

Speech Input to Infants: The Acoustic-phonetic Characteristics of Infant-directed Speech in Mandarin Chinese

Huei-Mei Liu[†], Feng-Ming Tsao[‡] and Patricia K. Kuhl[‡]

[†] Dept. of Special Education, National Taiwan Normal University, Taiwan

[‡] Center for Mind, Brain, and Learning, University of Washington, USA

E-mail: lium1m2@yahoo.com.tw

ABSTRACT

This study investigated acoustic-phonetic characteristics of infant-directed speech to examine the hypothesis that infant-directed speech provides exceptionally clear, exaggerated cues to language. Vowels and lexical tones in Mandarin Chinese were examined. Sixteen mothers were audio-recorded while speaking to their 6-8 month-old infants and to another adult. The results showed that the spectral features of /a/, /i/, and /u/ were altered to stretch the acoustic distances between them and resulted in a greater acoustic area encompassing the articulation space. For lexical tones, the results demonstrated that when talking with infants, Mandarin mothers raised the pitch level, exaggerated pitch contours, and produced a more complete contour at syllable level. Moreover, the critical temporal feature that perceptually distinguishes lexical tones, the relative timing of F0 turning point, was held constant across both speech conditions. Together, the results support the claim that infant-directed speech provides exceptionally well-specified acoustic-phonetic information, which could promote language learning.

1. INTRODUCTION

It has been well demonstrated that infant-directed speech has a unique acoustic signature, one characterized by a higher pitch, exaggerated intonation contours, and a slower tempo [1,2]. In addition, it is also suggested that adults modify their speech in such a way as to amplify important phonetic characteristics when addressing their infants compared with talking to adults [3,4]. However, unlike the prosodic modifications demonstrated near-universally across various languages, phonetic segmental modification has been only demonstrated in a limited number of languages (e.g., English, Russian, Swedish, and Australian English). The present study extended existing studies of acoustic-phonetic modifications in ID speech: 1) to a tonal language (i.e., Mandarin Chinese) that is phonetically different from non-tonal languages, such as English; and 2) to additional aspects of acoustic-phonetic characteristics at both segmental and suprasegmental levels, including vowels and lexical tones to investigate whether Mandarin-speaking mothers modified these important phonetic units and provide exceptionally clear and exaggerated cues when addressing infants compared with

talking to another adult.

To examine segmental modifications of ID speech, the “corner” vowels, /i/, /a/, and /u/, were selected because these vowels are the most common in human languages and they are perceptually and acoustically exceptional [5]. For each vowel, vowel duration, formant frequencies, vowel space area, and within-category vowel formant dispersion were measured instrumentally to indicate the acoustic-phonetic modifications made in ID speech compared to AD speech. For ID speech, lengthened vowel duration [6], less overlapping formant locations of vowels [7] and the expanded vowel articulation space [3] has been reported when compared to AD speech across several non-tonal languages. In addition to the expanded vowel space area indicating greater acoustic distance between vowel categories in ID speech, the dispersion of within-category vowel formant locations is also of interest. Greater within-category vowel formant dispersions in ID speech have been observed, but no quantitative data were reported in previous studies [3,8]. The hypothesis is that the acoustical exaggerations of vowels, e.g., longer vowel duration and expanded vowel space, revealed in previous non-tonal language studies, would generalize to Mandarin Chinese. Furthermore, the quantitative measure of within-category variations would provide more detailed information of the vowel modifications made in ID speech.

A suprasegmental dimension in ID speech, lexical tone, the pitch fluctuation patterns of which convey meanings in Mandarin Chinese, was also examined. As well defined, in Mandarin Chinese, four different types of pitch contours manifest four lexical tones, and these tones serve to differentiate the meanings of words. Pitch height and pitch contour excursion at syllable level are the crucial acoustic components for perceiving the lexical tones [9,10].

