An articulatory account of rhythm, prominence, and phrasal organization

Donna Erickson

Showa Music University, Kawasaki City, Japan,
EricksonDonna2000@gmail.com

Abstract

This paper examines some articulatory and acoustic characteristics of American English. The results suggest that the jaw may be the articulatory organizer of phrasal rhythm, manifested acoustically through the F2-F1 pattern. Utterance prominence, such as contrastive emphasis, is additionally manifested by increased F0 along with increased duration on the prominent word. The rhythmical organization of the utterance, based on strong-weak jaw opening patterns, may be different from the intonational organization involving pitch accents/ boundary strengths. American English prosody might be best described using a parallel system involving both a rhythm system based on articulation, and an intonational system involving pitch notations.

1. Introduction

Prosody affects articulation, e.g., [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16]. The amount of jaw articulation (excursion) is related to prosodic changes in American English, both in terms of lexical stress and phrasal prominence (contrastive emphasis, e.g., [6,8]) and syllable, word, phrase initial position, e.g., [2,5,12]. It has been shown that for virtually all American English vowels, the jaw opens more with greater phrasal prominence—for low vowels (e.g., [4,6,8]), high vowels [7,8,9] and mid vowels [9].

A complicating factor is that in addition, the height of the vowel also affects the amount of jaw opening: high vowels involve a less open jaw and low vowels, more (e.g., [9]), and it is difficult to separate the vowel articulation component from the prominence articulation component.

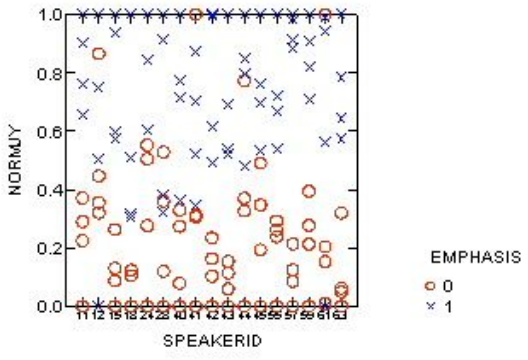

Figure 1. Normalized jaw positions shown on y-axis for 34 /i/ vowels (17 male speakers) in the word “these”, the o-symbol indicates unemphasized /i/ in the utterance “Put these two BACK”, the x symbol indicates emphasized /i/ in the utterance, “Put THESE two back.

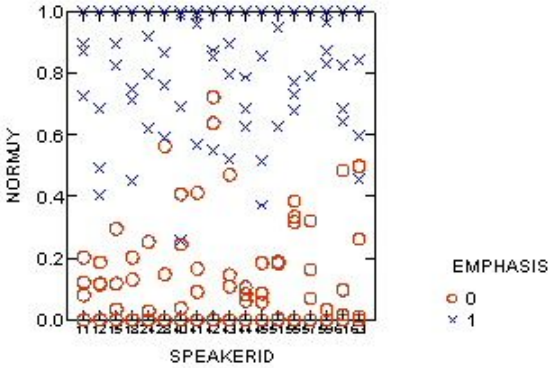

Figure 2. Normalized jaw position shown on y-axis for 36 /ae/ vowels (18 male speakers) in the word “back”, o indicates unemphasized / ae /, “Put these two back”, x indicates emphasized / ae / in the utterance, “Put these two BACK”

One solution proposed by [10] is to “normalize” the amount of jaw opening for vowel height so that the prominence component is separate from the vowel height component. For each vowel (in a set of emphasized & unemphasized vowels), the amount of jaw opening was rated between 0 to 1,with 0 being the least open, 1 being the most open, and in between jaw openings given values accordingly, such that if there were 8 instances of /i/ (emphasized and unemphasized /i/), there would be 8 normalized jaw positions. This was done for 333 /i/ vowels and 334 /ae/ vowels for 44 speakers, using the XRMB database, courtesy of John Westbury and the Waisman Center, U. Wisc. Figures 1 and 2 show the normalized jaw-y values for the male speakers.

Table 1. Averaged normalized jaw opening values for speakers for emphasized and unemphasized vowels

Vowel	Non-emphasis	Emphasis
i	0.25	0.72
ae	0.16	0.78

Table 1 shows that the normalized jaw opening values for high vowel /i/ and low vowel /ae/ are comparable. In this way, theoretically, it is possible to normalize for jaw opening across different vowel heights, so that jaw opening for articulating vowel quality and for articulating phrasal prominence can be separated. Also, it will be necessary to factor in effects of position in sentence, in order to account for such phenomena as phrase-elongation effects, etc. To date, this type of normalization has not been done, but needs to be.

