MANDIBULAR CONTRIBUTION TO VOWEL-INTRINSIC F0

Wei-Rong Chen[1], D. H. Whalen[1,2,3], Mark K. Tiede[1]

[1]Haskins Laboratories, [2]CUNY Graduate Center, [3]Yale University
chenw@haskins.yale.edu; whalen@haskins.yale.edu; tiede@haskins.yale.edu

ABSTRACT

The vowel-intrinsic fundamental frequency (IF0) is a universal tendency for high vowels to have higher F0 than low vowels. The "tongue pull" hypothesis is the most successful account of IF0, but other factors seem to play a role as well. Few studies have investigated the articulatory correlates of IF0, and their results are somewhat inconsistent. Here we extended such investigation and analyzed the data from two large articulatory corpora with 40 speakers and 124,341 vowel samples. Our results showed that both tongue height and jaw height significantly correlate with F0. However, after we removed the tongue height effect from F0, the jaw effect remains the same degree of correlation with the residual, while the reverse was not true for the tongue height effect. Thus our results support the hypothesis that the underlying mechanism for IF0 is at least as much controlled by mandibular position as it is by tongue height.

Keywords: Intrinsic F0, vowel articulation, tongue, jaw

1. INTRODUCTION

Vowel-intrinsic fundamental frequency (IF0), the tendency for high vowels to be produced with higher F0 than low vowels, has been reported on since the early 1900s, e.g. [1, 18, 21]. Despite the fact that the mechanism behind IF0 is still in debate (see [7] for a comprehensive review), it does appear that IF0 is a language universal (e.g., [24]) and F0-dependent (IF0 difference is larger in higher pitch range but smaller or disappears in lower pitch) (e.g., [9, 19, 20, 24]). Many hypotheses can be found in the literature to account for the underlying mechanisms for IF0. A commonly accepted account is the so-called "tongue pull hypothesis" (some researchers call it "physiological hypothesis" or "mechanical account"), which proposes a physical link between tongue articulation and the tension of the vocal folds. It was first proposed by Ladefoged in [10] and has then been revised (e.g., [5, 6, 7, 11, 14]). With the evidence of EMG data, Honda [6] proposed that the contraction of posterior genioglossus may simultaneously support tongue raising and a forward pull of the hyoid bone, which in turn rotates the thyroid cartilage and lengthens the vocal folds. However, debate about the exact details of this biomechanical link between supraglottal articulation and phonation continues.

The tongue pull hypothesis also implies that IF0 can be seen as a positive correlation between tongue height and F0 and thus is gradient in nature tracking vowel height. This gradient view of IF0 has been challenged in some studies. For example, German tense /e:/ has higher tongue position but lower F0 than the lax /ɪ/ (e.g., [8]).

To further explain IF0 in terms of vowel articulation, a few other studies have looked at the articulatory correlates of IF0. Zawadzki and Gilbert [26] found that the vertical position of the mandible was more closely related to IF0 than tongue height in three of five American English speakers, using cineradiography. Fisher-Jørgensen [4], measuring jaw and lip opening with video and tongue height with palatography, also found that jaw (and lip opening) were in better agreement with IF0 than tongue height in five German speakers. Conversely, Pape and Mooshammer [16] measured three German speakers with EGG and EMMA; two speakers showed the highest correlation of tongue height with F0, while for the third speaker, the articulator that correlated highest with F0 was the jaw.

Given these inconsistent results and limited amounts of data reported, the aim of this study is to carry out a more thorough investigation on the articulatory correlates of IF0, with two large articulatory corpora.

