Spanish articulatory rhythm

Donna Erickson (Kanazawa Medical University), Julián Villegas, Ian Wilson, Yuki Iguro
(The University of Aizu)

1 Introduction

All spoken languages show rhythmic patterns. Work on English shows that these rhythmic patterns can be described in terms of Metrical Phonology, such that stress levels are assigned to hierarchically organized syllables, outputting each syllable with a numerical value of stress determined by counting the number of metrical grids assigned according to the hierarchical prosodic layers [1, 2, 3]. Recent work by Erickson et al. [4] shows that the metrically assigned stress levels of the utterance correlate significantly with the amount of jaw displacement, and corresponding F1 values. Ongoing research by Kawahara et al. [5] shows that Japanese also has jaw displacement patterns that reflect the metrical structure of Japanese. Similarly, work by Iwata et al. [6] suggests that Mandarin Chinese metrical structure can be seen in jaw movement as well. Erickson's work underway suggests that the same is true for languages such as French and Spanish. Moreover, for all these languages these patterns appear to be independent of pitch accent or tonal contrasts. These findings also suggest that it is the jaw that provides the "beat" of the language, while the vocal fold vibratory patterns provide the "melody," a viewpoint that is in keeping with the C/D model of articulation proposed by Fujimura [7]. Given this understanding, it follows that changing the native "beat" when learning a second language is a challenging task, but these patterns can be changed as one becomes more fluent in the language [8, 9].

This paper addresses Spanish articulatory rhythm. Preliminary work with Spanish suggests that the initial syllable of a phrase has the strongest phrasal stress while the last one, the weakest. This contrasts with French and Mandarin Chinese, where the last syllable of a phrase has the strongest stress. For Japanese, it may be that both the first and last syllables of the phrase have the strongest stress [5].

The focus of this paper is on Spanish phrasal stress patterns and their articulatory correlates.

2 Method

2.1 Speakers

We recorded utterances from three paid Salvadorian female participants (s1, s2, and s3), who happened to be siblings. Their ages at the time of data collection were 28, 23, and 34, respectively. Although their native language is Salvadorian Spanish, they have different exposure to other languages: while s3 reported ten years of English learning and three years of Japanese (she had lived the last six years in Japan), s1 and s2 had five years and 10 years of English learning, respectively. S2 had also lived in the USA for one year, immediately preceding the data collection whereas s1 has lived mainly in El Salvador. With the exception of s2, the speakers reported to have still a neutral Salvadorian Spanish accent, as acknowledged by their Salvadorian acquaintances.

2.2 Stimuli

The speakers read sixteen Spanish, three Japanese, and seven English sentences. These sentences were selected so the same vowel was used in all the constituent words (i.e., the sentences were vowel normalized). In this report we limit the analysis to one or two sentences of each language, for which we have previously collected data, so that we can draw comparisons between languages. These sentences are summarized in Table 1: u08 and u09 are Spanish sentences of twelve syllables each, u01 and u03 are Japanese sentences of six and fifteen syllables, respectively, and u06 is an English sentence of eleven syllables.

2.3 Procedure

Speakers were recorded in two sessions: first without and second with an ultrasound probe under their chin. The ultrasound probe sessions were recorded so that we could analyze tongue movements during speech, but for the sake of brevity, in this report we limit the analysis to those sentences uttered without the probe. Each recording session was comprised of three blocks corresponding to three languages, recorded in this order: Spanish, English, and then Japanese. Within each block, each utterance was repeated seven times, each whole block of utterances was randomly presented from a laptop computer located in front of the speaker. We prevented head tilting by changing the height of the display for each participant. Errors (mainly coughs, reading errors, and ultrasound probe misalignments) were marked visually and aurally in the video and audio recordings, prior to having the speaker repeat the token. Speakers were able to take short breaks between blocks and sessions. The two sessions were recorded in about one hour. Permission for performing these recordings was obtained following the University of Aizu ethics procedure.

Table 1 Utterances used in this study

ID	Utterance
u08	Mamá valsará “Casablanca” mañana.
u09	Ana valsará “Casablanca” mañana.
u01	赤パジャマだ (Aka pajama da)
u03	だからまなは頭がサラサラだ (Dakara Mana wa atamaga sarasara da)
u06	I saw five bright highlights in the sky tonight.

After instructing the speakers about the experiment and querying them about their language background, they were asked to sit straight in a well-lit room, in front of a white background. The experimenters (two in each session) assisted them with putting on a lapel microphone. Subjects also wore a glasses frame (without lenses) with a blue circle of about 8 mm in diameter, located at the center of the frame, above the participant’s nose. A second marker was placed by the experimenters on the chin of the speaker and perpendicular to the frame line.

