2 Jaw-Opening Patterns and Their Correspondence with Syllable Stress Patterns

DONNA ERICKSON, MALIN SVENSSON LUNDMARK, AND TING HUANG

2.1 Introduction

This chapter approaches the topic of language rhythm from an articulatory point of view. It is divided into six sections where we present and discuss data on articulation, acoustics, and perception, with a focus on the jaw. The proposal is that articulatory patterns, specifically how much the jaw opens for each syllable, reflect the abstract hierarchical metrical structure of a spoken language (see, for example, work by Liberman and Prince, 1977; Selkirk, 1982; Hayes, 1995); it is these metrically derived patterns of syllable prominence and phrasing, observed in the articulatory patterns, that provide a basis for language "rhythm." In an English utterance, the metrical hierarchy manifests itself such that each syllable has an n-ary degree of prominence/stress, depending on its position in the hierarchy, with the nuclear stress syllable (or the emphasized/focused syllable) having the largest stress value, then subsequently the phrasally stressed and foot-stressed syllables, respectively, with the reduced syllables having relatively little jaw displacement.

2.1.1 Utterance Prominence and Acoustic Characteristics

In English, there is lexical stress, such that a word with more than one syllable has one syllable that receives more stress than others. The position of the stress is lexically fixed such that with the word *linguistics*, for example, the stressed syllable is the second one. Phrasal stress/prominence, in a sense, is like word stress: that is, with word stress, one syllable in the word receives more stress/prominence than the others; with phrase stress, one word (syllable) within the phrase receives more stress/prominence than the others.[1]

In a simple utterance, such as *I like dates*, the word *dates* probably receives the largest amount of prominence. In English, the syllable/word in an utterance with the largest

[1] Note we use the terms "stress" and "prominence" interchangeably; the term "phrase" refers to a grouping of words in an utterance.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

prominence is referred to as having utterance stress, also referred to as nuclear stress, with the default nuclear stress/utterance stress on the last content word of the utterance (e.g., Cole et al., 2019). In a more complex utterance, such as *I saw five bright highlights in the sky tonight*, there are perhaps three phrases: (*I saw*)(*five bright highlights*)(*in the sky tonight*). The second phrase (*five bright highlights*) consists of two smaller units (*five bright*) and (*highlights*), which we refer to as foot units. We suggest that for this utterance, the first member of each foot is more prominent than the second member (e.g., Erickson et al., 2012, 2014).

In English, speakers generally have choices about which words to group together in a phrase, and which word in that phrase gets the most stress. But the utterance prominence rule seems to be that no matter how the syllables are grouped into phrases, there will be only one syllable in the phrase that will be the most prominent. It is this pattern of prominence/stress that underlies English "rhythm."

A hypothesis is that the abstract metrical hierarchical organization of an English utterance is realized such that each syllable in an utterance has an n-ary degree of prominence/stress; the nuclear stress syllable has the largest stress value, and then subsequently the phrasally stressed and foot-stressed syllables, respectively. The acoustic consequences of stress/prominence in English tend to be increased duration, increased intensity, increased or decreased fundamental frequency (F0) (e.g., H* or L* patterns), and more extreme formants (e.g., Fry, 1955; Lehiste, 1970; Cooper et al., 1985; Beckman, 1986; Turk and Sawusch, 1996; Kochanski et al., 2005).

2.1.2 Utterance Prominence Patterns and Articulation

In terms of articulation, the proposal in this chapter is that the amount of jaw displacement for each syllable (i.e., how much the jaw opens for each syllable) is commensurate to the amount of prominence/stress for that syllable (e.g., Erickson et al., 2012). The hypothesis explored is that the patterns of varying amounts of jaw displacement provide a window into the metrical organization of spoken language; that is, we implement/translate the abstract metrical rhythm of our spoken language in terms of how much we open our jaw for each syllable in an utterance.

The jaw, thus, is a prosodic articulator, in addition, of course, to the larynx. Here we focus on the jaw. This chapter has the following sections: (2.2) A review of findings about jaw displacement and emphasis; (2.3) a review of jaw displacement and utterance prominence patterns; (2.4) the relation between segment articulation and syllable articulation; (2.5) new articulatory, acoustic, and perceptual findings about jaw displacement and how these relate to utterance prominence and phrase boundary patterns in American English (AE) utterances; and (2.6) jaw and phrasal stress in other languages and applications for language teaching, along with plans for future research.

2.2 Jaw Displacement and Emphasis

Jaw displacement increases with prominence, including contrastive emphasis, not only for low vowels (e.g., Kent and Netsell, 1971; Stone, 1981; Summers, 1987; Macchi, 1988; Westbury and Fujimura, 1989; Beckman and Edwards, 1994; de Jong, 1995; Erickson, 1998, 2002, 2004;

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Harrington et al., 2000; Menezes, 2003, 2004) but also for high vowels (Harrington et al., 2000; Erickson, 2002) and mid vowels (Erickson, 2002). Increased jaw displacement with increased prominence has also been reported for French (Loevenbruck, 1999; Tabain, 2003) and Japanese (Erickson et al. 2000).

An acoustic consequence of increased jaw displacement for emphasis is formant (F) changes. Jaw lowering changes the size and shape of the vocal tract; in order to produce the same phonological vowel with a larger jaw opening, the tongue must move accordingly. Erickson (2002) reported that when a vowel is emphasized, the jaw lowers and the tongue of necessity also changes position – more up and forward for high vowels, more low and back for low vowels. For emphasized high vowels, F2 becomes higher while F1 tends to lower; for emphasized low vowels, F2 lowers while F1 raises, resulting in the emphasized vowels positioned at the more extreme edges of the vowel triangle. Findings of increased F1 with increased jaw displacement for low vowels have also been reported by, for example, Menezes (2003).

Here we report on a pilot articulatory and perceptual study of emphasis in English (Erickson et al., 2015). Jaw displacement patterns were examined for two North American English speakers (one male, one female) for the utterance *Pam said bat that fat cat at that mat*, spoken in five different emphasis conditions: emphasis on *bat*, *that*, *fat*, *cat*, and no emphasis. Note that vowel quality also affects jaw displacement (i.e., a low vowel has about 4 mm lower jaw displacement than a high vowel) (Menezes and Erickson, 2013; Williams et al., 2013); thus, all vowels in the utterance must be phonologically the same in order to see the effects of utterance prominence. For tracking the jaw movement, we used electromagnetic articulography (EMA), where a sensor was glued to the middle of the two lower incisor teeth just above the gumline. MView (Tiede, 2010) was used to measure the point in each vowel where the jaw was at its maximally lowest position from the bite plane.

Figure 2.1 shows the average amount of jaw displacement for each of the syllables in the five different emphasis conditions. The height of each bar represents the jaw displacement (mm) for each syllable. Notice that the emphasized word (underlined in capital letters) always has the largest jaw displacement in the utterance. As reported in Kim et al. (2015), t-tests with Bonferroni post hoc tests showed that the difference was significant for *that* and *fat* for both speakers, but not for *bat* for A05, while neither of the two speakers showed a significant difference for *cat*.

The bottom panel of Figure 2.1 shows the jaw displacement pattern for the utterance type spoken without emphasis on any word. Notice that in the no-emphasis utterance, there is still one word, *cat*, which shows the largest jaw displacement in the utterance, suggesting that nuclear stress is on *cat*. Notice also the strong–weak pattern of jaw displacement for the first two pairs of words (*Pam said*, *bat that*) in the utterance; if the second member is emphasized, we then see a weak–strong pattern. But, for the third pair of words, *fat cat*, unless fat is emphasized, we see a weak–strong pattern of jaw displacement such that *cat* has more jaw displacement than *fat*. We suggest that this is because for this speaker, the nuclear stress is on *cat*. These jaw displacement patterns suggest that in some inherent way, jaw articulation patterns form a framework for English rhythmic prominence patterns.

