10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Temporal organization of spoken utterances from an articulatory point of view

Donna Erickson[1], *Ting Huang*[2], *Caroline Menezes*[3]

[1]Haskins Laboratories, U.S.A., Federal University of Rio de Janeiro, Brazil, Kanazawa U., Japan
[2]Haskins Laboratories, U.S.A. [3]University of Toledo, Ohio, U.S.A.
ericksondonna2000@gmail.com, funting.huang@gmail.com, Caroline.Menezes@utoledo.edu

Abstract

We view speech as organized articulatorily in terms of units that result from the speaker opening and closing the mouth, e.g., mandible. How much the mandible (aka jaw) opens during each syllable is determined by the characteristics of the vowel phoneme, and also the prosodic relationship of the syllable to other syllables in the utterance. Regression analyses of jaw displacements in an utterance containing all mid front /e/ vowels with prominence ratings (Rapid Prosodic Transcription (RPT)) by English listeners shows significant correlations between jaw and perceived prominence. The results suggest increased jaw displacement results in changes in acoustic information, which listeners may use to recognize nuclear stress in English. The results can be interpreted to support a view of temporal organization of spoken utterances as a series of syllables in which the speaker opens the jaw more for the important items in the utterance; specifically, in English, the syllable with the largest magnitude in the utterance (in terms of jaw displacement) is perceived as having the nuclear stress.

Index Terms: articulatory organization, jaw displacement, tongue blade, formants, nuclear stress

1. Introduction

The approach taken here is that speech is organized articulatorily in terms of units that result from the speaker opening and closing the mouth. Thus, spoken utterances are a series of mouth (jaw) openings and closings; each opening/closing is a syllable-unit; groupings of openings/closings constitute words, feet, phrases, and sentences. Maximum opening of the jaw is referred to as vowel, maximum closings (onset, offsets), as consonants. How much the jaw opens during each syllable is determined by the characteristics of the vowel phoneme, i.e., for low vowels (in English) the jaw is approximately 2 mm lower than for mid vowels, and 4 mm lower than for high vowels [1, 2].

In addition, how much the jaw opens for each syllable is prescribed by the relation of the syllable with regard to other syllables in the syllable groupings. For many languages, such as French, Japanese, Mandarin Chinese, the jaw opens more at the end (and optionally at the beginning) of each grouping unit, as if to mark that this series of syllables belong to this group (e.g., [4, 5, 6, 7]). For languages like English, the amount of jaw opening changes relative to other syllables within the grouping units. In this way, the amount of jaw opening can be seen to mark the hierarchical relationships among syllables within a group; specifically, within each grouping unit, one syllable will have more jaw opening than others. It is as if to mark which syllable within the grouping is "most important", the speaker opens his mouth more (e.g., [8, 9]). This is similar to [10], where stress is "one of the devices that a language can use to set up a hierarchical organization of its utterances" (p. 8), a concept formalized by [11, 12, 13], but from the point of view of temporal articulatory organization of speech.

Over an entire utterance, the syllable with the nuclear stressed unit is articulated with the largest jaw opening (e.g. [8, 9, 14]). To emphasize or especially focus a word, the jaw will open even more (e.g., [15, 16, 17]), marking that this particular item is very important. (Note that increased jaw opening for increased prominence occurs regardless of the vowel height, but it is greater for low vowels compared with high vowels (e.g., [22])). Increased jaw opening for emphasized/focused words has also been reported in other languages, suggesting that opening the jaw more is a "universal" method of signaling which item is important (e.g., [18] for French, [19] for Japanese).

In this approach to temporal organization of speech, we equate amount of jaw opening with "syllable magnitude," as proposed by Fujimura (e.g., [3]). Each syllable in a given utterance has a measurable amount of magnitude, based on the maximum jaw displacement at the nucleus of the syllable and its position in the prosodic structure of that utterance. In acoustic terms, increased jaw displacement/syllable magnitude results in changes in resonant frequencies of the vocal tract (i.e., formants). Depending on the vowel characteristics, the first (F1) and second formant (F2) vary with increased jaw displacement, as the tongue moves to preserve the phonemic identity of the vowel ([20, 21]. Specifically, F1 was found to increase with increased jaw displacement for contrastively emphasized syllables, regardless of vowel height [20], and the separation of F1 and F2 increased or decreased depending on whether the vowel was high or low, respectively. Often other acoustic changes, such as increased intensity, duration and F0, may co- occur with increased jaw displacement but not necessarily (e.g., [10, 21, 22]).