2. METHOD

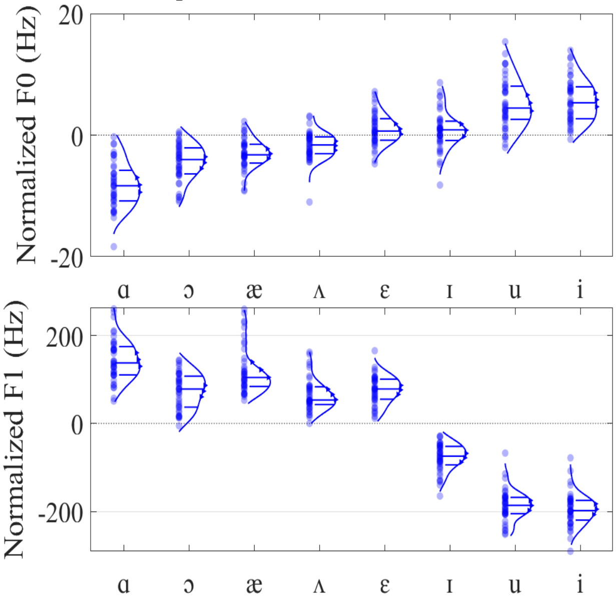
We selected 8 monophthong vowels of American English (/ɑ, ɔ, æ, ʌ, ɛ, ɪ, u, i/) (with primary stress) from 32 speakers (17 females) in the University of Wisconsin x-ray microbeam database (XRMB, [23]), and from 8 speakers (4 females) in the Haskins IEEE Rate Comparison Database (HIRCD, [22]) (total number of vowel samples = 124,341). Both corpora simultaneously recorded acoustic and mid-sagittal articulatory data of tongue, jaw and lip movements in running speech. The tongue measurements in XRMB were four points on the tongue (T1: ~1cm posterior to tongue apex; T2: ~1.5cm from T1; T3: ~3cm from T1; T4: ~4.5cm from T1), whereas those in HIRCD were three points (T1: ~1cm from tongue apex; T2: ~1.7cm from T1; and T3: ~3.4cm from T1). In order to carry out comparable analyses across two corpora, we defined the tongue height '*maxTy*' as the highest vertical position on the tongue (from any tongue sensor), and the jaw height '*JH*' as the first principal component of jaw position with the positive sign set to indicate upward movement. We also created a new variable "*-F1*" as the negative first formant, in order to have the same sign of correlation with the articulatory parameters. Formants were measured by LPC (45ms window, 2ms step, 14 poles, pre-emp. from 50 Hz) and tracked by the Viterbi algorithm, and F0 values were calculated by the autocorrelation method with the F0 range properly set for each individual speaker in PRAAT [2] (Ver:6.0.43). We carried out a series of Pearson correlation analyses. For each correlation analysis, outliers were removed by the 'elbow method', as described in [25].

3. RESULTS

3.1. Acoustic correlates of IF0

Figure 1 presents an overview of IF0 by plotting the normalized F0 (upper) and F1 (lower) frequencies by vowels. Each data point represents the median value of a vowel produced by one speaker. Each curved line represents the distribution (probability density function) of 40 speakers for each vowel (see the legend of Figure 2 for the definitions of distribution). F0 and F1 values were normalized by subtracting the median across all vowels separately for each speaker.

Figure 1: Normalized F0 (upper) and F1 (lower) for eight vowels. Each point indicates the median of a vowel for a speaker.

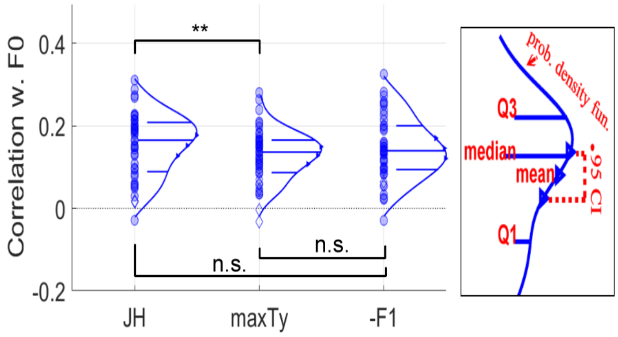


An IF0 effect can be observed as an increasing trend of F0 from low to high vowels, parallel to a general decreasing trend of F1 for the same order of vowels. However, an exception to the pattern of IF0 was observed for the /ɛ, ɪ/ pair, where a difference in nominal tongue height and a large difference in F1 did not correspond to a difference in F0. In general, despite changes in prosodic context, IF0 was observed, and the mean difference in F0 between /a/ and /i/ is 13.8 Hz, comparable to the previously reported ranges of IF0 for American English (e.g., [24]).

