PAPER

Interaction of jaw displacement and F0 peak in syllables produced with contrastive emphasis

Donna Erickson[1], Kiyoshi Honda[2] and Shigeto Kawahara[3]

1 Haskins Laboratories, 300 George Street, Suite 900, New Haven, CT, U.S.A.

2 School of Computer Science and Technology, Tianjin University, 135 Yaguan Road, Jinnan District, Tianjin 300350, China.

3 The Institute of Cultural and Linguistic Studies, Keio University, 2-15-45 Mita Minato-ku, 108-8345, Japan.

Abstract: In order to explore the articulatory nature of contrastive emphasis, this study compares contrastively emphasized and non-emphasized syllables in terms of mandible position and F0 peaks. The stimuli were English mono-syllabic words with /aɪ/, spoken in short utterances as part of read dialogues. Articulatory and acoustic data obtained by the University of Wisconsin x-ray microbeam facilities from six American English speakers were analyzed. The results show that for emphasized syllables, the jaw is lower and generally more front, and F0 is higher, compared to non-emphasized syllables. In addition to corroborating previous observations about larger jaw opening and higher F0 for emphasized syllables, our new finding is protrusion of the jaw in emphasized syllables. A possible hypothesis that we entertain in this paper is that fronting of the jaw may allow large jaw opening with high F0 target. We offer a tentative, yet concrete, hypothesis about the biomechanical interaction between F0 control and jaw opening mediated by anatomical connections between the jaw and the larynx.

1. INTRODUCTION

What is the articulatory mechanism behind the production of contrastive emphasis? How are the jaw and the larynx coordinated to make emphatically produced syllables? It has been observed that emphasized syllables tend to involve large jaw opening and high F0. Do these two features present a conflict between the two articulators, the jaw and the larynx? If so, how do speakers resolve that conflict? To address these questions, this study compares contrastively emphasized and non-emphasized syllables in terms of mandible position and F0 peaks. By studying how emphasis is implemented by our articulatory gestures, we hope to better understand the articulatory organization of speech. This investigation is guided by the theory in which the syllable is the organizational concatenative unit of speech, with jaw opening as the direct articulatory correlate of syllable prominence, e.g., [1] (although our descriptive finding is independent of this theoretical framework).

In this framework, the extent of jaw opening for the syllable nucleus is determined by several factors, [1-10]. One factor that affects jaw displacement is its vowel quality. That is, a low vowel has a low jaw position, and a high vowel, a high jaw position, [11-13]. The other factor, which is the main focus of this paper, is the effect of prosody, such as contrastive focus. Segmental and prosodic effects have been assumed to be independent in some previous studies, and we follow that assumption here [10, 14, 15] (although the current study controls for the vowel quality). In this framework, the prosodically-determined jaw setting affects the articulatory positions of the tongue and lips for producing specific phonological vowels through anatomical connections

What about control of F0 associated with an emphasized syllable? F0 control and jaw opening control seem to be at first sight independent functions—changes in F0 do not entail changes in jaw movement, or vice versa, - [6, 7, 10, 14, 16, 17]. Jaw opening for speech is performed by a complex set of articulatory muscles,

whereas F0 control is accomplished primarily by a combination of respiratory and laryngeal muscles. Generally, it is thought that increased cricothyroid muscle activity results in a raised F0, and relaxing the cricothyroid results in a lowered F0, [18-24]. Lowering F0 beyond a speaker's mid F0 range may involve the extrinsic strap muscles as well [20- 23].

However, considering the anatomical complexity of the whole speech production system, there may be a certain interaction between F0 control and jaw position. Such an interaction between phonation and articulation can be understood as our control strategy for facilitating speech production, [25-33]. Jaw opening for low vowels may impose a certain mechanical influence on the vocal folds to reduce their tension, resulting in lower F0. Further, Hirai et al. [34] and Honda et al. [35] proposed an extra-laryngeal physiological mechanism of F0 lowering by downward movement of the larynx along the curved cervical spine. The MRI images in [35] show that for vowel /a/, rotation of the cricoid cartilage can be facilitated by lowering of the jaw, hyoid bone and larynx, which results in lower F0. Overall, low jaw position seems to be by default associated with low F0, possibly because of the anatomical connection between the jaw and the larynx.

In contrastively emphasized speech, however, an opposite effect is found, with low jaw opening occurring with high F0. That is, emphasized syllables have more jaw opening than non-emphasized syllables, [2, 3, 4, 8, 9, 36, 37] and also, generally, higher F0 [26, 38-40] (although, emphasis can occur with low F0 as well [41-44]). High F0 with large jaw opening may present an articulatory challenge, if large jaw opening by default results in lower F0 for anatomical reasons. Do speakers resort to a special articulatory maneuver to cope with this apparent conflict? This is one of the main questions addressed in the current study.

Generally, jaw position, viewed on a sagittal plane, reflects a combined movement of rotation and vertical/horizontal translation, [45-51]. In phonetic descriptions, vertical jaw position has generally been used as a good indicator of jaw opening, representing a major function of the jaw gesture mainly by the rotational component of the movement [11, 52, 53]. The horizontal position of the jaw, as an indicator of the translational component of jaw rotation, has been treated as a minor element of speech gesture [53]. Despite this lack of focus on the horizontal dimension in the previous studies, translation is inevitable as the jaw opens wider [54]. Indeed, the results of our paper suggest that looking at the translational component of jaw rotation is crucial for

understanding the implementation of emphasis. This new finding is in and of itself a fresh contribution to the field.

The main question examined in this study is the interaction between jaw position and F0, when the speaker produces contrastively emphasized syllables, given that a high F0 must be achieved within the setting of a low jaw. To answer this question, this study examines jaw setting and peak F0 in the context of emphasized vs. non-emphasized mono-syllabic words in English.