This paper addresses nuclear stress in English. According to work by e.g., [23], nuclear stress often occurs on the "right-most content noun of the utterance". But, nuclear stress can be "elusive" to non-first language speakers of English, and frequently, even to first language speakers; sometimes, it is not necessarily clear which of the items in the utterance has nuclear stress. Thus, a speaker could produce several tokens of the same utterance with nuclear stress on a different content word. In previous studies by Erickson and colleagues [9], in which the amount of jaw displacement was considered to indicate which item has nuclear stress, it was found that nuclear stress could shift in the same sentence due to the speaker's affective judgement. For the utterance, *I saw five bright highlights in the sky tonight*, three of the four speakers tended to choose *<u>high</u>lights*, while one chose *five* [9]; in another study [24], two speakers emphasized *five*, one *<u>high</u>lights*, and one *sky*.

10.21437/SpeechProsody.2020-3

This paper is a small pilot study in which we examine perceptibility for strong syllable magnitude/jaw displacement on nuclear-stressed syllables in English. The question we address is whether indeed the largest jaw displacement occurs with nuclear stress, and, do listeners perceive the word with largest jaw displacement as having nuclear stress. Articulatory recordings of jaw displacement were made using electromagnetic articulographic recordings; perception of nuclear stress was done using the Rapid Prosodic Transcription (RPT) methods (e.g., [23]).

With regard to RPT, prosody is not codified with a canonical spelling system as are consonants and vowels. Thus, ascertaining which words in an utterance are perceived as being prominent is a challenging task for linguists. Rapid Prosodic Transcription (RPT), a method developed by Jennifer Cole and colleagues, is one approach to addressing this challenge. In this approach, naïve listeners are asked to specify which syllable/word in an utterance receives prominence, the more times a word/syllable is marked as being prominent by listeners is interpreted as representing the amount of perceived prominence on that word. Listeners may mark more than one word as being prominent. The number of prominence markings assigned to each of the syllables can theoretically range from 0 to n, where n is the total number of listeners in the experiment; the number of markings of prominence on each syllable indicate the perceived stress level of each syllable, which theoretically, then corresponds to the descriptions of metrical grids/prosodic hierarchy of the utterance along the lines proposed by e.g., [11, 12, 13]. The word/syllable with the most markings of prominence is concluded to have the nuclear stress in the utterance; the next largest markings will indicate where phrasal and foot stresses occur.

An earlier study by Erickson and colleagues using RPT for assessing perceived prominence in an utterance showed that that patterns of jaw displacement matched perceptual judgments of listeners [14]. But their paper did not specifically address nuclear stress; also, the sentence examined contained all low vowels /a/, while the sentence examined in this study contains all mid front vowels /e/. Some of the articulatory data presented here were previously reported in Huang et al. [24] and Smith et al. [7]. The new information reported here has to do specifically with production and perception of nuclear stress, and some of the acoustic correlates of nuclear stress.

2. Method

2.1. Data Recording & Analysis

Articulatory and acoustic data for two American English speakers (1 M, 1 F) were recorded at JAIST, Japan, using 3-D EMA (Carstens AG500). For all recordings, one sensor was placed on the lower medial incisors to track jaw opening, and four additional sensors (upper incisors, bridge of the nose, left and right mastoid processes behind the ears) were used as references to correct for head movement. Additional sensors were placed on the upper and lower lips, and three on the tongue: Tongue Tip (TT), Tongue Blade (TB) and Tongue Dorsum (TD). Here, we report on the lowest position of the jaw during each target vowel, as measured from the occlusal plane, and the TB at the time the jaw is maximally low. 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. Acoustic measures of maximum F0 and average formant frequencies (F1 & F2) for each target syllable were made using the software PRAAT [25].

