Articulatory Prosody: A comparison of Mandarin Chinese and Japanese

Donna Erickson

Haskins Laboratories, U.S.A., Kanazawa Medical University, Japan
ericksondonna2000@gmail.com

Abstract

This paper is a synopsis of a series of research studies previously reported about articulatory prosody, based on jaw displacement data collected from Mandarin Chinese and Japanese speakers using electromagnetic articulography (EMA). On one hand, Mandarin Chinese has tone contrast while Japanese has pitch accents. However, the experimental data suggest that in terms of articulatory prosody, they are similar. The term “articulatory prosody” as discussed here is used to describe prosodically-motivated jaw displacement patterns. Both Mandarin Chinese and Japanese have final and initial phrasal stress, implemented by increased jaw displacement, and the jaw displacement patterns seem to be independent of the F0 contours of the tones/pitch accents. Acoustically, both Mandarin Chinese and Japanese show F1 increases with increased jaw displacement; Mandarin Chinese also shows increased duration. The results suggest that for both Mandarin Chinese and Japanese, it is the pattern of jaw displacement that provides the underlying prosodic framework of phrase-initial and phrase-final stress, resulting in increased F1 for (the low vowel /a/) in both languages, and increased duration for Mandarin Chinese.

Index Terms: Articulatory prosody, jaw displacement, duration, F1, Mandarin Chinese, Japanese

1. Introduction

This paper is a synopsis of a series of research studies previously reported about articulatory prosody, based on jaw displacement data collected from Mandarin Chinese and Japanese speakers using electromagnetic articulography (EMA), see e.g. [1, 2, 3]. Articulatory prosody is used here to refer to how we articulate stress/prominence patterns in a language; specifically, how the jaw works as a prosodic articulator. The jaw sets the framework for the movement of the other articulators (tongue, lips, velum). Increased jaw displacement results in acoustic changes signaling stress/prominence changes, such as duration and formant changes. Signaling of prominence is also implemented by our Laryngeal Articulator by increasing/decreasing the fundamental frequency (F0). A Thoracic Articulator, involved in providing air flowing through the vocal folds, might also be considered.

The viewpoint here is that the Jaw Articulator works in tandem with the Laryngeal Articulator (and the Thoracic Articulator) for bringing about the various acoustic changes, but it is the Jaw Articulator that sets the framework for the rhythmic patterns of stress/prominence. The reasoning is that for each syllable, the jaw must open and close. The working hypothesis is that the amount of jaw opening is prescribed by the phonologically-determined prominence patterns of the language, e.g., [4, 5, 6], which are learned in infancy (see [16] for discussion of importance of visual information when learning a first language.) Indeed, vowel specifications affect jaw displacement, with low vowels having more jaw displacement than high vowels [17]. The claim here, however, is that the jaw-displacement framework is set by the language-specific prominence patterns; once the vowel-specific requirements have been factored out, what remains is the prosodically-determined jaw displacement pattern of an utterance [5, 6, 7, 8]. At this point, however, in order to observe prosodically-determined jaw displacement patterns, utterances containing the same vowel are examined.

This paper examines articulatory prosody, i.e., jaw displacement patterns, in two seemingly disparate languages, Mandarin Chinese and Japanese. Mandarin Chinese is a tonal language, where each syllable is associated with a specific tone, associated with a specific F0 contour, conveying specific lexical information. The classic example is /ma/, which can mean among other things, ‘mother’, ‘horse’, ‘curse’, depending on the tone. Japanese also uses F0 to denote lexical differences; it is referred to as a pitch accent language. In the standard Tokyo dialect, pitch accents are assigned to moras. Moras with pitch accents are characterized by a high F0 followed, by a sudden drop, e.g., [9]). An example of pitch accents is /hashi/ : /hashi/ (with a pitch accent on the final mora) meaning bridge, and /ha’shi/ (with a pitch accent on the first mora) meaning chopsticks.

Thus, both languages use F0 to convey lexical meaning. In addition, both languages are edge strengthening languages, see e.g. [10]. In these languages, prominence is used at the phrase level to mark boundaries, at the end of the phrase/utterance and optionally, at the beginning (see also [18]). This is similar to French, but very different from English, where each syllable has an n-ary amount of prominence, determined by its position in the phrase/intended meaning of the speaker, e.g., [6, 11]. In English, the nuclear stress syllable in the utterance receives the largest amount of prominence; English prominence specifications result in decidedly different jaw displacement patterns, e.g. [5, 6], compared to edge-strengthening languages.

The findings reported here are about jaw displacement patterns seen in Mandarin Chinese and Japanese, e.g., [1, 2, 3]. Specifically, the findings are that (1) speakers of both Mandarin Chinese and Japanese implement phrasal stress by increased jaw displacement, both at the end and optionally at the beginning of each phrase/utterance, (2) the amount of jaw displacement is independent of tone/pitch accent, and (3) the acoustic results are increases in F1 and duration.