2. METHODS

2.1. Data Recordings

Acoustic and articulatory recordings were made using the x-ray microbeam (XRMB) facilities at the University of Wisconsin [55-58]. Spherical gold pellets (2.5-3 mm in diameter) were affixed to selected points on the tongue, lips, and jaw of the speakers (Fig. 1). Two pellets were attached to the mandible, one at the lower incisor, and another on a molar tooth (MANm), and they were sampled at a rate of 40 samples/sec. Only the pellet attached to the mandible incisor was used for the analysis of the x-y movement of the jaw. In addition, reference pellets were affixed midsagittally to the nose bridge and to the anterior surface of the maxillary incisor. These references were used to project the articulatory pellet data onto the standard midsagittal coordinate system based on the maxillary occlusal plane (see [59] for more detailed descriptions).
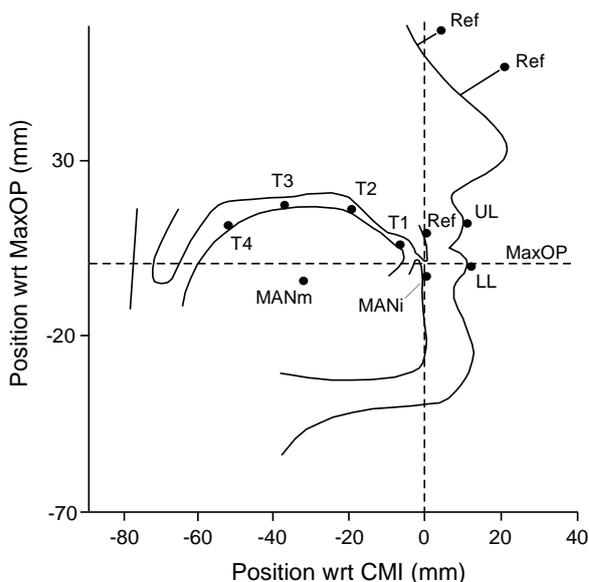

Fig. 1 Placement of pellets on tongue, lips, jaw, and reference points, based on [54. p. 37]. MaxOP is Max(illary) Occlusal Plane; CMI, Central Maxillary Incisors. The x-axis corresponded to the intersection of the midsagittal plane and the maxillary occlusal plane (MaxOP), and the origin of

the coordinate system was the lowermost edge of the maxillary incisor (CMI). The y-axis was normal to the maxillary occlusal plane, intersecting the plane at the origin. The coordinate value *y* represented the vertical distance from the maxillary occlusal plane to the center of the pellet sphere attached to the mandible incisor (MANi), and it is always negative

. In this paper, jaw opening is defined in terms of vertical mandible position; thus a maximum jaw opening corresponds to a minimum of vertical mandible position. Horizontal coordinate values represent the center of the lower incisor pellet relative to the tip of the upper incisor. Negative *x*-values indicate horizontal positions of the mandible pellet, which remain on the oral side of the tip of the upper incisors, and positive values indicate the amount by which the lower incisor pellet protrudes beyond the tip of the upper incisor.

2.2. Stimuli

Six American English speakers, (three male, three female) produced the question-answer sentence pairs like "Is it 599 Pine Street? No, it's 59FIVE Pine Street." The data-sets were collected in two stages: data for S1 (male), S2 (female), S3 (female) are from experiments by [36], and for S4 (male), S5 (female) and S6 (male) are from data collected in 1996 (a subset of which was

reported on in [3]—see also [8]). The same speech materials, shown in Table I, were used in these experiments.

Table I The stimuli, consisting of two digit-sequence types. NE indicates No Emphasis; E, Emphasis.

Digit Sequence Type 1:

Is it 5 9 5 Pine Street? Yes, it's 5 9 5 Pine Street.	(NE)
Is it 9 9 5 Pine Street? No, it's FIVE 9 5 Pine Street.	(E)
Is it 5 5 5 Pine Street? No. it's 5 NINE 5 Pine Street.	(E)
Is it 5 9 9 Pine Street? No, it's 5 9 FIVE Pine Street.	(E)

Digit Sequence Type 2:

Is it 9 5 9 Pine Street? Yes, it's 9 5 9 Pine Street.	(NE)
Is it 5 5 9 Pine Street? No, it's NINE 5 9 Pine Street.	(E)
Is it 9 9 9 Pine Street? No, it's 9 FIVE 9 Pine Street.	(E)
Is it 9 5 5 Pine Street? No, it's 9 5 NINE Pine Street.	(E)

The digit sequence in the answer was either "595" or "959," both of which contain /aɪ/. The utterances were randomized and read from a monitor screen, with the digit to be emphasized in capital letters, as in Table I. The emphasized digit occurred in initial, middle, or final position of the digit sequence. The utterances were also read with no emphasis when they constituted affirmative answers such as, "Is it 595 Pine Street? Yes, it's 595
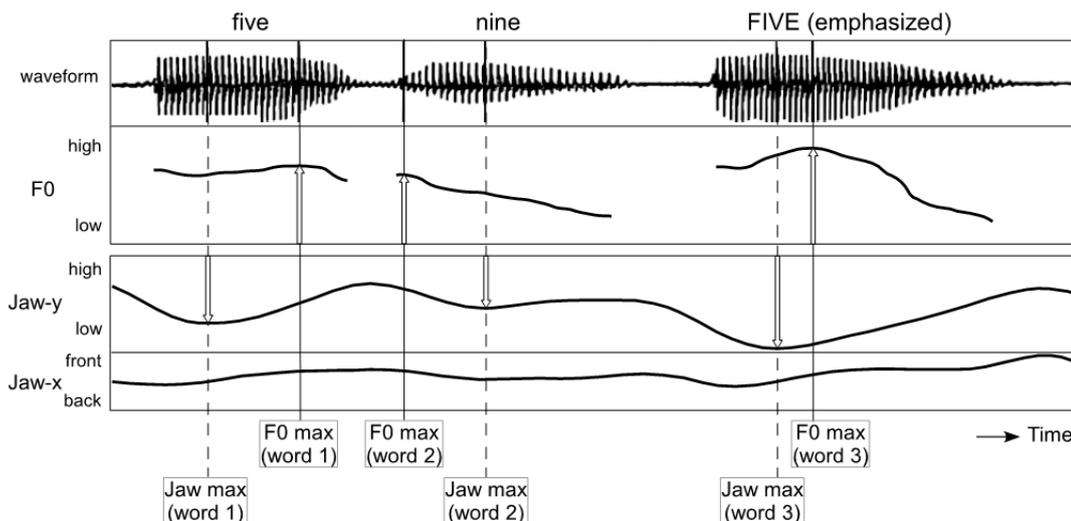