What about perception? Do jaw displacement patterns affect listeners' perception of prominence? In order to test the hypothesis that the patterns of jaw displacement

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

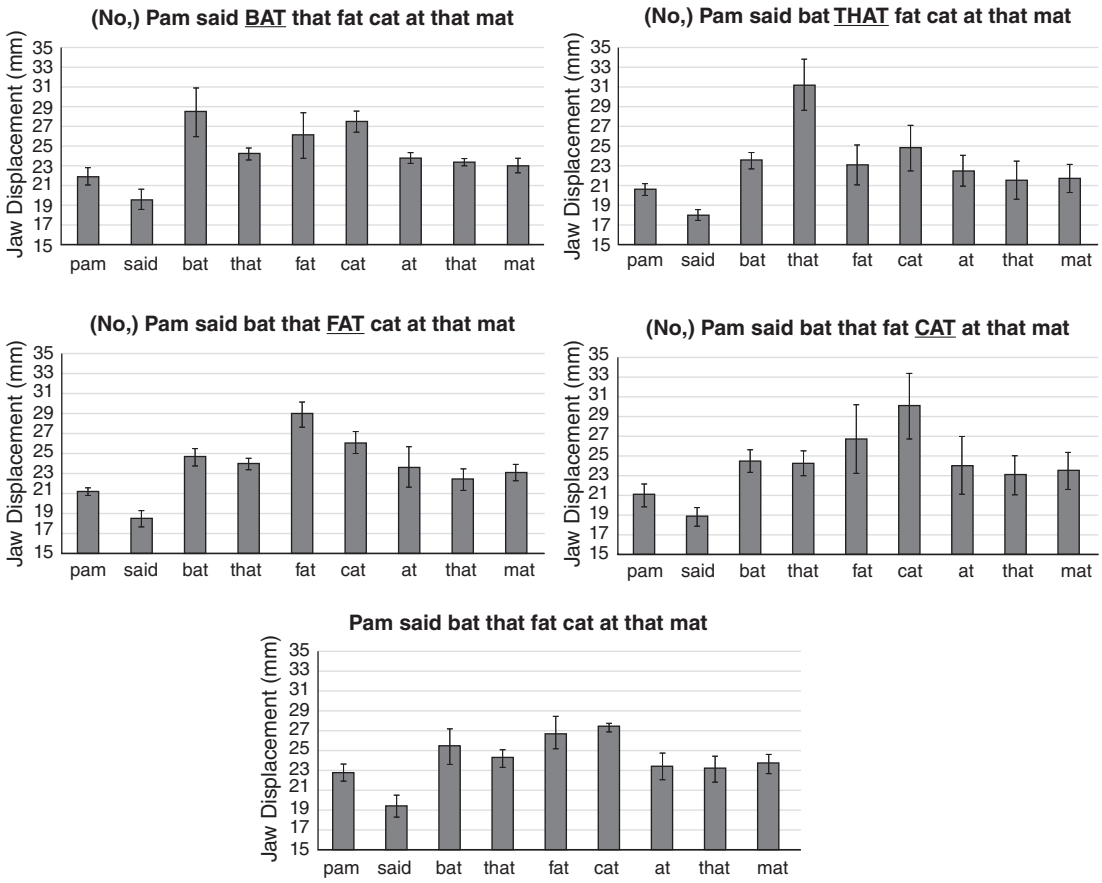


Figure 2.1 Jaw displacement values for different emphasis conditions.
Jaw displacement values (mm) for each syllable in the utterance *Pam said bat that fat cat at that mat*, spoken in five utterance conditions by a single speaker. The emphasized words, from top to bottom, are BAT, THAT, FAT, CAT, and no emphasis.
Erickson et al. (2015)

"match/reflect" the prominence patterns of an utterance, an online rapid prosodic transcription (RPT) listening test was done (Cole et al., 2019). A total of 50 listeners (18 for the 26 tokens of A03, and 32 for the 24 tokens of A05) were asked to evaluate each token twice: first to mark with a vertical line between each word where they heard phrase breaks, and second to underline which word or words "stood out" more than the others, that is, which word seemed louder, longer, or higher pitched. Listeners could listen to each token as many times as they wished. Results of the RPT perception test indicated a significant relation ($p <. 001$) between jaw displacement and perceived prominence for both speakers ($r = 0.60$ for A03; $r = 0.68$ for A05). We also see a significant, but less strong, relation between jaw displacement and perceived boundaries, with one speaker (A03, $r = 0.43$, $p < 0.001$) showing a slightly stronger relation than the other (A05, $r = 0.18$, $p < 0.05$).

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

The interaction between phrase boundaries and prominence, both in terms of perception and articulation, is explored further in the following sections.[2]

Given that jaw displacement increases for emphasized words, in the next section, we explore further how jaw displacement patterns vary with utterance prominence patterns. The hypothesis is that we will see patterns of increased jaw displacement that correspond to, for example, foot stress, phrasal stress, and utterance nuclear stress.

2.3 Jaw Displacement and Utterance Prominence Patterns

As seen in the previous sections, not only does jaw displacement increase for emphasis, it also increases for utterance (nuclear) stress. Here we report on findings from earlier studies (e.g., Erickson et al., 2012, 2014; Huang and Erickson, 2019; Erickson and Niebuhr, 2023). Articulatory and acoustic recordings by a number of AE speakers were made for the utterance *I saw five bright highlights in the sky tonight*. This utterance was chosen because the vowels in the content words are all /aɪ/ diphthongs (except for the word *saw* in the phrase *I saw*, and this phrase was excluded from the analysis); also, as briefly described above, the utterance has the two phrases *five bright highlights* and *in the sky tonight*. The first phrase consists of two two-word feet: *five bright* and *highlights*. (The first phrase could be referred to as an accent phrase, or an intermediate phrase, but here we refer to it simply as a phrase.) Maximum jaw displacement was measured during the vowel /a/ of the diphthong.

Figure 2.2 (top) shows jaw tracings of an AE speaker producing the utterance (*I saw*) *five bright highlights in the sky tonight*. The valleys indicate when the jaw is lowered/open, as measured from the occlusal (bite) plane. We observe that for the top jaw tracing (gray arrows), (1) there are valleys (jaw openings) for each syllable; (2) the depth of the valleys varies, even though the vowel is always /aɪ/; and (3) the biggest valley (jaw displacement) occurs for *sky*, then for *high* of *highlights*, and then for *five* of *five bright*.

The jaw displacement patterns for the top jaw tracing in Figure 2.2 (seen for the four English speakers reported in Huang and Erickson, 2019), with nuclear stress on ***sky***, can be depicted in terms of a metrical arrangement of syllable stress patterns along the lines of, for example, Liberman and Prince (1977), Selkirk (1982), and Hayes (1995), as shown in the top part of the metrical grid displayed in Table 2.1. The word *sky* has utterance nuclear stress (level 5 stress), *high* has phrasal stress (level 4 stress), and *five* has foot stress (level 3 stress). Notice that for this speaker, nuclear stress occurs in the default position, that is, the last content word of the utterance. (Note that *tonight* is not a content word but rather an adverb.) The stress pattern in this utterance would be 3-2-4-1/5-2 (where a slash indicates a syntactic phrasal break).

When AE speakers are asked to read a sentence with no prior context, they have choices of where to put nuclear stress, as indicated by the bottom jaw tracing of Figure 2.2. In the bottom tracing, the pattern of jaw displacement shows the largest stress in the utterance on *high (lights)*, perhaps because the word *highlights* seemed to carry a certain amount of salience

[2] See also more details in Erickson et al. (2014, 2015).

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Table 2.1 Metrical grids for two types of nuclear stress

Metrical grids for two types of nuclear stress productions for the utterance *(Yes, I saw) five bright highlights in the sky tonight* (along the lines of, for example, Hayes, 1995). Both productions show foot stress on *five*, but the top grid shows default type nuclear stress on ***sky*** and phrasal stress on *highlights*; the bottom grid shows intent nuclear stress on ***high****lights*, and phrasal stress on *sky*.

Utterance					x	
Phrase			x		x	
Foot	x		x		x	
Word	x	x	x		x	x
Syllable	x	x	x	x	x	x
Stress level	3	2	4	1	5	2
(Yes, I saw)	five	bright	high	lights	**sky**	night
Utterance			x			
Phrase			x		x	
Foot	x		x		x	
Word	x	x	x		x	x
Syllable	x	x	x	x	x	x
Stress level	3	2	5	1	4	2
(Yes, I saw)	five	bright	**high**	lights	sky	night

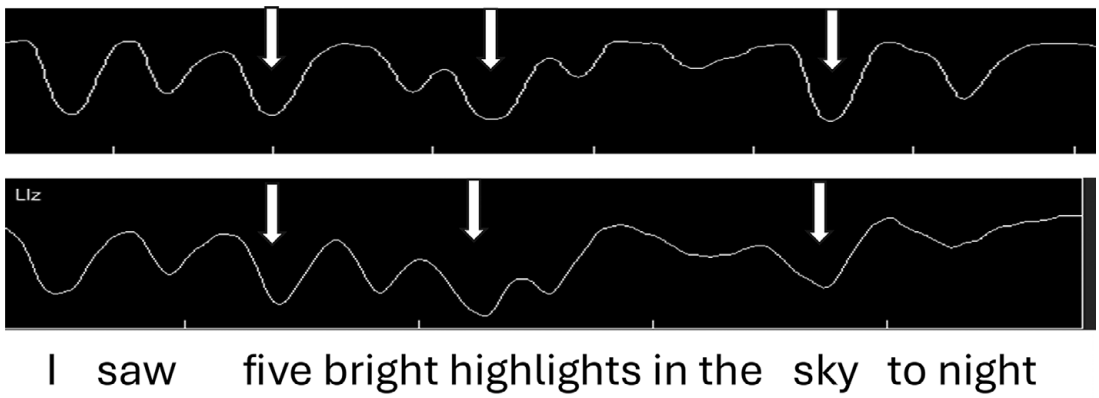


Figure 2.2 Samples of default and intent nuclear stress.
Jaw tracings for AE speakers for the utterance *I saw five bright highlights in the sky tonight*. For the top figure, gray arrows from left to right point to foot stress on *five*, phrase stress on the first syllable of the compound word *highlights*, and utterance (nuclear) stress on ***sky***. For the bottom figure, nuclear stress is instead on ***high****(lights)*, phrasal stress is on *sky*, and foot stress is on *five*, as indicated by the white arrows.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

for these speakers. Thus, the stress pattern in this utterance would be 3-2-5-1/4-2 (where a slash indicates a syntactic phrasal break).