2. Methods

EMA data was collected using the 3-D EMA (Carstens AG500 Electromagnetic Articulograph) at Professor Jianwu Dang’s lab, Japan Advanced Institute of Science and Technology. Figure 1
shows the EMA setup and illustrates where the sensors were placed on the tongue and lower incisor; only the sensor on the lower incisor to measure jaw displacement is reported here. Correction for head movement was done from four sensors attached to the upper incisors, bridge of the nose, and left and right mastoid processes behind the ears, respectively. The articulatory and acoustic data were digitized at sampling rates of 200 Hz and 16 kHz, respectively. The occlusal plane was estimated using a biteplate with three additional sensors. In post processing, the articulatory data were rotated to the occlusal plane and corrected for head movement using the reference sensors after low-pass filtering at 20 Hz. The lowest vertical position (maximum displacement) of the jaw with respect to the bite plane was located for each target syllable of the utterance using the MATLAB-based custom software mvview (Haskins Laboratories). Acoustic duration F1 measurements were made using Praat software [12]. Approximately 6 repetitions of each utterance were successfully recorded and analyzed.

Figure 1: EMA setup and sensor placements

Six speakers of Mandarin Chinese and five speakers of standard Tokyo Japanese were recorded. For more details of these experiments, see e.g., [1, 2, 3]. In order to illustrate some common articulatory prosodic features of these two languages, here are reported some highlights from these experiments.

3. Results

3.1. Jaw displacement and phrasal stress

Mandarin Chinese and Japanese implement phrasal stress by increased jaw displacement, at both the end and beginning of the utterance. This is illustrated in Figure 2 (Mandarin Chinese) and Figure 3 (Japanese).

Figure 2: Mandarin Chinese sample jaw opening pattern for 'Mother curses the horse' (ma1 ma ma4 ma3). The top panel is the acoustic wave form, the middle, the spectrogram and the bottom, the jaw displacement pattern. The x-axis is time in ms., the y-axis, jaw position (mm) relative to the occlusal plane. The lower the jaw curve, the greater the amount of jaw displacement.

In figure 2, all syllables contain the same low vowel /a/, but the amount of jaw displacement varies for each of them. The final /ma/ syllable has the largest jaw displacement, and the initial /ma/ syllable has the next largest jaw displacement. This suggests that increased jaw lowering seems to be a method Mandarin Chinese speakers use to mark the ending, as well as the beginning of an utterance. Similar patterns of jaw displacement for this utterance type were found for all six Mandarin Chinese speakers examined.

Figure 3: Japanese sample jaw opening pattern for 'It’s a red umbrella' (aka gasa da). The top panel is the acoustic wave form, the middle, the spectrogram and the bottom, the jaw displacement pattern. The x-axis is time in ms., the y-axis, jaw position (mm) relative to the occlusal plane. The lower the jaw curve, the greater the amount of jaw displacement.

Notice that all the vowels are the low vowel /a/, yet there are different amounts of jaw displacement for each of the syllables, with the largest jaw displacements at the beginning and end of the utterance. Thus, increased jaw lowering seems to be a method Japanese speakers also use to mark the ending, as well as the beginning of an utterance. Similar patterns were found for all five of the Japanese speakers examined.

3.2. Jaw displacement independent of tone/pitch accent

Edge strengthening jaw displacement in both Mandarin Chinese and Japanese is independent of the tone/pitch accent.

Figure 4: Jaw displacement patterns for four Mandarin Chinese speakers (columns) for two utterances (rows), with the height of each bar representing the amount of jaw displacement per syllable. The x-axis represents the syllables, the y-axis, the jaw position (mm) relative to the occlusal plane for each syllable. The jaw measurements were made at the
point in the syllable where the jaw had the lowest vertical position (illustrated in Figures 2 and 3, and explained in Methods.) In order to enhance the patterns of relative jaw displacement per syllable for each speaker, the minimum and maximum y-values are not constant across the graphs. The top row shows jaw displacement patterns for ‘Jia1na2da4’ (Canada) and the bottom row for ‘ba1na2ma3’ (Panama). The circled tones indicate tonal differences between the two utterances. The arrows point to the syllables with the largest jaw displacement, always at the end of the utterance, approximately 4-5 mm larger than the syllable with the smallest amount of jaw displacement (from [1]).