Fig. 2 A schematic figure to illustrate the identification of measurement points. The figure is created based on the utterance "five nine FIVE (emphasis)", but is schematized for the sake of exposition. The top panel: acoustic waveform; the second panel: F0 contour (50-150 Hz); the third panel: vertical jaw movement; the fourth panel: horizontal jaw movement. The measurement points are shown with vertical lines. Jaw measurement points are shown with dotted lines. F0 measurement points are shown with solid lines. *3*

Pine Street." The digit sequences only in the answer part of the question-answer utterances were analyzed in the current study. The six speakers read each of the question-answer sequences 71 to 113 times. The number of utterances in the data sets varied due to repeated collection for records with pellet mistracking (in most cases, mistracking took place for tongue pellets, not for the mandible pellet). Mistracking resulted in the different numbers of valid utterances for mandible data across the speakers. *N*s for each speaker are summarized in Table IV.

2.3. Measurements

Speech signals were recorded simultaneously with the articulatory data at a sampling rate of 22,000 samples/sec. The F0 contour was extracted using the auto-correlation-based F0 extraction program in WAVES+. Fig 2. provides a schematic figure to illustrate the identification of measurement points. The figure is created based on the utterance "five nine FIVE (emphasis)", but is schematized for the sake of exposition. Measurements were made for each of the digits, each containing the same vowel nucleus /aɪ/.

We measured two landmark points of the syllable: the points at which the jaw was maximally open (uisng the MATLAB-based algorithm check_beam, and illustrated here in Fig. 2 with the dashed vertical lines), and at which the F0 was maximally high during the voiced portion of the syllable (per the digital readout on the WAVES+ display program, and shown with the solid vertical lines)[1]. This measurement protocol was in keeping with the purpose of investigating the interaction of the prominence-determined jaw setting of the syllable with the peak F0 of the syllable. We did not measure mandible height and F0 at the same point in time, given the complexity of the relation between onset of muscle

---

[1] An anonymous reviewer asked whether it would be informative to measure F0 at the maximum jaw opening points. While we agree that this analysis is interesting, as shown in Figure 1, F0 maxima and jaw opening maxima do not necessarily coincide, and therefore it is not clear what F0 values at jaw opening maxima would represent. We believe, however, that exploring synchronization between F0 maxima and jaw opening maxima is an interesting and important question, although it is beyond the task of the present investigation. From Figure 1, at least, we do not observe a consistent synchronization pattern, and a more systematic study would be warranted to fully address this question.

activity and F0 event; often, as in Fig. 2, maximum jaw displacement and F0 displacement do not occur simultaneously—these two articulatory events often do not occur simultaneously, although it would not be surprising if they are synchronized to some extent.

2.4. Statistical Tests

Statistical differences between the non-emphatic condition and the emphatic condition were assessed by an independent-sample *t*-test. All the differences turned out to be statistically significant except for one condition. For the sake of exposition, the details of all the statistical results are collectively reported in Table IV in section 3.4.

3. RESULTS

3.1. Peak F0

Fig. 3 illustrates the average peak F0 across repetitions of the non-emphasis condition and the emphasis condition. The error bars represent standard errors. It shows that for all speakers, emphasized syllables have higher peak F0 than non-emphasized syllables. This pattern is in line with other studies of the relationship between F0 and emphasis, [26, 38, 39, 40]. See Table IV for detailed statistical comparisons, which show that the differences between the two conditions are significant for all the speakers.
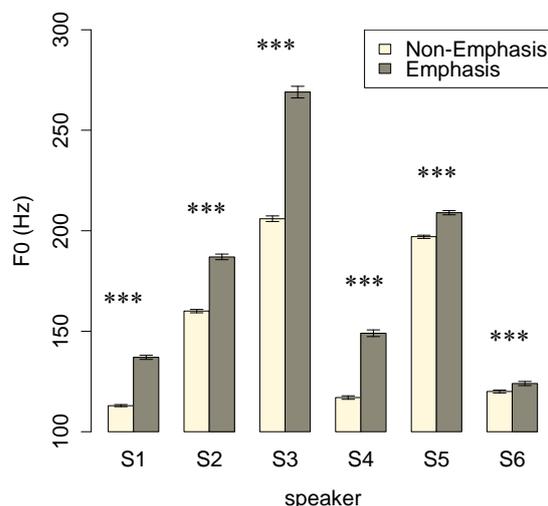

Fig. 3 The average peak F0 of the non-emphasis condition and the emphasis condition for all the speakers. Error bars represent standard errors (SE). All differences are significant at p<.001 level.

3.2. Vertical Jaw Displacement

Table II shows data for maximum and minimum jaw displacement among all the emphasized and non-emphasized target syllables. These values indicate

the range of low mandible position. It shows that there is some inter-speaker variation in terms of the degree of jaw opening. For example, Speaker 6 opens his jaw only 2.6 mm at his minimum, whereas Speaker 2 opens her jaw as much as 13.2 mm. Speaker 4 opens his jaw 11.8 mm at his maximum, whereas Speaker 5 can open her jaw up to 19.0 mm.

Table II Range (maximum and minimum) for each speaker of lowest mandible position, measured as vertical distance of the center of the lower incisor pellet in mm from the occlusal plane. The values are converted to absolute values for the sake of exposition.