Speakers read a randomized list from a PowerPoint display of the sentences, with 5 to 6 randomizations of each utterance. The corpus consisted of a number of utterances; the utterances examined here are the tokens of *I gave dates to Kay today*. It was important to keep all the vowels as similar as possible, in order to control for vowel quality, since vowel height affects jaw opening. For this sentence, measurements were made at the maximum jaw displacements for the [e] part of the /ei/ diphthongs for the four words *gave*, *dates*, *kay*, and *day*. The first word of the sentence *I*, as well as the reduced syllables for *to* are not measured, since the vowels are different.

MView [26] was used for marking boundaries of vowel gestures (see Figure 1). The articulatory parameters (jaw and Tongue Blade (TB) positions relative to occlusal plane) were measured at the lowest jaw position (marked with vertical thick lines in the figure) within the vowel for each of the target words.

Figure 1 shows jaw and tongue blade tracings for the four words, *gave*, *dates*, *kay*, and *day*. Upper panels show the acoustic wave form and spectrogram and bottom panels show TB and jaw tracings (mm), respectively. The vertical thick lines demarcate the point of maximum jaw opening in the vowel part of each word, the point at which the jaw and tongue measurements were made. A comment: Although each syllable is produced by opening the jaw, the jaw openings are then re-organized within the hierarchical structure of English such that the reduced syllables are often produced with little to no jaw opening, as shown in Figure 1.

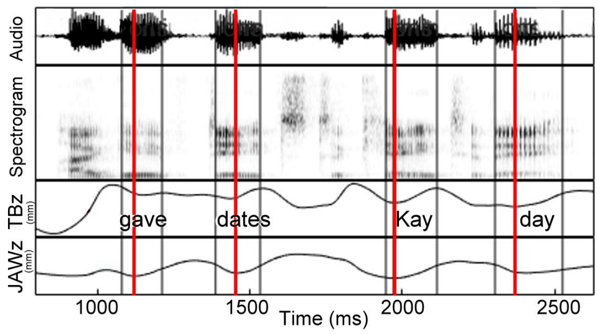

Figure 1: *A typical articulatory pattern for a sample utterance of (I) gave dates to Kay today by L1 speaker, A10.*

2.2. RPT perception tests

The listeners were presented with two sentence types (*Yes, I saw five bright highlights in the sky tonight* and *I gave dates to Kay today*) produced by the two English speakers (A09 and A10) for a total of 22 utterances. There were 5 tokens of each of the sentences by A09 and 6 by A10; however, one of the tokens by A09 was not included in the RPT test, due to technical difficulties. The tests were presented by computer over headphones to nineteen listeners from a Midwestern university in the United States. The instructions were to mark with a line underneath the word if some words stood out more than others, if they were louder, longer or higher pitched. Only the results of the *I gave dates to Kay today* were analyzed in this paper.

The number of times listeners heard a word “stand out” is interpreted as the level of perceived prominence of that word.

3. Results

Figure 2 shows the patterns of jaw displacement and prominence ratings for each of the tokens of the *I gave dates to Kay today* utterance types for each of the speakers (4 tokens by A09 and 5 tokens by A10). Red dotted lines here refer to standardized jaw opening values and green solid lines indicate perceived prominence, with green band representing standard deviation of perceived scores. Absolute values of jaw displacement (in mm) was converted to z-score by speaker (subtracting mean and then divided by the SD within speaker). Green shaded area for the perceptual judgements indicates 95% confidence interval for the mean (+/- 1.96 * standard error). Correlation analyses show a significant correlation ($p<0.05$) with jaw and prominence for both speakers.

Notice that for all tokens, maximum jaw displacement occurs on the word, *Kay*, and this is the word that is usually perceived by listeners to be the most prominent (i.e., had the largest number of times the word was underlined). These results suggest that indeed speakers open the mouth more to indicate a word in the sentence which is both intended by the speaker as well as perceived by the listener to have nuclear stress.

