channels. Each of these models is compatible with our results that low-frequency words and words produced under noise are strengthened in tone. However, in order to account for the differences between the two types of hyperarticulation and especially the change in phonemic contrasts, further specifications of the models are needed.

A final possible model, due to Pierrehumbert (2002) and Myers (pers. commun.) among others, is that high-frequency words are stored in a lenited form in their lexical representation. An extension of Pierrehumbert's model to tone, assuming that lenition in tone implies lower *f*0, would suggest that high-frequency words are stored with lower *f*0, and that our result would be modeled as an effect of hypoarticulation rather than hyperarticulation. The idea that exemplars of words are stored with their prosodic attributes has proved to be a successful model in engineering applications like speech synthesis (Nenkova et al., 2007; Strom et al., 2007). It is less clear how an exemplar model like Pierrehumbert's would predict the differences we see in the effect of usage frequency on tone, or the adjustment of tone space, but perhaps the model could be extended with some notion of phonemic space.

In summary, whether hyperarticulation in tone production is due to prosodic strengthening, due to perceptual factors, due to their influence on long-term representation of exemplars, or due to some combination of these remains an important open question. Our study nevertheless shows important basic results on lexical production in tone languages that these models must account for, and offers suggestive evidence that extends previous work showing speakers' manipulation of production space to accommodate functional needs. To continue this line of research, future studies will need to investigate further lexical factors, such as phonological neighborhood density, to see if neighborhood effects influence tone production and whether they are sensitive to the phonemic contrasts of lexical tones. In addition, an important goal is the development of a unified production model that can account for strength-oriented "phonetic hyperarticulation" effects due to Lombard reflex as well as the contrast-oriented "phonemic hyperarticulation" effects due to usage frequency.

Acknowledgement

The work was partially supported by the Edinburgh-Stanford Link and by the NSF via award NSF IIS-0624345. We would also like to thank the editor Alice Turk and three anonymous reviewers who made suggestions leading to significant improvements of the work.

Appendix

Tone	High- frequency		Low- frequency	
Tone 1	Orthography	Romanization	Orthography	Romanization
	中	zung1	髒	zong1
	加	gaa1	疤	baa1
	高	gou1	篝	gau1
	都	dou1	羔	gou1
	將	zeong1	錚	zaang1
	機	gei1	浇	giu1
	多	do1	菠	bo1
	因	jan1	瘟	wan1
	教	gaau1	苞	baau1
	公	gung1	鬆	zung1
Tone 2				
	九	gau2	妞	nau2
	等	dang2	耿	gang2
	本	bun2	惋	wun2
	比	bei2	杞	gei2
	品	ban2	癮	jan2
	展	zin2	匾	bin2
	果	gwo2	跺	do2
Tone 3				
	到	dou3	灶	zou3
	最	zeoi3	踞	geoi3
	意	ji3	亞	aa3
	進	zeon3	鑽	zyun3
	正	zeng3	柄	beng3
	建	gin3	濺	zin3
	計	gei3	庇	bei3
Tone 4				
	來	lai4	桅	wai4
	年	nin4	眠	min4
	能	nang4	萌	mang4
	然	jin4	憐	lin4
	民	man4	綿	min4
	員	wan4	仁	jan4
	明	ming4	擰	ning4
	名	meng4	淩	ling4
	如	jyu4	壺	wu4
Tone 5				
	有	yau5	舀	jiu5
	老	lou5	侮	mou5
	兩	loeng5	痒	joeng5
	眼	ngaan5	挽	waan5
	滿	mun5	缈	miu5
Tone 6				
	第	dai6	陛	bai6
	電	din6	辨	bin6
	問	man6	燉	dan6
	樣	joeng6	晾	long6
	位	wai6	荔	lai6
	大	daai6	寨	zaai6
	動	dung6	頌	zung6
	道	dou6	哺	bou6

References

Al-Nassir, A. (1993). *Sibawaih the phonologist: A critical study of the phonetic and phonological theory of Sibawaih in his treatise "Al-Kitab"*. London: Kegan Paul.

Aylett, M. (2000). *Stochastic suprasegmentals: Relationships between redundancy, prosodic structure and care of articulation in spontaneous speech*. Ph.D. thesis, University of Edinburgh.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence and duration in spontaneous speech. *Language and Speech*, *47*(1), 31–56.

Aylett, M., & Turk, A. (2006). Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. *Journal of the Acoustical Society of America*, *119*(5), 3048–3059.

Academia Sinica: Balanced Corpus of Chinese. 〈http://godel.iis.sinica.edu.tw/CKIP/engversion/20corpus.htm〉.

Bell, A., Brenier, J., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111.

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America*, *113*(2), 1001–1024.