The metrical grid for this utterance, with the largest jaw displacement on *highlights*, typical for three of the four AE speakers reported in Erickson et al. (2012), might be as represented in the bottom metrical grid portrayed in Table 2.1. The largest jaw displacement is hypothesized to be the utterance nuclear stress *high(lights)*, while the phrasal stress is *sky* and the foot stress is *five*, with a stress pattern of 3-2-5-1/4-2. Another speaker in the Erickson et al. (2012) study, however, put the most jaw displacement/prominence on *five*, and for this speaker, there would be yet another different pattern of utterance stress: 5-2-3-1/4-2 (Figure 2.1; Erickson et al., 2012). Our findings suggest that nuclear stress in AE may be of two types: default type, which occurs on the last content word of the utterance, and intent type, which occurs on the content word that seems to be most salient to the speaker.

Erickson et al. (2012) reported that increased jaw displacement showed a significant relation with F1 for three of the four speakers; moreover, both jaw displacement and F1 showed a significant relation with the metrical grid as shown in Table 2.1 (bottom part). These findings lend support to the jaw as being the metrical prosodic organizer of English utterances, with increased F1 being one of the acoustic changes. The Huang and Erickson (2019) study investigating utterances with mid front vowel /ɛ/ also addresses possible effects on jaw displacement due to interactions among the tongue body together with formant changes and F0 (see also Chen et al., 2019, as well as Erickson, 2002).

A question is to what extent jaw displacement patterns affect/influence listeners' perceptions of utterance prominence. Pilot studies suggest that the word with the largest jaw displacement is also perceived as having the most prominence (e.g., Erickson et al., 2015, 2020a). What in the acoustic signal is related to increased jaw displacement and prominence perception? These questions are explored in Section 2.5, reporting on new articulatory, acoustic, and perceptual data. In addition to jaw data, we report on vowel acoustic measurements (F1, F2, F0, intensity, and duration) along with listeners' assessment of prominence and phrase boundaries. However, first we wish to report on some recent findings concerning the relationship between segmental articulation and syllabic articulation, and to present ideas on how this relationship might be affected by prominence.

2.4 Relation between Segment Articulation and Syllable Articulation

The pattern of jaw displacement reflects the abstract metrical prosodic hierarchy of a spoken utterance (see, for example, Erickson et al., 2012). These rhythmic closed and opened periods of the jaw underlying each syllable must synchronize with the alternating degrees of constriction of consonants and vowels for the syllable. In this section we examine segmental articulation and how it might time with syllabic articulation. Here, we review recent work by Svensson Lundmark (2023), who describes segmental articulation in terms of de/acceleration peaks of the articulators (here the slash refers to both deceleration and acceleration) where the acceleration and deceleration peaks of the segmental articulators align with acoustic segment boundaries. This approach is centered around the concept of acceleration which is a movement

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

characteristic of a mass (= an articulator). Both acceleration (movement initiation) and deceleration (movement "braking", approaching a target) involve positive muscle forces—generated by agonist or antagonist muscles, respectively. The amount and type of force applied determine articulatory speed. The moments of maximal acceleration or deceleration (de/acceleration peaks) divide the articulatory movement into postures/steady states, which are delimited by very fast intervals. It is specifically the postures of the crucial segmental articulators of the consonant production, that is, the lips for a bilabial segment, that align with the acoustic consonant segment. For more details concerning this approach, we refer to Svensson Lundmark (2023) and Svensson Lundmark and Erickson (2024).

As for vowel segment duration, the relationship to the de/acceleration peaks is different: as the consonantal production is of an instantaneous nature, layered on top of vowels (Öhman, 1966), the vowel segment duration is the result of when the consonant segments are produced. More specifically, the acceleration peak at the end of the syllable onset consonant forms the vowel segment onset, and the deceleration peak of the coda consonant forms the vowel segment offset, as indicated in the schematized kinematic trajectories of a CVC (consonant-vowel-cononant) sequence in Figure 2.3. Note that the de/acceleration peaks of the tongue body for vowel targets do not align with the segment boundaries, but instead shape a much shorter tongue body posture (Svensson Lundmark, 2023). In other words, the duration of the acoustic

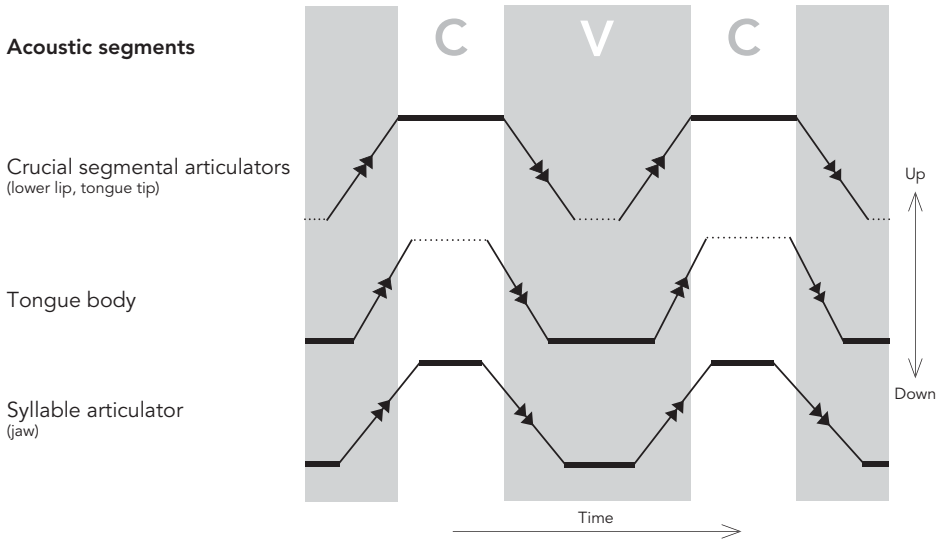


Figure 2.3 Schematized steady stades and fast transit intervals.
A schematized figure on steady states (thick solid lines) and fast transit intervals (solid lines with arrows) of a syllable, as divided by de/acceleration peaks. The areas marked CVC are the duration of the acoustic segments. The steady states of the crucial segmental articulators (here, vertical positions of the tongue tip or lower lip) form the CVC segments. The syllable articulator (the jaw) displays shorter steady states than both C and V segments. Note that the steady state of the tongue body (a low vowel) does not align with the V segment. Still unresolved questions are marked with dotted lines. The schematized figure is based on findings of Svensson Lundmark (2023) and Svensson Lundmark and Erickson (2024).
Reproduced with permission from M. Svensson Lundmark and D. Erickson, JSLHR, https://doi.org/10.1044/2024_JSLHR-23-00092, accepted; licensed under a Creative Commons Attribution (CC BY) license.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

vowel segment is a reflection of when in time the first crucial segmental articulator (for the syllable onset) leaves its position, be it the tongue tip or the lips, and when the second crucial segmental articulator (for the syllable coda) arrives at its target destination.

The duration of the vowel segment is connected to jaw displacement by the nature of the jaw cycle; the vowel appers during its open state, while the closed steady states of the jaw (when the jaw stays up) are when the onset and coda appear, as shown in Figure 2.3 (Svensson Lundmark and Erickson, 2024). Furthermore, the steady states of the jaw are shorter than the steady states of the crucial segmental articulators, as also indicated in Figure 2.3 (Svensson Lundmark and Erickson, 2024). We suggest that these timing differences between the steady states/postures on the jaw cycle and the ones on the crucial segmental articulators could be because the syllabic jaw movements and the segmental articulators have inherently different rhythmic patterns. We intend to follow up this approach in future research.

Although the nature of the timing of the de/acceleration peaks is still unclear, it's related to how fast the articulators are moving. We know that the speed of the segmental articulators is affected by the distance to the target and what type of target is made (the task difficulty) (Bootsma et al., 2004). While task difficulty is related to how the constriction is manifested (and therefore possibly related to deceleration), the distance to the target is manifested by the size of the oral cavity, that is, how long/far the articulator has to travel. If the tongue is producing a low vowel, the tongue tip might need to travel further to make the following coda constriction at the palate, whereas for a high vowel, the tongue tip is presumably already near the target constriction. We presume these conditions affect the timing of the de/acceleration peaks, hence directly having an impact on the posture and the resulting acoustic segment duration. Furthermore, as the size of the oral cavity changes according to jaw displacement and longer distances for the articulators to travel for a more prominent syllable, this would possibly affect tongue body positions, leading to even more variants of travel distances for the tongue tip between, for example, a prominent /a/ and a prominent /i/. However, this relationship still needs to be investigated.

