Some acoustic and articulatory correlates of phrasal stress in Spanish

Donna Erickson[1], Julián Villegas[2], Ian Wilson[2], Yuki Iguro[2], Jeff Moore[3], Daniel Erker[4]

[1] Kanazawa Medical University, Japan
[2] The University of Aizu, Japan
[3] Sophia University, Japan
[4] Boston University, U.S.A

ericksondonna2000@gmail.com, julian@u-aizu.ac.jp, wilson@u-aizu.ac.jp, s1200249@gmail.com, jeffmoore@sophia.ac.jp, danny.erker@gmail.com

Abstract

All spoken languages show rhythmic patterns. Recent work with a number of different languages (English, Japanese, Mandarin Chinese, and French) suggests that metrically (hierarchically) assigned stress levels of the utterance show strong correlations with the amount of jaw displacement, and corresponding F1 values. This paper examines some articulatory and acoustic correlates of Spanish rhythm; specifically, we ask if there is a correlation between phrasal stress values metrically assigned to each syllable and acoustic/articulatory values. We used video recordings of three Salvadoran Spanish speakers to measure maximum jaw displacement, mean F0, mean intensity, mean duration, and mid-vowel F1 for each vowel in two Spanish sentences. The results show strong correlations between stress and duration, and between stress and F1, but weak correlations between stress and both mean vowel intensity and maximum jaw displacement. We also found weak correlations between jaw displacement and both mean vowel intensity and F1.

Index Terms: Spanish rhythm, duration, intensity, F0, F1, jaw displacement

1. Introduction

Rhythmic patterns are seen in all languages. 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 for English the metrically assigned stress levels of the utterance correlate significantly with the amount of jaw displacement, and corresponding F1 values; similar findings are reported by Kawahara et al. [5] that Japanese jaw displacement patterns/F1 reflect the metrical/hierarchical structure of Japanese, specifically, with increased jaw displacement/higher F1 at phrase-initial and phrase-final positions; and also a similar pattern is reported for Mandarin Chinese [6] . Moreover, these patterns appear to be independent of pitch accent or tonal contrasts [4, 5, 6].

Our working hypothesis is that in terms of articulation, 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, 10].

This paper examines some acoustic and articulatory correlates of Spanish phrasal stress. Acoustic cues of lexical stress have been reported by e.g., Llisterri et al. [11], Ortega-Llebaria and Prieto [12], and Hualde [13] to include changes in F0, intensity and duration, but no work specifically on acoustic or articulatory cues of phrasal stress has been done that we know of, except for earlier pilot studies by us [14].

About stress in Spanish, according to Ladd and Roca [15], Spanish has three types of stress: "primary word stress, corresponding to the highest prominence in the lexical word, main phrasal stress, which signals the accentual peak in the phrase of phonic group, and secondary stress, which includes all remaining discernible stresses" with phrasal stress having the largest amount of prominence. See also Harris [16] for a discussion of secondary stress being assigned post-lexically. Spanish is also said to have a trochaic metrical structure, such that the initial syllable of a foot receives the stress [17, 18]. Pilot studies on articulation of Spanish phrasal stress [14] show a pronounced articulatory declination over the sentence. More details about Spanish metrical structure are discussed below in the results section.

In this paper, we investigate whether there is a correlation between phrasal stress values as assigned by hierarchical layers of metrical phonology and measured values of jaw displacement, F0, duration, intensity and F1.

2. Method

We videotaped utterances from three paid Salvadoran female participants (s1, s2, and s3), who were siblings. Their ages at the time of data collection were 28, 23, and 34, respectively. The speakers read sixteen Spanish, three Japanese, and seven English sentences. In this report we limit the analysis to two Spanish sentences. The sentences were selected so that the same vowel was used in all the constituent words (i.e., the sentences were vowel normalized, since jaw displacement varies according to vowel height, e.g., [19]. The sentences are shown in Table 1: u08 and u09 are Spanish sentences of twelve syllables each. Notice that the two Spanish utterances differ only in terms of the first word: u08 has lexical stress on the second syllable whereas u09 has it on the first one.

Table 1. *Utterances used in this study*

ID	Utterance
u08	Mamá valsará "Casablanca" mañana.
u09	Ana valsará "Casablanca" mañana.

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 [20], but 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. After instructing the speakers about the experiment and querying them about their language background, they were asked to sit in a well-lit room, in front of a white background. The experimenters (two in each session) assisted them with putting on a DPA 4080 cardioid 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, as shown in Figure 1.

Speakers were recorded with a Panasonic HDC-TM750 digital video camera in 1080p video at 29.97 frames per second and at 44.1 kHz/16 bits in audio.