To date, it has been noted that Mandarin-speaking mothers exhibit the prosodic modification at sentence and phrase levels as demonstrated in other nontonal languages [6], but there are no available studies showing whether or how Mandarin-speaking mothers modify the crucial acoustic characteristics of lexical tones in ID speech to convey meaning. To further understand how adults modify the important suprasegmental characteristics in speech directed to infants in a tonal language, the microanalyses of pitch contours on a syllable-by-syllable basis in Mandarin Chinese is essential.

The hypothesis was that Mandarin mothers might modify elements of pitch contours, such as raising their average pitch level and exaggerating the range of pitch contours, to enhance the lexical tone distinctiveness in speech directed to infants. However, in order to convey the correct lexical meaning, mothers would have to preserve the appropriate and similar pitch contour of each lexical tone, such as the appropriate timing of “turning point”, in both speech styles. The results of this study filled a gap in the literature by documenting and providing a more comprehensive picture of how mothers in a tonal language modify important phonetic elements in speech directed to infants.

2. METHODS

A. Participants Sixteen mother-infant dyads with infants aged 6-8 months (mean = 7.3 m; 8 boys, 8 girls) participated in this study. Infants were recruited from the infant list of the House Registry Office of a metropolitan area in Taiwan. Inclusion criteria for mothers participating in the speech recording were: 1) Mandarin-Chinese was the mothers’ only or the dominant language, 2) the mother was the primary caregiver for the infant, and 3) the mother did not have any known physical, sensory, or mental handicap.

B. Stimuli The target words selected for speech recording were 21 Mandarin Chinese bisyllabic words (3 vowels × 7 tokens) containing the target vowels, /i/, /a/, and /u/, and target tones (Chinese tone 1,2,3,4) in the first syllable. All target words were constructed as (C)VCV(V) and the tone numeral was attached to the end of each syllable, such as pa1ʂə4 (*bus*) and tɕ^h14p^hau4 (*bubble*). The vowel and tone in the first syllable were the targets for acoustic analysis.

C. Recording Procedure Each of 16 mother was audio-recorded when she spoke to a native-Mandarin-speaking adult and to her infant in a sound-attenuated booth. To help mothers use the target words, they were provided with toys or pictures that corresponded to target words in a naturalistic face-to-face conversational situation.

D. Acoustic Analysis Tokens of target words, with the exception of those that overlapped with conversation from the experimenter, infant vocalization, or toy noise, were acoustically analyzed. Acoustic analysis was conducted using KAY Elemetrics’ Computerized Speech Laboratory (CSL) 4300 software. Target words were sampled at 20 kHz, 16-bit resolution rate, and low-pass filtered at 10 kHz. Each vowel’s duration was measured from onset to offset of the segment. The first three formant frequencies (i.e., F1, F2, and F3) were measured at the onset, middle, and offset of each vowel segment. Narrow-band spectrograms, FFT spectra, and autocorrelation LPC spectra, were used to judge the locations of formants. The F1 and F2 of vowels were viewed as Descartes’ coordinates on the x-y plane and the area of vowel space compassing /i/, /a/, and /u/ was equal to the triangular area constructed from the three (F1, F2) pairs of each corner vowel in the x-y plane.

Vowel space area = $ABS \{ [F1i*(F2a - F2u) + F1a*(F2u - F2i) + F1u*(F2i - F2a)]/2 \}$

where “ABS” is absolute value, “F1i” is the F1 value of vowel /i/, “F2a” is the F2 value of vowel /a/,..., and so on.

A quantitative measure of the within-category vowel formant dispersion is of interest to indicate possible vowel articulation differences between ID speech and AD speech. A Principal Component Analysis (PCA) was used to calculate the area of the oval surface that represents the two-dimensional spectral variations of multiple tokens of a vowel. The area of this oval surface for each vowel category can be seen as an index of the within-category variability or dispersion. A greater surface area corresponded to greater variability of vowel formants. Therefore, the average surface areas of three corner vowels was used to compare the within-vowel articulation difference between AD and ID speech conditions.