The vowel segment differences are also dependent on what type of syllable onset or coda constriction, that is, manner, place, and voicing, the segmental articulator is making, as the de/acceleration peaks of the segmental articulators determine the acoustic vowel segment boundaries. Thus, where a segmental articulator travels from and where it is going next affects the distance to the target and its velocity, which ultimately affects the acceleration and the deceleration phase, and the timing of those de/acceleration peaks. In other words, any correlation between vowel segment duration and jaw displacement is dependent on the timing of the segmental articulators (= the consonants) and the context in which they are produced.

2.5 Articulatory, Acoustic, and Perceptual Study of AE Utterance Prominence Patterns: New Data

In Section 2.3, we established that the greatest jaw displacement corresponds to the perception of contrastive emphasis in utterances containing the diphthong /aɪ/. This section continues to investigate utterance prominence in terms of articulation, perception, and acoustic cues.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Here we look at the utterance *Pam had a chance to chat and nap* as spoken by five AE speakers (three male, two female). This utterance was selected because (1) the content words, *Pam*, *chance*, *chat*, and *nap*, all contain the phonological AE /æ/ vowel, and (2) it has two phrases, *Pam had a chance* and *to chat and nap*. Looking at jaw displacement across the same vowels allows us to see prominence effects; looking at an utterance with two phrases allows us to see possible differences in articulation of phrasal stress and nuclear stress.

2.5.1 Articulatory and Acoustic Study

The acoustic and articulatory recordings were done with EMA at Professor Jianwu Dang's laboratory at the Japan Advanced Institute of Science and Technology, Nomi, Japan (see Section 2.2 for an explanation of sensor placement). The target sentence was read in five different randomizations for four of the speakers, and for S4 there were eight randomizations of the target sentence. MView (Tiede, 2010) was used to measure the amount of maximum jaw displacement from the occlusal plane during each vowel. The average F1, F2, F0, intensity, and duration for each vowel was estimated using Praat (Boersma and Weenink, 2025) to mark TextGrids, and a customized Praat script. Perception tests to assess listeners' perceptions of prominence and phrase break/boundary were conducted using an online interface.

Figure 2.4 shows the mean values of jaw displacement for each of the five speakers. In the bar graphs, the height of each bar represents the average amount of jaw displacement for each of the one-syllable content words, *Pam*, *chance*, *chat*, and *nap*. As discussed in the previous sections, bar graphs convey relative syllable size, that is, *big* (more prominent) syllables will have larger y-values (mm) than *smaller* ones. Thus, the y-axis values reflect the relative values of jaw displacement (syllable stress) within the utterance.

As discussed in the above paragraph, the bar graphs suggest that in terms of an abstract metrical prosodic hierarchy, the word with the largest jaw displacement in the utterance carries the nuclear stress of the utterance; when there is more than one phrase, the word in each phrase that has the largest jaw displacement is referred to as having phrasal stress, and often this is the word with the second largest jaw displacement. These data suggest that for this sentence, we can have at least two types of nuclear stress intent type – that is, either *chat* or *Pam* can have nuclear stress.

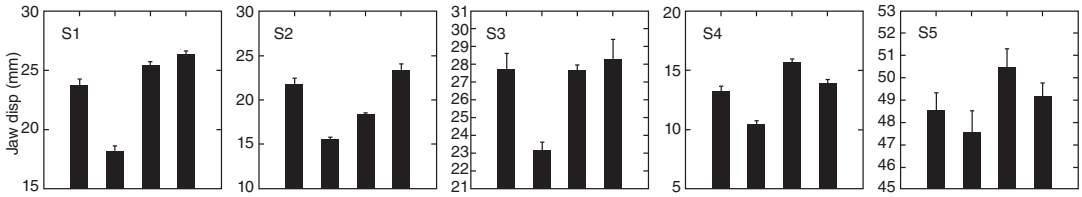


Figure 2.4 Jaw displacement patterns for five speakers.
Bar graph displays of average amount of jaw displacement (mm) shown on y-axis with content word on the x-axis (*Pam*, *chance*, *chat*, *nap*), for each of the five speakers (S1, S2, S3, S4, and S5). Jaw displacement ranges from 15 to 30 mm for S1, 10 to 30 mm for S2, 21 to 31 mm for S3, 5 to 20 mm for S4, and 45 to 53 mm for S5.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Table 2.2 shows acoustic measurements related with the average jaw displacements for each of the content words. The word with the largest jaw displacement also has the largest F1 value; peak F0 always occurs on the initial word, with the final word having the lowest F0; intensity is greatest on the initial word of the utterance for three of the speakers, but on the initial word of the second phrase for the other two speakers; acoustic duration is longest on the initial word of the utterance for two of the speakers, while the other three speakers show the longest duration on *chance*, *chat*, or *nap*. In Section 2.5.3, we discuss some of the acoustic cues listeners may be listening to in order to hear increased prominence on a syllable in an utterance.

Linear correlation results with jaw displacement and acoustic measurements indicate that four of the five speakers showed a significant relation between the amount of jaw displacement and F1 ($p < 0.001$), indicating that as jaw displacement increases, F1 raises; three other speakers showed significant relations between jaw displacement and F2 minus F1, indicating that as the distance between F2 and F1 decreases, the low vowel became even more extreme and compact, as discussed in Erickson (2002).

No significant relations for the five speakers were found between amount of jaw displacement and the acoustic measures of F0 and duration. Only one speaker (S4) showed a significant positive relation between jaw displacement and intensity ($p < 0.001$); another speaker (S1) showed a significant negative relation between jaw and vowel duration. When we look at each speaker separately (Figure 2.5), four of the speakers (S1, S2, S4, and S5) show negative linear correlation between vowel duration and jaw displacement for each of the content words, while only S3 shows a positive correlation. These findings are surprising given that increased F0 and duration are what has been reported for increased prominence/stress (e.g., Turk and Sawusch, 1996).

In the next section, we examine how listeners hear prominence/stress in English utterances – specifically, the probability of stress perception as a function of both prosodic articulation (i.e., jaw displacement) and prosodic acoustic measurements (F0, F1, F2, intensity, and vowel duration). Since, in addition, prosodic boundaries also affect utterance prominence (e.g., Fujimura, 2000; Erickson et al., 2014; Cole et al., 2019), we also include boundary perception in the next section.

2.5.2 Pilot Study of Perception of Prominence and Boundaries in Utterances

Here we introduce some preliminary results about the extent to which perception of nuclear (utterance) and phrasal stress and phrase boundaries relate to jaw displacement. Twenty-eight AE listeners participated in our two experiments designed respectively for perceiving nuclear stress and phrasal boundaries in a neutral context. For each experiment, we recruited 14 participants (aged 18–78), who voluntarily participated or attended for course credit at the University of Arizona.

As described in Section 2.5.1, five native speakers of AE produced the utterance *Pam had a chance to chat and nap*. We selected one token from each speaker (S1 token 55; S2 token 115; S3 token 16; S4 token 64; S5 token 001) as the stimuli investigated in our perception tests. The reason for doing this was because we felt having listeners try to hear differences in prominence and boundaries within a single speaker as well as across speakers might be too difficult, given

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Table 2.2 Average jaw displacement and acoustic values for content words

Average jaw displacement and average acoustic values of the content words for the five speakers. Bold numbers indicate the largest value for each measurement across the content words for each of the speakers.