Boersma, P., & Weenink, D. (2000). *Praat: A system for doing phonetics by computer (Version 4.0.30). [Computer software]*. Amsterdam: Institute of Phonetic Sciences. Retrieved from 〈http://www.praat.org〉.

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*, 255–272.

Bybee, J. (2000). Lexicalization of sound change and alternating environments. In M. Broe, & J. Pierrehumbert (Eds.), *Papers in laboratory phonology V: Acquisition and the lexicon* (pp. 250–268). Cambridge: Cambridge University Press.

Bybee, J. (1996). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow, & S. Kemmer (Eds.), *Usage-based models of language* (pp. 65–86). Stanford: CSLI.

Caramazza, A., Costa, A., Miozzo, M., & Bi, Y. (2001). The specific-word frequency effect: Implications for the representation of homophones in speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *2*(6), 1430–1450.

Carter, M. G. (2004). *Sibawayhi*. London: I.B. Tauris.

Cheung, K. H. (1986). *The phonology of present day Cantonese*. Unpublished doctoral thesis, University of London.

Diehl, R., Lindblom, B., Hoemeke, K., & Fahey, R. (1996). On explaining certain male–female differences in the phonetic realization of vowel categories. *Journal of Phonetics*, *24*, 187–208.

Draegert, G. L. (1951). Relationships between voice variables and speech intelligibility in high level noise. *Speech Monograph*, *18*, 272–278.

Fidelhotz, J. L. (1975). Word frequency and vowel reduction in English. In *CLS 11* (pp. 200–213).

Fosler-Lussier, E., & Morgan, Nelson (1998). Effects of speaking rate and word frequency on conversational pronunciation. In *Proceedings of ESCA tutorial and research workshop on modeling pronunciation for variation for automatic speech recognition*, Kerkrade, Netherlands (pp. 1–6).

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. In *CLS 99* (pp. 151–166).

Hanley, T. D., & Steer, M. D. (1949). Effect of level of distracting noise upon speaking rate, duration and intensity. *Journal of Speech and Hearing Disorders*, *14*, 363–368.

Henton, C. (1995). Cross-language variation in the vowels of female and male speakers. In *Proceedings of the XIIIth ICPhS*, Stockholm (pp. 420–423).

Hombert, J. M., Ohala, J. J., & Ewan, W. G. (1979). Phonetic explanations for the development of tones. *Language*, *55*, 37–58.

Hooper, J. B. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In Christie, W. (Ed.), *Current progress in historical linguistics*, Amsterdam (pp. 96–105).

Hsu, H., & Huang, C. (1995). Design criteria for a balanced modern Chinese corpus. In *Proceedings of ICCPOL'95*, Hawaii.

Huang, C., & Chen, K. (1992). A Chinese corpus for linguistics research. In *Proceedings of the 1992 international conference on computational linguistics (COLING-92)*, Nantes, France (pp. 1214–1217).

Huang, S. L. (1997). A Chinese talking syllabary of the Cantonese dialect. Available online at: 〈http://humanum.arts.cuhk.edu.hk/Lexis/Canton2/〉.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *20*, 824–843.

Junqua, J. C. (1993). The Lombard reflex and its role on human listeners and automatic speech recognizer. *Journal of the Acoustical Society of America*, *93*, 510–524.

Jurafsky, D., Bell, A., Gregory, M., & Raymond, W. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. Bybee, & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins.

Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In Rens Bod, Jennifer Hay, & Stefanie Jannedy (Eds.), *Probabilistic linguistics* (pp. 39–96). Cambridge, MA: MIT press.

Kochanski, G., & Shih, C. (2003). Prosody modeling with soft templates. *Speech Communication*, *39*(3–4), 311–352.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, *35*, 104–117.

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

LDC. (2000). *Hong Kong news parallel text*. Linguistic Data Consortium.

Letowski, T., Frank, T., & Caravella, J. (1993). Acoustical properties of speech produced in noise through supra-aural headphone. *Ear and Hearing*, *14*, 332–338.

Liberman, P. (1963). Some effects of the semantic and grammatical context on the production and perception of speech. *Language and Speech*, *6*, 172–175.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle, & A. Marchal (Eds.), *Speech production and speech modeling* (pp. 403–439). Netherlands: Kluwer Academic Publishers.

Lombard, E. (1911). Le signe de l'elevation de la voix. *Annales de Maladies de L'oreille et du Larynx*, *37*, 101–119.

Lu, Y., & Cooke, M. (2008). Lombard speech: Effects of task and noise type. *Journal of the Acoustical Society of America*, *123*(5), 3072.

Ma, Z. (2004). *Xiandai Hanyu Xuci Yanjiu Fangfalun (Research methodology for the study of functional words in modern Chinese)*. Beijing: Commercial Publishing House.