SPEAKER	MAXIMUM (MM)	MINIMUM (MM)
S1 (M)	16.8	12.5
S2 (F)	14.9	13.2
S3 (F)	17.6	11.5
S4 (M)	11.8	5.0
S5 (F)	19.0	11.2
S6 (M)	13.1	2.6

Fig. 4 illustrates the mean vertical jaw displacement in the non-emphasis and emphasis conditions. It shows that, for all the speakers examined, emphasized syllables have a greater jaw opening than non-emphasized syllables. The results thus add to the body of studies that emphasis can result in larger jaw opening [2, 3, 4, 8, 9, 60-68].
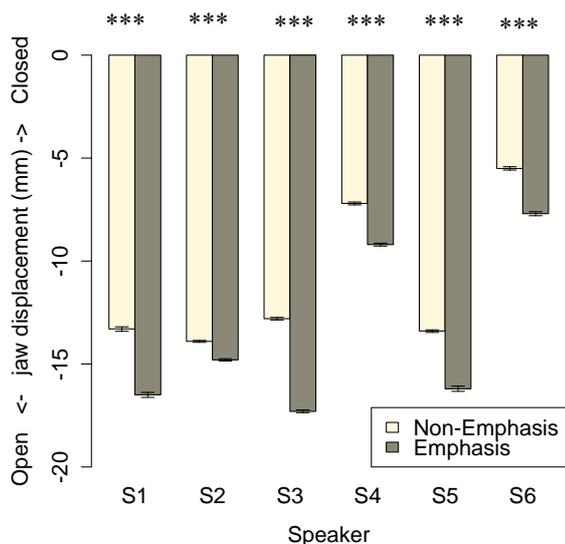

Fig. 4 Vertical jaw displacement for the non-emphasis and emphasis conditions for each speaker. The error bars are standard errors. All comparisons are significant at p<.001 level.

3.3. Horizontal Jaw Position

The horizontal jaw position was measured at the moment of the lowest jaw position for each of these syllables. The extreme values in the data for horizontal position are shown in Table III.

Table III The range of horizontal jaw displacement, measured at the time of maximum vertical jaw displacement

SPEAKER	MAXIMUM (MM)	MINIMUM (MM)
S1 (M)	0.00	-6.53
S2 (F)	-5.21	-7.52
S3 (F)	-0.72	-3.52
S4 (M)	0.57	-2.04
S5 (F)	1.08	-4.28
S6 (M)	5.68	0.34

The maximum and minimum in Table III refer to the extreme front and back values, respectively, for all the samples (including emphasized and non-emphasized) of horizontal jaw position. The more negative the number, the more retruded the position. (The values in the table may include the x-component of jaw rotation; when the jaw opens wider by rotation, the jaw pellet moves backward on the coordinate system shown in Fig. 1.) For five speakers, the lower incisor pellet was found to be located at or behind the origin (see the minimum values in Table 3). One speaker, S6, showed the mandible pellet moved maximally forward by 5.7 mm from the origin, the most retracted position being 0.3 mm in front of the upper incisor, indicating Class I occlusion. (Class I occlusion is such that "the maxillary first molar is slightly posterior to the mandibular first molar." (www.dentalcare.com)). Interestingly, this individual was among those exhibiting large horizontal mandibular movements, and is the one who shows a pattern that is different from the others in Fig 5.

The comparison between the non-emphasized and emphasized conditions is shown in Fig. 5. All the average values were negative (see Table IV). However, for the sake of illustration, it shows absolute values of the mean jaw displacement so that larger values indicate jaw fronting. Fig 5. shows that Speakers 1-4 have more front jaw position in the emphasized condition than in the non-emphasized condition. Speaker 5 shows the same trend, but the difference did not reach significance (see Table IV). Speaker 6 shows a significant reversal. The majority of the pattern is, thus, that speakers move their jaw more front for more emphasized condition—this

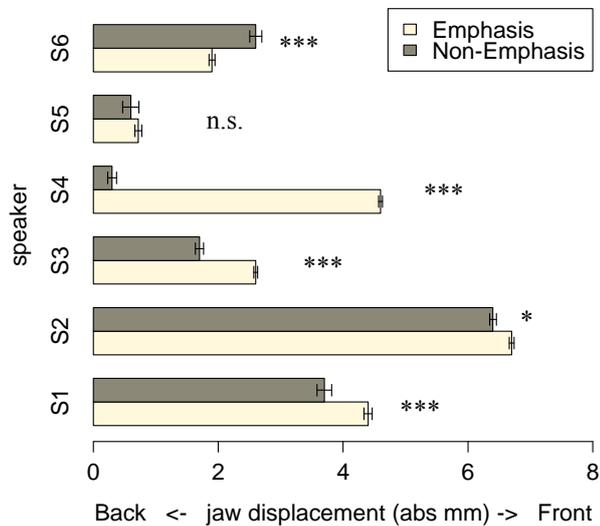observation, as far as we know, is a new descriptive finding.[2]

Fig. 5 Horizontal jaw displacement (in absolute mm). The actual mean values were all negative. For the sake of visual clarity, the values are shown in absolute values, in such a way that larger values represent more front jaw position. * = p < .05, *** = p < .001

3.4. Statistical Summary

Table IV shows the summary of statistical comparisons between the emphasized and non-emphasized conditions for each measurement discussed in this section. All the differences are assessed via an independent-sample *t*-test, which reveals that the two conditions are significant, except for the horizontal jaw displacement for Speaker 5.

4. DISCUSSION

---

[2] Since the data were collected in the 1990's, the original raw data are no longer available. In order to test the reliability of this new finding, which we believe is the core value of this work, we restored data points based on random sampling using normal distribution based on the means and standard deviations that were available (see Table IV), using R [69]. Based on these restored data, we ran a linear mixed model with speaker as a random variable, including both its intercept and slope [70-73], using the `lmer` function of the `lme4` package [74]. The resampling based on Speakers 1-4 turned out to be significant, but the resampling based on all the speakers turned out to be non-significant. These resampling exercises show that the conclusion that speakers move their jaw forward for the emphasis condition is secure for Speakers 1-4.