The acoustic information occurring during the series of jaw openings is shaped by the characteristics of the voice source passing through the vocal tract filter. Regression analyses (Table 1) show significant relationships between changes in articulation and formants for both speakers. All measurements are made at the point of maximum jaw displacement. Significant relations were found for F1 and F2-F1 with jaw and tongue blade (TB), when the jaw position is subtracted from the tongue blade position.

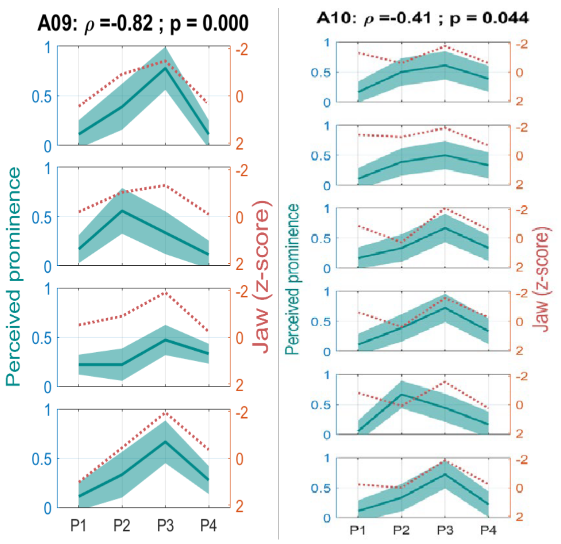

Figure 2: *Patterns of jaw displacement and prominence ratings for each of the tokens of the “I gave dates to Kay today) utterance for speaker A09 (left) and speaker A10 (right). The x-axis indicates each content word in utterance; give (P1), dates (P2), kay (P3), day (P4). Notice that increased jaw displacement indicated by higher z-score.*

Table 1: *Regression analyses (r2 values) with formants and jaw & tongue (Tongue Blade) articulation. R2*

Speaker	F1 x TB-jaw	F2-F1 x TB-Jaw
A09	0.35 (p= 0.01)	0.43 (p=0.00)
A10	0.27 (p=0.01)	0.37 (p=0.00)

Table 2: *Regression analyses (r2 values) with F0, formants, jaw & tongue (Tongue Blade) articulation*

Speaker	F1-F0 x TB-jaw
A09	0.43 (p=0.00)
A10	0.76 (p=0.00)

Significant relations related to F0 and formants and jaw were also found, specifically, between F1-F0 and TB-jaw, although only modestly so for A09 (Table 2). Thus, for the mid-front vowel, we see an interaction among jaw, tongue body, F1, F2, and F0.

4. Discussion

These results suggest that increased jaw displacement results in changes in acoustic information, which the listeners can use to recognize nuclear stress in English. The results can be interpreted to support a view of temporal organization of spoken utterances as a series of syllables in which the speaker opens the mouth more for the important items in the utterance; specifically, in English, the syllable with the largest magnitude in the utterance (in terms of jaw displacement) is perceived as having the nuclear stress.

It is interesting that for the mid front vowel we do not see the strong correlations between F1 and jaw displacement we reported for the low vowel /a/ (e.g., [9]), but we do see significant correlations between formants and jaw-tongue related measures (also see [24]). This can be understood better in terms of the findings of [20]: increased jaw displacement for increased prominence (contrastive emphasis) occurs for high, mid and low front vowels, and also low back vowels. As the jaw lowers more, the tongue moves more in the phonological direction of the vowel. For /a/, it moves down and back, for /æ/, it moves down and forward, for /e/, it moves forward, and for /i/ it moves forward and up. As a result, F1 tends to increase for all emphasized vowels due to jaw lowering, and F2 follows the tongue pattern: F2 is lowest for low vowels and highest for high front vowels [20]. Thus, it makes sense for mid front vowel /e/, that we find the tongue body is involved along with the jaw in producing increased syllable magnitude. Because the tongue body is involved, we see significant correlations of tongue-jaw movement with both F1 and F2. In addition, due to the linking of jaw and tongue with the larynx [27, 28, 29, 30, 31], we also see significant correlations with F1-F0 values.

