Metrical Structure and Jaw Displacement: An Exploration

Donna Erickson [1], *Shigeto Kawahara* [2], *J.C. Williams* [3], *Jeff Moore* [4], *Atsuo Suemitsu* [5], *Yoshiho Shibuya* [1]

[1] Kanazawa Medical University, Japan; [2] Keio Institute of Language & Cultural Studies, Tokyo;
[3] Independent, Kamakura, Japan; [4] Sophia University, Tokyo;
[5] Japan Advanced Institute of Science & Technology, Ishikawa Pref., Japan

EricksonDonna2000@gmail.com, kawahara@icl.keio.ac.jp, catitawms@gmail.com,
jeffmoore.personal@gmail.com, sue@jaist.ac.jp, yshibuya@kanazawa-med.ac.jp

Abstract

Building on Erickson et al. [9], the current electromagnetic-articulography (EMA) experiment proposes that the amount of jaw displacement—or mandible movement—may reflect the metrical organization of English sentences. The experiment also supports F1 as a reliable acoustic correlate of jaw displacement, hence metrical organization. On the other hand, the study also demonstrates that F0 does not have a similar relationship to mandible movement.

Index Terms: metrical organization, jaw movement, EMA, F1, F0

1. Introduction

Work in Metrical Phonology [1] (*et seq.*) posits that our speech exhibits patterns of organization that are akin to the metrical organization of poetry [2], [3]. Metrical structure consists of rhythmic categories, within which strong-weak patterns are assigned. Different theories propose different levels of rhythmic categories, but they generally include such units as syllable, (prosodic) word, minor phrase, major phrase (and utterance), as in Figure 2 [4], [5], [6:384]. Each syllable, based on its metrical constituency, is assigned a level of stress or prominence relative to other syllables in the utterance [4], [5], and overall prominence may be represented as a number of grid marks [7]. The acoustic prominence of each unit is often described in terms such as quality, duration, loudness, and pitch [8].

Recent X-ray Microbeam and electromagnetic-articulography (EMA) studies, inspired by Fujimura [10]-[12], with four American English speakers [9], suggest that an articulatory correlate of syllable stress in English is the amount of jaw (mandible) displacement. As a speaker opens and closes the mouth to produce each syllable, the jaw moves. It opens more for low vowels than high vowels *ceteris paribus* [13], [14]; if all the vowels in an utterance are the same, the pattern of jaw opening should mirror syllable stress levels and correspond to the metrical structure of the speaker's utterance. The greater the jaw displacement, the higher the level of syllable stress. Also, the first resonant formant of the vowel (F1) exhibits a significant correlation with jaw displacement and metrical structure, suggesting that F1 is an important acoustic correlate of English metrical structure. Figures 1 and 2, based on [9], illustrate the pattern of jaw displacement for three American English speakers uttering: "[Yes, I saw] five bright highlights [in the] sky [to]night". The greater the jaw displacement, the greater the syllable stress (syllable "magnitude" according to [10]), and this magnitude is displayed vertically in bars, making it easier to visualize how jaw displacement mirrors the metrical structure of the utterance, shown in Figure 2. F1 values are not reproduced here, but they present essentially the same pattern as jaw displacement [9]. For these three speakers, nuclear stress—the most prominent stress in a sentence [15]—is on "high" of "highlights". For the fourth speaker (not shown here), nuclear stress is on "five"; this difference of nuclear stress placement is also observed in our current experiment, as discussed in the Results section.

