HHS Public Access

Author manuscript

J Phon. Author manuscript; available in PMC 2022 July 01.

Published in final edited form as:

J Phon. 2021 July ; 87: . doi:10.1016/j.wocn.2021.101063.

A dual mechanism for intrinsic f0

Wei-Rong Chen[a,*], **D. H. Whalen**[a,b,c], **Mark K. Tiede**[a]

[a]Haskins Laboratories, 300 George Street #900, New Haven, CT 06511

[b]City University of New York, 205 E 42nd Street, New York, NY 1001

[c]Yale University, New Haven, CT 06520

Abstract

Vowel-intrinsic fundamental frequency (IF0), the phenomenon that high vowels tend to have a higher fundamental frequency (f0) than low vowels, has been studied for over a century, but its causal mechanism is still controversial. The most commonly accepted "tongue-pull" hypothesis successfully explains the IF0 difference between high and low vowels but fails to account for gradient IF0 differences among low vowels. Moreover, previous studies that investigated the articulatory correlates of IF0 showed inconsistent results and did not appropriately distinguish between the tongue and the jaw. The current study used articulatory and acoustic data from two large corpora of American English (44 speakers in total) to examine the separate contributions of tongue and jaw height on IF0. Using data subsetting and stepwise linear regression, the results showed that both the jaw and tongue heights were positively correlated with vowel f0, but the contribution of the jaw to IF0 was greater than that of the tongue. These results support a dual mechanism hypothesis in which the tongue-pull mechanism contributes to raising f0 in non-low vowels while a secondary "jaw-push" mechanism plays a more important role in lowering f0 for non-high vowels.

Keywords

intrinsic f0; mandibular contribution; tongue-pull hypothesis; articulatory correlation

1. Introduction

In vowel production, an important intrinsic property known as vowel-intrinsic fundamental frequency (IF0), also termed intrinsic pitch, is the general tendency for high vowels to have higher fundamental frequency (f0) than low vowels. IF0 was first discovered in German (Meyer, 1896/7) and has been studied for more than a century (e.g., Black, 1949; Shadle,

[*]Corresponding author. chenw@haskins.yale.edu (Chen) whalen@haskins.yale.edu (Whalen), tiede@haskins.yale.edu (Tiede).

Wei-Rong Chen: Conceptualization, Methodology, Formal analysis, Investigation, Software, Validation, Writing - Original Draft, Writing - Review & Editing, Visualization

D.H. Whalen: Conceptualization, Validation, Resources, Writing - Review & Editing, Investigation, Supervision, Funding acquisition

Mark K. Tiede: Conceptualization, Software, Validation, Investigation, Data Curation, Writing - Review & Editing

1985; Taylor, 1933; Whalen & Levitt, 1995). It appears that IF0 is a language-universal phenomenon[1] that is more likely to be physiologically driven than deliberately introduced, and the effect size of IF0 is f0-dependent (i.e., the IF0 effect was observed in higher voice frequency ranges but not in lower ones) (e.g., Ladd & Silverman, 1984; Shadle, 1985; Steele, 1986; Whalen & Levitt, 1995). Many causal explanations have been proposed and tested, only to be discarded due to lack of support from experimental data. For example, the "acoustic coupling account" is a hypothesized aerodynamic effect proposing that vocal tract resonance can attract the frequency of the vocal fold vibration when the two frequencies are close to each other; it was tested by Beil (1962) and Ewan (1979) but subsequently failed to be replicated (see Hoole & Honda, 2011 for a comprehensive review). The debate about the underlying mechanism of IF0 continues, and there are three main streams of hypotheses that have been proposed to account for the phenomenon: 1) biomechanical linkage, 2) perceptual enhancement, and 3) phonologization. These are separately discussed below.

1.1. Biomechanical linkage account

The most representative biomechanical account for IF0 is the tongue-pull hypothesis, first suggested by Ladefoged. He noted that "the tongue is attached to the superior part of the hyoid bone, and some of the laryngeal muscles are attached to the inferior part. When the tongue is raised these laryngeal muscles are stretched, and the tension of the vocal cords is increased" (Ladefoged, 1968, p. 41). This view was challenged by the fact that the height of the larynx-hyoid structure (commonly believed to be a correlate of f0)) for the vowel /u/ is lower than that for the vowel /ɑ/, opposite of the expected order for IF0 given the lower acoustic height (higher F1) for the latter vowel (e.g., Ewan & Krones, 1974; Ladefoged, DeClerk, Lindau, & Papçun, 1972; Sundberg, 1969). However, Ohala (1973) argued that this fact does not necessarily preclude the association of tongue height and vocal fold tension, and proposed a modified version of tongue-pull in which tongue raising acts to elevate the ventricular fold and concomitantly increases the vertical tension of the vocal fold. On the one hand this account was supported in studies that showed correlation of ventricle size with IF0 (Shimizu, 1960, 1961; Van den Berg, 1955). On the other hand, Rossi and Autesserre (1981) presented xeroradiographic measurements to argue that the advancement of the tongue root may increase f0 by stretching the soft tissue (mucous membrane) between the epiglottis and vocal fold, resulting in increased size of the laryngeal vestibule. Supported with electromyographic (EMG) data, Honda (1983) offered an explanation that the contraction of the posterior genioglossus (GGP), an action associated with the production of high vowels, pulls forward the hyoid bone and rotates the thyroid cartilage, thus increasing the longitudinal tension of the vocal fold. In a review of the data Hoole and Honda (2011) suggest that Honda (1983)'s explanation may be the most widely accepted version of the tongue-pull hypothesis. However, one piece of evidence against all biomechanical link hypotheses comes from the reported IF0 effect in the esophageal speech of laryngectomy patients (Gandour & Weinberg 1980). Since any possible link between the vocal tract and larynx is removed in laryngectomy patients, the existence of an IF0 pattern in those patients

[1] To the best of our knowledge, there is only one exception in the literature. Connell (2002) reported that Mambila, a tonal language in Africa, showed little or no evidence for IFO differences, based on the data of four native speakers. Results of Whalen & Levitt (1995) indicate that this is too small a sample from which to draw firm conclusions.

cannot be a “direct” consequence of a biomechanical link. However, Whalen, Gick, Kumada, and Honda (1999) explained that this fact may be due to a deliberate reintroduction of IF0 of the esophageal speakers that was formerly automatically driven, probably for the purpose of increasing naturalness, and it is not incompatible with the biomechanical account for IF0.

1.2 Perceptual enhancement account

Researchers have also proposed that IF0 might be due to deliberate enhancement in production for the purpose of distinguishing vowels, instead of being caused by an automatic process such as tongue-pull or acoustic coupling (e.g., Diehl & Kluender, 1989; Kingston, 1992). Diehl and Kluender (1989) suggested that the covariation of f0 and tongue height may be an instance of speakers deliberately enhancing the phonological contrasts, based on the perceptual results of Traunmüller (1981), in which the judgements of vowel openness depended on the distance between f0 and F1, instead of just F1. This perceptual enhancement hypothesis received the most direct support from EMG studies reporting higher cricothyroid (CT) activity in high vowels (e.g., Autesserre et al., 1987; Dyhr, 1990; Honda & Fujimura, 1991). CT activity is known to be the primary muscle responding for the active control of f0 (Löfqvist, Baer, McGarr, & Story, 1989), and so higher CT activity in high vowels entails a greater degree of deliberate f0 raising. However, those studies had limited sample sizes and did not single out the vowel height effect from other factors that might increase CT activity, such as prosodic function. To mitigate this issue, Whalen et al. (1999) designed EMG experiments with tone matching conditions to control for active pitch control other than the effect of vowel categories. They reported that only one of four American English speakers showed a pattern of CT activity conforming to the (deliberate) perceptual enhancement hypothesis for IF0, and that the magnitude of CT activity only poorly explained observed changes in f0 for all speakers.

1.3 Phonologization account and remaining issues

Honda and Fujimura (1991) proposed that while the effects of IF0 likely originate from biomechanical links, these effects may then undergo phonologization and thus become deliberate enhancements produced by some speakers. This account seems to be the only hypothesis that is compatible with all pieces of evidence observed so far. However, while the consensus is that IF0 is best explained by some kind of biomechanical link at least for its origin, the current biomechanical hypotheses are not completely satisfactory. One problem for the widely accepted tongue-pull hypothesis is that it predicts that IF0 differences can be observed between high and low vowels and among high vowels due to different degrees of GGP pull, but not among low vowels, since posterior genioglossus retraction is not necessary for producing low vowels (Honda & Fujimura, 1991). However, this prediction (i.e., no IF0 in low vowels) is not fulfilled as IF0 differences between low vowels have been observed in many previous studies (e.g., Honda & Fujimura, 1991; Hoole & Mooshammer, 2002; Iivonen, 1989; Peterson & Barney, 1952; Turner & Verhoeven, 2011) as well as in our own data. To the best of our knowledge, this issue for the tongue-pull hypothesis has not been previously raised in the literature.