Tongue dorsum position also changes as a function of phrasal prominence [8, 9]. According to these studies, for emphasized low vowels, the tongue dorsum moves more back

and down. For emphasized high vowels, it moves more up and forward. For emphasized low front vowels (/ae/), it moves more forward. For emphasized low vowels, the tongue dorsum moves more back and down. Thus, as the jaw lowers more for increased prominence, the tongue dorsum moves more in the direction of the phonological specification of the vowel.

The acoustic consequences of joint jaw-tongue dorsum movement are that formant frequencies change. Low vowels become more compact (the distance between F2 and F1 decreases) and high vowels become more diffuse (the distance between F2 and F1 increases) [8,9].

The articulation of American English rhythm is connected to the pattern of jaw movement: Jaw control (not necessarily the resultant position) sets the framework for the prominence characteristics of the syllable. The tongue gesture is controlled appropriately in order to produce the desired vowel. The primary acoustic percept of this is the F2-F1 pattern.

The working hypothesis is that the rhythm, prominence and phrasing patterns of the utterance are reflected in the pattern of jaw opening, acoustically manifested in its F2-F1 patterns. Utterance prominence, in addition, is manifested by increased F0 along with increased duration on the prominent word.

2. Methods

Articulatory and acoustic data were recorded at the University of Wisconsin X-Ray Microbeam Facilities, Madison, Wisconsin [17] from one American English male college student (Midwest dialect, Wisconsin.) Spherical gold pellets (2.4-3 mm in diameter) were glued onto selected points on th tongue, lips, and mandible of the speaker. Two pellets wei attached to the mandible, one at the lower incisors, and one c the buccal surface of a lower molar tooth. For this da analysis, only the y-movement of the pellet attached to th mandible incisor (45 samples/s) was measured. In additio reference pellets were affixed midsagitally to the nose brid and to the center anterior surface of the maxillary incisor These were used to determine the head coordinate system an set the maxillary occlusal plane at each instance durin utterances. The x-axis corresponds to the intersection of th midsagittal plane and the maxillary occlusal plane, with th origin of the x-axis corresponding to the lowermost edge c the maxillary incisor. The y-axis is normal to the maxillar occlusal plane, intersecting the plane at the origin. Ja opening is measured in terms of the lowest vertical position c the mandibular pellet in the syllable from the maxillar occlusal plane. It is referred to as maximum jaw displacemer or simply jaw opening for the syllable.

Measurements of jaw x-y positions and LPC-Cepstrum method formant extraction using a MATLAB-based program were made at the time of maximum jaw opening during the target word. Acoustic durations (measured from articulatory release of initial consonant to that of final consonant) and peak F0 of the syllables were estimated using WaveSurfer (www.speech.kth.se/wavesurfer). Intensity is not reported due to the fact that the microphone placement was not constant throughout the experiment.

The sentences were "No, I saw five bright highlights in the sky tonight" in which the contrastive nuclear accent was assigned to either "five," "bright," or "highlights," and "Yes, I saw five bright highlights in the sky tonight" with no contrastive nuclear accent. Each utterance was repeated 10-12 times. A total of 36 utterances were analyzed.

The analysis in this paper will focus on the how the five words (six syllables) which contain /aJ/ vowels are chunked into rhythmical units, and to what extent this chunking changes as a function of increasing the prominence (contrastive emphasis) on one of the words.

In this paper, I adhere to a rough hierarchical organization involving syllable, word, foot, accentual phrase (containing one pitch accent, and marked with a 2 or 3 BI in the ToBI system[see e.g., [18]), and intermediate phrase (marked with a 3 BI), and an intonational phrase: (marked with a 4BI). The intonational phrase is "'(Yes/No, I saw) five bright highlights in the sky tonight.", and has 3 intermediate phrases "'(yes/no, I saw), "five bright highlights", "in the sky tonight". I will suggest, based on the data analysis, that the accentual phrases vary depending on the location of the emphasized word, but the foot structure remains invariant throughout the utterance "five bright", "highlights" (and "sky (to)night"). I also suggest that the foot structure is a function of the articulation and provides the basic rhythmic structure of the utterance.

3. Results

The results of the acoustic and articulatory measurements are shown in Figure 3a-d. Fig. 3a shows the average amount of jaw opening for each of the target words in the sentence; 3b, the pattern of average difference between F2-F1 for each of the target words; 3c, average acoustic syllable duration measurements for each of the target words; and 3d, average peak F0 values for each of the target words.