For each tone contour, the F0 values in the initial, middle, and final syllable positions, along with the highest and lowest (valley) points of the contour, were measured to track the overall pitch height and contour. A number of critical attributes, such as F0 average, F0 range, and the turning point of pitch excursion, were used to characterize the acoustic features of Mandarin lexical tones and compare the difference between speech styles. In addition, the relative timing of the turning point where the slope of the contour changed from falling to rising was equal to the relative temporal position between the onset and the valley of the pitch contour to the whole vowel-segment. This measure was used to examine whether appropriate pitch contour patterns for individual tones were preserved in both ID and AD speech. The inter-rater reliability of overall acoustic measures was high (Cronbach $\alpha = .90$).

3. RESULTS AND DISCUSSION

A. Vowel characteristics in ID vs. AD speech

The vowel duration of ID speech ($M = 238.44$ ms, $SE = 14.06$) is significantly longer than that of AD speech ($M = 181.95$ ms, $SE = 12.09$), $F(1, 15) = 18.54$, $p < .001$. This result is consistent with the previous findings that mothers tend to slow speaking rate when addressing infants [6,11]. Since the slower speaking rate is arguably an important cue for clear speech [12], the vowel lengthening observed in ID speech suggests that ID speech would be clearer and more intelligible.

Figure 1 displays the average vowel space area encompassing the corner vowels /i/, /a/, and /u/ for 16 speakers in the AD and ID speech conditions. The results clearly show that ID speech exhibits significantly larger vowel space area when compared to AD speech, $F(1, 15) = 65.22$, $p < .001$. The ratio of ID vowel triangular area to AD vowel triangular area is about 1.37, which is significantly greater than one, $t(15) = 6.24$, $p < .001$. These results clearly reveal that Mandarin-speaking mothers produced acoustically more extreme vowels when addressing infants compared to addressing with adults, resulting in an expanded vowel space area.

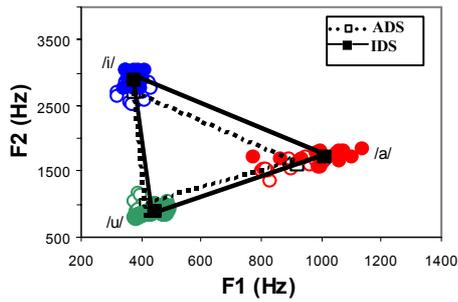


Figure 1. Vowel space area comprising the corner vowels, /i/, /a/, and /u/, in AD (open circles) and ID (solid circles) speech.

The results also show that the within-vowel dispersion of ID speech ($M = 7228.91 \text{ Hz}^2$) is significantly larger than for AD speech ($M = 4814.79 \text{ Hz}^2$), as shown in a one sample t -test, $t(15) = 5.51$, $p < .001$. The pattern of greater dispersion in ID speech has been shown for all subjects ($p < .001$, Binomial test). Figure 2 illustrates an example of the greater within-category vowel dispersion of multiple tokens of vowels /i/, /a/, and /u/ from one speaker's AD and ID speech conditions. This suggests that mothers produce a greater range of formant values and variety of vowel instances when addressing their infants.

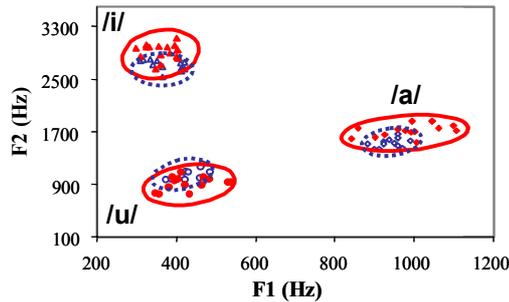


Figure 2. An example of the within-category vowel dispersion of multiple tokens of vowel /i/, /a/, and /u/ of one speaker's AD (solid oval) and ID (dashed oval) speech.

B. Lexical tone characteristics in ID vs. AD speech

The results show that Mandarin-speaking mothers raised their overall pitch level for each lexical tone in infant-directed speech compared to AD speech $F(1, 15) = 56.41$, $p < .001$. The mean F0s among 4 Mandarin lexical tones are distinctive with the order: Tone 1 \approx Tone 4 > Tone 2 > Tone 3 for both AD and ID speech.