1.4 Mandibular contribution to f0

An explanation for residual effects not accounted for by the tongue-pull hypothesis may be found in the jaw contribution to f0 variability. Co-variation of the mandible position with fundamental frequency has long been noted in the literature. Using cineradiography, Zawadzki and Gilbert (1989) reported that three of five American English speakers had higher correlations of f0 with the vertical position of the mandible (lower mandible position associated with lower f0) than with tongue height. (Fischer-Jørgensen 1990) measured jaw opening as the distance between the upper and lower incisors in video recordings and tongue height as the minimal distance between the two lateral contacts on the palate in static palatography in five German speakers. She found that jaw opening was in better agreement with f0 (wider jaw opening associated with lower f0) than tongue height; however, tongue height was measured only for the four non-low front vowels [iː, ɐ, ɪ, eː] in German. Pape and Mooshammer (2006) measured the articulatory positions of the tongue, lip, and jaw using Electromagnetic Midsagittal Articulography (EMMA) and f0 with electroglottography (EGG) in three German speakers, with a special focus on the IF0 differences between the German tense-lax vowel opposition. They entered the tongue, jaw, and lip positions in regression models and showed that tongue height explained the greatest amount of variance of f0 for two of three speakers; this is somewhat at odds with Zawadzki and Gilbert (1989) and Fischer-Jørgensen (1990)'s results. However, their results may not be directly comparable because the extent of mechanical linkage between articulators and vocal fold tension may be underestimated due to likely active control of f0 for German lax vowels, thus overlaying the mechanical IF0 differences (Hoole & Honda, 2011; Pape & Mooshammer, 2006)[2]. Most recently, Erickson, Honda, and Kawahara (2017) studied articulatory and acoustic correlates of contrastive emphatic conditions in American English using X-ray microbeam data from six speakers taken from Erickson (2002) and (Westbury and Fujimura 1989). They reported that emphasized syllables consistently showed a higher f0 and lower jaw position for all six speakers, with a more anterior jaw position for five of the six. To explain the association between a lower jaw position and higher f0 (contradicting the previously found positive correlations of jaw height and f0), Erickson et al. (2017) suggested that while mandible lowering (rotational movement) pushes back the hyoid bone and thyroid cartilage and slackens the vocal folds, mandible protrusion (translational movement) pulls forward the hyoid bone and thyroid cartilage and stretches the vocal folds; thus, the coexistence of jaw fronting in emphasized syllables may compensate for the effect of a reduced f0 through jaw lowering.

1.5 Interim summary and research goals

To summarize, the contribution of the jaw to IF0 remains uncertain, partly due to low sample sizes, inconsistent results, and lack of thorough and large-scale physiological measurements in the related studies. More importantly, since tongue height and jaw height are also highly correlated in vowel production (e.g., low vowels are usually produced with both lower tongue and lower jaw positions), the positive correlations of f0 with tongue and jaw height reported in previous studies might have been due to collinearity effects. Therefore, in this

[2] We thank an anonymous reviewer for reminding us this possibility.

study, we aimed to distinguish between the contributions of the tongue and jaw articulators to the magnitude of IF0, compensating for collinearity if present. Specifically, we employed two techniques: 1) data subsetting and 2) stepwise regression to maintain the relative independence of tongue and jaw height. Using data subsets in which the tongue and jaw are determined *a priori* to be relatively uncorrelated, we can test whether the tongue articulator is the only (or dominant) causal factor for IF0: We expect to see a positive correlation of f0 with tongue height and no (or little) correlation of f0 with the jaw, and vice versa if the jaw is the only (or dominant) factor in IF0. Alternatively, if the uncorrelated tongue and jaw heights are *both* positively correlated with f0, a hypothesis of dual mechanism is supported, such that tongue-pull accounts for increasing f0 in high vowels while jaw-push accounts for decreasing f0 in low vowels.

2. Method

2.1 Data and measurements

The data were taken from two existing corpora of simultaneous articulatory and acoustic recordings: 1) the X-ray microbeam database (XRMB) (Westbury, 1994) and 2) the Haskins IEEE rate comparison database (HIRCD)[3] (Tiede et al., 2017). Both corpora comprise simultaneous recordings of articulatory movements (sampling rates = 145.6 Hz and 100 Hz in XRMB and HIRCD, respectively) and acoustic signals (sampling rates = 21.7 kHz and 44.1 kHz in XRMB and HIRCD, respectively), produced by American English speakers, labeled using forced alignment of the acoustics (P2FA; Yuan & Liberman, 2008). We selected 8 monophthong vowels of American English (/ɑ, ɔ, æ, ʌ, ɛ,ɪ, u, i/) produced with primary stress from 36 speakers (21 female) in XRMB and 8 speakers (4 female) in HIRCD, totaling 43 890 vowel samples. Any tokens shorter than 50 ms and vowels followed by a nasal were excluded. The average number of samples per vowel for each speaker was 120. The speech materials in XRMB comprised citation words, short and long sentences, and paragraphs. Those in HIRCD were 720 phonetically balanced short sentences (“Harvard sentences”, IEEE, 1969) produced at normal and fast rates (however, only the sentences produced at a normal rate were used in this study).

The articulatory movements in XRMB were tracked by rasterized focused X-ray sweeps following gold pellets midsagittally placed at four points on the tongue (T1~T4), upper lip (UL), lower lip (LL), lower incisor (JAW), and left molar. For HIRCD, kinematic data were acquired by using an electromagnetic articulography (EMA) system (NDI WAVE) that tracks three-dimensional motions of coil sensors attached to three points on the tongue (T1~T3), upper lip (UL), lower lip (LL), lower incisor (JAW), and canine tooth (JawL). Figure 1 illustrates the locations of kinematic tracking on the mid-sagittal plane for XRMB and HIRCD. Both corpora were aligned to each speaker’s occlusal plane.

The most anterior points on the tongue (T1) in both corpora were located at approximately one cm posterior to the tongue apex; the most posterior points on the tongue (T4 in XRMB and T3 in HIRCD) were attached as far back as possible given speaker tongue protrusion, and the remaining tongue point(s) were placed roughly equidistant between them. Figure 2

[3] Available from http://bit.ly/2s4mtOq

displays the relative distances from the tongue tip (T1) to the other tongue points for both corpora. Each circle indicates the average distance of a tongue point to T1 for a speaker, and the short horizontal bar represents the mean across speakers.

To combine data from the four tongue pellets of the XRMB corpus with that of the three tongue sensors of the HIRCD corpus, we identified the highest vertical position reached across all the three or four tongue points as our operational tongue height measurement *maxTy*. We then calculated the first principal component (PC1) of the vertical and horizontal coordinates of *JAW* across all samples, normalized such that positive values indicate upward movement. For each sample, the projection of the *JAW* signal onto PC1 was defined as the jaw height measurement *JH*, representing a function of jaw opening and closing with the positive sign assigned to the closing direction, as shown in Figure 3. For ease of visualization, we created a new variable $-F1$ as the negated first formant, so that intrinsic f0 could be shown as positive correlations of f0 with both acoustic and articulatory variables.

Articulatory positions and acoustic features were measured at the midpoint of the acoustically labeled vowel tokens. Fundamental frequencies (f0s) were calculated using the Boersma (1993) autocorrelation method implemented in PRAAT (Boersma & Weenink, 2019) with a manually determined acceptable f0 range (the variables 'pitch floor' and 'pitch ceiling' in PRAAT) for each speaker. The analysis window for f0 tracking varied with the preset minimum value of f0 (window size = 3000 / 'pitch floor' in ms). Formant frequencies were measured by the 'seeding' method (Chen, Whalen, & Shadle, 2019), in which static resonance frequencies were estimated by the Burg LPC method (window size = 45 ms; step size = 2 ms; number of LPC coefficients = 14; pre-emp from 50 Hz; cut-off frequency = 5000 Hz for male and 5500 Hz for female) in PRAAT, and then tracked by the Viterbi algorithm with vowel-specific references (i.e., seeds) based on the mean first three formant frequencies reported in Peterson and Barney (1952).

2.2 Statistics

The most common parametric measure of association between variables is the Pearson correlation. However, because the Pearson correlation coefficient is well-known for being easily distorted by outliers (Rousseeuw, 1984), we adopted a "robust correlation" approach, specifically, the 'skipped-correlation' in the 'Robust correlation toolbox' (Pernet, Wilcox, & Rousselet, 2013), which determines a robust center of data and outliers based on a minimum covariance determinant (MCD) estimator.

We fitted a series of linear mixed effect (LME) models by using the *lme4* (Bates, Mächler, Bolker, & Walker, 2015) and *lmerTest* (Kuznetsova, Brockhoff; & Christensen, 2017) packages in R (R Core Team, 2020). Post-hoc comparisons were carried out using the *multcomp* (Hothorn, Bretz, & Westfall, 2008) package and marginal effects were estimated by the *effects* (Fox & Weisberg, 2019) package. Details of models are described in the following sections.