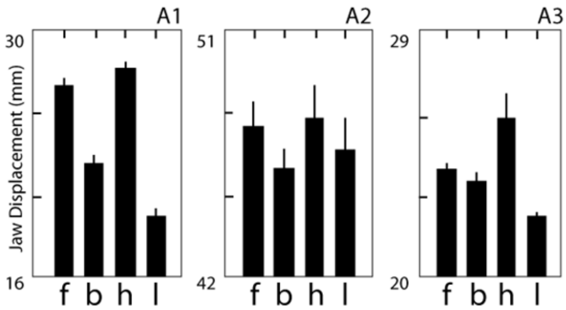

Figure 1: *Bar graphs showing the amount of jaw displacement (mm) for 3 American English speakers, for the sentence "[Yes, I saw] 5 bright highlights [in the sky tonight]." Error bars indicate standard error of the mean. Ordinate scaling is by speaker, to better display individual jaw displacement patterns.*

Major Phrase			x	
Minor Phrase	x		x	
Word	x	x	x	
Syllable	x	x	x	x
Stress level	3	2	4	1
	five	bright	high	lights

Figure 2: *Metrical structure of the utterance in Figure 1.*

One question addressed in this paper is how robust is the correlation between metrical prominence and jaw opening and between F1 and maximum jaw displacement. Additionally, do these patterns generalize beyond the single sentence tested in [9]? Is metrical prominence the only factor that affects jaw movement patterns? Does this correlation generalize to other vowels or do specific consonantal or vocalic gestures affect jaw displacement patterns? In addition, we examined F0 in

10.21437/SpeechProsody.2014-47

order to see what role it plays in implementing metrical stress patterns in American English.

2. Method

To address these questions, we report on jaw displacement patterns and corresponding acoustic measurements of data recorded by 3D-EMA (Carstens AG500 Electromagnetic Articulograph) at the Japan Advanced Institute for Science and Technology (JAIST, Ishikawa Prefecture, Japan) for three English sentences, as produced by two speakers of American English, one male (Speaker 1, fourth author), and one female, (Speaker 2, third author, an English-Spanish bilingual). Custom software (mview, Haskins Laboratories) was used to analyze the articulatory data. The lowest vertical position (maximum displacement) of the jaw with respect to the bite plane was located for each syllable of each utterance using a velocity-based criterion. For a more detailed description of EMA, including placement of sensors and recording procedures, see [9]. The sentences were all presumed to have the same metrical pattern; in the first and second sentences, the vowel of interest is /a/ and in the third, /e/, all followed by the palatal glide. The sentences are shown below, with the target words in the midsentence underlined.

(1) [Yes, I saw] five bright highlights [in the sky tonight].

(2) [Yes, I saw] nine nice bike fights [on the dyke tonight].

(3) [Yes, I saw] eight great playmates [in the bay today].

Sentence (1) is the same as reported in [9]; Sentence (2) contains the same diphthong /aɹ/ but different words, and Sentence (3) contains the diphthong /eɹ/. The participants practiced the sentences with pictures illustrating a scenario to match each sentence. For data recording, each sentence was presented (without parentheses or underlining) six times in random order using a PowerPoint display. The underlined syllables in each sentence were analyzed for jaw displacement, F1, and peak F0. Acoustic measurements were made using Praat [16]. Maximum F0 was measured in Hz for the steady state portion of the vowel, before the transition to the glide, and F1 was measured at that same point in time. F1 measurements in [9] were taken at the time of maximum jaw displacement. However, sometimes maximum jaw displacement occurred before the onset of voicing, so that F1 was not measurable. Measuring F1 at the time of maximum F0 has its own drawbacks, in that physiologically, lowering jaw would result in low F0 since the larynx gets lowered as well. However, this measuring method is at least consistent and obvious across all the tokens recorded.

3. Results and Discussion

Figure 3 shows a sample jaw tracing of Sentence (1) "[Yes, I saw] five bright highlights [in the sky tonight]." for the male speaker. Notice in the bottom panel of Figure 3 that the jaw opens and closes for each of these four syllables, and even though the vowel plus glide is always /aɹ/, the amount of jaw opening varies systematically.