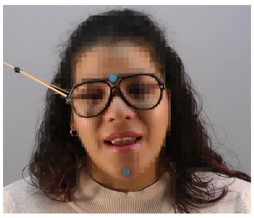

Figure 1: *Marker positions*

End points of each utterance were located from the audio of the video recordings in Praat [21] by visual inspection. These end-points were used to extract the videos using ffmpeg routines (https://ffmpeg.org). From the extracted videos, the blue dots were traced using the marker tracker program described in [22]. These trajectories were used to compute the Euclidian distance between the markers. Note that in this research the term "jaw displacement" is the distance measured between the two markers, as opposed to the jaw position relative to the occlusal bite plane as in previous electromagnetic articulatory studies. The y-axes of the graphs displayed in this study are reversed, in order to show increased jaw displacement (i.e., increased distance between the two markers) in the downward direction.

3. Results

3.1. Jaw displacement and F0

Typical jaw displacement, F0 and spectrograms of Spanish speakers (s1, s2, s3) while uttering Spanish sentence u08 *Mamá valsará "Casablanca" mañana* ("Mama dances 'Casablanca' tomorrow") are shown in Figure 2 and u09 *Ana valsará "Casablanca" mañana ("Ana dances 'Casablanca'* tomorrow") (for s2 and s3) in Figure 3. Looking first at the F0 contours (middle panels) in Figure 2, we see a consistent F0 declination across the sentence but also variability among the speakers in terms of peak F0: peak F0 for s1 occurs on syllable 5 (*rá*), for s2, on syllable 2 (*val*), and for s3, on syllable 3 (*má*). Notice that only for s1 do we see peak F0 on the lexically stressed syllable (*rá*). As for the jaw displacement pattern (bottom panels), for each speaker we see a jaw opening for each syllable. S1 shows a slightly different pattern than s2 and s3—a gradual increase of jaw displacement throughout the utterances with the final syllable showing the largest jaw displacement; however, s2 and s3 show overall an articulatory declination where the jaw aperture is initially more open and ends more closed. A similar pattern of articulatory declination was reported for a pilot study for another speaker [14]. Notice also that we see jaw opening for each syllable in the utterance. Moreover, even though the vowels are all the same, the amount of jaw displacement changes as a function of the position of the syllable in the word, foot, phrase, and sentence.

Looking now at Figure 3, for both s2 and s3 the peak of F0 occurs on the second syllable of *Ana*. S1 is not shown here, since there seemed to be no typical pattern for this speaker for this utterance. As for jaw displacement, we see a pattern very similar to that of the previous utterance, except that we see a reversed pattern of jaw displacement for the first word: for *Mamá*, we see increased jaw displacement on the second syllable, but for *Ana*, we see increased jaw displacement on the first syllable.