3. Results

3.1 Initial test of IF0

Traditionally, the existence of lF0 has been reported as a non-negligible difference in f0 between high and low vowels, and the degree (effect size) of IF0 most commonly measured as the mean f0 for /i/ (or /u/, or the average of /i/ and /u/) minus that of /a/ (or /ɑ/) (see Whalen & Levitt, 1995). We followed this approach as an initial test to confirm the presence of IF0 in our dataset. Specifically, for each speaker, we assembled samples from /i/ and /u/ as the group of high vowels, and from /a/ as the group of low vowels, and then compared the f0 values between the two groups. A two-sample *t*-test was carried out to test if the mean f0 of the high vowel group was significantly higher than that of the low vowel group. The *p*-values were then adjusted by controlling the familywise type-I error to be less than 5% with the False Discovery Rate (FDR) (Benjamini & Hochberg, 1995). Among all (44) speakers, 42 had a higher mean f0 for the high vowel group than for the low vowel group, and 39 of the 42 comparisons were significant (FDR-adjusted $p < 0.05$).

Figure 4 presents the overall pattern of f0 and F1 for the eight vowels across the 44 speakers, sorted roughly from low (left) to high (right) vowels. The f0 and F1 values presented in this figure were speaker-normalized by subtracting the median value of all data within-speaker; this was done only for visualization purposes. Each datapoint indicates the median of the samples for each vowel produced by one speaker. For the correlation analyses, the original values of f0 were used (not normalized). The curved lines represent the probability density functions fitted by kernel density estimation. As shown in Figure 4, there is a general trend of increasing f0 (top panel) from low to higher vowels, and roughly the opposite trend in the F1 values (bottom panel), demonstrating the effect of IF0.

We further fitted an LME model by entering vowel category and gender as the fixed effects and a random intercept of speaker. The model-predicted marginal effects of vowels are plotted in Figure 5, which resembles the raw by-speaker means in Figure 4a. Pairwise post-hoc comparisons reveal that the F0 means for all vowel categories all significantly differ from each other, except for three pairs: /ɔ, æ/, /ʌ, ɛ/, and /u, i/. The model predicts that the mean F0 for English vowels has the following order: /ɑ/ < /ɔ, æ/ < /ʌ, ɛ/ < /ɪ/ < /u, i/.

3.2 Articulatory and acoustic correlates of IF0

To explore the articulatory correlates of IF0, we calculated the correlations of f0 values with the articulatory variables *JH* and *maxTy* and the acoustic variable *−F1* separately for each speaker. The horizontal components of all tongue measurements did not show appreciable correlations with f0 and are thus subsequently ignored here. In Figure 6, each datapoint indicates a coefficient for one speaker, and the curved lines represent the distributions of the correlation coefficients for 44 speakers. Shaded circles indicate the individual correlation coefficient for that speaker was significantly different from zero (FDR-adjusted $p < 0{,}05$); unfilled diamond symbols indicate non-significant correlation coefficients. We henceforth abbreviate the correlation of f0 with jaw height *JH* as *Cor(f0, JH)*, with tongue height *maxTy* as *Cor(f0, maxTy)*, and with *−F1* as *Cor(f0, −F1).* As shown in Figure 6, *Cor(f0, JH)*, *Cor(f0, maxTy)*, and *Cor(f0, −F1)* are all significantly positive and with similar slopes. The

positive correlation of *Cor(f0, −F1)* is consistent with the traditional view of IF0: high vowels (low F1) have higher f0. Both *JH* and *maxTy* are positively correlated with f0, suggesting that both jaw height and tongue height contribute to IF0. The means of *Cor(f0, JH)*, *Cor(f0, maxTy)*, and *Cor(f0, −F1)* were 0.25 (SD = 0.13), 0.21 (0.10) and 0.24 (0.12), respectively, and 29 of 44 speakers had higher *Cor(f0, JH)* than *Cor(f0, maxTy)*. We ran a series of two-tailed paired sample *t*-tests (df = 43) to compare the means of these distributions (*p*-values were adjusted by FDR). The results showed that both *Cor(f0, JH)* and *Cor(f0, −F1)* were significantly higher than *Cor(f0, maxTy)* (FDR-adjusted $p < 0.05$). The effect size for the *t*-tests as measured by Cohen's *d* was 0.5 for comparing *Cor(f0, JH)* and *Cor(f0, maxTy)*, and 0.52 for comparing *Cor(f0, −F1)* and *Cor(f0, maxTy)*; both are considered to be medium effect sizes, according to Cohen (1988).

Three separate LME models were fitted by entering f0 as the dependent variable, one of the three variables *JH*, *maxTy* and *−F1* as the main effect and Gender as another fixed effect. By-speaker random intercept and random slope were also included in the models[4]. In order to make the comparisons of regression coefficients meaningful, values of *JH*, *maxTy* and *−F1* were by-speaker normalized in z-scores. The model formulas are as follows:

$$f0 \sim JH + Gender + (1 + JH|Speaker)$$

$$f0 \sim \max Ty + Gender + (1 + \max Ty|Speaker)$$

$$f0 \sim (-F1) + Gender + (1 + (-F1)|Speaker)$$

Note that it would be problematic to enter *JH* and *maxTy* (or *−F1*) in the same model due to the strong collinearity between the two variables. As Tomaschek, Hendrix, and Baayen (2018) demonstrated in a computer simulation, two strongly correlated fixed effects in a linear model can result in a fitted coefficient opposite to the true value. Figure 7 summarizes the results of the three models (each model represented in a column). Each shaded circle represents the random slope of one speaker superimposed on the main effect; the curved line is the probability density function fitted to the random slopes of all speakers. The square symbols and error bars indicate the coefficient (slope) and 95% confidence interval (+/− 1.96 x standard error) of the main effect for each model. The patterns shown in Figure 7 are very similar to those in Figure 6. Pairwise comparisons with FDR correction revealed that the slopes of *JH* were significantly higher than those of *maxTy* ($p = .02$). The regression coefficients for *JH*, *maxTy*, and *−F1* were 5.08 (SE = 0.63), 4.2 (SE = 0.48), and 4.54 (SE = 0.51), respectively.

3.3 Subset with uncorrelated tongue and jaw heights

In the above section, we have shown that both jaw and tongue height were positively correlated with f0, with the mean of *Cor(f0, JH)* significantly higher than that of *Cor(f0,*

[4] We thank an anonymous reviewer suggesting fitting by-speaker random slope models.

maxTy). However, previous results have demonstrated that jaw height also highly correlates with tongue height in speech (e.g., Shaw, Chen, Proctor, & Derrick, 2016). Here, all 44 speakers in our dataset showed significantly positve correlations between *JH* and *maxTy* (mean correlation coefficient = 0.55, SD = 0.09). Therefore, from the correlation analyses above, we do not know how much of the variance of f0 explained by jaw position is independent of the tongue and vice versa (i.e., degree of collinearity).

Our first attempt to distinguish the contributions of jaw and tongue height to IF0 was to subset the data such that the *JH* and *maxTy* components were relatively uncorrelated. Our procedure was as follows. For each speaker, a scatterplot of all samples was mapped with *JH* on the x-axis and *maxTy* on the y-axis, centered on their medians as shown in Figure 8. Then the data were apportioned by the four quadrants of this scatterplot. The first and third quadrants contain the samples for which higher tongue height is correlated with higher jaw height, and vice versa. The second quadrant contains the samples with higher tongue height and lower jaw height, and the fourth quadrant, the samples with lower tongue height and higher jaw height, that is, only those samples in which *maxTy* and *JH* are no longer positively correlated (in fact, they were negatively correlated). Separating the samples by the median has the benefit that exactly 50% of samples are distributed above or below the median boundary and are unaffected by outliers. This subset retained around 20~40% (mean = 26%; SD = 4%) of the data for each speaker. Table 1 summarizes the proportions of vowel tokens being allocated to each quadrant, averaged across all speakers (the numbers in parentheses indicate standard deviations). If the hypothesis that *JH* contributes more to f0 then *maxTy* is true, we would expect to see a lower f0 in the second quadrant than in the fourth quadrant. Therefore, for each speaker, we defined a new variable $dF0_{q4-q2}$ as the mean f0 in the 2nd quadrant subtracted from the mean f0 in the 4th quadrant. The hypothesis predicts that $dF0_{q4-q2}$ will be positive. The alternative outcome, that *maxTy* contributes more to f0 than *JH*, predicts a negative $dF0_{q4-q2}$. Figure 9 presents the result of this analysis; 31 out of 44 speakers had positive values for $dF0_{q4-q2}$. A one sample *t*-test showed that the mean of $dF0_{q4-q2}$ was significantly greater than zero ($p < 0.001$); the effect size estimated by Cohen's *d* was 0.58 (medium effect). Therefore, the hypothesis that *JH* contributes more to f0 than *maxTy* is supported.

3.4 Correlation analysis with subsets

A follow-up subsetting analysis repeats the same correlation analyses of Sec. 3.2 using only the subset of the 2nd and the 4th quadrants in the *JH*:*maxTy* space (as the subsetting described in Sec. 3.3). Since the 2nd and 4th quadrants contain the data in which the *JH* and *maxTy* components are relatively uncorrelated (the 2nd quadrant containing samples with higher tongue position and lower jaw position, and the 4th quadrant, samples with lower tongue position and higher jaw position), if the jaw height contribution is dominant over the tongue height contribution to f0, then the correlation *Cor(f0, JH)* calculated with this subset will remain positive and *Cor(f0, maxTy)* will decrease, as compared to those calculated with the full dataset (shown in Figure 6), and vice versa. Figure 10 displays the results of this analysis. The significance symbols displayed vertically indicate whether the correlation of each variable with f0 is significantly different from zero. Significance symbols displayed horizontally indicate whether the two variables are significantly different from each other.