4.1. Summary of Findings

Analysis of mandible position based on the x-ray microbeam data pertaining to a low vowel indicates that emphasized syllables involve (i) larger jaw opening, (ii) more forward jaw (4 out of 6 speakers), and (iii) higher F0. The findings of more open jaw with emphasis have been reported in previous studies. The higher F0 in emphasized syllables has also been known. In addition to corroborating these observations, what is particularly new in this study is that, at least for 4 out of 6 speakers, protrusion of the jaw appears to occur with high F0 to realize open vowels on emphasized syllables. Specifically, the results show different effects of rotation and translation of the jaw: the jaw tended to advance for vowel emphasis in emphasized syllables with high F0. A possible mechanism for counterbalancing F0 fall by jaw translation is proposed as below.

4.2. Biomechanical Account of Mandible-F0 Interaction in Emphasized Syllables

This section offers a tentative, yet concrete and testable hypothesis about how to account for the pattern of jaw position and F0 in terms of certain anatomical constraints of the jaw and larynx that underlie the observed phenomenon. Biomechanical connections by the soft tissues exist between the jaw-hyoid and the hyoid-larynx complexes as shown in Fig. 6(a).

The mandible and the hyoid bone are interconnected by several muscles as well as other soft tissues: the digastric (anterior belly) (not shown in Fig 6a), mylohyoid, and geniohyoid muscles directly, and other extrinsic lingual muscles indirectly (not shown). The thyroid cartilage is suspended from the hyoid bone by the thyrohyoid muscle and other membranes or ligaments. These soft tissue connections (shown in the figure) form the thyrohyo-mandibular chain [75, 76] and cause a mandible-larynx interaction in speech articulation. When the jaw opens, certain biomechanical effects are expected to occur with respect to the relations among these structures. The jaw opening by rotation alone (filled arrow) can cause backward translation of the hyoid bone (shaded grey) and consequent rotational separation of the thyroid cartilage from the cricothyroid around their joint (clockwise open arrow) because of the passive elasticity of the cricothyroid muscles and other tissues in the anterior portion of the larynx, as well as the horizontal friction in the more posterior area around the thyrohyoid muscle and other tissues (Fig. 6b). This change of state of the cricothyroid joint is opposite to the (vocal-fold lengthening) action of the cricothyroid muscle, and results in shortening of the vocal folds, as-

suming that the cricothyroid is not contracting at the same time. All other things being equal, jaw opening by rotation around the temporomandibular joint (TMJ) can lower F0 by means of the thyro-hyo-mandibular chain. This explains the tendency for jaw lowering for low vowels to accompany low F0 as described in the literature, [77-80]. For example, a cineradiographic study [77] reported a correlation between low jaw position and low F0 (with vowels with intrinsically low F0, e.g., /a/) as opposed to high jaw and high F0 (with vowels with intrinsically high F0, e.g., /i/).

The findings from this study may suggest a deliberate use of jaw advancement to enable an increase F0 when the jaw is low. If the jaw translates forward as the jaw opens by rotation, the effect of jaw rotation on the hyoid bone can be counteracted, and the consequent F0 lowering effect can be minimized or reversed. A possible—though admittedly tentative—mechanism that we propose for this compensatory adjustment is shown in Fig. 6c. When a wide opening of the jaw and a high F0 are both required for syllable prominence, the jaw translates forward, probably by contraction of the lateral pterygoid, so that forward hyoid position for high F0 and open jaw position for enhanced magnitude of the

when the jaw is more open for the low vowel to begin with. In other words, one way to counteract the tendency for jaw opening to accompany F0 lowering would be to protrude the jaw. The average distance between the instances of the most protruded and most retruded positions as measured at the time of maximum jaw opening, observed for the six speakers in this study, was 3.6 mm, ranging from 6.5 mm (SI) to 1.5 mm (S4). In dental studies, the maximum range of human jaw translation is about 10 mm [46]. This value is obtained by a measurement with a closed jaw from the maximally retruded to maximally protruded position; in our study, range of horizontal jaw position measurements are made with an open jaw. Although the measurements in our study and the dental measurements are not equivalent, the ranges of displacement of the jaw observed for our speakers seem to occupy a fair portion of the physiologically feasible range of jaw translation.

Some questions of course remain—recall that we found that only 4 out of 6 speakers showed the significant fronting. Therefore, there is a question of how the proposed mechanism generalizes to other English speakers (or speakers of other languages, for that matter). And why do we observe a reversal for Speaker 6? This

(a) Thyro-hyo-mandibular chain

(b) CT-joint rotation to lower F0 by jaw rotation

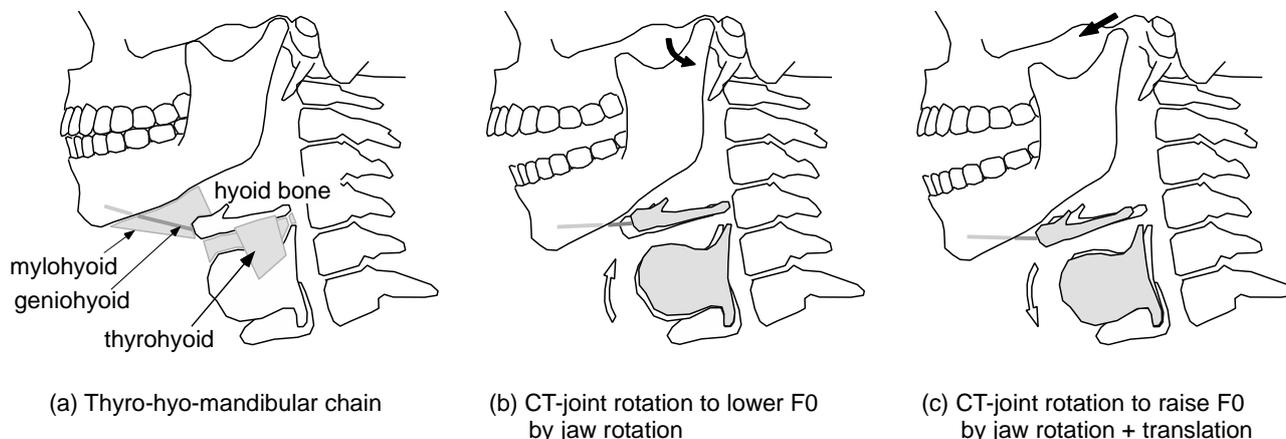(c) CT-joint rotation to raise F0 by jaw rotation + translation

Figure 6. Possible effects of jaw rotation and translation on the larynx, (a) Anatomical relationship of the jaw, hyoid bone, and laryngeal cartilages, (b) Jaw opening by rotation can cause a reverse rotation of the cricothyroid joint, effectively shortening the vocal folds, (c) Jaw translation pulls the hyoid bone forward to counteract the F0-lowering effect of jaw opening.

syllable are simultaneously attained.