3.2. Articulatory correlates of IF0

In the following text, we will abbreviate the correlation of F0 with jaw height as *Cor(F0, J)*, with tongue height as *Cor(F0, T)*, and with *–F1* as *Cor(F0, -F1)*. Figure 2 summarizes the results of *Cor(F0, J)*, *Cor(F0, T)* and *Cor(F0, -F1)*. Each symbol indicates a correlation coefficient calculated separately for a speaker. Shaded circles indicate significant correlation coefficients while non-significant ones are marked with unfilled diamonds. Positive values of *Cor(F0, -F1)* represent the degree of IF0 for each speaker. The positive correlations of *Cor(F0, J)* and *Cor(F0, T)* are of similar degree across speakers, while paired *t*-tests show that *Cor(F0, J)* is significantly higher than *Cor(F0, T)* (p = .004; t = 3.1), although the effect size is medium (Cohen's d = 0.48). The horizontal positions of all articulators did not show appreciable correlations with F0 and are thus not reported here. And, as expected, *maxTy* and *JH* were also significantly correlated for all speakers (mean correlation coefficient = 0.46).

Figure 2: Articulatory and acoustic correlates of IF0. Each blue circle indicates the correlation of F0 with each articulatory or acoustic parameter (*JH*, *maxTy*, and *–F1*) for one speaker. Unfilled diamond markers indicate the individual correlation is not significant.



3.3. Subset with uncorrelated tongue and jaw heights

To further distinguish the contributions of tongue and jaw heights to F0, we created two subsets of the data such that tongue and jaw heights were uncorrelated. Specifically, these subsets contain vowels produced with higher (above median) tongue heights and lower (below median) jaw heights, or with lower tongue heights and higher jaw heights. Figure 3 demonstrates the scatter plot of *maxTy* against *JH* for the speaker M04. The origin indicates the medians of both dimensions. These uncorrelated subsets are the tokens within the 2nd and 4th quadrants in Figure 3; these retain around 20~40% of data for each speaker. In effect, the 2nd quadrant consists of high vowels produced with low jaw positions, and the 4th quadrant low vowels with high jaw positions.

Figure 3: Scatter plot of jaw and tongue positions for the speaker M04. The origin indicates the speaker medians in both dimensions.

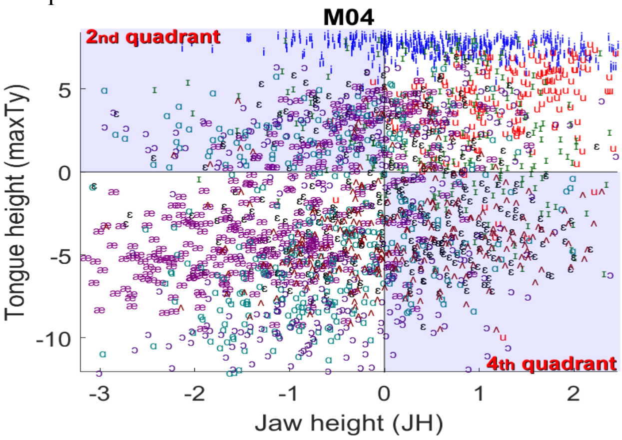
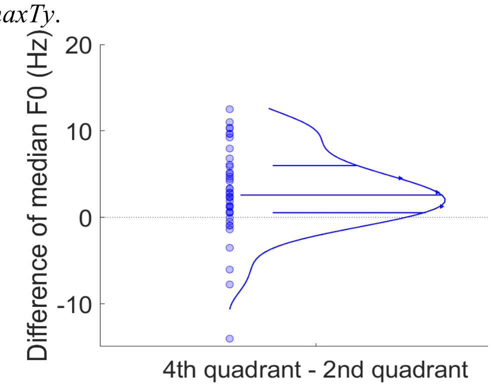


Figure 4: Difference of median F0 between vowels in the 4th quadrant and 2nd quadrant, schematized in Figure 3. Positive values support the hypothesis that *JH* contribution to F0 is more prominent than *maxTy*.