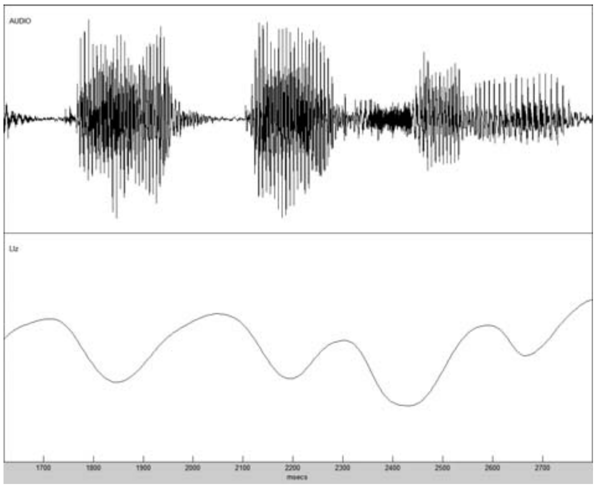

Figure 3: *Acoustic waveform (top) and jaw displacement tracing in mm (bottom) for Speaker 1 saying: "[Yes, I saw] 5 bright highlights [in the sky tonight.*]"

The patterns of jaw displacement averaged over 6 repetitions for Speakers 1 and 2 are displayed in Figure 4 such that the taller the bar, the larger the jaw displacement/syllable stress. Speaker 2 (at right in Figure 4) shows patterns similar to those reported for three of the four speakers in [9] (see Figure 1), supporting our hypothesis that metrical structure is reflected in jaw displacement patterns.

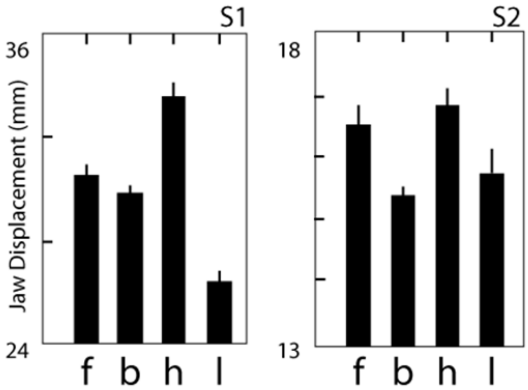

Figure 4: *Speaker 1 (L) & Speaker 2 (R) "[Yes, I saw] 5 bright highlights [in the sky tonight]" (Sentence 1).*

Figure 5 shows jaw displacement measures for Sentence (2) "[Yes, I saw] nine nice bike fights [on the dyke tonight]." Both speakers produce patterns similar to those for sentence (1), except that "nine" (the first syllable of this phrase) receives the nuclear stress. The pattern of jaw displacement for these speakers, especially for Speaker 1, is similar to that reported for the fourth speaker in [9] for "[Yes, I saw] five bright highlights [in the sky tonight]", putting nuclear stress on "five".

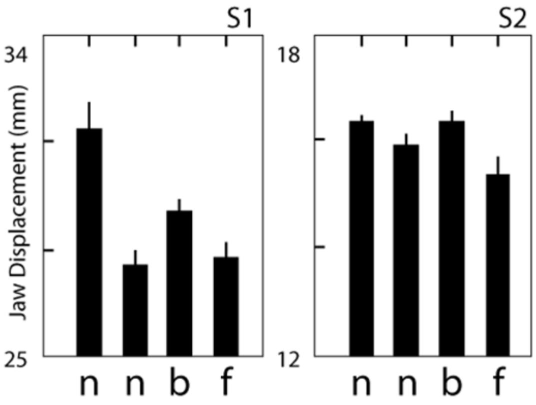

Figure 5: *Speaker 1 (L) & Speaker 2 (R) "[Yes, I saw] 9 nice bike fights [on the dyke tonight.]" (Sentence 2).*

Major Phrase	x			
Minor Phrase	x		x	
Word	x	x	x	x
Syllable	x	x	x	x
Stress level	4	2	3	2
(Yes, I saw)	nine	nice	bike	fights

Figure 6: Metrical structure of Sentence 2.