Sub	*N*	*word*	*Jaw*		*F0Av*		*F1*		*F2*		*IntAv*		*vowel dur (s)*	
S1	5	Pam	23.7	1.1	**244.8**	5.3	726.8	10.4	1229.8	78.8	**52.8**	0.3	0.17	0.01
	5	chance	18.1	0.9	191.2	3.3	661.8	7.8	1401.5	167.0	49.0	0.9	**0.19**	0.02
	5	chat	25.4	0.6	189.8	5.4	890.4	40.9	1429.9	140.0	49.5	1.2	**0.19**	0.02
	5	nap	**26.3**	0.5	152.4	16.0	**914.8**	61.8	1547.8	93.8	42.3	1.6	0.16	0.01
S2	5	Pam	21.9	1.1	**153.3**	6.6	666.9	20.9	1367.7	189.4	33.9	2.2	0.13	0.01
	5	chance	15.5	0.5	116.0	2.6	612.6	24.2	1443.8	153.0	33.9	1.0	**0.14**	0.01
	5	chat	18.3	0.5	110.7	1.9	695.5	12.1	1357.3	38.1	**39.9**	1.5	**0.14**	0.02
	5	nap	**23.3**	1.4	98.2	10.3	**769.7**	30.6	1416.0	103.5	28.2	1.9	0.13	0.01
S3	5	Pam	27.7	1.8	**137.6**	8.7	560.6	16.8	1746.5	32.2	**59.9**	1.4	**0.18**	0.01
	5	chance	23.2	0.9	110.7	7.9	475.4	34.0	1708.1	48.0	56.7	0.9	0.15	0.01
	5	chat	27.6	0.7	105.4	9.0	583.8	30.1	1600.9	31.4	52.8	1.4	0.16	0.02
	5	nap	**28.3**	2.2	81.5	1.4	**600.6**	45.0	1718.1	52.3	48.8	1.4	0.16	0.02
S4	8	Pam	13.3	0.9	**119.1**	4.4	728.6	15.6	1632.2	73.9	54.6	0.5	**0.23**	0.02
	8	chance	10.5	0.6	106.7	4.0	659.6	19.5	1775.0	30.0	51.3	0.6	0.17	0.01
	8	chat	**15.7**	0.8	102.6	2.1	690.2	11.2	1470.9	22.5	**55.3**	1.2	0.16	0.01
	8	nap	13.9	0.9	91.8	2.4	**752.6**	16.9	1510.0	13.6	49.0	1.4	0.17	0.02
S5	5	Pam	48.5	1.6	**247.9**	9.6	682.6	34.8	1593.9	224.3	**46.0**	1.1	0.16	0.03
	5	chance	47.6	1.9	196.8	8.4	699.2	20.2	1653.6	124.1	39.9	1.8	**0.19**	0.01
	5	chat	**50.5**	1.6	184.5	6.1	**804.4**	5.1	1190.7	39.9	44.3	1.5	0.16	0.01
	5	nap	49.2	1.1	161.6	8.2	621.9	34.6	1040.6	88.9	34.7	0.9	**0.19**	0.0

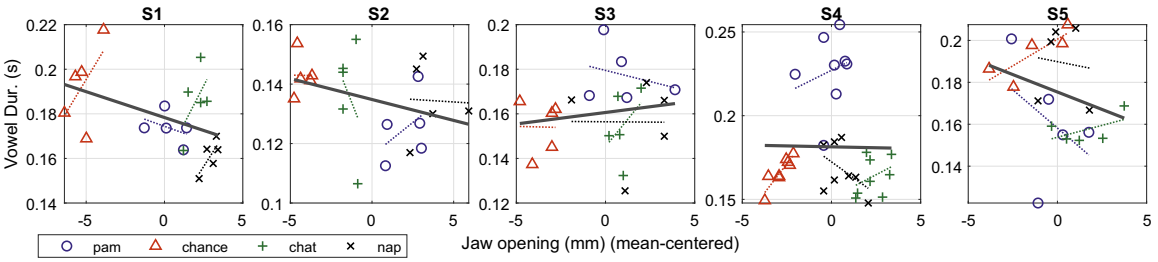


Figure 2.5 Correlation of vowel duration and jaw displacement.
Correlation plot on acoustic vowel duration (s) and jaw displacement (mm) for each speaker. Jaw displacement is mean-centered for reasons of clarity. The four content words are represented by different symbols.

that most listeners are usually not familiar with listening to stress patterns. Also, based on the intra-speaker variations in jaw displacement, it seemed that speakers may have been inconsistent in their stress patterns. The present study employed two small-scaled RPT experiments. Sixteen stimuli (= (5 target sentences + 3 filler sentences) × 2 repetitions) were presented randomly online to each participant. An inter-stimulus interval (ISI) of 200 ms was used. Because perceiving utterance stress or boundaries is not something that most listeners are consciously aware of, before starting each of the tests, in order to help the subject "tune their ears" for prominent words, or for boundaries, subjects were asked to listen to a brief one-minute training video prepared by the first author. In the video for prominence testing, the listeners were asked to hear the difference between *I like red apples*, where *apples* had more stress, and *I like RED apples*, where *red* had more stress. After the training session, the participants were instructed to do the online RPT over their headphones in a quiet room. In the nuclear stress experiment, the listeners (2 males, 12 females) were asked to choose which word bore nuclear stress after hearing the given utterance. For the phrasal boundary experiment, the listeners (3 males, 11 females) were told to determine after what words they considered a boundary occurred.

2.5.2.1 Perceived Prominence and Boundaries

Not surprisingly, our results revealed that stress is perceptible on content words (i.e., *Pam*, *chance*, *chat*, *nap*), but not on function words (i.e., *a*, *to*, *and*). Listeners reported nuclear stress on *Pam* for three of the speakers, and *chat* and *chance* for the other two speakers. However, it seems that the relation between articulation of prominence (i.e., amount of jaw displacement) and perception of prominence (listener ratings) is not necessarily straightforward. For example, Figure 2.6 shows the relation of jaw displacement with vowel duration, and the probability of listeners hearing nuclear stress and boundaries. Figure 2.6b shows S2 had a preference for nuclear stress on *Pam* (35.7%), even though the jaw displacement was only the second largest (22.63 mm), and the phrasal stress was on *chat* (28.6%), even though *nap* had the largest jaw displacement. In the case of S2, thus, listeners' perceptions of stress and speakers' articulations of stress (i.e., jaw displacement) do not align.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

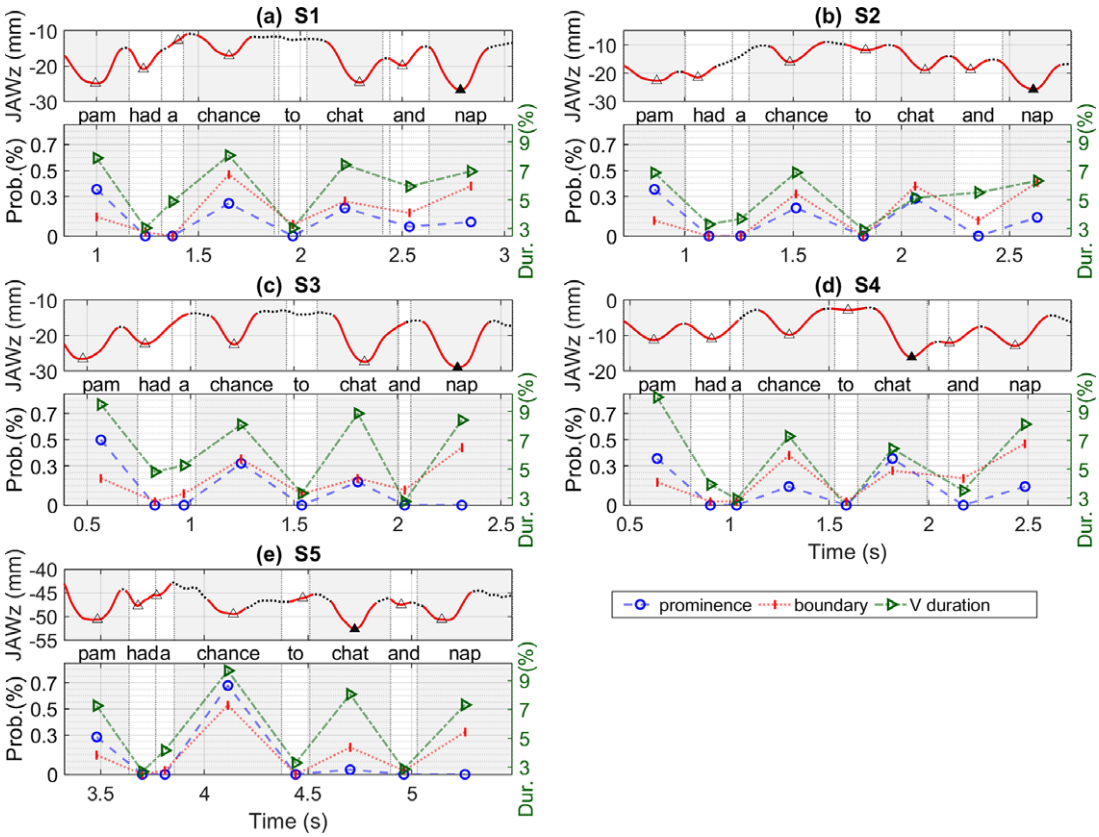


Figure 2.6 Probability of perceived prominence and boundaries.
(Top) jaw displacement for the utterance; (bottom) probability of nuclear prominence (dash lines with circle symbols) and phrase boundaries for each word produced by (a) S1, (b) S2, (c) S3, (d) S4, and (e) S5.