Recall that the correlation coefficients of *Cor(f0, JH)*, *Cor(f0, maxTy),* and *Cor(f0, −F1)* calculated with the full dataset were all significantly higher than zero (Sec. 3.2, Figure 6). When calculated with this relatively uncorrelated subset, *Cor(f0, JH)* was still significantly higher than zero ($p < 0.001$), whereas *Cor(f0, maxTy)* and *Cor(f0, −F1)* were not ($p = 0.2$ and 1.0, respectively), as shown in Figure 10 and supported by one-sample *t*-tests. The means of *Cor(f0, JH), Cor(f0, maxTy),* and *Cor(f0, −F1)* with this subset were 0.11 (SD = 0.16), −0.04 (0.16) and −0.00 (0.14), respectively. Two-tailed paired sample *t*-tests with FDR correction revealed that *Cor(f0, JH)* was significantly higher than *Cor(f0, maxTy)* ($p <$ 0,01, Cohen's $d = 0.49$) and *Cor(f0, −F1)* ($p < 0.01$, Cohen's $d = 0.47$).

To summarize, when we subset the data such that the jaw and tongue heights are relatively uncorrelated, the effect of jaw height on f0 remains, while the effect of tongue height on f0 disappears, suggesting a greater contribution from jaw height to IF0 than from tongue height.

3.5 Stepwise regression analysis

Our results of data subsetting all pointed to greater contribution of the jaw to IF0. However, the unven distributions of vowels in the subsetted data (as shown in Table 1) warranted a further analysis to separate the contribution of the jaw on f0 from that of the tongue (and the other way around), namely stepwise regression. This approach is similar to the directed factor analysis in Maeda (1990), in which the total variance of the vocal tract shape during speech was modeled as a linear combination of independent factors associated with articulators, such that their effects could be added to or subtracted from the total variance. We adopted this approach to model the variance of f0 as a linear combination of the jaw height and tongue height components. Specifically, for each speaker, the tongue height effect on f0 was first estimated by fitting a linear regression model, with f0 as the dependent variable and *maxTy* as the independent variable. The estimated tongue height effect was then removed (partialed out) from the observed f0 values to obtain the residual variance of f0. Subsequently, we fit the jaw height (*JH*) to the residual variance of f0 to obtain the remaining jaw height effect on f0 after the removal of the tongue height effect, denoted as $JH_{\text{-}maxTy}$. Similarly, the remaining tongue height effect on f0 independent of the jaw, denoted as $maxTy_{\text{-}JH}$, can be calculated by fitting *maxTy* to the residual variance of f0 after the removal of the jaw height effect. The hypothesis that the jaw height contributes more to f0 than tongue height predicts a higher $JH_{\text{-}maxTy}$ than $maxTy_{\text{-}JH}$. Alternatively, if the tongue height contributes more to f0 than the jaw height, then $maxTy_{\text{-}JH} > JH_{\text{-}maxTy}$ is predicted.

Figure 11 presents the distributions of the remaining jaw height effect $JH_{\text{-}maxTy}$ (left) and the remaining tongue height effect $maxTy_{\text{-}JH}$ (right) for all 44 speakers (each circle or diamond symbol indicates one speaker); 29 of 44 speakers had higher $JH_{\text{-}maxTy}$ than $maxTy_{\text{-}JH}$. A two-tailed paired sample *t*-test showed that the mean of $JH_{\text{-}maxTy}$ (0.15, SD = 0.1) was significantly higher than that of $maxTy_{\text{-}JH}$ (0.065, SD = 0.09) ($p < 0.001$). Again, the hypothesis that jaw height contributes more to f0 was supported.

4. Discussion and Conclusions

4.1 Prevalence and universality of IF0

Our study showed that IF0 is prevalent in American English, with the majority of speakers (42 of 44) showing a higher mean f0 in high vowels than in low vowels. In our correlation analysis, 42 of 44 speakers showed significant positive correlations of f0 with $-F1$ (Figure 6); the two male speakers (M01 and M04; mean f0 = 117 Hz and 134 Hz, respectively) with non-significant correlations of f0 and $-F1$ coincide with the two speakers who did not have a higher f0 in higher vowels than in low vowels. It is tempting to hypothesize that these two exceptional speakers did not show IF0 contrast due to lower F0 range because, as Whalen and Levitt (1995) pointed out, the effect of IF0 disappears in the lower pitch range. However, there were other speakers who have even lower F0 ranges (than M01 and M04) and yet still exhibit significant IF0 contrasts. Also, an LME model including speaker's pitch range (coding each token as either higher or lower than median F0 within each speaker) as a fixed factor did not show significant effect of interaction of speaker pitch range and jaw height or tongue height. A closer examination of individual data did not find a good reason accounting for the absence of IF0 in these two speakers either. A possible account is that both XRMB and MRCD recorded a large amount of continuous speech over a range of prosodic contexts, and our analyses of IF0 treated all the uncontrolled factors, such as speech rate, coarticulation, prosodic influence, etc., as unexplained variance (i.e., noise); if the effect size of IF0 is lower than that of the noise, it will not be reflected in our measurements. As shown in Ladd and Silverman (1984), the effect size of IF0 was smaller in connected speech than in the carrier phrase condition. Shadle (1985) reported the IF0 effect differed in different positions of a sentence. Erickson et al. (2017) hypothesized that, at prosodic prominent positions, the f0 lowering caused by jaw opening may be (partially) cancelled out by the f0 raising effect of jaw protrusion. Perhaps the range of (uncontrolled) prosodic effects shadowed the smaller effect of IF0 for the two exceptional speakers. However, the fact that the majority of our speakers showed significant IF0 effects even over variegated prosodic conditions reveals the prevalence of IF0 in American English. Better controlled prosodic contexts employed in future studies may sharpen the findings in this study.

4.2 Articulatory correlates of IF0

Our results showed that both tongue height and jaw height positively correlated with f0 for almost all speakers, with jaw height slightly but significantly more correlated with f0 than tongue height (Figure 6). We then subset the data such that tongue height and jaw height were relatively uncorrelated (Figure 8) and showed that the mean f0 of the samples with higher jaw position and lower tongue position was higher than the mean f0 of the samples with lower jaw position and higher tongue position for 31 of 44 speakers (Figure 7). In correlation analyses calculated only with the subset of data where tongue and jaw heights were uncorrelated, the correlation of jaw height with f0 remained significantly positive, whereas the positive correlation of tongue height (as well as F1) with f0 previously seen in the full dataset disappeared in this subset (Figure 10). Our additional stepwise regression analysis confirmed the findings in the data subsetting analysis that jaw height contribution to IF0 was significantly greater than tongue height contribution (Figure 11).

4.3 Biomechanisms for IF0

Researchers have described physiological links between the vocal tract and the larynx as likely sources of IF0 effects. Such links have been predominantly ascribed to tongue movement in the literature, and related hypotheses such as tongue-pull and tongue-compress mechanisms for IF0 have been proposed. Besides tongue-pull, some studies have noted the mandibular contribution to f0, but reached conservative conclusions about the role that the jaw plays in IF0; two previous studies (Fischer-Jørgensen 1990; Zawadzki & Gilbert, 1989) reported a higher positive correlation of jaw height with f0 than tongue height, while one study (Pape & Mooshammer, 2006) reported the opposite. Yet, none of the previous studies have excluded the possibility that the observed agreements of the jaw with f0 could be actually due to collinearity effects between jaw and tongue height; i.e., that the observed tongue effect could be partly due to the jaw. We employed three analyses to separate the effects of jaw and tongue (on IF0) from each other; all our results pointed to a conclusion that the mandible plays a more important role in IF0 than the tongue. Our finding supports the jaw-push mechanism (i.e., jaw lowering reduces f0) recently proposed in Erickson et al. (2017): The complex muscular connections between the mandible, hyoid bone, and thyroid cartilage form the thyrohyo-mandibular chain. Thus, a biomechnical consequence of jaw lowering by rotation is backwards movement of both the hyoid bone and thyroid cartilage (rotating around the cricothyroid joint), with an effect of slackening the vocal folds, whereas jaw protrusion imposes the opposite effect (lengthening the vocal folds). The interaction of these processes and their complexity open the possibility that idiosyncratic speech patterns may lead to exceptions to the general IF0 pattern. Erickson et al. (2017) also noted that a more direct test of their hypothesis could be accomplished through decomposing jaw movement into translational and rotational components, which requires tracking the jaw at two locations. While both corpora adopted in this study do have tracked jaw movements at two locations, the tracking of the molar (see Figure 1) in XRMB was unreliable due to its frequent intersection with lingual pellets (Westbury, 1994: 39). The test of the Erickson et al. (2017) hypothesis should be addressed in future studies.