It is interesting that each speaker seems to show the jaw displacement pattern previously reported for Sentence 1, even though the words in the sentence have changed; i.e., the way they implement the metrical patterns of utterances is consistent across different sentences.

Figure 7 shows jaw displacement for a sentence with seemingly the same stress patterns, but the diphthong is /eɪ/: "[Yes, I saw] <u>eight</u> <u>great</u> <u>playmates</u> [in the bay today.]"

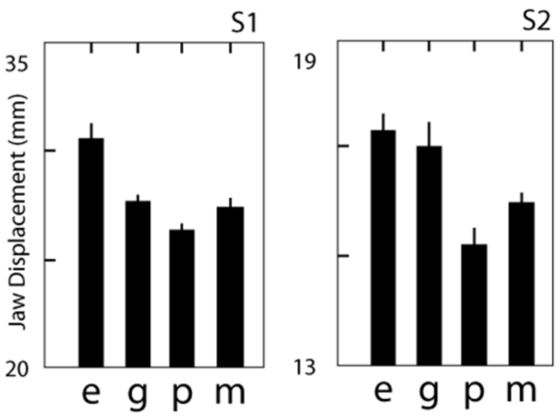

Figure 7: *Speaker 1 (L) & Speaker 2 (R). "[Yes, I saw] 8 great playmates [in the bay today]." (Sentence 3).*

Speakers 1 and 2 both show stronger jaw displacement for "eight" compared with "great", but "mates" is stronger than "play". Thus, the first phrase, "eight great", is similar to the first phrase of "nine nice bike fights" but the second phrase, "play mates", is different. It appears that the speakers treated "play mates" as a two-word phrase rather than a compound; i.e., as in the classic "<u>blaˈck</u>boaˌrd" vs. "blaˌck <u>boaˈrd</u>" minimal pair, where the second element in a phrasal compound receives more prominence. Thus, this sentence may have a different metrical structure than the previous two sentences.

Major Phrase	x			
Minor Phrase	x			x
Word	x	x	x	x
Syllable	x	x	x	x
Stress level	4	2	2	3
(Yes, I saw)	eight	great	play	mates

Figure 8: *Metrical structure for Sentence 3.*

What are the acoustic consequences of a speaker using the jaw to articulate the metrical structure of an utterance? Figure 9 below is a scatter plot of jaw displacement and F1 for each target word in the three sentences, produced by Speaker 1, grouped by the hypothesized stress levels.

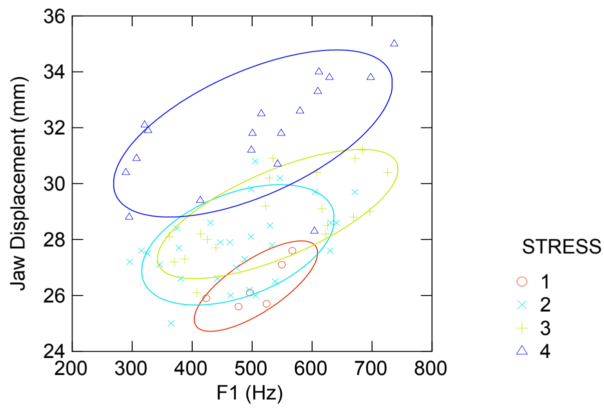

Figure 9: *Jaw Displacement and F1 (Speaker 1) as a function of stress level.*

Note that for Speaker 1, the greater the syllable stress, the larger the jaw displacement, and within each stress group, there is a significant correlation between jaw displacement and F1 for the three highest stress levels; the correlation is also high for the lowest stress group, but did not reach significance due to small N (See Table 1 for statistical results).

Table 1. *Pearson Correlation Analysis of Jaw Displacement with F1, within stress levels.*

stress level	*r*	Bartlett Chi-square statistic	df	*p*	N
1	0.737	2.743	1	0.098	6
2	0.433	5.512	1	0.019	29
3	0.694	10.823	1	0.001	19
4	0.600	6.923	1	0.009	18

Figure 10 shows a scatter plot of jaw displacement and peak F0 for each target word in the three sentences, produced by Speaker 1, grouped by the hypothesized stress levels.