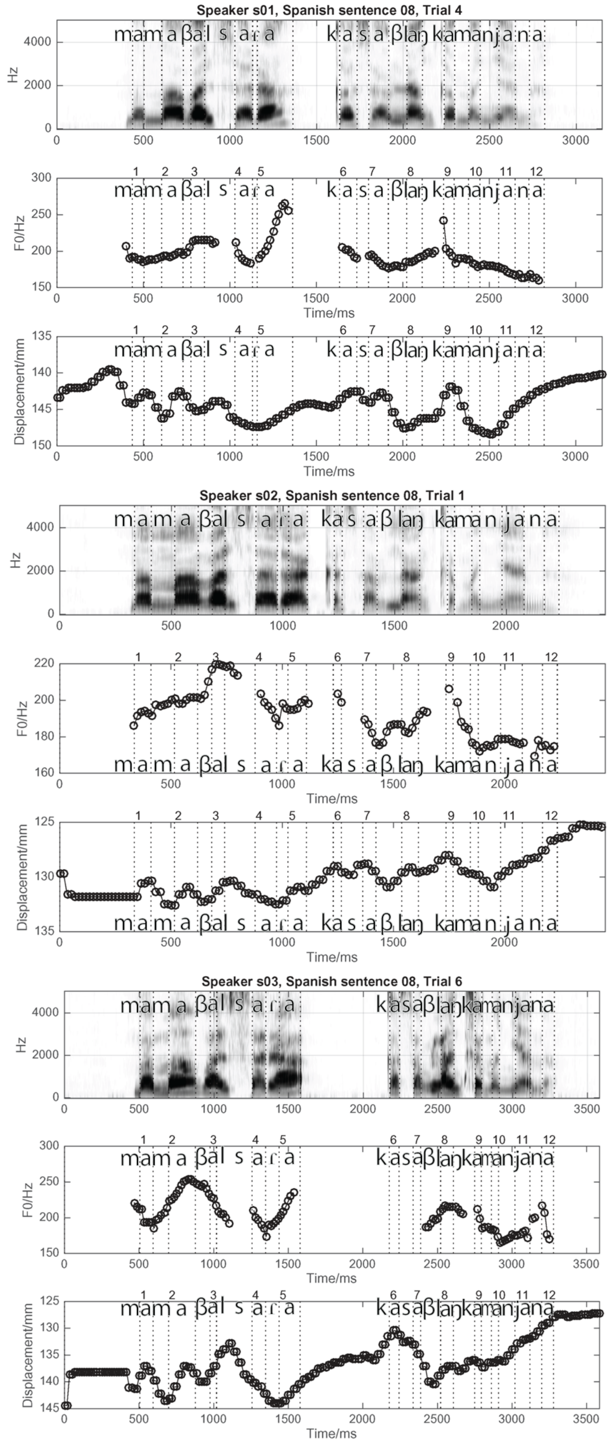

Figure 2: *Spectrograms, F0 and jaw displacement of Spanish speakers (s1, top, s2, middle, s3, bottom) while uttering Spanish sentence u08 Mamá valsará "Casablanca" mañana. The syllables are numbered above the panels showing F0 and jaw displacement.*

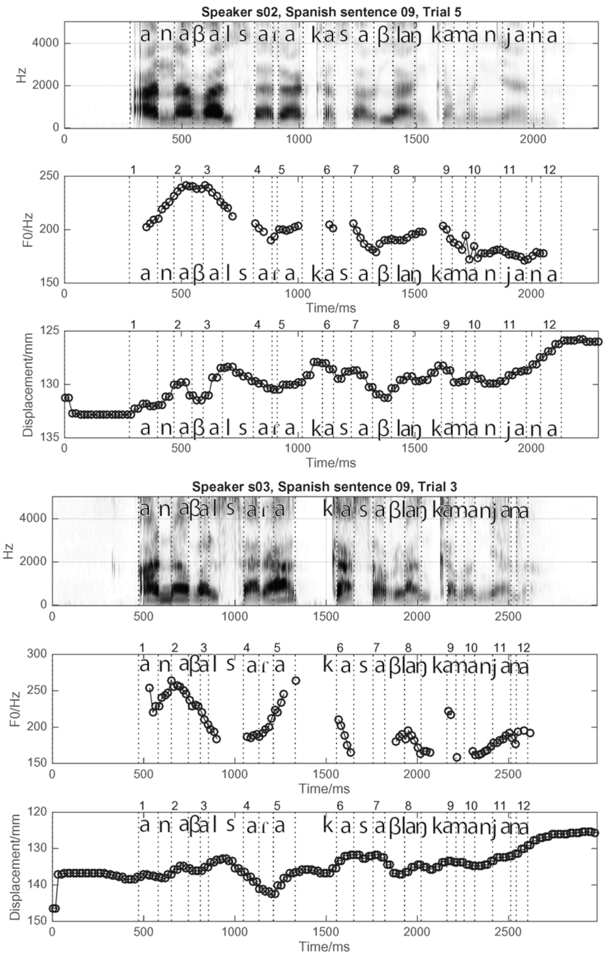

Figure 3: *Spectrograms, F0 and jaw displacement of Spanish speakers (s2, top, s3, bottom) while uttering Spanish sentence u09 Ana valsará "Casablanca" mañana.*

3.2. Metrically-assigned phrasal stress

Inspired by e.g., Hualde [13] and Buckley [18], the metrical structure of the Spanish sentences u08 and 09 is posited to be as shown in Figure 4. The syllable stress levels are assigned by counting the number of black circles above each syllable.

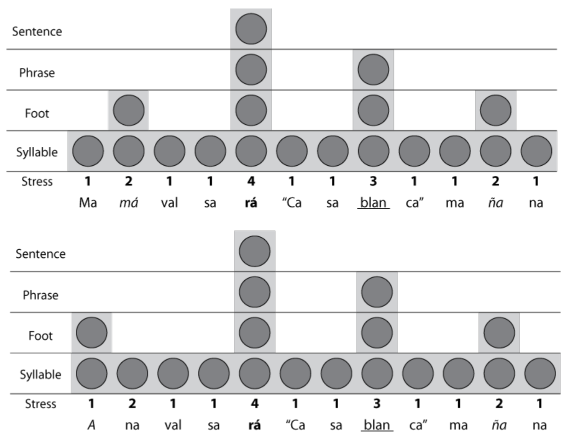

Figure 4. Metrical stress patterns of the two sentences, *Mamá valsará "Casablanca" mañana* and *Ana valsará "Casablanca" mañana.*