Intuitively, one might think of a straightforward way of separating the tongue and jaw effects by means of a bite-block. With a bite-block that fixes the opening of the jaw, the observed tongue effect on IF0 can be free from the influence of the jaw. In fact, bite-block experiments in Ohala and Eukel (1987) and Mooshammer, Hoole, Alfonso, and Fuchs (2001) showed that when a high vowel was produced with a bite-block (i.e., wider jaw opening), the f0 was even higher than the same high vowel in normal conditions; this seems in conflict with the jaw-push hypothesis. However, Ohala and Eukel (1987) explained that when high vowels are produced with a forced jaw opening, the tongue stretches upwards more than it normally would to compensate, as evidenced in Lindblom and Sundberg (1971)’s tracings, resulting in a higher f0 than in non-manipulated high vowels due to a greater tongue-pull effect. Therefore, bite-block experiments do not reflect natural interactions between the tongue and the jaw in their effects on IF0. On the other hand, if the dataset is large enough, our data-subsettting method can achieve the purpose of evaluating the tongue effect with a ‘virtually’ fixed jaw position and vice versa, without unwanted articulatory compensation. Using this method, our results concluded that high vowels with a

lower jaw position were produced with a lower f0 than low vowels with a higher jaw position for the majority of the speakers (31 out of 44) (Figure 9).

It should be acknowledged that our results do not entail an argument against the tongue-pull hypothesis. Although we observed a greater jaw effect than tongue effect on IF0 in the majority of our speakers, there was still a sizable portion of speakers (13 of 44) who showed the opposite phenomenon. The residual correlation of tongue height with f0 after removing the jaw effect *(maxTy*$_{-JH}$*)* was weak, but the mean (across speakers) was still significantly greater than zero (mean = 0.07; $t = 4.6$; $p < 0.000$) (Figure 11, right). Individually, significantly positive residual correlations of tongue height with f0 were observed in 20 of 44 speakers. Note that tongue-pull and jaw-push mechanisms are not mutually exclusive. Here we offer a dual mechanism hypothesis for IF0 such that: tongue elevation facilitated by the posterior genioglossus raises f0 by a tongue-pull mechanism (Honda, 1995), whereas jaw-opening lowers f0 by a jaw-push mechanism (Erickson et al., 2017). Assuming Honda's version of tongue-pull hypothesis is correct, we expect the tongue-pull mechnism to be more active in non-low vowels, where genioglosus contraction is more relevant. On the other hand, the jaw push mechanism may apply across the whole vowel height range, but we do not know at which point in a continuous jaw lowering the vocal fold tension is affected; neither do we know whether the effect of jaw push on f0 is linear or not. A reasonable speculation is that the jaw push mechanism may be less effective for high vowels such as /i/ and /u/ where jaw lowering is much smaller in magnitude.

The dual mechanism hypothesis predicts gradient IF0 differences across the whole range of vowel height, which is supported by previous observations (e.g., Honda & Fujimura, 1991; Hoole & Mooshammer, 2002; Iivonen, 1989; Peterson & Barney, 1952; Turner & Verhoeven, 2011; this study). This view is also in line with the findings of Shaw et al. (2016), who observed that an "inverse effect" of IF0 (the same vowel produced with lower tongue or jaw positions in low tones) was observed only in the vowel /ɑ/ among /i, ɑ, u/ in Mandarin, and such tone-conditioned vowel variations were in better agreement with jaw than tongue position.

4.4 Caveats and limitations

Of the several limitations that exist in this study, the most critical is that both the EMA and X-ray microbeam techniques are limited in their measurements of the anterior vocal tract. At least three hypotheses in the literature imply a correlation of pharyngeal width with IF0, such as Rossi and Autesserre (1981)'s advanced tongue root account, Ewan (1979)'s tongue compression hypothesis, and Whalen and Gick (2001)'s tongue depth report, but such information (e.g., back cavity pharyngeal width) is not represented in our data. More holistic measurements of the vocal tract, such as ultrasound imaging or real-time MRI, are needed in future studies. Further, the selected variable of the highest point on the tongue (i.e., *maxTy)* may not reflect the true highest point of the tongue due to under-sampling of the tongue. However, to evaluate the effectiveness of *maxTy*, we manually traced 815 tongue images in another corpus of head-corrected ultrasound with simultaneous EMA recordings (Tiede, Chen, & Whalen, 2019). We found that the median absolute difference (across 815 frames) between *maxTy* and the highest point on the ultrasound tongue contour *(maxUSy)* was only

2.06 (SD=1.1) mm, and the correlation coefficient between *maxTy* and *maxUSy* was 0.94 ($p < .0001$).

Lastly, we hypothesized that while the tongue-pull mechanism may account for the IF0 differences in non-low vowels, the jaw contribution plays a more important role in IF0 in non-high vowels. This dual mechanism hypothesis, although supported by our current results, could potentially be further tested were we to divide the data into non-high and non-low vowels after subsetting the data, such that jaw and tongue are uncorrelated; however, the number of tokens in the two corpora we adopted was not sufficient to warrant such further subsetting.

Acknowledgments

This work was supported by NIH grant DC-002717 to Haskins Laboratories. A preliminary version of this study was presented at the International Congress of Phonetic Sciences 2019 (Chen, Whalen, & Tiede, 2019), for which we received useful feedback from the audience.

References

Autesserre D, Roubeau R, Di Cristo A, Chevrie-Muller C, Hirst D, Lacau J, & Maton B. (1987). Contribution du cricothyroidien et des muscles sous-hyoidiens aux variations de la fréquence fondamentale en françgais: Approche électromyographique. Paper presented at the Proceedings XIth International Congress of Phonetic Science.

Bates D, MÄchler M, Bolker B, & Walker S. (2015). Fitting Linear Mixed-Effects Models Using lme4.2015, 67(1), 48. doi:10.18637/jss.v067.i01

Beil RG (1962). Frequency Analysis of Vowels Produced in a Helium - Rich Atmosphere. The Journal of the Acoustical Society of America, 34(3), 347–349. doi:10.1121/1.1928124

Benjamini Y, & Hochberg Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological), 57(1), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x

Black JW (1949). Natural Frequency, Duration, and Intensity of Vowels in Reading. Journal of speech and hearing disorders, 14(3), 216–221. doi:10.1044/jshd.1403.216

Boersma P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Paper presented at the Proceedings of the Institute of Phonetic Sciences.

Boersma P, & Weenink D. (2019). Praat: doing phonetics by computer (version 6.0.49) [computer program] (Version 6.0.49): http://www.praat.org. Retrieved from www.praat.org

Chen W-R, Whalen DH, & Shadle CH (2019). F0-induced formant measurement errors result in biased variabilities. The Journal of the Acoustical Society of America, 145(5), EL360–EL366. doi:10.1121/1.5103195 [PubMed: 31153348]

Chen W-R, Whalen DH, & Tiede MK (2019, 8). Mandibular contribution to vowel-intrinsic F0. Paper presented at the International Congress of Phonetic Sciences (ICPhS), Melbourne, Australia.

Cohen J. (1988). Statistical Power Analysis for the Behavioral Sciences (Second ed.): Routledge.

Connell B. (2002). Tone languages and the universality of intrinsic F 0: evidence from Africa. Journal of Phonetics, 30(1), 101–129. doi:10.1006/jpho.2001.0156

Diehl RL, & Kluender KR (1989). On the objects of speech perception. Ecological psychology, 1(2), 121–144. doi:10.1207/sl5326969eco0102_2

Dyhr N. (1990). The Activity of the Cricothyroid Muscle and the Intrinsic Fundamental Frequency in Danish Vowels. Phonetica, 47(3–4), 141–154. doi:10.1159/000261859 [PubMed: 2130379]

Erickson D. (2002). Articulation of extreme formant patterns for emphasized vowels. Phonetica, 59(2–3), 134–149. [PubMed: 12232464]

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(3), 137–146.

Ewan WG (1979). Can intrinsic vowel F0 be explained by source/tract coupling? The Journal of the Acoustical Society of America, 66(2), 358–362. doi:10.1121/1.383669 [PubMed: 512198]

Ewan WG, & Krones R. (1974). Measuring larynx movement using the thyroumbrometer. Journal of Phonetics, 2(4), 327–335. doi:10.1016/S0095-4470(19)31302-6

Fischer-Jørgensen E. (1990). Intrinsic f0 in tense and lax vowels with special reference to German. Phonetica, 47(3–4), 99–140. doi:10.1159/000261858 [PubMed: 2130383]

Fox J, & Weisberg S. (2019). An R Companion to Applied Regression (Third ed.). Thousand Oaks, CA.

Gandour J, & Weinberg B. (1980). On the relationship between vowel Height and fundamental frequency: evidence from esophageal Speech. Phonetica, 37(5–6), 344–354. doi:10.1159/000260002

Honda K. (1983). Relationship between pitch control and vowel articulation. Haskins Laboratories Status Report on Speech Research, SR 73, 269–282.