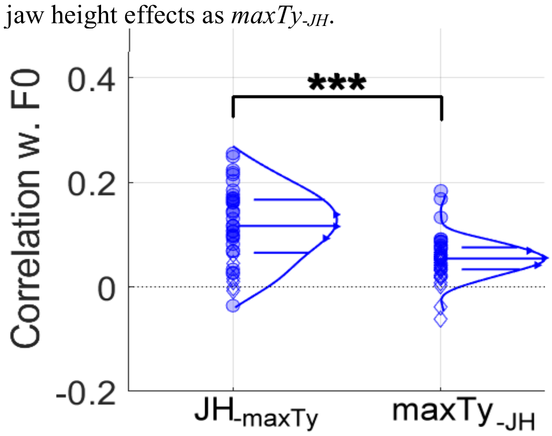
One hypothesis is that if *JH* contributes more to F0 than *maxTy*, then we expect to see low F0 in the 2nd quadrant and high F0 in the 4th quadrant. We defined '*DiffF0(4q-2q)*' as the median F0 in the 4th quadrant minus the median F0 in the 2nd quadrant, and the hypothesis predicts that *DiffF0(4q-2q)* should be positive. And conversely, the opposite hypothesis is that if *maxTy* contributes more than *JH*, then *DiffF0(4q-2q)* should be negative.

The results of such analysis are presented in Figure 4. 31 out of 40 speakers have positive values of *DiffF0(4q-2q)*. A paired *t*-test shows that the mean of *DiffF0(4q-2q)* is significantly higher than 0 ($p = .001$; $t = 3.5$), with a medium effect size (Cohen's $d = 0.55$). Thus the hypothesis that *JH* contribution to F0 is more prominent than *maxTy* is supported.

3.4. Stepwise regression analysis

We further carried out stepwise regression analyses on the articulatory effects on F0. The tongue height (*maxTy*) effects were first fitted and removed from F0 (subtracting the predicted F0 by *maxTy*), and then jaw height (*JH*) effects were subsequently fitted to the residuals to obtain the residual effects of *JH* on F0, coded as "JH_{-maxTy}" here, and the reverse was done to obtain "$maxTy_{-JH}$", the residual tongue height effect on F0 by removing the jaw height effect. As shown in Figure 5. the residual jaw height effects retain the same degree of positive correlations with F0 as those jaw height effects seen in Figure 2. On the other hand, the residual tongue height effects were reduced from the *maxTy* effects in Figure 2. A two-tailed *t*-test revealed that the correlation coefficients of JH_{-maxTy} are significantly higher than those of $maxTy_{-JH}$ ($p < .001$; $t = 4.4$).

Figure 5: Stepwise regression analyses. The residue jaw height effects on F0 by removing tongue height effects are coded as JH_{-maxTy}, and the residue tongue height effects on F0 by removing jaw height effects as $maxTy_{-JH}$.



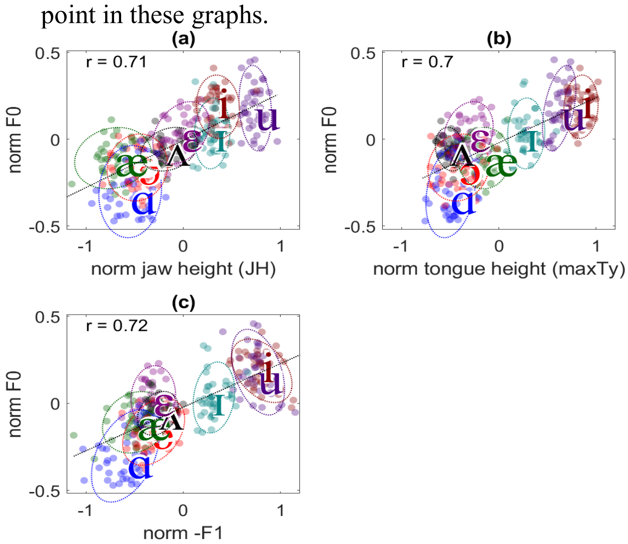
3.5. Discrete or continuous nature of IF0

Lastly, we present the pooled view of the articulatory correlates of IF0. For each speaker, the acoustic (F0 and F1) and articulatory parameters (*JH* and *maxTy*) were first normalized by subtracting the speaker median and divided by the interquartile range. The medians for each vowel category in each parameter were calculated and the medians of normalized F0 were plotted against those of *JH* (Figure 6a), *maxTy* (Figure 6b) and –*F1* (Figure 6c).