3.3 Articulatory and acoustic correlates of metrically-assigned stress

Correlation analyses, with the averaged values of all the repetitions of each utterance type pooled across the three speakers, for jaw displacement and mean vowel F0, mean vowel intensity and F1 at the vowel midpoint showed no significant correlation between jaw displacement and mean F0 (r = .000, p=1), and significant but weak correlations between jaw displacement and intensity (r = -.169, p<.001), and jaw displacement and F1 (r = .192, p<.001). As for correlations with the metrically assigned stress values shown above, since stress pattern is not normally distributed, a Kendall correlation was applied in this case (alternative hypothesis: the two samples are different). Significant but weak correlations were found between stress and jaw displacement (r = .107, p=.030) and between stress and intensity (r = .178, p<.001). The strongest correlations were found between stress and vowel duration (r = .44, p<.001) and between stress and F1 (r = .443, p<.001).

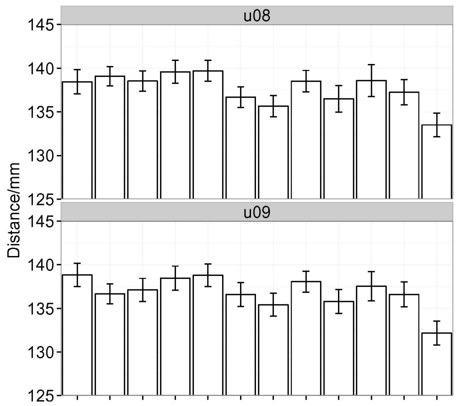

Figure 5*: Bar graphs showing average jaw displacement of each syllable for the three Spanish speakers for Spanish sentence u08 Mamá valsará "Casablanca" mañana and u09 Ana valsará "Casablanca" mañana. The height of the bar represents the amount of jaw displacement for each syllable (i.e., the higher the bar, the lower the jaw).*

Visual inspection of the two sentences in Figures 5, 6 and 7, seems to show a match with the metrical structure described in Figure 4. Also, note the similarity in the patterns of jaw displacement, F1 and duration. For the lexically stressed syllables, *A/má, rá, blan, and ña,* we see increases in jaw opening, F1 and duration. For *ña,* note however, jaw displacement is a bit lower than might be expected, perhaps because of the palatalization on the syllable initial consonant, causing the tongue and jaw to raise slightly. A curious observation is on utterance final syllable *na* an increase in F1 with no increase of jaw displacement. Reduction of jaw opening may be related to the fact that one of the speakers tended to produce this final syllable as a schwa. Also, this may be because it is the final word in the sentence; further work is needed to understand to what extent phrase-final phenomena may be involved. In English, for instance, phrase-final lengthening occurs, without concomitant increase in jaw displacement, see e.g., [23]. One hypothesis we entertain is that for Spanish there may be independent ways of manifesting lexical stress, phrasal stress and phrase finality. One approach to resolving this last possibility is to examine the same sentences, but put *mañana* in another place in the utterance, either the first word of the sentence, or some place in the middle. We also plan to analyze other sentences with different lexical stress patterns.

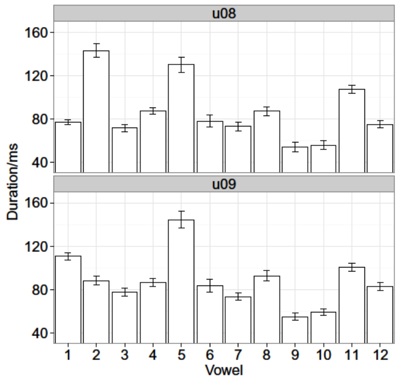

Figure 6*: Bar graphs showing average vowel duration of each syllable for the three Spanish speakers for Spanish sentence u08 Mamá valsará "Casablanca" mañana and u09 Ana valsará "Casablanca" mañana.*

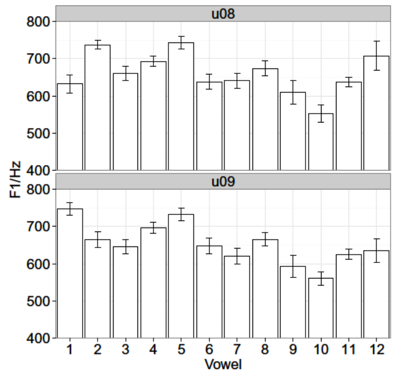

Figure 7*: Bar graphs showing average F1 a vowel midpoint of each syllable for the three Spanish speakers for Spanish sentence u08 Mamá valsará "Casablanca" mañana and u09 Ana valsará "Casablanca" mañana.*

4. Conclusions

This paper examines some articulatory and acoustic correlates of Spanish rhythm; specifically, we ask if there is a correlation between phrasal stress values metrically assigned to each syllable and acoustic/articulatory values. The results show strong correlations between stress and duration, and between stress and F1, but weak correlations between stress and both mean vowel intensity and maximum jaw displacement. We also found weak correlations between jaw displacement and both mean vowel intensity and F1.