Honda K. (1995). Laryngeal and extra-laryngeal mechanisms of F0 control. In Bell-Berti F & Raphael LJ (Eds.), Producing speech: contemporary issues for Katherine Safford Harris (pp. 215–232). New York: American Institute of Physics.

Honda K, & Fujimura O. (1991). Intrinsic vowel F0 and phrase-final F0 lowering: phonological vs. biological explanations. In Gauffin & Hammarberg (Eds.), Vocal fold physiology (pp. 149–158).

Hoole P, & Honda K. (2011). Automaticity vs. feature-enhancement in the control of segmental F0. In Clements GN & Ridouane R (Eds.), Where do phonological features come from (pp. 131–171). Amsterdam/Philadelphia: John Benjamins B. V.

Hoole P, & Mooshammer C. (2002). Articulatory analysis of the German vowel system. In Auer P, Gilles P, & Spiekermann H (Eds.), Silbenschnitt und Tonakzente (pp. 129–152). Tügingen: Niemeyer.

Hothorn T, Bretz F, & Westfall P. (2008). Simultaneous Inference in General Parametric Models. Biometrical Journal, 50(3), 346–363. [PubMed: 18481363]

IEEE. (1969). IEEE recommended practice for speech quality measurements. IEEE Transactions on Audio and Electroacoustics, 17(3), 225–246. doi:10.1109/TAU.1969.1162058

Iivonen AK (1989). Regionally determined realization of the standard German vowel system. Mimeographed Series of the Department of Phonetics, University of Helsinki, 15, 21–28.

Kingston J. (1992). The phonetics and phonology of perceptually motivated articulatory covariation. Language and speech, 35(1–2), 99–113. doi:10.1177/002383099203500209 [PubMed: 1287395]

Kuznetsova A, Brockhoff PB, & Christensen RHB (2017). lmerTest Package: Tests in Linear Mixed Effects Models. 2017, 82(13), 26. doi:10.18637/jss.v082.i13

Ladd DR, & Silverman KEA (1984). Vowel intrinsic pitch in connected speech. Phonetica, 41(1), 31–40. doi:10.1159/000261708

Ladefoged P. (1968). A phonetic study of West African languages: an auditory-instrumental survey (Second ed.). Cambridge: Cambridge University Press.

Ladefoged P, DeClerk J, Lindau M, & Papçun G. (1972). An auditory-motor theory of speech production. UCLA Working Papers in Phonetics, 22(48), 48–76.

Lindblom B, & Sundberg J. (1971). Neurophysiological representation of speech sounds. Paper presented at the 15th World Congress of Logopedics and Phoniatrics, Buenos Aires, Argentina.

Löfqvist A, Baer T, McGarr NS, & Story RS (1989). The cricothyroid muscle in voicing control. The Journal of the Acoustical Society of America, 85(3), 1314–1321. doi:10.1121/1.397462 [PubMed: 2708673]

Maeda S. (1990). Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model. In Speech Production and Speech Modelling (pp. 131–149): Springer.

Meyer E. (1896/7). Zur Tonbewegung des Vokals im gesprochenen und gesungenen Einzelwort. Phonetische Studien (Beiblatt zu der Zeitschrift Die neueren Sprachen), 10, 1–21.

Mooshammer C, Hoole P, Alfonso P, & Fuchs S. (2001). Intrinsic pitch in German: A puzzle? Paper presented at the The 142nd Meeting of the Acoustical Society of America, Ft Lauderdale, Florida. 10.1121/1.4777648

Ohala JJ (1973). Explanations for the intrinsic pitch of vowels. Monthly Internal Memorandum, Phonology Laboratory, University of California at Berkeley, 9–26.

Ohala JJ, & Eukel BW (1987). Explaining the intrinsic pitch of vowels. In Channon R & Shockey L (Eds.), In honor of Ilse Lehiste: Ilse Lehiste Pühendusteos (pp. 207–215). Dordrecht, The Netherlands: Foris Publications.

Pape D, & Mooshammer C. (2006, Dec-13 ~ 15). Intrinsic F0 differences for German tense and lax vowels. Paper presented at the Proceedings of the 7th International Seminar on Speech Production, Ubatuba, Brazil.

Pernet C, Wilcox R, & Rousselet G. (2013). Robust correlation analyses: false positive and power validation using a new open source MATLAB toolbox. Frontiers in Psychology, 3(606). doi:10.3389/fpsyg.2012.00606

Peterson GE, & Barney HL (1952). Control methods used in a study of the vowels. The Journal of the Acoustical Society of America, 24(2), 175–184. doi:10.1121/1.1906875

R Core Team. (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from URL http://www.R-project.org/

Rossi M, & Autesserre D. (1981). Movements of the hyoid and the larynx and the intrinsic frequency of vowels. Journal of Phonetics, 9(2), 233–249.

Rousseeuw PJ (1984). Least median of squares regression. Journal of the American Statistical Association, 79(388), 871–880. doi:10.1080/01621459.1984.10477105

Shadle CH (1985). Intrinsic fundamental frequency of vowels in sentence context. The Journal of the Acoustical Society of America, 78(5), 1562–1567. doi:10.1121/1.392792 [PubMed: 4067070]

Shaw JA, Chen W-R, Proctor MI, & Derrick D. (2016). Influences of tone on vowel articulation in Mandarin Chinese. Journal of speech, language, and hearing research, 59(6), S1566–S1574. doi:10.1044/2015_JSLHR-S-15-0031

Shimizu K. (1960). On the motions of the vocal cords in phonation studied by means of the high voltage radiograph movies. Oto-rhino-laryng. Clinic, Kyoto, 53, 446–461.

Shimizu K. (1961). Experimental studies on movements of the vocal cords during phonation by high voltage radiograph motion pictures. Studies in Phonology, 1, 111–116.

Steele SA (1986). Interaction of Vowel F0 and Prosody. Phonetica, 43(1–3), 92–105. doi:10.1159/000261763

Sundberg J. (1969). Articulatory differences between spoken and sung vowels in singers. STL-QPSR, KTH, 10(1), 33–46.

Taylor HC (1933). The fundamental pitch of English vowels. Journal of Experimental Psychology, 16(4), 565–582. doi:10.1037/h0070672

Tiede M, Chen W. r., & Whalen DH (2019, 8). Taiwanese Mandarin sibilant contrasts investigated using coregistered EMA and Ultrasound. Paper presented at the International Congress of Phonetic Sciences (ICPhS), Melbourne, Australia.

Tiede M, Espy-Wilson CY, Goldenberg D, Mitra V, Nam H, & Sivaraman G. (2017). Quantifying kinematic aspects of reduction in a contrasting rate production task. The Journal of the Acoustical Society of America, 141(5), 3580–3580. doi:10.1121/1.4987629

Tomaschek F, Hendrix P, & Baayen RH (2018). Strategies for addressing collinearity in multivariate linguistic data. Journal of Phonetics, 71, 249–267. doi:10.1016/j.wocn.2018.09.004

Traunmüller H. (1981). Perceptual dimension of openness in vowels. The Journal of the Acoustical Society of America, 69(5), 1465–1475. doi:10.1121/1.385780 [PubMed: 7240581]

Turner P, & Verhoeven J. (2011). Intrinsic vowel pitch: a gradient feature of vowel system? Paper presented at the 17th International Conference of Phonetic Sciences (ICPhS XVII), Hong Kong.

Van den Berg J. (1955). On the role of the laryngeal ventricle in voice production. Folia Phoniatrica et Logopaedica, 7(2), 57–69.

Westbury JR (1994). X-ray Microbeam Speech Production Database User's Handbook. Madison, WI: University of Wisconsin.

Westbury JR, & Fujimura O. (1989). An articulatory characterization of contrastive emphasis in correcting answers. The Journal of the Acoustical Society of America, 85(S1), S98–S98. doi:10.1121/1.2027241

Whalen DH, & Gick B. (2001). Intrinsic F0 and tongue depth in ATR languages. The Journal of the Acoustical Society of America, 110(5), 2761–2761. doi:10.1121/1.4777643

Whalen DH, Gick B, Kumada M, & Honda K. (1999). Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0. Journal of Phonetics, 27(2), 125–142. doi:10.1006/jpho.l999.0091

Whalen DH, & Levitt AG (1995). The universality of intrinsic F0 of vowels. Journal of Phonetics, 23(3), 349–366. doi:10.1016/S0095-4470(95)80165-0

Yuan J, & Liberman M. (2008). Speaker identification on the SCOTUS corpus. The Journal of the Acoustical Society of America, 123(5), 3878.

Zawadzki PA, & Gilbert HR (1989). Vowel fundamental frequency and articulator position. Journal of Phonetics, 17(3), 159–166.