The focus of this paper is mainly on relating syllable magnitude articulation with perception, not F0, duration and intensity. Auditory impressions of the utterances indicated that *Kay* was always produced as a high stressed syllable (H*), but did not always have the highest F0 in the utterance, as shown in figures 3 and 4. It is interesting to note that although *Kay* was produced with increased jaw displacement (Figure 2), listeners gave more prominence ratings to *dates*. But for both A09 (Figure 3, second token) and A10 (Figure 4, fifth token) *dates* did not have the highest F0; *gave* did. These observations are compatible with the study by [32], reporting that F0 is not the main acoustic cue to signal prominence; also see [33] for a description of the relationship of laryngeal source information and supralaryngeal filter settings within his proposal of prosodic temporal organization of speech. Future work will pursue the various acoustic cues listeners rely on for perceiving a syllable to have nuclear stress.

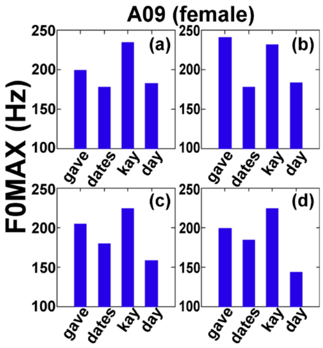

Figure 3: *Patterns of peak F0 for each syllable for each of the tokens of the "I gave dates to Kay today) utterance for speaker A09.*

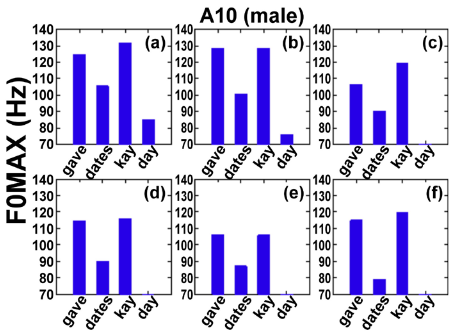

Figure 4: *Patterns of peak F0 for each syllable for each of the tokens of the "I gave dates to Kay today) utterance for speaker A10.*

Also, an important question to explore is why some speakers, e.g., A09, show stronger correlations with jaw and prominence than do others, e.g., A10.

Future work will explore the integration of vocal tract and phonation types for the implementation of prosodic changes. One approach suggested by Osamu Fujimura in the C/D model [3, and also, 33] is that the temporal organization of speech forms a scaffold or base on which the segmental layer and melodic layer are aligned as different step functions. Further, the temporal/rhythmic organization of speech is determined from the articulatory pattern of jaw openings at the realization of the syllable. Thus, there are two components to prosody in this model: rhythm (pattern of jaw displacement) and melody (pattern of intonational F0 contours).

An interesting application of this type of approach to temporal organization of spoken utterances from an articulatory point of view is with second language teaching of prosody. Using jaw tracings to teach Japanese learners of English seems to be an effective way to help them know where the nuclear stress item is and to actually change their jaw and formant patterns appropriately [34].

5. Conclusion

Our results indicated that the word with the highest number of perceived prominences is the word that had the largest amount of jaw displacement, e.g., *Kay*. We interpret these results to indicate that *Kay* was the word with nuclear stress in the utterance. That some of the listeners marked prominence on *gave* suggests that this word also received some degree of prominence but it was not the word with nuclear stress, since it didn't have the largest amount of jaw opening in the utterance. These results are in accordance with previous research with RPT-derived perceptions of prominence with two English speakers, but with a different utterance, "Pam said bat that fat cat at that mat", uttered with 5 different prosodic conditions: neutral, and contrastive emphasis on each of the four words in the second intermediate phrase, BAT, THAT, FAT, CAT [14]. The results showed a significant correlation between jaw displacement and amount of perceived prominence by listeners. In both this study and our previous study with RPT and jaw displacement, we see that in some cases, listeners were more in agreement that in others. This makes sense, since prominence is not codified, neither from the point of view of perception nor production.