For S4 (Figure 2.6d), the words *Pam* and *chat* are perceived as equally prominent (35.7%), but note that jaw displacement is largest for *chat* (16.15 mm), followed by *Pam* (11.27 mm). The suggestion here is that the nuclear utterance stress is on *chat*, with the phrasal stress on *Pam*, but in terms of perception it is somewhat ambivalent. Nevertheless, we see an alternating strong–weak pattern of jaw displacement within each of the phrases, as well as an alternating pattern of perceived stress. For S5 (Figure 2.6e) the likelihood of perceiving stress on *chance* was highest (67.9%), but jaw displacement was smallest (49.50 mm). We discuss this pattern later.

As for boundary perception, the amount of jaw displacement can account for both syllable prominence as well as phrasal boundaries (see also, for example, Fujimura, 2000; Erickson et al., 2015).

In order to further investigate the relationship between the amount of jaw displacement with the probability of perceived prominence and that of phrase boundaries for each syllable across the five speakers, a linear correlation analysis was applied. By-speaker z-score normalization of articulatory measurements were done before entering correlation analyses to resolve inter-speaker

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

physiological differences. We found a significant difference between jaw displacement and phrase boundaries ($p < .05$). For instance, for S3 (Figure 2.6c), the largest amount of jaw displacement corresponds with the greatest number of perceived breaks. These results suggest that the more open the jaw, the larger the number of perceived boundaries; in contrast, there was only a marginally significant difference observed for prominence in relation to jaw displacement ($r = -0.281$).

2.5.3 Acoustic Cues for Perception

In this section, we explore some acoustic cues that might affect listeners' perception of prominence and boundaries. We measured duration, mean F0, mean intensity, and the first two formants (F1, F2) for the vowel /æ/ of the content words (*Pam*, *chance*, *chat*, *and nap*). A linear correlation analysis was taken to assess the relation of each acoustic measure with perceived prominence and boundaries. Acoustic measurements underwent by-speaker z-score normalization before entering correlation analyses.

Table 2.3 presents the correlation results for the acoustic cues related to perceived prominence. Prominent syllables are significantly longer, have significantly higher maximum F0, have marginally significantly higher mean F0, and are significantly louder. That syllables that are longer, louder, and higher in F0 were rated as more prominent is not surprising, given the literature about acoustic cues of stress/prominence, for example, Lehiste (1970) and Erickson and Niebuhr (2023); also see Section 2.2.

As for formant values, perceived prominent syllables have significantly higher F2 values ($p < .05$), implying the vowel /æ/ tends to be more fronted when it is perceived as having nuclear stress. *Pam* produced by three of the speakers has higher F2 (see Table 2.5) and is perceived as having nuclear (utterance) stress for four of the speakers. (Note that for S2, *Pam* is perceived has having nuclear stress, but *nap* has the highest F2.) Interestingly, however, F1 values in this current study are not remarkably raised for the perceived prominent syllable, even though we see significant correlation with increased F1 and articulatory prominence, that is, jaw displacement, both in this study as well as previous studies.

Table 2.3 Correlation analysis for perceived prominence and acoustic measures

The correlation analysis for perceived prominence and vowel acoustic measurements – mean duration, mean F0, mean intensity, F1, and F2. *** indicates significance < 0.001, ** indicates significance < 0.01, * indicates significance < 0.05, and † indicates significance < 0.1 (marginally significant).

Variables	*R*	*p-value*
Prominence and duration	0.740	(< .001) ***
Prominence and mean F0	0.302	(0.065) †
Prominence and max F0	0.442	(0.005) **
Prominence and intensity	0.381	(0.015) *
Prominence and F1	0.053	(0.743)
Prominence and F2	0.353	(0.026) *

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Table 2.4 Correlation analysis for perceived boundaries and acoustic measures

The correlation analysis for perceived boundary and mean duration, mean F0, mean intensity, F1, and F2. *** indicates significance < 0.001, * indicates significance < 0.05, and † indicates significance < 0.1 (marginally significant).

Variables	*R*	*p-value*
Boundary and duration	0.740	(< .001) ***
Boundary and mean F0	−0.316	(0.053) †
Boundary and mid F0	−0.371	(0.022) *
Boundary and intensity	−0.193	(0.233)
Boundary and F1	0.310	(0.052) †
Boundary and F2	0.240	(0.136)

Table 2.5 Average F2 values for four syllables of five speakers

Average F2 values (in Hz) for the four syllables for each of the five speakers' utterances. F2 values in bold indicate the highest value among the four syllables in the utterance. Asterisks mark perceived nuclear stress.

	Pam	*chance*	*chat*	*nap*
S1	***2284.8**	1985.7	2002.2	2086.1
S2	1634.7	1592.6	1646.1	***1671.1**
S3	***1889.6**	1612.4	1515.5	1875.0
S4	***1851.4**	1843.7	1463.8	1514.7
S5	2043.8	***2109.8**	1995.9	2023.3

For perceived boundaries, as shown in Table 2.4, vowel duration is also significantly correlated with phrase boundaries. Long syllables tend to be followed by a phrase boundary. This is confirmed in our study. In addition, perceived phrase boundaries have marginally significantly higher mean F0, and significantly higher vowel mid-point F0, indicating listeners' sensitivities to changes in F0 when a boundary is perceived. Intensity is nonsignificant. Also, the F1 value tends to be increased for syllables with a perceived boundary; specifically, it is relatively high in *chance* or *nap*, which are perceived as the final syllables in the phrases. To summarize, based on this small data set, increased duration, heightened F0, and raised F1 tend to cue phrase boundaries. Similar findings about duration and F0 as cues to phrase boundaries have been reported by, for example, Yang et al., 2014, but as far as we know, no one has reported raised F1 as a cue.

Our data suggest that when one's jaw is open more, the acoustic cues including duration, F0, and intensity seem to increase, which in turn affects our judgment of prominence and boundaries. However, the interaction between perception of prominence, phrase boundaries, and articulation of prominence is complex and needs to be explored further.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

2.6 Jaw and Phrasal Stress in Other Languages and Applications for Language Teaching

This section is a brief review of jaw studies (e.g., Vatikiotis-Bateson and Kelso, 1993; Erickson et al., 2014, 2016; Kawahara et al., 2014a; Erickson and Kawahara, 2016; Smith et al., 2019; Erickson, 2021) showing how patterns of jaw displacement are language-specific, reflecting the prominence patterns (metrical organization) of the language; being aware of these language-specific differences has applications for teaching second language prosody. The above sections discussed jaw displacement patterns of English utterances. Here we look at jaw displacement patterns of languages that are purported to be edge-strengthening languages (e.g., Jun 2005, 2014), such as French, Japanese, and Mandarin.

As shown in Figure 2.7, jaw displacement increases at the end of each phrase, with the largest amount of jaw displacement at the end of the utterance – which is in accord with these languages being called edge-strengthening. For the Japanese and Mandarin utterances, we also see increased jaw displacement at the beginning of the utterance, as if the speaker uses increased jaw displacement to indicate "I'm starting to speak" and "I've finished speaking."

In learning a second language, beginners tend to carry over their first language jaw displacement patterns to their second language (e.g., Wilson et al., 2020; Erickson and Niebuhr, 2023) – see Figure 2.8.

In Figure 2.8, the L1 English speaker (bottom row) shows clear patterns of strong–weak jaw displacement, with the largest jaw displacement on nuclear utterance stress *sky*, the next largest on phrasal stress *highlights*, and then on foot stress *five*. The speakers of Japanese (top) and French (middle), both L1 speakers of edge-strengthening languages, show different patterns of jaw displacement when speaking English. The L1 Japanese speaker shows more jaw

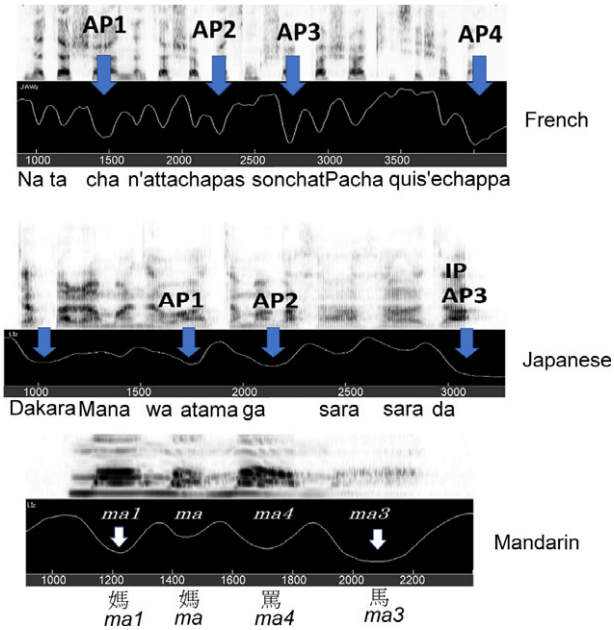


Figure 2.7 Jaw displacement patterns for French, Japanese, and Mandarin.
Jaw displacement patterns from top to bottom for French (*Natacha didn't attach her cat, Pasha, who escaped*), Japanese (*That's why Mana's hair is smooth*), and Mandarin (*Mother curses the horse*). The term AP refers to *accent phrase*, and the final AP is referred to as an IP (*intonational phrase*).