Figure 4 shows jaw displacement patterns in terms of bar graphs for two Mandarin Chinese utterances: ‘Jia1na2da4’ (Canada) and ‘ba1na2ma3’ (Panama) as produced by four speakers. The height of the bar indicates the amount of jaw displacement in mm. Notice that for both utterances, we see similar patterns of marked increase of jaw displacement on the final syllable, yet, the final syllables have different tones. For Canada, the final syllable is tone 3, a low rising tone. Similar patterns of jaw displacement for these utterances were found for the six Mandarin Chinese speakers examined. Thus, phrasal stress in Mandarin Chinese appears to be independent of tone. Further work will address whether and to what extent tonal contrasts affect quantitative amounts of jaw displacement; the point here is that the pattern of jaw displacement seems to be impervious to tonal contrasts.

Figure 5 shows jaw displacement patterns (top) and F0 patterns (bottom) in terms of bar graphs for three Japanese phrases: ‘hasi’ ga’, ‘ha’shi ga’, ‘hashi ga’ (there is bridge, there are chopsticks, there is an edge). Each of the words is followed by the nominative particle ga. Looking at the bottom graph first, we see that each of the words has a different F0 pattern: for bridge, with a pitch accent on the middle mora, F0 increases on the middle mora of the utterance; for chopsticks, with a pitch accent on the first syllable, it increases on the first mora, and for edge, with no pitch accent, it increases on the final particle ga. However, notice that the jaw displacement patterns are very similar for the three utterances: increased jaw displacement on the first and last mora, regardless of the pitch accent patterns. Similar patterns for these utterances were found for all five Japanese speakers examined. A regression analysis showed little correlation (r=0.177) between F0-max and jaw displacement in these three words whose pronunciations differ only by pitch accents (Kawahara et al. 2016). Thus, it seems, similar to Mandarin Chinese where phrasal stress is independent of tone, phrasal stress in Japanese is independent of pitch accent.

3.2. Jaw displacement and acoustic correlates

In the case of Mandarin Chinese, increased jaw displacement was found to result in increased duration as well as increased F1. Figure 6 shows vowel durations for four Mandarin Chinese speakers (each column represents data from one speaker) for two utterance types: Mā1ma mà4 mā3 (Mother curses the horse) (top row) and dā4bá3 dà4bá3 (‘a lot of’, literally ‘big grasp big grasp’) (bottom row). Notice that the vowel duration patterns are very similar to the jaw displacement patterns: increased duration on the final vowel, and optional increased duration on the initial vowel. Similar findings of increased duration for phrase-initial and phrase-final syllables have been reported by [18].

Regression analyses (Figure 7) show correlations between syllable duration and jaw displacement (bottom graph), as well as F1 and jaw displacement (top graph) for two speakers (in columns). This suggests that increased jaw displacement has acoustic consequences of both increased duration and increased F1 for Mandarin Chinese speakers.
As for Japanese speakers, Figure 8 shows that F1 increases with jaw displacement. Jaw displacement (top row) and F1 (bottom) data for two Japanese speakers for the utterances ‘aka pajama da’ (they are red pajamas’ (left column) and ‘aka kasa da’ (right column). The patterns of jaw displacement and F1 are very similar; the arrows point to the moras with increased jaw displacement and increased F1 in the utterance. Similar patterns for these sentences were seen for all the Japanese speakers examined.

As increased jaw displacement changes vocal tract area functions, consequently, formant frequencies also change. For low /a/ vowels, as the jaw lowers more, the front oral cavity increases, which leads to increased F1, e.g., [15].

Both Mandarin Chinese and Japanese show a significant relation between jaw displacement and F1. A similar finding is reported for English, e.g., [4]. As increased jaw displacement changes vocal tract area functions, consequently, formant frequencies also change. For low /a/ vowels, as the jaw lowers more, the front oral cavity increases, which leads to increased F1, e.g., [15].

4. Discussion & Conclusion

On one hand, Mandarin Chinese and Japanese are different languages, one is tonal, one has pitch accents. However, the experimental data suggest that in terms of articulatory prosody, they have some interesting similarities. For one thing, Mandarin Chinese and Japanese have both final and initial phrasal stress, implemented by increased jaw displacement. Moreover, the jaw displacement patterns are independent of the F0 contours of the tones/pitch accents. In terms of acoustic manifestations of increased jaw displacement for phrasal stress, both Mandarin Chinese and Japanese show F1 increases. Mandarin Chinese, in addition, shows increased duration. Thus, it seems that for both of these languages, it is the Jaw Articulator that is responsible for the rhythmic patterns of utterance stress/prominence. This work is based on utterances with the low vowels /a/; future work will explore other vowels to assess how formants change due to increased jaw displacement. Ultimately, it is hoped an easy to use algorithm can be developed to factor out vowel quality, and thus observe the underlying jaw displacement patterns due to utterance prosody.

5. Acknowledgments

Thanks to my many colleagues, Jianwu Dang, Atsuo Suemitsu, Ray Iwata, Shigeto Kawahara, Chunyue Zhu. Also thanks to the subjects who helped make this research possible. This work was supported by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C) #25370444 and (A) #25240026.

6. References