As shown in Figure 6, the correlations of F0 with *JH*, *maxTy* and –*F1* increased substantially from those calculated separately for each speaker (Figure

2) to around .7. Moreover, the discreteness of IF0 can be seen in the /ɛ, ɪ/ pair in Figure 6(b-c) such that the vowel /ɪ/ has noticeably higher tongue height and lower F1 than the vowel /ɛ/, but both have similar F0. Such discreteness is less clear in Figure 6(a); it appears that the correlation between normalized F0 and *JH* can be more successfully explained by a linear function than those in Figure 6(b-c) except for the vowel /a/.

Figure 6: Scatter plots of normalized F0 against normalized jaw height (a), tongue height (b), and negative F1 (c). The median of each vowel produced by one speaker contributes to one data point in these graphs.



4. DISCUSSION

4.1. Robustness of IF0

Although we have known that IF0 is language universal for some time [24], a large scale survey of the robustness of IF0 within a language is less often reported. Here we demonstrate such robustness of IF0 in American English in terms of both F0 differences between high and low vowels and the positive correlations of F0 with *–F1*. In Figure 1, 39 out of 40 speakers showed positive increasing trends of F0 from low vowels to high vowels, whereas in Figure 2, 39 speakers have positive correlations of F0 with *–F1* and 37 of them are significant. Thus, our results confirm that IF0 is very robust in American English, with only one speaker failing to show IF0 effects in our analyses. Note that the correlation analysis we performed in this study is a single linear regression, which treats all other factors, such as prosodic effects, speech rate, coarticulation, etc., as unexplained variability (i.e., noise). It is possible that the magnitude of the unexplained variability for that speaker is greater than that of the main effect (IF0) and thus IF0 did not appear. It is also important to note that the effect was present for almost all speakers despite these uncontrolled factors.

4.2. Articulatory correlates of IF0

One of the main purposes of this study is to explore the articulatory correlates of IF0. Pearson correlation results showed that both tongue height and jaw position have significant correlations with F0, while our two further analyses, data subsetting and stepwise regression, support the hypothesis that jaw position contributes more to F0 than the vertical position of the highest point on the tongue. One limitation of this study is that both EMA and x-ray microbeam technologies can only measure flesh points on the front part of the tongue. As suggested by [17], IF0 can be explained by adjustment of the hypopharynx, which has a direct effect on the tension of vocal folds, and may have interactions with the horizontal movements of the tongue root. Unfortunately, the data we used do not provide measurements in the posterior region of vocal tract, leaving direct tests to the future.

4.3. Jaw as an active or passive role in IF0

Our main finding that the jaw rather than the tongue has higher correlations with IF0 is consistent with [4] and [26]. Both studies reached conservative conclusions about the role that jaw plays in IF0. One argument against an active account of a jaw component is that the same or increased IF0 differences were found in vowels produced in a bite-block condition [13, 15]. However, [12] presented the opposite results: that two of three speakers used lower F0 for all vowels produced with the jaw propped open. More recently, [3], based on previous anatomical studies, proposed that the muscular chain between jaw, hyoid bone and cricothyroid joint (larynx) may be the biomechanical linkage for the tendency of jaw lowering to accompany low F0. While we do not have direct evidence for the underlying mechanism of the jaw effect on IF0, our results show that high vowels with lower jaw position were produced with lower F0 than low vowels with higher jaw position (Figure 4), which supports the hypothesis that jaw position plays an active role in F0 control, as proposed by [3].

5. CONCLUSIONS

In this study, we revisited the phenomenon of IF0 in American English with extensive physiological evidence. A series of our analyses support the hypothesis that jaw height contributes more to the magnitude of IF0 than does tongue height. Future studies need to examine the role of the pharynx.

6. ACKNOWLEDGMENTS

This work was supported by US NIH grant DC-002717 to Haskins Laboratories.

7. REFERENCES

[1] Black, J. W. 1949. Natural Frequency, Duration, and Intensity of Vowels in Reading. *Journal of Speech and Hearing Disorders* 14, 216-221.