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

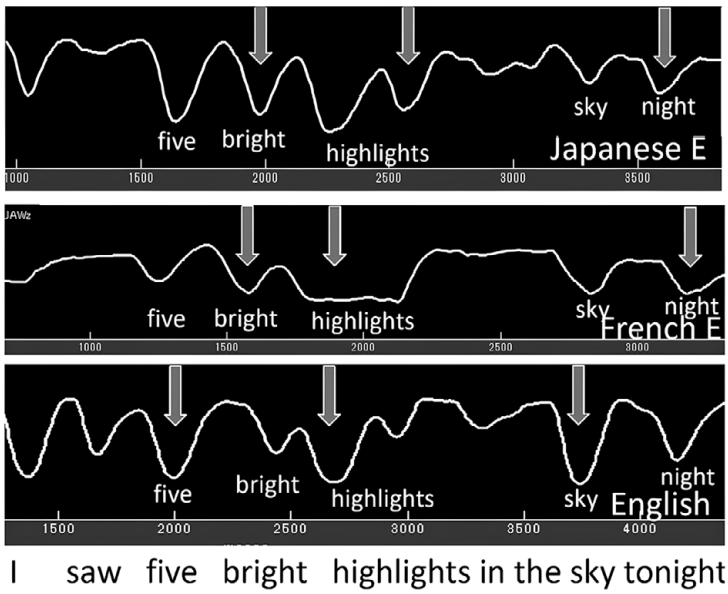


Figure 2.8 Jaw displacement for L2 Japanese and French speakers, and an L1 English speaker.
Jaw displacement patterns for L1 Japanese (top row), L1 French (middle row) and L1 English (bottom row) speakers for the utterance *I saw five bright highlights in the sky tonight*.

displacement on *highlights* and on *five*, but no clear reduction on *bright* and *lights*, as seen for the L1 English speaker. This makes sense since Japanese does not reduce vowels. For the L1 French speaker, the largest jaw displacement is at the end of the phrase, *highlights*, produced with one sustained jaw lowering; for the foot, *five bright*, more jaw displacement is on the final member of this foot, *bright*. These patterns suggest that the French speaker may be carrying over her L1 edge-strengthening pattern of increased jaw displacement for phrase-final positions. As for the final phrase, *sky tonight*, both L1 Japanese and French speakers carry over their phrase-final jaw patterns to produce the utterance-final word *night* with more jaw displacement than on *sky* – a jaw pattern not seen with L1 AE speakers. All AE speakers put reduced jaw movement on *night*, resulting in a more shwa-like vowel for this word.

Recent work with teaching English jaw displacement patterns to Japanese learners of English (Wilson et al., 2019, 2020) suggests that when a Japanese speaker acquires an English pattern of jaw displacement, not only does the stressed word have larger jaw displacement but the following word has less jaw displacement. The acoustic results were increased F1 and decreased F2 on the stressed word (with the low vowel /a/) and decreased F1/increased F2 on the following word (also a low vowel /a/), thus making the second vowel reduced, and resulting in a more English-type strong–weak rhythm. These results encourage teaching second language prosody using a kinematic approach.

The results also encourage thinking about jaw mechanics; perhaps the jaw cannot make two consecutive large jaw displacements without pause or resetting. To produce two or three consecutive syllables, one jaw opening will be stronger than the adjacent ones. In an oft-times called stress-timed language such as English, where increased prominence/jaw displacement occurs within each prosodic structure (e.g., word, foot, phrase, utterance), the other syllables in the prosodic structure will be produced with less displacement. Studying displacement patterns of various languages and how to kinematically teach these to second language learners can lead to an increased

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

understanding of the biophysical constraints of our articulators, and also the segmental articulators, for example, the lips, tongue, and velum, since syllable prominence affects both the vowel nucleus as well as the onset and coda segmental articulators (e.g., Heldner, 2001; Giavazzi, 2010; Fletcher, 2013; Svensson Lundmark et al., 2017, for discussion of how prominence can affect segmental phonology; Kawahara et al., 2014b, for discussion of the interaction between jaw and segment acoustics in Japanese; de Jong et al., 1993, reporting that stressed syllables show less coarticulatory overlap than non-stressed syllables).

This chapter has focused on the jaw as a prosodic syllable articulator; work is underway for further exploration into the interactions between crucial segmental articulators and the syllabic articulator, as they work together to orchestrate the rhythm of a language.

Future work into the relation between utterance prominence, jaw displacement, and language rhythm requires, first and foremost, more articulatory data. Since articulatory data collection is still a relatively expensive undertaking both in terms of time and money – for example, EMA-type facilities – currently, some new approaches are being developed for recording articulatory data. One of them is being developed by Oliver Niebuhr at the University of Southern Denmark – the MARRYS hat – which registers jaw displacement via two bending sensors in the cheek straps on both sides, timed with the audible voice signal; it is cheaper and also requires reduced preparation and processing times, allowing the collection of jaw data from a large number of speakers – even in the field, given its mobility (Erickson et al., 2020b; Niebuhr and Gutnyk, 2021; Svensson Lundmark et al., 2023; Svensson Lundmark and Niebuhr 2025; Weston et al., 2023).

To summarize, we suggest that speech rhythm is based on an abstract metrical hierarchical organization of syllable stress patterns. These stress patterns are implemented by jaw articulation patterns, that is, how much the jaw lowers for each syllable within the utterance. For English, within each prosodic unit (word, foot, phrase, utterance), one syllable has incrementally the largest jaw displacement, such that the word/syllable with the most prominence in the utterance also has the largest jaw displacement. Languages such as French, Japanese, and Mandarin have different underlying metrical structures, and hence different patterns of jaw articulation from English. Previous studies of jaw displacement patterns for English utterances show how the patterns relate to metrical patterns of utterance stress, and how increased jaw displacement is always accompanied by formant changes – specifically, increased F1 (or decreased F2 minus F1) for low vowels. Since jaw patterns vary depending on the language, language learners often apply their first language jaw displacement patterns to produce the second language.

Box 2.1 Chapter Overview

Summary

Articulatory, acoustic, and perceptual data examined in this study suggests that spoken language rhythm is a dynamic system involving (a) an abstract metrical structure implemented by jaw articulation patterns, (b) segmentally modified syllable articulation, and (c) perceptual sensitivities of listeners to the segmental and syllabic articulations.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Box 2.1 (cont.)

Implications

Speech rhythm can best be understood by examining the interactions and constraints between segmental and syllabic articulation. The implications are invitations to further explore language rhythm from an articulatory framework, similar to the one introduced here.

Gains

The gain is a framework to explore language rhythm across numerous languages and speakers, leading to innovative ways to understand and mediate first and second language acquisition.

REFERENCES

Beckman, M. E. (1986). *Stress and Non-Stress Accent*. Berlin, Boston: De Gruyter Mouton.

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

Boersma, P., and Weenink, D. (2025). Praat: Doing phonetic analysis by computer [Computer program]. Version 6.4.30. www.praat.org

Bootsma, R. J., Fernandez, L., and Mottet, D. (2004). Behind Fitts' law: Kinematic patterns in goal-directed movements. *International Journal of Human-Computer Studies*, 61(6), 811–821. https://doi.org/10.1016/j.ijhcs.2004.09.004

Chen, W.-R., Whalen, D., and Tiede, M. K. (2019). Mandibular contribution to vowel-intrinsic F0. Conference: International Congress of Phonetic Sciences (ICPhS), pp. 152–156. www.researchgate.net/publication/340618853_MANDIBULAR_CONTRIBUTION_TO_VOWEL-INTRINSIC_F0.

Cole, J. I., Hualde, I., Smith, C. L., et al. (2019). Sound, structure and meaning: The bases of prominence ratings in English, French and Spanish. *Journal of Phonetics*, 75, 113–147.

Cooper, W. E., Eady, S. J., and Mueller, P. R. (1985). Acoustic aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, 77, 2142–2155.

de Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97, 491–504.

de Jong, K., Beckman, M. E., and Edwards, J. (1993). The interplay between prosodic structure and coarticulation. *Language and Speech*, 36, 197–212.

Erickson, D. (1998). Effects of contrastive emphasis on jaw opening. *Phonetica*, 55, 147–169.

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

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

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Erickson, D. (2021). Articulatory prosody: A comparison of Mandarin Chinese and Japanese. Proceedings of the Second International Conference on Tone and Intonation.