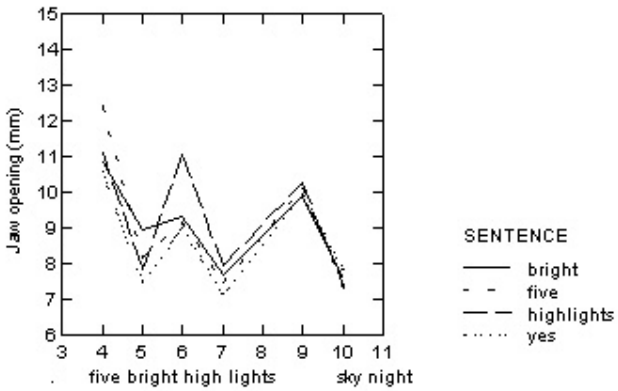

Figure 3a. The x- axis indicates the target word in the sentence (pos.4 is "five;" pos.5 is "bright;" pos.6 is "high;" pos.7 is "lights;" pos.9 is "sky;" pos.10 is "night"). The y-axis indicates the amount of jaw opening for each of the syllables.The legend indicates which target word was emphasized. "yes" indicates the utterances with no contrastive emphasis.

Looking at Fig. 3a, we see a striking pattern of alternating strong-weak jaw opening for each of the target words in the utterance. Strong refers to "more extreme articulations" and weak, to "less extreme ones" [5]. In addition, notice that the first intermediate phrase-initial word "five" has the largest jaw opening, and the next largest jaw opening, is on the next

intermediate phrase-initial (content) word "sky". This is similar to what was previously reported by [2, 5, 12] about phrase initial positions. The exception to this is when the word "highlights" is emphasized, and is discussed shortly Fig. 3a suggests that the pairs of words "five bright" and "highlights" and "sky night" make up a "two-foot bar" each, in which the first member is stronger than the second, and has the larger jaw opening. Thus, this utterance has an alternating pattern of s-w jaw opening, with the largest jaw opening on the first intermediate phrase-initial, then on the next intermediate phrase-initial, and then on the 2-foot bar-initial. The effect of contrastive emphasis is to increase the amount of jaw opening for the word within the framework of its position in this hierarchy, such that an emphasized word in an s-position becomes the strongest (in terms of jaw opening) in the phrase, and an emphasized word in w-position becomes stronger than in the neutral utterance, but not strongest in the phrase (i.e., see the jaw opening for the emphasized word "bright"). Also, when "highlights" is emphasized, the "high" of "highlights" has the largest jaw opening in the utterance. Perhaps this is because there becomes only one accent phrase in the intermediate phrase (as discussed in the section about intonation below), and consequently there is an hierarchical reordering of the strong-weak patterning such that prominence goes to the strong component of the accent phrase. It is important to note that in this analysis, always the 2-foot bar-units, "five bright", "highlights", "sky (to)night" remain intact across the emphasis conditions, i.e., they do not form part of an adjoining bar if the word emphasis changes. Thus, there are two phrases, two 2-foot bars in the first phrase, and one 2-foot bar in the second phrase, which are invariant.

The pattern of F2-F1 (Fig. 3b) is similar to that of the jaw opening pattern, with the s-syllable being more compact (as a measure of increased prominence for low vowels) and the w-syllable, less compact. However, note that the formant pattern does not increase (as it did for the jaw opening) when the emphasized word occurs in the w-position, i.e., "bright". Thus the formant pattern changes as a function of rhythm, but not emphasis.

Peak F0 measurements (Fig. 3d) indicate the largest prominence i.e., contrastively emphasized or nuclear-accented syllables, in the utterance, as do acoustic duration measurements (Fig. 3c) to some extent. But they do not reflect the foot structure/rhythm of the utterance.

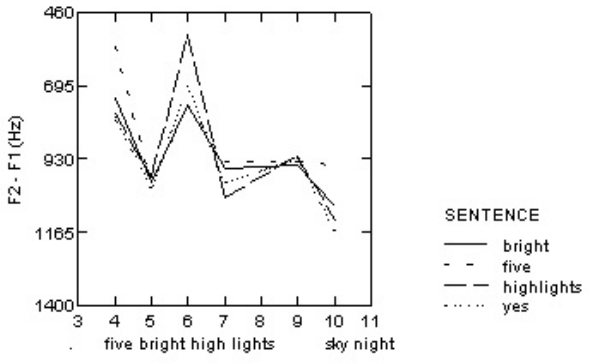

Figure 3b. Same as 3a, except that the y-axis indicates the difference between F2 and F1of each of the syllables