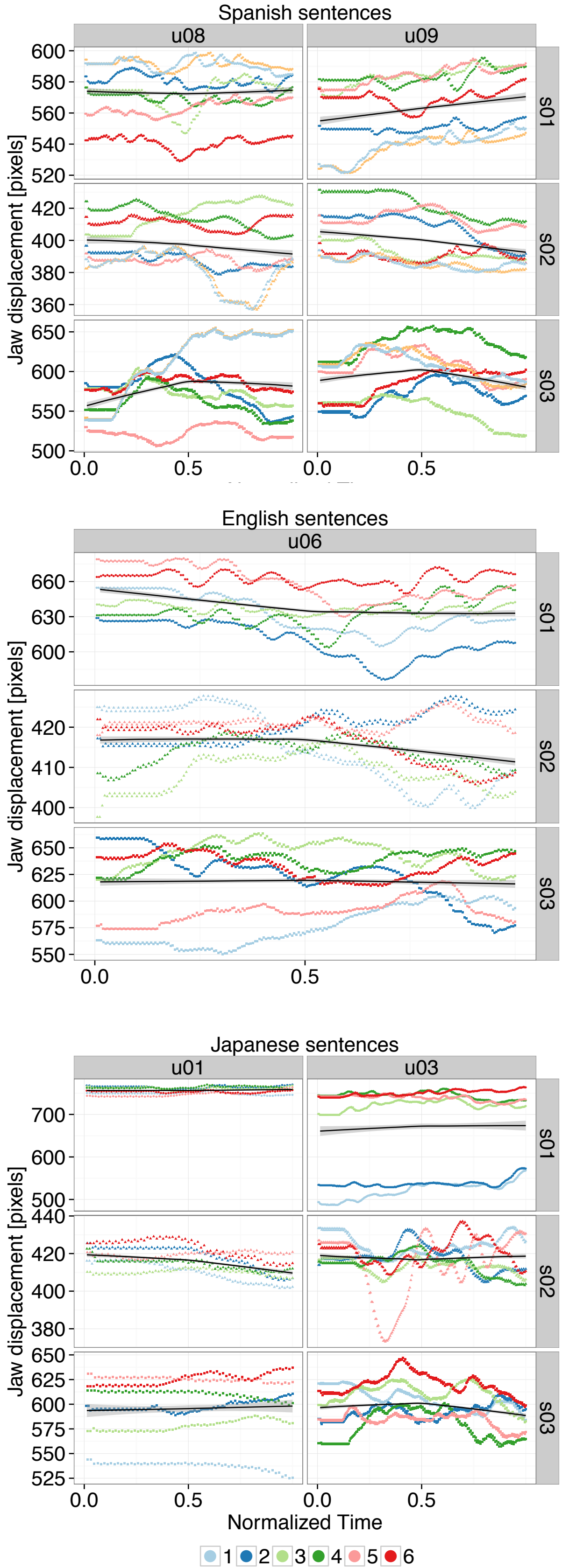

Figure 1. Jaw displacement of the Spanish speakers uttering sentences in Spanish (top), English (middle), and Japanese (bottom). Trials are color-coded; local linear regression fitting and its 95% CI are shown in black and gray.

Speakers were recorded in video at 29.97 frames per second and at 44.1 kHz/16 bits in audio.

2.4 Post-processing

End points of each utterance were located from the audio of the video recordings in Praat [10] by visual inspection. These end-points were used to extract the videos using ffmpeg[1] routines. From the extracted videos, the blue dots were traced using the marker tracker program described in [11]. These trajectories were used to compute the Euclidian distance between the markers. Jumps of over five pixels between consecutive frames were considered tracking errors and were automatically corrected by subtracting such distance from all the following points in the trace. The five pixels threshold was found by visually inspecting the trajectories and their corresponding videos. Note that in this research we present results in terms of the distance between the two markers as opposed to the absolute jaw position as in some previous reports. So, when comparing graphs of this work to those of previous work, the vertical scale of the graphs is inverted (i.e., as the displacement increase and the curve gets higher on these graphs, the jaw is lowering—the opposite display of previous work).

3 Results

The individual trajectories (per speaker and trial) are presented in **Error! Reference source not found.**. As a way of minimizing the effect of different speech rates across trials, trajectories are presented in normalized time. The solid black line represents the average value at a given time (for that speaker) across all trials of that sentence.