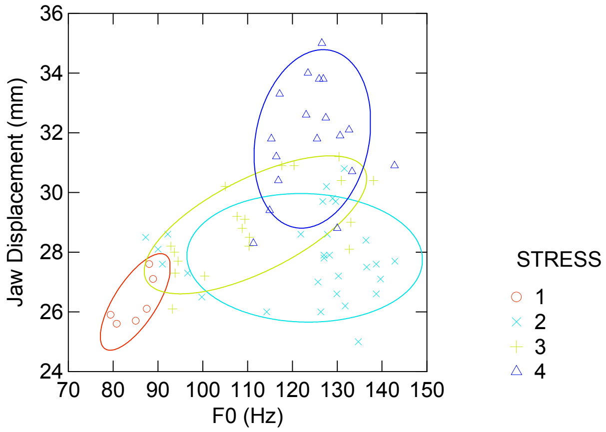

Figure 10: *Jaw Displacement and F0 (Speaker 1) as a function of stress level.*

The scatter plot does not show a consistent distribution of F0 according to the stress groups, thus suggesting that F0 is not a good indicator of stress. Stress level 2 shows a bimodal distribution, which indicates that metrical prominence cannot be the only factor that determines F0. For instance, sentence-final lowering may be responsible for the left cluster of stress level 2 [19]. In addition, F0 overlaps with a cluster of stress levels 2 and 4; and those of stress level 3 are a subset of stress level 2. From our data, it appears that F0 is not a reliable acoustic measure of metrical prominence.

For Speaker 2, we see that patterns of jaw displacement by F1 and jaw displacement by F0, when grouped by stress level, are somewhat similar to those of Speaker 1. Speaker 2's grouped stress data for jaw displacement by F1 show that the greater the stress, the larger the amount of jaw displacement (except for stress level 1); however, there is no correlation between jaw displacement and F1 within each level, as we saw for Speaker 1. Additionally, Speaker 2 does not exhibit any consistent distribution of jaw displacement by F0 according to stress groups, failing to confirm F0 as a reliable measure of metrical prominence for English sentences. The data scatter for Speaker 2 appears somewhat messy and not as clear as that seen for Speaker 1, which may reflect Speaker 2's bilingualism. The effect of bilingualism on the articulation of metrical stress is an interesting area for future exploration.

4. Conclusions

This is a preliminary report on an initial research study to investigate articulatory correlates of metrical structure in American English. Clearly, more speakers are needed, as well as more sentences. Further research, including constructing syllable triangles and boundaries according to the C/D model [10], [11], [20] will provide additional information on syllable magnitudes and differences in boundary strengths. Given that rhythm is often perceived as a pattern of "beats", the "beats" of jaw opening/closing may well be related to the rhythm/metrical structure of spoken language. Obviously, the listener does not hear jaw movement, but the resulting changes in formant frequencies, particularly in F1, may cue the listener to these "jaw beats". In our analysis of the data, peak F0 patterns fail to distinguish metrical stress levels. The results of this exploratory articulatory study reinforce the observation that metrical structure may best be reflected in jaw displacement patterns [9].

In closing, we address an issue that was raised by a reviewer -- that both of our subjects are authors. Trained speakers, including phoneticians and professional speakers, are probably able to consciously control the "rhythm" (metrical organization) of their utterances; however, it is unlikely that even these speakers can manipulate their jaw displacement patterns, since these are probably "hard-wired" from infancy [21]. Research is needed to explore this issue.

This process of deriving speech organization from mandible movement raises many questions for future investigation and provides a possible method for exploring metrical structure of language in general as well as specific languages, as demonstrated here for American English. It is hoped that this approach may offer a means to mesh the phonological abstractness of language with the phonetic instantiation of speech, resulting in new views and insights.