As the hyoid bone moves forward, the thyrohyoid muscle and membrane are stretched. Consequently, the elasticity of the tissue tends to rotate the thyroid cartilage around the cricothyroid joint in the same direction as that of the rotation brought about by cricothyroid muscle contraction. Thus, the jaw protrusion, in effect, can stretch the vocal folds by a chain of mechanisms, resulting in a higher F0 than otherwise, particularly

speaker behaves differently in terms of minimal and maximal displacements from other speakers as well (see Table III). Specifically, his jaw seems to protrude slightly more compared with the other speakers, which might account for his larger vertical and horizontal mandibular movements (i.e., a more protruded jaw can open wider, in that mandibular protrusion makes the post-mandibular space wider and facilitates wider jaw opening, e.g., [54]).

What roles do individual and genetic differences in occlusal class/jaw morphology play in this story, e.g., [81]? Since jaw protrusion makes the post-mandibular space wider, this facilitates wider jaw opening, and vice versa. Thus, a person with a protruded jaw is able to open the jaw wider when talking, whereas one with a retruded jaw opens their jaw less widely in speaking.

Unfortunately, the x-ray microbeam no longer functions to address these questions, but a similar follow-up experiment is possible using EMA. In order to accurately decompose the rotation and translation components at the condyle center for rotation, two pellets are needed to be placed on the jaw. Other innovative techniques, such as 3D fluoroscopy [82], might be helpful for assessing the components of jaw rotation.

On the one hand, the mechanism that we propose, especially (6c) remains tentative. However, we would like to make clear that (i) our descriptive finding that the jaw is fronted for emphasized syllables seems clear, and that (ii) our findings open up several research projects that need to be conducted in order to understand how human speakers express emphasis.

5. SUMMARY

This paper examined how peak F0 in a syllable interacts with the prosodically-determined jaw setting of emphasized low vowels. The results showed that the mandible moves forward, in addition to being lowered. Timing relationship between maximum jaw displacement and F0 targets is a topic we hope to explore in the future. This forwarding gesture may be required to achieve both large jaw opening and high F0. The biomechanics of the two biomechanical systems, the jaw and the larynx, under the anatomical constraints of their mutual interaction, accounts for the horizontal jaw movement when the jaw setting is extremely low, as required for a contrastively emphasized syllable on a low vowel. The findings from this study about the biomechanical interaction between the articulatory and phonatory organs can also be used toward developing a realistic biological model of speech production, e.g., [80, 83, 84], which is a necessary ingredient of a comprehensive model of phonetic organization [7, 85, 86].

ACKNOWLEDGMENTS

REFERENCES

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

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

[3] D. Erickson, "Articulation of extreme formant patterns for emphasized vowels," *Phonetica,* **59**, 134-159 (2002).

[4] D. Erickson, "The jaw as a prominence articulator in American English," *Acoust. Soc. Jpn Fall Meeting*, 311-312 (2003).

[5] D. Erickson, O. Fujimura and B. Pardo, "Articulatory correlates of prosodic control: Emotion and emphasis.," *Language Speech,* **41**, 399-417 (1998).

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

[7] O. Fujimura, "Stress and tone revisited: Skeletal vs. melodic and lexical vs. phrasal," In S. Kaji (ed.) *Cross-Linguistic Studies of Tonal Phenomena: Historical Development, Phonetics of Tone, and Descriptive Studies*, 221–236. Tokyo University of Foreign Studies, Tokyo (2003).

[8] C. Menezes, B., Pardo, D. Erickson and O. Fujimura, "Changes in syllable magnitude and timing due to repeated correction," *Speech Communication,* **40**, 71-85 (2003).

[9] C. Menezes, "Comparing the metrical structure of a digit sequence in American English: Articulatory vs. acoustic analysis," *J. Phonetic Soc. Jn,* **19**, 70–77 (2015).

[10] D. Erickson and S. Kawahara, "Articulatory correlates of metrical structure: Studying jaw movement patterns," *Linguistic Vanguard ,* 2 (2016).

[11] P. Keating, A. B. Lindblom, J. Lubker and J. Kreiman, "Variability in jaw height for segments in English and Swedish VCVs," *J.of Phonetics,* **22**, 407–422 (1994).

[12] S. Kawahara, H. Masuda, D. Erickson, J. Moore, A. Suemitsu, and Y. Shibuya, " Quantifying the effects of vowel quality and preceding consonants on jaw displacement: Japanese data," *J. Phonetic Soc. Jn,* **18**, 54-62 (2014).

[13] S. Kawahara, D. Erickson, D. and A. Suemitsu, "A quantitative study of jaw opening: An EMA study of Japanese vowels" (Submitted).

[14] C. Menezes and D. Erickson, "Intrinsic variations in jaw deviation in English vowels," *Proceedings of International Congress of Acoustics, POMA* **19**, #060253 (2013).

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

[16] S. Kawahara, D. Erickson, J. Moore, A. Suemitsu and Y. Shibuya, "Jaw displacement and metrical structure in Japanese: The effect of pitch accent, foot structure, and phrasal stress," *J. Phonetic Soc. Jpn* **18,** 77-87 (2014).

[17] D. Mücke, M. Grice, and T. Cho, " More than a magic moment - Paving the way for dynamics of articulation and prosodic structure," *J. of Phonetics* **44**, 1-7 (2014).