Erickson, D., and Kawahara, S. (2016). Articulatory correlates of metrical structure: Studying jaw displacement patterns. *Linguistic Vanguard*, 2, 102–110. https://doi.org/10.1515/lingvan-2015-0025

Erickson, D., and Niebuhr, O. (2023). Articulation of prosody and rhythm: Some possible applications to language teaching. Proceedings of the 13th International Conference of Nordic Prosody, pp. 1–45. https://doi.org/10.2478/9788366675728-001

Erickson, D., Hashi, M., and Maekawa, K. (2000). Articulatory and acoustic correlates of prosodic contrasts: A comparative study of vowels in Japanese and English. Proceedings of the Acoustical Society of Japan, vol. 56, pp. 265–256.

Erickson, D., Huang, T., and Menezes, C. (2020a). Temporal organization of spoken utterances from an articulatory point of view. Speech Prosody 2020.

Erickson, D., Iwata, R., and Suemitsu, A. (2016). Jaw displacement and phrasal stress in Mandarin Chinese, Tonal Aspects of Languages 2016, May 24–27, 2016, Buffalo, New York.

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

Erickson, D., Kawahara, S., Shibuya, Y., Suemitsu, A., and Tiede, M. (2014). Comparison of jaw displacement patterns of Japanese and American speakers of English: A preliminary report. *Journal of Phonetic Society of Japan*, 18, 88–94.

Erickson, D., Niebuhr, O., Gu, W., Huang, T., and Geng, P. (2020b). The MARRYS cap: A new method for analyzing and teaching the importance of jaw movements in speech production. International Seminar of Speech Production.

Erickson, D., Kim, J., Kawahara, S., et al. (2015). Bridging articulation and perception: The C/D model and contrastive emphasis. International Congress of Phonetic Sciences 2015.

Fletcher, J. (2013) The prosody of speech: Timing and rhythm, in W. J. Hardcastle, J. Laver, F. E. Gibbon (eds.), *The Handbook of Phonetic Sciences*, Second Edition, pp. 523–602. Chichester, West Sussex, UK; Malden, MA: Wiley-Blackwell.

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America*, 27, 765–768.

Fujimura, O. (2000). The C/D model and prosodic control of articulatory behavior. *Phonetica*, 57, 128–138.

Giavazzi, M. (2010). The phonetics of metrical prominence and its consequences for segmental phonology. MIT dissertation, Massachusetts Institute of Technology.

Harrington, J., Fletcher, J., and Beckman, M. (2000). Manner and place conflicts in the articulation of accent in Australian English, in M. B. Broe and J. B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V: Language acquisition and the lexicon*, pp. 40–51. Cambridge: Cambridge University Press.

Hayes, B. (1995). *Metrical Stress Theory: Principles and case studies*. Chicago: University of Chicago Press.

Heldner, M. (2001). On the non-linear lengthening of focally accented Swedish words. Nordic Prosody VIII, Trondheim. Proceedings of the VIIIth Conference. W. van Dommelen and T. Fretheim (eds.), pp. 103–112.

Huang, T., and Erickson, D. (2019). *Articulation of English "Prominence" by L1 (English) and L2 (French) speakers*. Melbourne: International Congress of Phonetic Sciences.

Jun, S.-A. (2005). Prosodic typology, in S.-A. Jun (ed.), *Prosodic Typology: The phonology of intonation and phrasing*, pp. 430–458. Oxford: Oxford University Press.

Jun, S.-A. (2014). 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.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Kawahara, S., Erickson, D., Moore, J., Suemitsu, A., and Shibuya, Y. (2014a). 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), 77–87.

Kawahara, S., Masuda, H., Erickson, D., et al. (2014b). Quantifying the effects of vowel quality and preceding consonants on jaw displacement: Japanese data. *Journal of Phonetic Society of Japan*, 18(2), 54–62.

Kent, R. D., and Netsell, R. (1971). Effects of stress contrasts on certain articulatory parameters. *Phonetica*, 24, 23–44.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Liberman, M., and Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, 8, 249–336.

Kim, J., Erickson, D., and Lee, S. (2015). More about contrastive emphasis and the C/D model. Special Issue on the C/D model. *Journal of Phonetic Society of Japan*, 19(2), 44–54.

Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (2005). Loudness predicts prominence: Fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118, 1038. https://doi.org/10.1121/1.1923349

Loevenbruck, H. (1999). An investigation of articulatory correlates of the accentual phrase in French. Proceedings of the 14th International Congress of Phonetic Sciences, San Francisco, 1999, I, pp. 667–670.

Macchi, M. (1988). Labial articulation patterns associated with segmental features and syllable structure in English. *Phonetica*, 45, 109–121.

Menezes, C. (2003). Rhythmic pattern of American English: An articulatory and acoustic study. PhD dissertation, Ohio State University.

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

Menezes, C., and Erickson, D. (2013) Intrinsic variations in jaw deviation in English vowels. Proceedings of the International Congress of Acoustics. Proceedings of Meetings on Acoustics, 19, 060253.

Niebuhr, O., and Gutnyk, A. (2021). Pronunciation engineering: Investigating the link between jaw-movement patterns and perceived speaker charisma using the MARRYS cap. Proceedings of the 3rd IEEE International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Kuala Lumpur, Malaysia, pp. 1–6.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America*, 39(1), 151–168.

Selkirk, E. O. (1982). *The Syllable*. Dordrecht: Foris.

Smith, C., Erickson, D., and Savariaux, C. (2019). Articulatory and acoustic correlates of prominence in French: Comparing L1 and L2 speakers. *Journal of Phonetics Special Issue*, 77. https://doi.org/10.1016/j.wocn.2019.100938

Stone, M. (1981). Evidence for a rhythm pattern in speech production: Observations of jaw movement. *Journal of Phonetics*, 9, 109–120.

Summers, W. V. (1987). Effects of stress and final-consonant voicing on vowel production: Articulatory and acoustic analyses. *Journal of the Acoustical Society of America*, 82, 847–863.

Svensson Lundmark, M. (2023). Rapid movements at segment boundaries. *Journal of the Acoustical Society of America* , 153(3), 1452–1467.

Svensson Lundmark, M., and Erickson, D. (2024). Segmental and syllabic articulations: A descriptive approach. *Journal of Speech, Language, and Hearing Research*, 67(10S), 3974–4001. https://doi.org/10.1044/2024_JSLHR-23-00092

Svensson Lundmark, M., and Niebuhr, O. (2025). A practical guide to recording and analyzing jaw movements in speech research using MARRYS Proceedings from FONETIK 2025. Växjö: Linneaus University Press, pp. 51–56.

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press

Svensson Lundmark, M., Ambrazaitis, G., and Ewald, O. (2017). Exploring multidimensionality: Acoustic and articulatory correlates of Swedish word accents. Proceedings of Interspeech 2017, Stockholm, pp. 3236–3240.

Svensson Lundmark, M., Erickson, D., Niebuhr, O., Tiede, M., and Chen, W.-R. (2023). A new articulatory tool: Comparison of EMA and MARRYS, pp. 33–34. Abstract from Phonetics and Phonology in Europe 2023, Nijmegen, Netherlands.

Tabain, M. (2003). Effects of prosodic boundary on /aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America*, 113, 2834–2849.

Tiede, M. (2010). MVIEW: Multi-channel visualization application for displaying dynamic sensor movements. Unpublished manuscript.

Turk, A. E., and Sawusch, J. R. (1996). The processing of duration and intensity cues to prominence. *Journal of the Acoustical Society of America*, 99, 3782–3790.

Vatikiotis-Bateson, E., and Kelso, J. A. S. (1993). Rhythm type and articulatory dynamics in English, French and Japanese. *Journal of Phonetics*, 21, 231–265.

Westbury, J. R., and Fujimura, O. (1989). An articulatory characterization of contrastive emphasis in correcting answers. *Journal of the Acoustical Society of America*, 85(S1), S98. https://doi.org/10.1121/1.2027241

Weston, H., Svensson Lundmark, M., Erickson, D., and Niebuhr, O. (2023). Jaw movements in speech during physical activity. Poster session presented at Phonetics and Phonology in Europe 2023, Nijmegen, Netherlands.

Williams, J. C., Erickson, D., Ozaki, Y., et al. (2013). Neutralizing differences in jaw displacement for English vowels. Proceedings of the International Congress of Acoustics. POMA 19, 060268.

Wilson, I., Erickson, D., Kawahara, S., and Monou, T. (2019). Acquiring jaw movement patterns in a second language: Some lexical factors. Phonetic Society of Japan, Fall Meeting.

Wilson, I., Erickson, D., Vance, T., and Moore, J. (2020). Jaw dancing American style: A way to teach English rhythm. 10th International Conference on Speech Prosody 2020, pp. 556–560. https://doi.org/10.21437/SpeechProsody.2020-114

Yang, X., Shen, X., Li, W., and Yang, Y. (2014). How listeners weight acoustic cues to intonational phrase boundaries. *PLoS ONE*, 9(7), e102166. https://doi.org/10.1371/journal.pone.0102166

https://doi.org/10.1017/9781009295888.004 Published online by Cambridge University Press