The results also show that Mandarin-speaking mothers use greater pitch fluctuation in terms of greater F0 range when addressing their infants than when addressing adults, $F(1, 15) = 39.02$, $p < .001$. This suggests that Mandarin-speaking mothers tend to use exaggerated pitch excursions that emphasize the tone distinctions in speech directed to infants. In addition, the F0-range order among 4 tones is consistent with previous reports on Mandarin tones, showing the greatest F0 range in the high-falling tone (Tone 4) and the smallest F0 range in the high-level tone (Tone 1)

[13].

Figure 3 exemplifies the F0 contours of 4 lexical tone patterns in AD and ID speech. As shown in Figures 3, the shape of the F0 contour of each of the four Mandarin lexical tones is strikingly similar in ID and AD speech conditions, but ID speech exhibits more distinct lexical tone patterns than AD speech. The lexical tone patterns revealed in this study are basically consistent with previous reports on the Mandarin tones: a high-level pitch contour for Tone 1, a mid-rising contour for Tone 2, a low-dipping contour (or mid-falling-rising) for Tone 3, and a high-falling contour for Tone 4 [9,14]. In brief, the results of similar F0 contours in both speech styles suggest well-preserved tone patterns across both AD and ID speech.

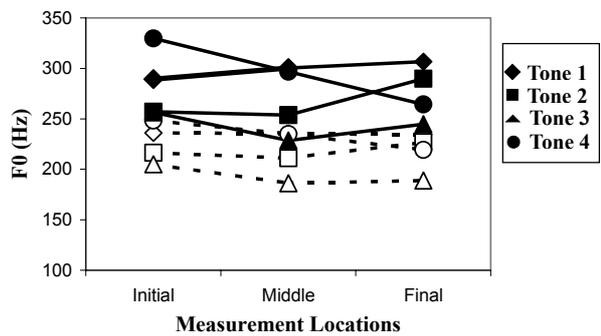


Figure 3. Schematic pitch contours of 4 Mandarin tones in the AD (dashed line) and ID (solid line) speech.

A quantitative measure, the timing of F0 turning point of each lexical tone contour was used to examine the preservation of appropriated pitch contour patterns in both ID and AD speech. The results show that there is no significant difference of F0 turning point between two speech conditions, $F(1, 15) = 4.47$, $p > .05$ (see Figure 4). This suggests that Mandarin-speaking mothers tend to preserve the critical temporal feature of lexical tones when addressing both infants and adults, while the spectral features of lexical tone (i.e., mean F0 and F0 range) are exaggerated in ID speech. Moreover, the result is consistent with the previous findings that the tuning point was critical in differentiating Tones 2 and 3, with the pitch excursion of Tone 2 changing direction earlier than Tone 3 [15].

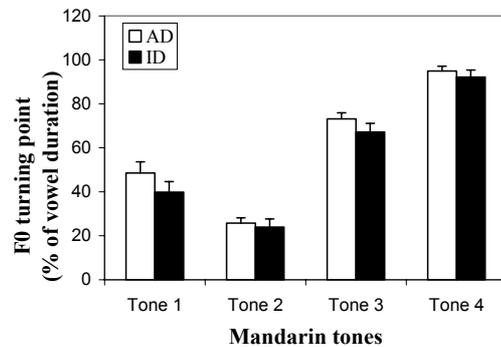


Figure 4. The timing of F0 turning points of 4 Mandarin tones in AD and ID speech.

Taken together of the tone analysis, the results support the hypothesis that critical temporal cue for distinguishing lexical tones, such as the appropriate shape and timing of the pitch contour, is well preserved in both ID and AD speech conditions while mothers tend to raise the overall pitch and exaggerate the pitch contour at syllable level when talking with infants

4. CONCLUSIONS

The results of acoustic analysis of corner vowels clearly demonstrate that Mandarin-speaking mothers produce acoustically more extreme vowels when addressing their infants. Compared to vowel characteristics in AD speech, vowel segments in ID speech are characterized by longer duration, expanded vowel space area, greater inter-vowel distance, and larger within-vowel formant variations. Data from the present study support the claim that the acoustic-phonetic characteristics of important phonetic units (i.e., vowels) are adjusted to increase the distinction between phonetic categories in speech addressed to infants. These modifications may contribute to the perceptual salience of phonetic units when infants perceive speech sounds.