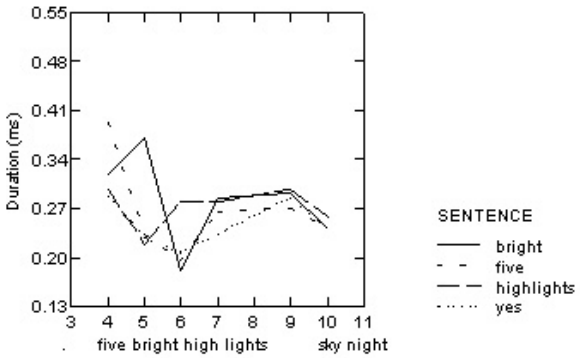

Figure 3c. Same as 3a, except that the y-axis indicates the acoustic syllable duration.

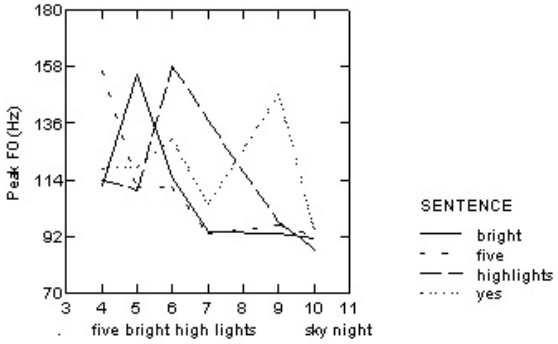

Figure 3d. Same as 3a, except that the y-axis indicates the peak F0 of the syllables.

These results suggest that the jaw may be the articulatory organizer of phrasal rhythm in an American English utterance, manifested acoustically through the F2-F1 pattern.

To summarize thus far, according to the pattern of jaw opening, F2-F1 differences, duration and peak F0, it seems that the *rhythmical structure* of this sentences is implemented by the strong-weak pattern of jaw movement, manifested acoustically by the F2-F1 differences to give a pattern of three 2-syllable feet:

[five bright] [high lights] [sky (to)night]

where the bracketed items indicate a 2-syllable foot. The *prominence* is implemented by increased jaw opening increased F0, and increased acoustic syllable duration. The striking thing is that according to this analysis, the foot structure remains invariant no matter which word is contrastive emphasized.

Now we compare this rhythmical pattern with the phrasing pattern derived from an intonational analysis of the same utterances, using ToBI analysis. In these utterances, the contrastively emphasized word always had L*+H pitch accent, and the nuclear accented syllable had H*. The word in bold indicates the emphasized word. Here, the (minor) accent phrases are indicated by square brackets (and contain a pitch accent), the intermediate phrases, by single slashes, and the

intonational phrases, by double slashes. The subscripts after each word indicate the boundary strength, ranging from 0 to 4, 0 is no boundary and 4 is an intonational phrase boundary.

Table 2. Pattern of accentual phrases according to intonational analysis (e.g., ToBI) for utterances with no contrastive emphasis and with contrastive emphasis

1.Neutral: [five$_1$ bright] $_2$ [high$_0$lights] $_3$ / [sky $_2$tonight]$_4$//.
2.**FIVE**: [**five**]$_{2+}$ [bright$_1$ high$_0$lights]$_2$ /[sky $_2$tonight]$_4$//.
3.**BRIGHT**: [five$_1$ **bright**] $_{2+}$ [high$_0$lights] $_2$ /[sky $_2$tonight]$_4$//.
4.**BRIGHT** [five]$_2$[**bright**]$_3$ [high$_0$lights] $_2$ /[sky $_2$tonight]$_4$//.
5.**HIGH**lights: [five$_1$ bright$_1$ **high**$_0$lights]$_3$ /[H*sky $_2$tonight]$_4$//.

Notice that the number of accent phrases in the first intermediate phrases (here we ignore the first intermediate phrase, "yes/no, I saw") changes as a function of the emphasis condition of the utterance. In terms of the first intermediate phrase, utterance 1(Neutral) has two accent phrases in the first intermediate phrase, and the constituents are [five bright] and [highlights], the same as in the analysis based on articulatory jaw rhythm described above in which there are two 2-foot bar units. Utterance 2 (contrastive emphasis on FIVE), also has two accent phrases but the constituents have changed so that [bright highlights] is in the second accent phrase. Utterances 3 & 4 (contrastive emphasis on BRIGHT) can be spoken with either two or three accent phrases, and utterance 5 (contrastive emphasis on HIGHlights) has only one accent phrase.