Highlights:

- The articulatory mechanism for intrinsic f0 (IF0) is still unclear
- Articulatory data of 44 American English speakers was examined for correlates of IFO
- The mandible was found to make contributions to IF0 independently of the tongue
- A hypothesis for a dual mechanism for IF0 is supported

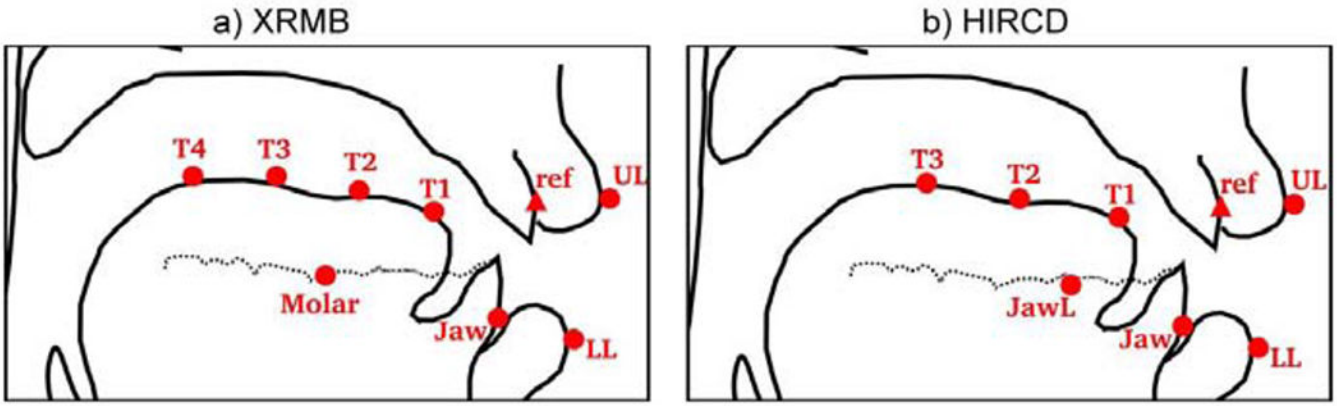

Figure 1.

Schematic plots on the mid-sagittal plane for the kinematic tracking points in (a) the X-ray microbeam database (XRMB) and (b) the Haskins IEEE rate comparison database (HIRCD). Dotted lines in both (a) and (b) represent the traces of lower teeth on the speaker's left-hand side.

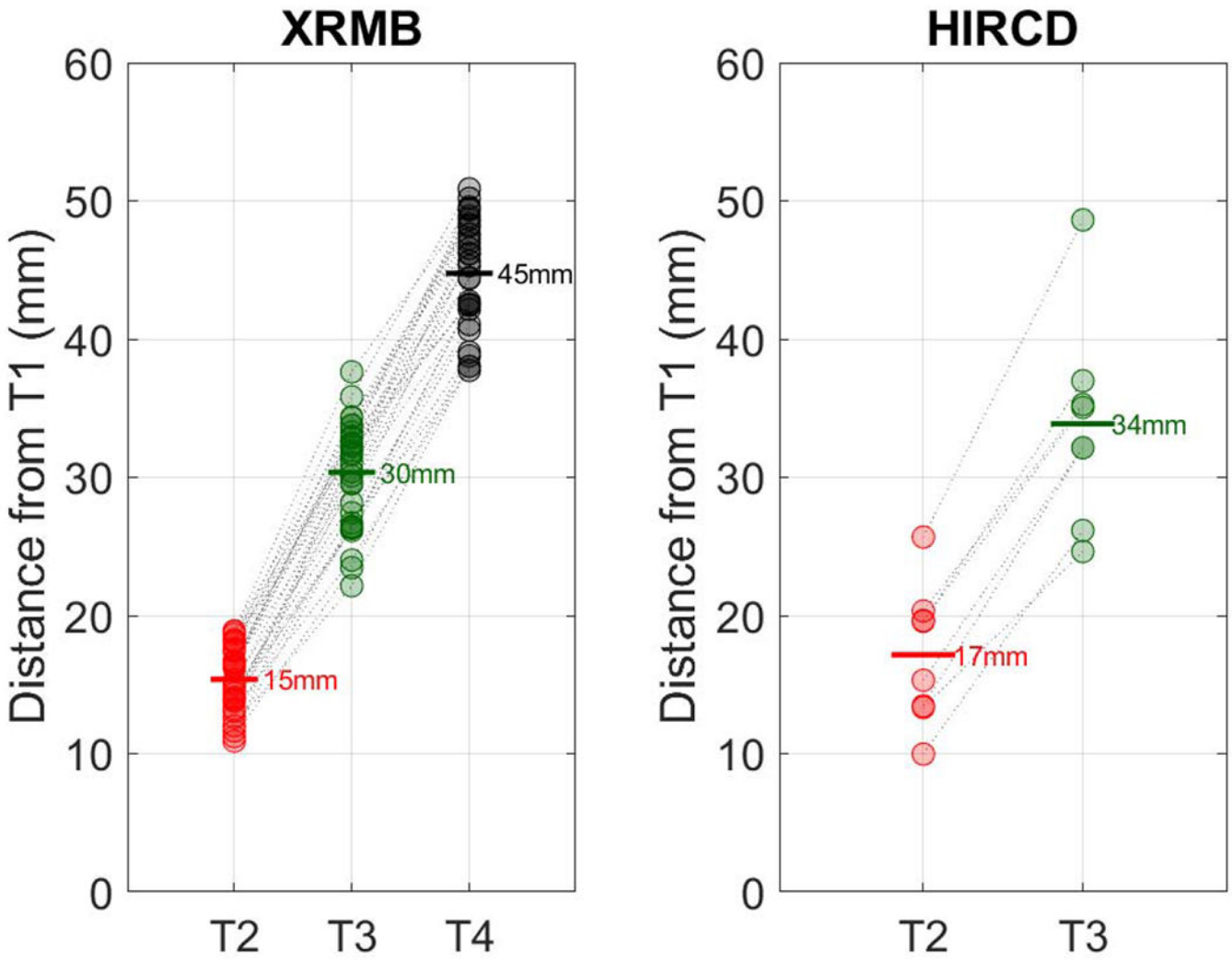

Figure 2.
Relative distances from the tongue tip (T1) to the other tongue points for XRMB (left) and HIRCD (right). Horizontal lines indicate the means.

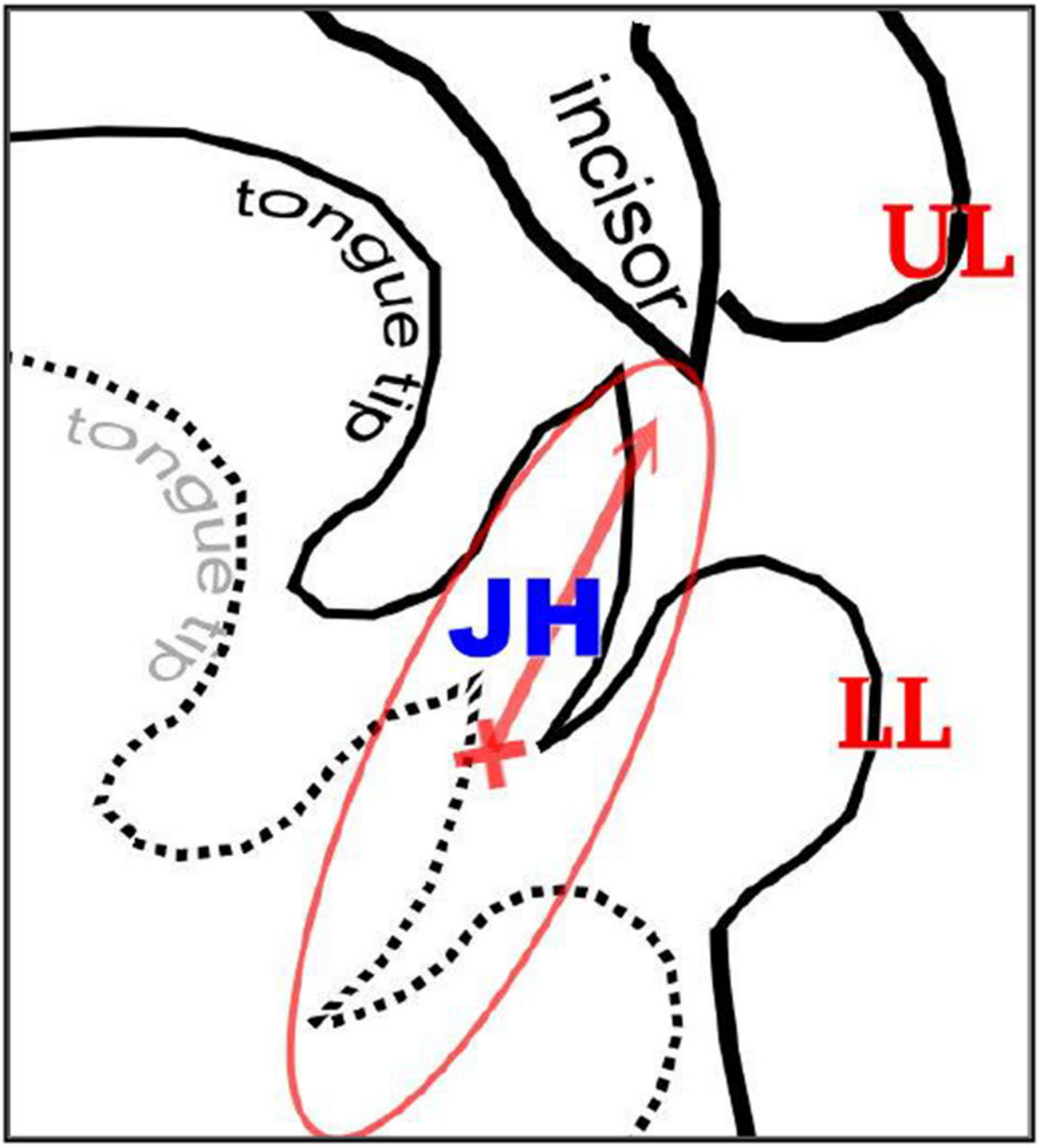

Figure 3.
Schematic plot of jaw height (JH) as the first principal component of the jaw position (measured by locations of the lower incisor). Dotted line represents a lower jaw position. UL and LL indicate upper lip and lower lip, respectively.