The results of various acoustic analyses of lexical tones clearly show that Mandarin-speaking mothers exaggerate the spectral characteristics of lexical tones but preserve the contour shape and the critical temporal feature to convey the correct lexical meaning when talking with their infants. These findings extend the previous results to further indicate that the prosodic modifications (e.g., higher pitch and exaggerated pitch contours) in Mandarin Chinese ID speech are found not only in sentences and phrases but also in syllables, and that these modifications serve to better specify appropriate lexical tone. These findings show that the acoustic exaggeration of prosodic patterns in ID speech can be generalized to lexical tone in a tonal language. Since lexical tone is a language-specific characteristic of tonal languages, the results of this study support for the hypothesis that mothers selectively enhance speech features that are important to the language. The well-preserved lexical tones in Mandarin ID speech suggests that ID speech may provide exceptionally specific information at a variety of levels to infants, and that this may assist learning.

ACKNOWLEDGMENT

Funding for the research was provided by grants to PKK from NIH (HD37954), the Human Frontiers Science Program (HF159). Additional support was provided by Chin-Hsing Tseng at National Kaohsiung Normal University, Taiwan.

REFERENCES

[1] A. Fernald, T. Simon. Expanded intonation contours in mothers' speech to newborns, *Developmental Psychology*, **20**(1), pp. 104-113, 1984.

[2] D. N. Stern, S. Spieker, R. K. Barnett, K. Mackain. The prosody of maternal speech: infant age and context related changes, *Journal of Child Language*, **10**, 1-15, 1983.

[3] P. K. Kuhl et al. Cross-language analysis of phonetic units in language addressed to infants, *Science*, **277**, 684-686, 1997.

[4] D. Burnham, C. Kitamura, Vollmer-Conna. What's new, pussycat? On talking to babies and animals, *Science*, **296**, 1435-1435, 2002.

[5] K. N. Stevens. *Acoustic phonetics*, Cambridge, MA: MIT Press, 1998.

[6] D. L. Grieser, P. K. Kuhl. Maternal speech to infants in a tonal language: Support for universal prosodic feature in motherese, *Developmental Psychology*, **24**(1), 14-20, 1988.

[7] N. B. Bernstein-Ratner. Phonological rule usage in mother-child speech, *Journal of Phonetics*, **12**, 245-254, 1984.

[8] B. L. Davis, B. Lindblom. "Phonetic variability in baby talk", In F. Lacerda, C. V. Hofsten & M. Heimann (Eds.), *Emerging cognitive abilities in early infancy*. Mahwah, NJ: Lawrence Erlbaum Associates, 2001, 135-171.

[9] C. Y. Tseng. *An acoustic phonetic study on tones in Mandarin Chinese*, Taipei, Taiwan: Institute of History & Philology Academia Science, 1990.

[10] D. W. Massaro, M. M. Cohen, C. Tseng. The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese, *Journal of Chinese Linguistics*, **13**, 267-290, 1985.

[11] C. E. Snow. Mothers' speech to children learning language, *Journal of Child Development*, **43**, 549-565, 1972.

[12] M. A. Picheny, N. I. Durlach, L. D. Braida. Speaking clearly for the hard of hearing III: an attempt to determine the contribution of speaking rate to difference in intelligibility between clear and conversational speech. *Journal of Speech and Hearing Research*, **32**, 600-603, 1989.

[13] J. Leather. Speaker normalization in perception of lexical tone, *Journal of Phonetics*, **11**, 373-382, 1983.

[14] Y. R. Chao. "*Mandarin Primer*". Cambridge: Harvard University Press, 1948.

[15] C. B. Moore, A. Jongman. Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, **102**, 1864-1877, 1997.