[18] G. E. Arnold, " Physiology and pathology of the cricothyroid muscle," *Laryngoscope,* **71**, 687-753 (1961).

[19] J. Ohala and H. Hirose, "The function of the sternohyoid muscle in speech,"*Acoustic. Soc. Jpn n Fall Meeting*, 359-360 (1969).

[20] Z. Simada and H. Hirose, "Physiological correlates of Japanese accent patterns," *An.Bull. Res.Inst. Logopedics and Phoniatrics Univ. Tokyo*, **5**, 41-49 (1971).

[21] D. Erickson, *A physiological analysis of the tones of Thai*. Ph.D. Dissertation, University of Connecticut (1976).

[22] D. Erickson, D. Laryngeal muscle activity in connection with Thai tones. *Festschrift in Honor of Professor Hajime Hirose, RILP, University of Tokyo,* **27**, 135-149 (1993).

[23] D. Erickson, "Thai tones revisited," *J. Phonetic Soc.Jpn.*,**15.2**, 1-9 (2011).

[24] J. E. Atkinson, " Correlation analysis of the physiological features controlling fundamental frequency," *J. Acoust. Soc. Amer*. **63**, 211-222 (1978).

[25] P. Ladefoged, "Some possibilities in speech synthesis," *Language and Speech,* **7**, 205-214 (1964).

[26] I. Lehiste, *Suprasegmentals*. (MIT Press, Cambridge, MA, 1970).

[27] P. Ladefoged, J. DeClerk, and R. Harshman, "Control of the tongue in vowels," *Proc. 7th International Congress of Phonetic Sciences,* 349-354 (1972).

[28] M. Rossi, and D. Autesserre, "Movements of the hyoid and the larynx and the intrinsic frequency of vowels, "*J. of Phonetics,* **9**, 233-249 (1981).

[29] K. Honda, "Relationship between pitch control and vowel articulation," In: Titze, I. and Scherer, R. (eds.) *Vocal Fold Physiology: Biomechanics, Acoustics, and Phonatory Control*, 113-126. Denver Center for the Performing Arts, Denver (1983).

[30] K. Honda and O. Fujimura, "Intrinsic vowel F0 and phrase-final F0 lowering: Phonological vs. Biological explanations," In W. J. Hardcastle and A. Marchal *(eds.) Speech Production and Speech Modelling*, 149-157. Kluwer Academic Publishers, Dordrecht (1990).

[31] S. Shapir, "The intrinsic pitch of vowels: Theoretical, physiological, and clinical considerations. "*J. Voice,* **3**, 44-51(1989).

[32] D. Walen and A. Levitt, "The universality of intrinsic F0 of vowels," *J. of Phonetics,* **23**, 349-366 (1995).

[33] D. Whalen, B. Glick, M. Kumada and K. Honda, "Cricothryroid activity in high and low vowels: Exploring the automaticity of intrinsic F0," *J. of Phonetics,* **27**, 125-142 (1988).

[34] H. Hirai, K. Honda, I. Fujimoto and Y. Shimada, "Analysis of magnetic resonance images on the physiological mechanisms of fundamental frequency control," *J. Acoust. Soc. Jpn,.* **50**, 296-304 (in Japanese) (1994).

[35] K. Honda, H. Hirai, S. Masaki and Y. Shimada , "Role of vertical larynx movement and cervical lordosis in F0 control," *Language Speech,* **42**, 401-411, (1999).

[36] J. R. Westbury and O. Fujimura, "An articulatory characterization of contrastive emphasis," *J. Acoust. Soc.Am.* **85**, (A), S98 (1989).

[37] D. Erickson and O. Fujimura, "Maximum jaw displacement in contrastive emphasis," *Proc. of ICSLP96 Philadelphia* , **1**, 141-144 (1996).

[38] S. J. Eady, W. Cooper, G. V. Klouda, P. R. Mueller and D. W. Lotts, "Acoustical characteristics of sentential focus: Narrow vs. broad focus and single vs. dual focus environments," *Language and Speech,* **29**, 233-251 (1986).

[39] Y. Xu and CX Xu, "Phonetic realization of focus in English declarative intonation.," *J. of Phonetics,* **33**, 159- 197 (2005).

[40] J. Katz and E. Selkirk, "Contrastive focus vs. discourse-new: Evidence from prosodic prominence in English," *Language,* **87**, 771-816 (2011).

[41] K. Silverman, M. E. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert and J. Hirschberg, "TOBI: a standard for labeling English prosody," *Proc. of 1992 International Conference on Spoken Language* Processing, 867-870 (1992).

[42] D. Erickson, K. Honda, H. Hirai, and M. Beckman, "The production of low tones in English intonation," *J. of Phonetics* ,**23**, 179-188 (1995).

[43] C. Menezes, D. Erickson, and O. Fujimura, "Contrastive emphasis: Comparison of pitch accents with syllable magnitudes,"*Proc. Speech Prosody 2002 Aix-en-Provence*, 495-497 (2002).

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

[45] B. G. Sarnat, *The temporo mandibular joint*. (Thomas, Springfield, IL, 1964).

[46] U. Posselt, *Physiology of occlusion and rehabilitation*. (Blackwell Scientific, Oxford, 1968).

[47] J. R. Westbury, "Mandible and hyoid bone movements during speech," *J. Speech Hear. Res.,* **31**, 405-416 (1988).

[48] J. Edwards and K. S. Harris, "Rotation and translation of the jaw during speech," *J. Speech Hear. Res.,* **33**, 550-562 (1990).

[49] E. Vatikiotis-Bateson and D. J. Ostry, "Analysis and modeling of 3D jaw motion in speech and mastication," *IEEE Xplore Conference: Systems, Man, and Cybernetics,* **2** DOI: 10.1109/ICSMC.1999.825301 (1999).

[50] E. Vatikiotis-Bateson and D. J. Ostry, "An analysis of the dimensionality of jaw motion in speech," *J. of Phonetics*, **23**,101-117 (1995).

[51] R. Laboissière, D. Ostry and A. Feldman, "The control of human jaw and hyoid movement," *Biological Cybernetics* ,**74**, 373-384 (1996).