With regard to possible segmental influences on jaw displacement, work by [1, 2] shows that combination of vowel height and prominence has the largest, most significant effects on jaw displacement; the extent to which onset and coda affects syllable jaw displacement seem to be minimal but has not been systematically investigated to our knowledge as of yet. Thus, since all syllables contained the same mid-front vowel, it seems reasonable to assume that increased jaw displacement found for "Kay" was significantly influenced by increased syllable prominence.

An additional finding is that articulation of prominence is different for mid front vowel /e/ compared to that for the low vowel /a/. For /a/, increased jaw displacement results in increased F1. For /e/, there is a linking between the jaw and tongue such that increased jaw displacement shows significant correlations for TB-jaw with F1, TB-jaw with F2-F1, and TB-jaw with F1-F0. Given that increased prominence on the mid front vowel requires the tongue blade to move more forward to compensate for the increased jaw lowering, it makes sense to see significant correlations with TB-jaw and the formants. Also, given the interaction between tongue, jaw and larynx, it makes sense to see significant correlations with F1-F0. In order to confirm these findings, future work will include more utterance types, more speakers, and more listeners, as well as acoustic analyses related to changes in both source and filter (e.g., F0, periodicity, spectral tilt, formants, intensity and duration).

Our main finding (more jaw opening, more perceived prominence) was expected, based on ongoing work with jaw displacement and prominence, including work with perceived prominence using RPT. We believe this work is an important addition to elucidate temporal structures of speech from a syllable point of view.

6. Acknowledgements

This work was supported in part by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C) No. 22520412 and (C) #25370444 to the first author.

7. References

[1] C. Menezes, and D. Erickson, D. "Intrinsic variations in jaw deviation in English vowels," Proc. of International Congress of Acoustics. Proceedings of Meetings on Acoustics, Vol. 19, 060253, 2013

[2] J. C. Williams, D. Erickson, Y. Ozaki, A. Suemitsu, N. Minematsu, O. Fujimura, O. "Neutralizing differences in jaw displacement for English vowels", Proc. of International Congress of Acoustics. POMA **19**, 060268, 2013.

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

[4] S.-A. Jun, "Prosodic typology," In S.-A. Jun (ed.) Prosodic typology: The phonology of intonation and phrasing, pp. 430–458, Oxford: Oxford University Press, 2005.

[5] S.-A. Jun, "Prosodic typology: By prominence type, word prosody, and macro-rhythm", In S.-A. Jun (ed.) Prosodic typology II: The phonology of intonation and phrasing, pp. 520– 539, Oxford: Oxford University Press, 2014.

[6] D. Erickson, and S. Kawahara, S. "Articulatory correlates of metrical structure: Studying jaw displacement patterns", Linguistic Vanguard 2, pp.102-110, De Gruyter Mouton. DOI 10.1515/lingvan-2015-0025, 2016.

[7] C. L. Smith, D. Erickson, and C. Savariaux. "Articulatory and acoustic correlates of prominence in French: Comparing L1 and L2 speakers", Journal of Phonetics Special Issue. Vol. 77, https://doi.org/10.1016/j.wocn.2019.100938, 2019.

[8] D. Erickson, "On phrasal organization and jaw opening," Proc. From Sound to Sense; in honor of Ken Stevens, MIT, CDRom, 2004.

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

[10] M. E. Beckman, and J. Edwards, "Articulatory evidence for differentiating stress categories." In P. Keating (ed.), Papers in laboratory phonology III, pp.7–33, Cambridge: Cambridge University Press, 1994.

[11] M. Liberman, and A. Prince, "On stress and linguistic rhythm," Linguistic Inquiry, vol. 8, pp. 249–336, 1977.

[12] B. Hayes, "Metrical stress theory: principles and case studies," University of Chicago Press, Chicago, 1995.

[13] E. O. Selkirk, "The syllable", Foris, Dordrecht, 1982.