The intonational pattern for utterances 2, 3, and 4, in which one word is emphasized is different from the articulatory jaw rhythm pattern, in which there is no change in grouping of 2-foot bars as a function of emphasis. Thus, an interesting observation arising from the data analysis is that intonational phrase patterning and articulatory rhythm patterning only overlap in the utterances with no contrastive emphasis; in utterances with contrastive emphasis, the intonational phrase patterning and the rhythm patterning are different.

4. Summary

This study suggests a hierarchy of organization of jaw opening related to the phrasal organization characteristics of the utterance, which may be different from the intonational pitch accent/ boundary strength organization of the utterance. It may be that American English prosody is best described by invoking a parallel system involving both an intonational system based on pitch notations and an articulatory rhythm system based on hierarchically-defined patterns of jaw opening, manifested acoustically as changes in formant patterns. This data is from only one speaker; however, relevant findings can be found using different databases and multiple speakers. Exploring the inter-relation among formant patterns, F0, duration, jaw opening, syntax, onset-coda characteristics, etc. requires developing a model to assess the recursive assignment of prominence within an utterance. Clearly, more work is needed to explore this further: more speakers, more utterances, more syllable types with varying onset/offset consonants and vowels, and more rhythm patterns, with both mono-syllabic and polysyllabic words.

Acknowledgements

I thank J. Westbury and the Waisman Center for the use of the x-ray microbeam system, J. Dang for the MATLAB-based analysis program; also D. Hirst for his suggestion of a 2-foot bar, and L. Selkirk, M. Liberman, and O. Fujimura for discussions about the paper. This paper is an elaborated version of my previous paper [13],

References

[1] Stone, M. (1981) Evidence for a rhythm pattern in speech production: observations of jaw movement. J. *Phonetics*, *9*, 109-120.

[2] Fujimura, O. (1990) Methods and goals of speech production research. *Language and Speech 33*, 195-258.

[3] Beckman, M.E., and Edwards, J. (1994) Articulatory evidence for differentiating stress categories. In *Papers in Laboratory Phonology,* P. Keating (ed.) Cambridge: Cambridge University Press, pp. 7-33.

[4] Jong, K. de, (1995) The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation. *J. acoust. Soc. Am*., *97*, 491-504.

[5] Fougeron, C. & Keating, P. (1997). Articulatory strengthening at edges of prosodic domains. *J. Acoust. Soc. Am*., 101, 3728-40.

[6] Erickson, D. (1998). Effects of contrastive emphasis on jaw opening. *Phonetica*, *55*, 147-169.

[7] Harrington, J., Fletcher, J., & Beckman, M. (2000) Manner and place conflicts in the articulation of accent in Australian English, *Papers in Laboratory Phonology* (edited by M. Broe and J. Pierrehumbert), Cambridge: Cambridge University Press, 40-51.

[8] Erickson, D. (2002) Articulation of extreme formant patterns for emphasized vowels. *Phonetica, 59*, 134-149.

[9] Erickson, D. (2003) Some effects of prosody on articulation in American English. In *A New Century of Phonology and Phonological Theory, A Festschrift for Prof. Haraguchi*, T. Honma, M. Okazaki, T. Tabata & S. Tanaka (eds.) Tokyo: Kaitakusha, pp. 473-491.

[10] Erickson, D. (2003) The jaw as a prominence articulator in American English, *Acoust Soc Jn, Fall Meeting,* 311-2.

[11] Fujimura, O. & Erickson, D. (2004). The C/D Model for prosodic representation of expressive speech in English. *Acoust Soc Jn, Fall Meeting*, 271-2.

[12] Menezes, C. (2004). Changes in phrasing in semi-spontaneous emotional speech: Articulatory evidences, *Journal of the Phonetic Society of Japan*, 8, pp 45-59.

[13] Erickson, D. (2004) On phrasal organization and jaw opening. *Proceedings of From Sound to Sense, June 13, MIT,*p. 24, CDRom publication.

[14] Cho**,** T. **(**2005**) P**rosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. Journal of the Acoustical Society of America 117(6), 3867-3878.

[15] Byrd, D. & Saltzman, E. (1998). Intragestural dynamics of multiple prosodic boundaries. J. of Phonetics, 26, 173-199.

[16] Cho, T. & Keating, P. (2009). Effects of initial position versus prominence in English. *J. Phonetics* doi:10.1016/j:wcon.2009.08.001

[17] Fujimura, O., Ishida, H., Kiritani, S. (1973) Computer controlled radiography for observation of movements of articulatory and other human organs. *Comp. Biol. Med*., *3*, 371-384.

[18] Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992) TOBI: A standard for labelling English prosody. *Proceedings of ICSLP, Banff, Canada*, 867-870.