[2] Boersma, P., Weenink, D. 2001. Praat, a system for doing phonetics by computer. *Glot International* 5, 341-345.

[3] Erickson, D., Honda, K., Kawahara, S. 2017. Interaction of jaw displacement and F0 peak in syllables produced with contrastive emphasis. *Acoustical Science and Technology* 38, 137-146.

[4] Fischer-Jørgensen, E. 1990. Intrinsic F0 in Tense and Lax Vowels with Special Reference to German. *Phonetica* 47, 99-140.

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

[6] Honda, K. 1983. Relationship between pitch control and vowel articulation. *Haskins Laboratories Status Report on Speech Research, SR* 73, 269-282.

[7] Hoole, P., Honda, K. 2011. Automaticity vs. feature-enhancement in the control of segmental F0. In: Clements, Ridouane (eds), *Where do phonological features come from.* Amsterdam/Philadelphia: John Benjamins B. V. 131-171.

[8] Hoole, P., Mooshammer, C. 2002. Articulatory analysis of the German vowel system. In: Auer, Gilles, Spiekermann (eds), *Silbenschnitt und Tonakzente.* Tügingen: Niemeyer. 129-152.

[9] Ladd, D. R., Silverman, K. E. A. 1984. Vowel Intrinsic Pitch in Connected Speech. *Phonetica* 41, 31-40.

[10] Ladefoged, P. 1964. *A phonetic study of West African languages: an auditory-instrumental survey.* University Press.

[11] Lehiste, I. 1970. *Suprasegmentals.* M.I.T. Press.

[12] Lubker, J., McAllister, R., Lindblom, B. 1977. Vowel fundamental frequency and tongue height. *J. Acoust. Soc. Am.* 62, S16-S17.

[13] Mooshammer, C., Hoole, P., Alfonso, P., Fuchs, S. 2001. Intrinsic pitch in German: A puzzle? The 142nd Meeting of the Acoustical Society of America Ft. Lauderdale, Florida, 2761-2761.

[14] Ohala, J. J., Eukel, B. W. 1987. Explaining the intrinsic pitch of vowels. In: Channon, Shockey (eds), *In honor of Ilse Lehiste: Ilse Lehiste Pühendusteos* Dordrecht, The Netherlands: Foris Publications. 207-215.

[15] Ohala, J. J., Eukel, B. W. 1976. Explaining the intrinsic pitch of vowels. *J. Acoust. Soc. Am.* 60, S44-S44.

[16] Pape, D., Mooshammer, C. Year. Intrinsic F0 differences for German tense and lax vowels. *Proc. Proceedings of the 7th International Seminar on Speech Production in Ubatuba, Brazil, December 13th to 15th.* 271-278.

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

[18] Ruederer, H. 1916. *Uber die Wahrnehmung des gesprochenen Wortes: eine experimentell-psychologische Untersuchung*. K. Ludwig-Maximilians-Universitat Munchen, Noske, Norna-Leipzig.

[19] Shadle, C. H. 1985. Intrinsic fundamental frequency of vowels in sentence context. *J. Acoust. Soc. Am.* 78, 1562-1567.

[20] Steele, S. A. 1986. Interaction of Vowel F0 and Prosody. *Phonetica* 43, 92-105.

[21] Taylor, H. C. 1933. The fundamental pitch of English vowels. *Journal of Experimental Psychology* 16, 565-582.

[22] Tiede, M., Espy-Wilson, C. Y., Goldenberg, D., Mitra, V., Nam, H., Sivaraman, G. 2017. Quantifying kinematic aspects of reduction in a contrasting rate production task. *J. Acoust. Soc. Am.* 141, 3580-3580.

[23] Westbury, J. R. 1994. *X-ray Microbeam Speech Production Database User's Handbook.* Madison, WI: University of Wisconsin.

[24] Whalen, D. H., Levitt, A. G. 1995. The universality of intrinsic F0 of vowels. *J. Phonetics* 23, 349-366.

[25] Whalen, D. H., Chen, W.-R., Tiede, M. K., Nam, H. 2018. Variability of articulator positions and formants across nine English vowels. *J. Phonetics* 68, 1-14.

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