[14] D. Erickson, J. Kim, S. Kawahara, I. Wilson, C. Menezes, A. Suemitsu, J. Moore, "Bridging articulation and perception: The C/D model and contrastive emphasis," International Congress Phonetic Sciences, 2015.

[15] J. Harrington, J. Fletcher, M.E. Beckman, "Manner and place conflicts in the articulation of Australian English,"; in Broe, Pierrehumbert, Papers in Laboratory Phonology, vol. 5, pp. 40–51, Cambridge University Press, Cambridge, 2000.

[16] D. Erickson, "Effects of contrastive emphasis on jaw opening," Phonetica, vol. 55, pp. 147–169, 1998.

[17] K. de Jong., "The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation", J. acoust. Soc. Am., vol.97, pp. 491–504, 1995.

[18] H. Loevenbruck, "An investigation of articulatory correlates of the accentual phrase in French", Proceedings of 14th International Congress of Phonetic Sciences, pp. 667–670, 1999.

[19] K. Maekawa, S. Kiritani, H. Imagawa & K. Honda, "Effects of focus on mandible movement," Journal of the Acoustical Society of America, vol. 54, pp. 259–260, 1998.

[20] D. Erickson, "Articulation of extreme formant patterns for emphasized vowels," Phonetica, vol. 59, pp.134–149, 2002.

[21] C. Menezes, "Rhythmic pattern of American English: an articulatory and acoustic study," PhD diss. Ohio State University, 2003.

[22] I. Wilson, D. Erickson, N. Horiguchi, "Articulating rhythm in L1 and L2 English: focus on jaw and F0," Acoust. Soc. Japan, Fall Meeting, 2012.

[23] J. Cole, J. I. Hualde, C. L. Smith, C. Eager, T. Mahrt, R. N., de Souza, R.N., "Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish," Journal of Phonetics, vol. 75, pp. 113–147, 2019

[24] T. Huang and D. Erickson, "Articulation of English 'prominence' by L1 (English) and L2 (French) speakers," In Canberra, Australia, S. Calhoun, P. Escudero, M. Tabain, and P. Warren (eds.), Proc. of the 19th International Congress of Phonetic Sciences, 2019: Australasian Speech Science and Technology Association Inc.

[25] P. Boersma, and D. Weenink, "Praat: doing phonetics by computer", 1992-2020.

[26] M. Tiede, "MVIEW: Multi-channel visualization application for displaying dynamic sensor movements", 2010.

[27] D, Erickson, K. Honda, and S. Kawahara, "Interaction of jaw displacement and F0 peak in syllables produced with contrastive emphasis," Acoust.Sci.&Tech., vol. 38.3, pp. 137-146, 2017.

[28] W-R, Chen, D. Whalen, M. K. Tiede, "Mandibular contribution to vowel-intrinsic F0", Proc ICPHS, pp. 152-156, 2019.

[29] K. Honda, "Laryngeal and extra-laryngeal mechanisms of F0 control," In: Bell-Berti, Raphael (eds), Producing speech: contemporary issues for Katherine Safford Harris, New York: American Institute of Physics, pp. 215-232, 1995.

[30] D. H. Whalen, A. G. Levitt, "The universality of intrinsic F0 of vowels," J. Phonetics, vol. 23, pp. 349-366, 1995.

[31] M. Rossi, and D. Autesserre, "Movements of the hyoid and the larynx and the intrinsic frequency of vowels," J. Phonetics, vol. 9, pp. 233-249, 1981.

[32] G. Kochanski, E. Grabe, J. Coleman and B. Rosner, "Loudness predicts prominence: fundamental frequency lends little," Journal of the Acoustical Society of America, vol. 118.2, pp. 1038–1054, doi:10.1121/1.1923349, 2005.

[33] O. Fujimura and D. Erickson, "The C/D Model for prosodic representation of expressive speech in English," *Acoustical Society of Japan*, Fall Meeting, Okinawa, pp. 271-2, 2004.

[34] I. Wilson, D. Erickson, T. Vance, and J. Moore, "Jaw dancing American style: A way to teach English rhythm," Speech Prosody, 2020.