5. Acknowledgements

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

6. References

[1] M. Liberman and A. Prince, “On stress and Linguistic rhythm,” *Linguistic Inquiry*, vol. 8, pp. 249–336, 1977.

[2] E.O. Selkirk, *The Syllable*, Foris, Dordrecht, 1982.

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

[4] D. Erickson, A. Suemitsu., Y. Shibuya, and M. Tiede, “Metrical structure and production of English rhythm,” *Phonetica,* vol. 69, pp. 180–190, 2012.

[5] S. Kawahara, D. Erickson, J. Moore, Y. Shibuya, and A. Suemitsu, “Jaw displacement and metrical structure in Japanese: The effect of pitch accent, foot structure, and phrasal stress,” *Journal of the Phonetic Society of Japan*, vol. 18, no. 2, pp. 77–87, 2014.

[6] R. Iwata, D. Erickson, Y. Shibuya, and A. Suemitsu, “Articulation of phrasal stress in Mandarin Chinese” *Proc. of the Acoustical Society of Japan*, autumn meeting, pp. 207–210, 2015.

[7] O. Fujimura, “The C/D model and prosodic control of articulatory behavior,” *Phonetica,* vol. 57, pp. 128–138, 2000.

[8] Y. Abe, I. Wilson, D. 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] I. Wilson, D. Erickson, and N. Horiguchi, “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] M. Young-Sholten, *The Acquisition of Prosodic Structure in a Second language*. Tübingen: Max Niemeyer Verlag, 1993.

[11] J. Llisterri, M. J. Machuca, C. de la Mota, M. Riera, and A. Ríos, "The role of F0 peaks in the identification of lexical stress in Spanish", in Braun, A. and Masthoff, H.R. (Eds.) Phonetics and its Applications. Festschrift for Jens Peter Köster on the Occasion of his 60th Birthday. Stuttgart: Franz Steiner Verlag (Zeitschrrift für Dialektologie und Linguistik, Beiheft 121). pp. 350–361, 2002.

[12] M. Ortega-Llebaria and P. Prieto, “Acoustic correlates of stress in Central Catalan and Castilian Spanish,” *Language and Speech*, vol. 54, no. 1, 2011.

[13] J.I. Hualde, “Chapter 8. Stress and rhythm,” In J. I. Hualde, A. Olarrea, and E. O’Rourke (eds.) *The Handbook of Hispanic Linguistics*. Hoboken, NJ: Wiley-Blackwell, pp. 153–172, 2012.

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

[15] D. R. Ladd and I. Roca, “Secondary stress and metrical rhythm,” *Phonology*, vol.3, no.1, pp. 341–370, 1986.

[16] J.W. Harris, “With respect to metrical constituents in Spanish” In Hector Campos, Fernando Martinez-Gil (eds.) *Current Studies in Spanish Linguistics* (pp. 447–473), Washington, DC: Georgetown University Press, 1991.

[17] N. Sebastian-Galles, “Speech perception in Catalan and Spanish” in M. Carreiras, J. E. Garcia-Albea, and N. Sebastian-Galles (eds.) *Language Processing in Spanish*, pp. 1–17, 1996.

[18] E. Buckley, “Foot alignment in Spanish secondary stress” to appear in J. Heinz, R. Goedemans, and H. van der Hulst (eds.) *Dimensions of Stress*. Cambridge University Press, To Appear.

[19] J.C. Williams, D. Erickson, Y. Ozaki, A. Suemitsu, N. Minematsu, and O. Fujimura, “Neutralizing differences in jaw displacement for English vowels.” *Proc. of International Congress of Acoustics*, 2013.

[20] J. Villegas, I. Wilson, Y. Iguro, and D. Erickson, “Effect of a fixed ultrasound probe on jaw movement during speech.” Proc. Ultrafest VII, 2015.

[21] P. Boersma and D. Weenink. Praat. Available [Mar. 2016] from http://www.praat.org

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

[23] C. Menezes, B, Pardo, D. Erickson, and O. Fujimura, “Changes in syllable magnitude and timing due to repeated correction,” *Speech Communication*, vol. 40, no. 1, 71–85, 2003.