For Spanish utterances, whereas for s2, the two sentences are in agreement with the overall pattern reported by Erickson [12]: an overall positive declination where the jaw aperture is initially more open and ends more closed; this pattern corresponds to the overall reduction of the distance between the markers shown in the middle panel of **Error! Reference source not found.**. This pattern is reversed for s1, and partially observed for s3 (although initially increasing, the distance between markers decreases towards the end of the sentence).

As for English, judging by the jaw movement patterns, s2 is probably the best speaker of English (probably because of her immersion for one year, even though she only had a total of five years of English study). S2 shows the typical pattern of no initial phrasal stress in English, and nuclear stress (increased jaw displacement) on the key word “five,” followed by an overall decrease in jaw displacement, in agreement with what was reported in [4], and [5]. In contrast, s1 and s3 have the poorest skills in English, and neither shows clear nuclear stress on “five.” In the case of s1, the largest jaw displacement (most open jaw) occurs at the beginning of the utterance (on “I”), and in the case of s3, the average displacement remains flat over the sentence.

The observed variability between speakers and within trials is manifested in the Japanese sentences as well. However, it seems that when confronted with an unknown language (as it was the case for s1 and s2), the speakers tried different articulations when the length of the sentence allowed it (i.e., for utterance u03 on the right side of the figure).

4 Final remarks

These results, especially the English results of speaker s2, suggest that immersion in a second language may produce changes to the rhythmic patterns observed in the native language of the speaker. Whereas s1 had lived all her life in El Salvador, s2 had just completed a year of living in an English environment, to the point that her relatives noted changes in her Spanish prosody, less noticeable was the effect for s3, who lived in Japan where English was spoken to her mostly with a foreign accent. These differences in second language immersion seem to be reflected in jaw displacement of the different languages studied, and this is especially evident in the fact that there was not a pattern that was common to

[1] https://ffmpeg.org

even any two of the speakers in their native Salvadorian Spanish.

This is especially remarkable, given the fact that the speakers are sisters who grew up together, presumably acquiring and speaking their native language using the same rhythm. Although our sample size is limited to only three speakers (albeit sisters), living environment seems to play a big role in the rhythm one uses, even in one's native language.

Acknowledgements

This work was partially supported by the Japan Society for the Promotion of Science (JSPS), Grants-in-Aid for Scientific Research (C) #25370444.

References

[1] Liberman, M. and Prince, A., "On stress and Linguistic rhythm," Linguistic Inquiry 8, pp. 249–336 (1977).

[2] Selkirk, E. O., "The Syllable," Foris, Dordrecht (1982).

[3] Hayes, B., "Metrical Stress Theory: Principles and Case Studies," The University of Chicago Press, Chicago (1995).

[4] Erickson, D., Suemitsu, A., Shibuya, Y., and Tiede, M., "Metrical structure and production of English rhythm," Phonetica 69, pp. 180–190. (2012).

[5] Kawahara, S., Erickson, D., Moore, J., Shibuya, Y., and Suemitsu, A., "Jaw displacement and metrical structure in Japanese: The effect of pitch accent, foot structure, and phrasal stress," Journal of the Phonetic Society of Japan, 18 (2), pp. 77–87, (2014).

[6] Iwata, R., Erickson, D., Shibuya, Y., and Suemitsu, A. "Articulation of phrasal stress in Mandarin Chinese" Proc. of the Acoustical Society of Japan, autumn meeting (2015).

[7] Fujimura, O. "The C/D model and prosodic control of articulatory behavior," Phonetica 57, pp. 128–138, (2000).

[8] Abe, Y., Wilson, I., and Erickson, D. "Video recordings of L1 and L2 jaw movement: effect of syllable onset on jaw opening during syllable nucleus," Proc. of the Acoustical Society of America, Fall Meeting, Kansas City, 132.3, Pt.2, p.2005 (2012).

[9] Wilson, I., Erickson, D., and Horiguchi, N. "Articulating rhythm in L1 and L2 English: Focus on jaw and F0," Proc. of the Acoustical Society of Japan, autumn meeting, pp. 319-322 *(2012)*.

[10] P. Boersma and D. Weenink. Praat. Available [Jul. 2015] from www.praat.org

[11] Barbosa, A. V., and Vatikiotis-Bateson, E. Video tracking of 2D face motion during speech. In *Signal Processing and Information Tech., IEEE International Symposium on* (pp. 791−796). (2006)

[12] Erickson, D. Articulation of rhythm: a Multilanguage perspective. Presentation at Haskins Laboratories, New Haven, CT. (2015).