Munson, B. (2007). Lexical access, lexical representation and vowel production. In *Papers in Laboratory Phonology* 10.

Munson, B., & Solomon, N. P. (2004). The effects of phonological neighborhood density on vowel articulation. *Journal of Speech, Language, and Hearing Research*, *47*, 1048–1058.

Nenkova, A., Brenier, J., Kothari, A., Calhoun, S., Whitton, L., Beaver, D., & Jurafsky, D. (2007). To memorize or to predict: prominence labeling in conversational speech. In *Proceedings of NAACL-HLT*.

Norris, D. (2006). The Bayesian reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review*, *113*(2), 327–357.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, *115*(2), 357–395.

Pan, S., & Hirschberg, J. (2000). Modeling local context for pitch accent prediction. In *Proceedings of the ACL*.

Patel, R., & Schell, K. (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech, Language and Hearing Research*, *51*, 209–220.

Patterson, D., & Connine, C. (2001). Variant frequency in flap production. *Phonetica*, *58*, 254–275.

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 and Hearing Research*, *29*, 434–446.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology, Vol. 7* (pp. 87–99). Berlin/New York: Mouton de Gruyter.

Pluymaekers, M., Ernestus, M., & Baayen, R. H. (2005). Articulatory planning is continuous and sensitive to informational redundancy. *Phonetica*, *62*, 146–159.

Raymond, W. D., Dautricourt, R., & Hume, E. (2006). Word-internal /t,d/ deletion in spontaneous speech: Modeling the effects of extra-linguistic, lexical, and phonological factors. *Language Variation and Change*, *18*, 55–97.

Rhodes, R. A. (1996). English reduced vowels and the nature of natural processes. In B. Hurch, & R. A. Rhodes (Eds.), *Natural phonology: The state of the art* (pp. 239–259). The Hague: Mouton de Gruyter.

Rhodes, R.A. (1992). Flapping in American English. In W. U. Dressler, M. Prinzhorn, & J. Rennison (Eds.), *Proceedings of the seventh international phonology meeting* (pp. 217–232).

Rostolland, D. (1982). Acoustic features of shouted voice. *Acoustica*, *50*, 118–125.

Scarborough, R. (2004). *Coarticulation and the structure of the lexicon*. Unpublished doctoral dissertation, UCLA.

Schuchardt, H. (1885). Uber die Lautgesetze: Gegen die Junggrammatiker. Excerpted with English translation. Reprinted in T. Vennemann, & T. H. Wilbur, (Eds.), 1972. *Schuchardt: The Neogrammarians, and the transformational theory of phonological change*. Frankfurt: Athenaum Verlag.

Strom, V., Nenkova, A., Clark, R., Vazquez-Alvarez, Y., Brenier, J., King, S., & Jurafsky, D. (2007). Modelling prominence and emphasis improves unit-selection synthesis. In *Proceedings of Interspeech*.

Summers, W. V., Pisoni, D., Bernacki, R., Pedlow, R., & Stokes, M. (1988). Effects of noise on speech production: Acoustic and perceptual analysis. *Journal of the Acoustical Society of America*, *84*(3), 917–928.

Van Son, R. J. J. H., & Pols, L. C. W. (2003). How efficient is speech? *Proceedings of the Institute of Phonetic Sciences*, *25*, 171–184.

Van Son, R. J. J. H., Koopmans-van Beinum, F. J. & Pols, L. C. W. (1998). Efficiency as an organizing principle of natural speech. In *Proceedings of ICSLP 98*, Sydney.

Vance, T. J. (1977). Tonal distinction in Cantonese. *Phonetica*, *34*, 93–107.

Wright, R. (2003). Factors of lexical competition in vowel articulation. In R. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology VI* (pp. 75–87). Cambridge: Cambridge University Press.

Whitehill, T. L., Ciocca, V., & Chow, T. Y. (2000). Acoustic analysis of lexical tone contrast in dysarthria. *Journal of Medical Speech and Language Pathology*, *8*, 337–344.

Xu, Y. (2001). Sources of tonal variations in connected speech. *Journal of Chinese Linguistics, Monograph Series*, *17*, 1–31.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61–83.

Xu, Y. (2006). Time-normalized *f*0 praat script. Available at 〈http://www.phon.ucl.ac.uk/home/yi/_TimeNormalizeF0.praat〉.

Zhang, B. (2003). *Xiandai Hanyu Xuci Cidian (Dictionary of functional words in modern Chinese)*. Beijing: Commercial Publishing House.

Zipf, G. (1929). Relative frequency as a determinant of phonetic change. *Harvard Studies in Classical Philology*, *15*, 1–95.