5. Acknowledgements

The impetus for this approach to the phonology-phonetics interface of language and its metrical structure comes from the insights of Osamu Fujimura and his life-long research in articulation, especially his theoretical framework of the C/D Model.

This work was supported by the Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (C)#22520412 and (C)#25370444. Special acknowledgement is made to Mark Tiede for help with mview, and thanks to Jianwu Dang for the use of the EMA Lab at JAIST.

6. References

[1] Liberman, M. and Prince, A., "On stress and linguistic rhythm", Linguistic Inquiry, 8(2):249-336, 1977.
[2] Kiparsky, P., "The rhythmic structure of English verse", Linguistic Inquiry, 8(2):189-247, 1977.
[3] Hayes, B., "A grid-based theory of English meter", Linguistic Inquiry, 14:357-393, 1983.
[4] Hayes, B., Metrical Stress Theory: Principles and Case Studies, University of Chicago Press, 1995.
[5] Selkirk, E., Phonology and Syntax: The Relation between Sound and Structure, MIT Press, 1984.
[6] Selkirk, E., "On derived domains in sentence phonology", in Phonology Yearbook 3, 371-405, Cambridge U. Press, 1986.
[7] Prince, A., "Relating to the grid", Linguistic Inquiry, 14(1):19-100, 1983.
[8] Laver, J., Principles of Phonetics, Cambridge U. Press, 1994.
[9] Erickson, D., Suemitsu, A., Shibuya, Y., and Tiede, M., "Metrical structure and production of English rhythm", Phonetica, 69:180-190, 2012.
[10] Fujimura, O., "The C/D model and prosodic control of articulatory behavior", Phonetica 57:128-38, 2000.
[11] Fujimura, O., "Stress and tone revisited: Skeletal vs. melodic and lexical vs. phrasal", in S. Kaji [Ed], Cross-Linguistic Studies of Tonal Phenomena, 221-234, Tokyo University of Foreign Studies, 2003.
[12] Bonaventura, P. and Fujimura, O., "Articulatory movements and phrase boundaries", in M.-J. Solé, P. S. Beddor, and M. Ohala [Eds], Experimental Approaches to Phonology, 209-227, Oxford University Press, 2007.
[13] Williams, J. C., Erickson, D., Ozaki, Y., Suemitsu, A. Minematsu, N., and Fujimura, O., "Neutralizing differences in jaw displacement for English vowels", Proc. of International Congress of Acoustics, POMA 19:060268, 2013.
[14] Menezes, C. and Erickson, D., "Intrinsic variations in jaw deviation in English vowels", Proc. of International Congress of Acoustics, POMA 19:060253, 2013.
[15] Chomsky, N. and Halle, M., Sound Pattern of English, MIT Press, 1968.
[16] Boersma, P. and Weenink, D., "Praat: Doing phonetics by computer. [Computer Software]", Department of Language and Literature, University of Amsterdam. Online: http://www.praat.org/, accessed on 15 Dec 2013.
[17] Erickson, D. and Honda, K., "Jaw displacement and F0 in contrastive emphasis", J. Acoust. Soc. Am., 99, 2494 (A), 1996.
[18] Edwards, J., Beckman, M, and Fletcher, J., "The articulatory kinematics of final lengthening", J. Acoust. Soc. Am., 89:369-382, 1991.
[19] Lieberman, P., Intonation, Perception, and Language, MIT Research Monograph 38, MIT Press, 1967.
[20] Erickson, D., Kawahara, S., Moore, J., Menezes, C., Suemitsu, A., Kim, J., Shibuya, Y., "Using the C/D Model to calculate articulatory syllable duration and prosodic boundaries", International Symposium on Speech Processing, Cologne, Germany, 2014.
[21] Riordan, J., "Anatomy and physiology of lactation", in J. Riordan and K. Wambach [Eds], Breastfeeding and Human Lactation, 79-116, Jones and Bartlett Publishers, 2010.