[52] R.-M. S. Heffner, *General Phonetics* (Univ. Wisconsin Press, Madison, 1964).

[53] J. Clark and C. Yallop, *An Introduction to Phonetics and Phonology*. (Blackwell, Cambridge, 2nd ed., 1995).

[54] https://en.wikipedia.org/wiki/Temporomandibular_joint (accessed 9 Sept. 2016).

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

[56] S. Kiritani, K. Itoh and O. Fujimura, "Tongue-pellet tracking by a computer-controlled x-ray microbeam system," *J. Acoust. Soc. Am.* **57**, 1516-1520 (1975)..

[57] R. D. Nadler, J. H. Abbs and O. Fujimura, "Speech movement research using the new x-ray microbeam system," *Proc. of International Congress Phonetic Sciences 87 Tallinn, Estonia,* **1**, 221-224 (1987).

[58] J. R. Westbury, *X-ray microbeam speech production database handbook user's handbook (1994).* Downloadable at://www.haskins.yale.edu/staff/gafos_downloads/ubdbman.

[59] J. R. Westbury, "The significance and measurement of head position during speech production experiments using the x-ray microbeam system," *J. Acoust Soc. Am.* **89**, 1782-1791 (1991)

[60] J. Perkell, *Physiology of speech production*. (Cambridge: MIT Press, 1969).

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

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

[63] W. V. Summers, "Effects of stress and final consonant voicing on vowel production: articulatory and acoustic analyses," *J. Acoust. Soc. Amer.*, **82**, 847–863 (1987).

[64] M. Macchi, *Segmental and suprasegmental features and lip and jaw articulations*. Doctoral dissertation, New York University (1985).

[65] K. Oshimat and V. L. Gracco, "Mandibular contributions to speech production," *Proc. of International Conference on Spoken Language Processing, Banff.* i92_0775 (1992).

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

[67] K. de Jong, "The supraglottal articulation of prominence in English: linguistic stress as localized hyperarticulation," *J. Acoust. Soc. Amer.*, **97**, 491–504 (1995).

[68] J. Harrington, J. Fletcher and M. E. Beckman, "Manner and place conflicts in the articulation of accent in Australian English," In M. Broe, J. Pierrehumbert (eds.) *Papers in laboratory phonology*

V, 40–51. (Cambridge University Press, Cambridge, 2000).

[69] R Development Core Team. 1993–2016. R: A language and environment for statistical computing. Vienna, Austria. R Foundation for Statistical Computing. Software, available at http://www.R-project.org.

[70] D. J. Barr, "Random effects structure for testing interactions in linear mixed-effects models," *Frontiers in Psychology* (2013).

[71] D. J. Barr, R. Levy, C. Scheepers, and H. J. Tily, "Random effects structure for confirmatory hypothesis testing: Keep it maxima," *J. of Memory and Language,* **68**, 255–278 (2013).

[72] D. Bates, "Fitting linear mixed models in R," *R News*, **5**, 27–30 (2005).

[73] F. T. Jaeger, "Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models," *J. of Memory and Language* **59**, 434–446 (2008).

[74] D. Bates, M. Maechler and B. Bolker, " lme4: Linear mixed-effects models using S4 classes," R package (2011).

[75] W. Zenker*, "Questions regarding the function of external laryngeal muscles",* I*n Research Potentials in Voice Physiology***,** 20-40**. (**State University of New York Press, New York, 1964).

[76] A. Sonninen, "The external frame function in the control of pitch in the human voice", *Ann. of the New York Academy of Sciences*, **155,** 68-90 (1968).

[77] P. A. Zawadzki and H. R. Gilbert, "Vowel fundamental frequency and articulator position," *J. of Phonetics,* **17**,159-166, (1989).

[78] M. Sawashima, H. Hirose, H. Yoshioka, and S. Kiritani, "Interaction between articulatory movements and vocal pitch control in Japanese word accent," *An.Bull. Res.Inst. Logopedics and Phoniatrics Univ. Tokyo,* **16**, 11-19, (1982).

[79] M. Sawashima, H. Hirose, H. Yoshioka, S. Horiguchi and S. Kiritani, "Interaction between jaw movements and vocal pitch control," In A. Cohen and M. P. R. v. d. Broecke *(eds.) Abstracts of the Tenth International Congress of Phonetic Sciences*, 454. (Foris Publications, Utrecht, 1983).

[80] H. Hirai, J. Dang, and K. Honda, "A physiological model of speech organs incorporating tongue-larynx interaction," *J. Acoust. Soc. Jpn.*, **51**, 918-928 (in Japanese) (1995).

[81] Y. Kamijo, *Oral Anatomy I* (Craniology), Anatom-sha, Tokyo (in Japanese)(1979).

[82] C.-C. Chen, C.-C. Lin, T.-W. Lu, H. Chiang, and Y.-J. Chen "Feasibility of differential quantification of 3D temporomandibular kinematics during various oral activities using a cone-beam computer tomography-based 3D fluroroscopic method," *Journal of Dental Sciences* (2013), http://dx.doi.org/10.1016/j.jds.2012.09.025

[83] J. Dang and K. Honda, "A physiological articulatory model for simulating speech production process," *Acoust. Sci. and Tech.* **22.6**, 415-425 (2001).

[84] X. Wu and J. Dang, "Control strategy of physiological articulatory model for speech production*," J. Chinese Linguistics* **43. 1B**, 337-363 (2015).

[85] O. Fujimura, "Phonology and phonetics~a syllable based model of articulatory organization," *J. Acoust. Soc. Jpn*., **13**, 39-48 (1992).

[86] O. Fujimura, "C/D model: A computational model of phonetic implementation,"In E. S. Ristad, E. S. (ed.) *Language Computations*, 1-20. (Am. Math. Soc, Providence, R.I., 1994).

[87] D. Erickson and K. Honda, "Jaw displacement and F0 in contrastive emphasis," *J. Acoust. Soc. Am.*, **99**, 2494 (A). (1996).