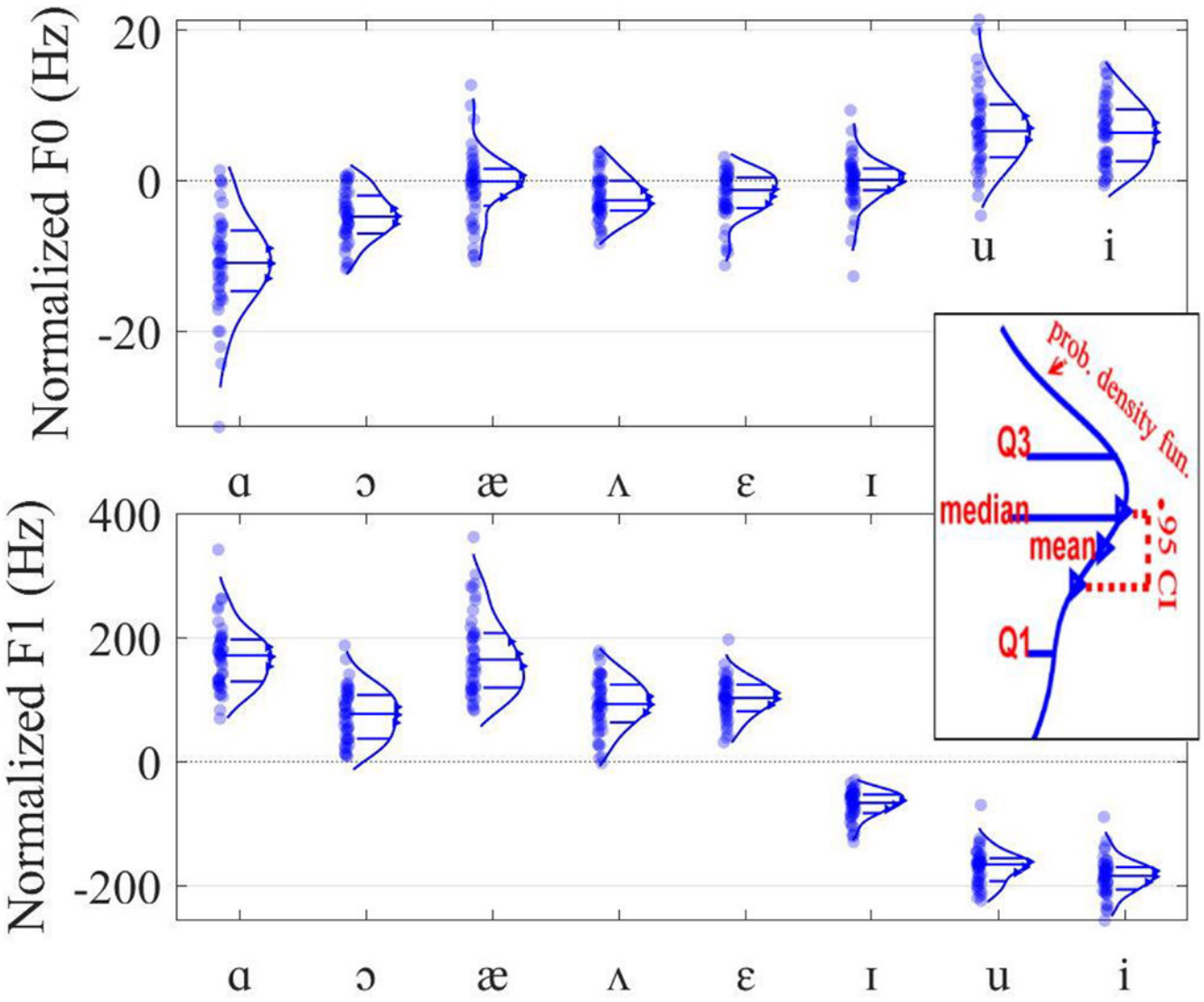

Figure 4.

Distributions of f0 (top panel) and F1 (bottom panel) for 44 speakers, sorted from low (left) to high (right) vowels. Normalization was done by subtracting the median value within-speaker. Each datapoint indicates the median of one speaker for a vowel. Each curved line represents the probability density function for a vowel. Horizontal lines delimit quartiles 1 ~ 3.

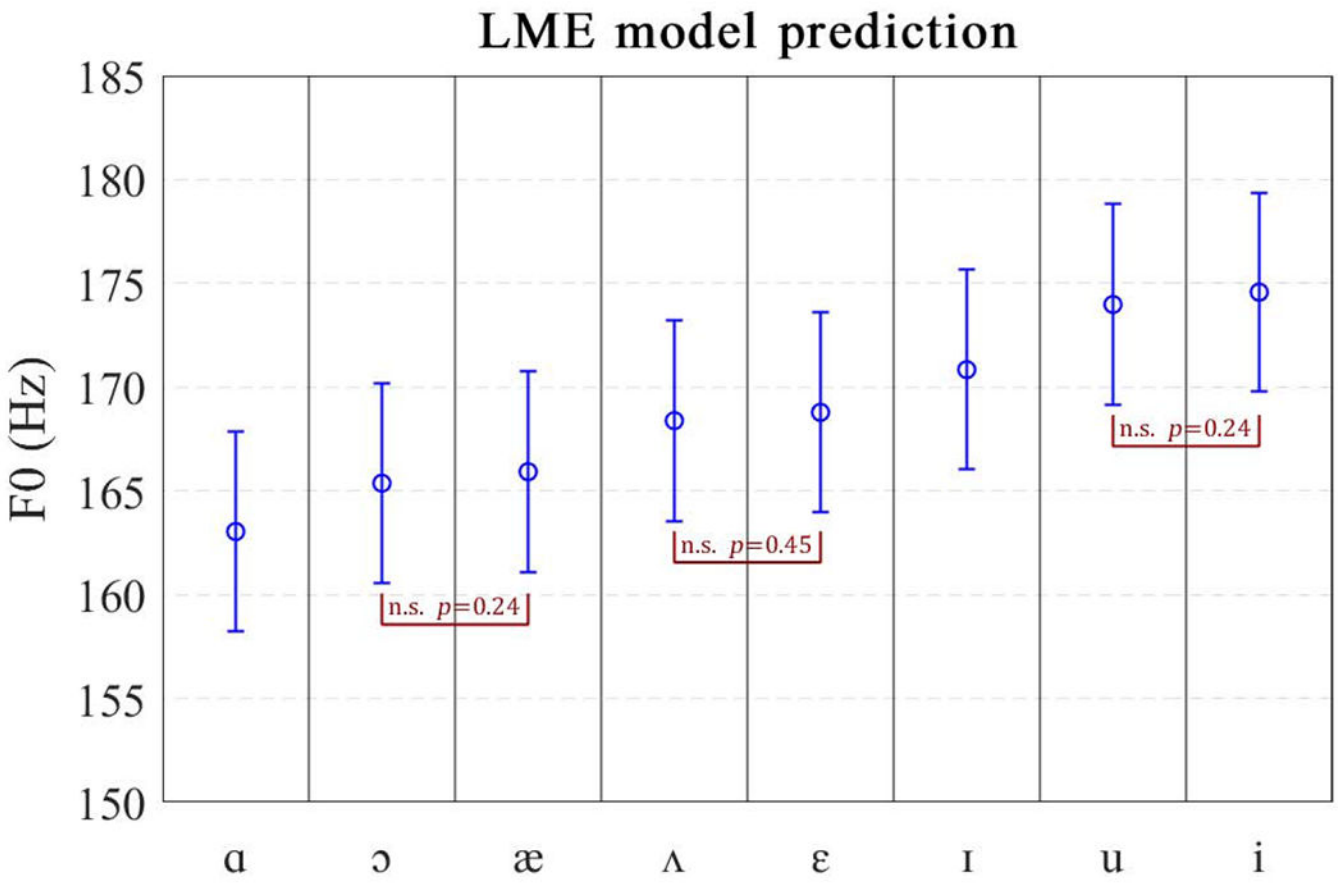

Figure 5.
Marginal effects of vowels predicted by an LME model (f0 ~ Vowel + Gender + (1| Speaker)). P values of pairwise post-hoc comparisons are displayed only for non-significant pairs.

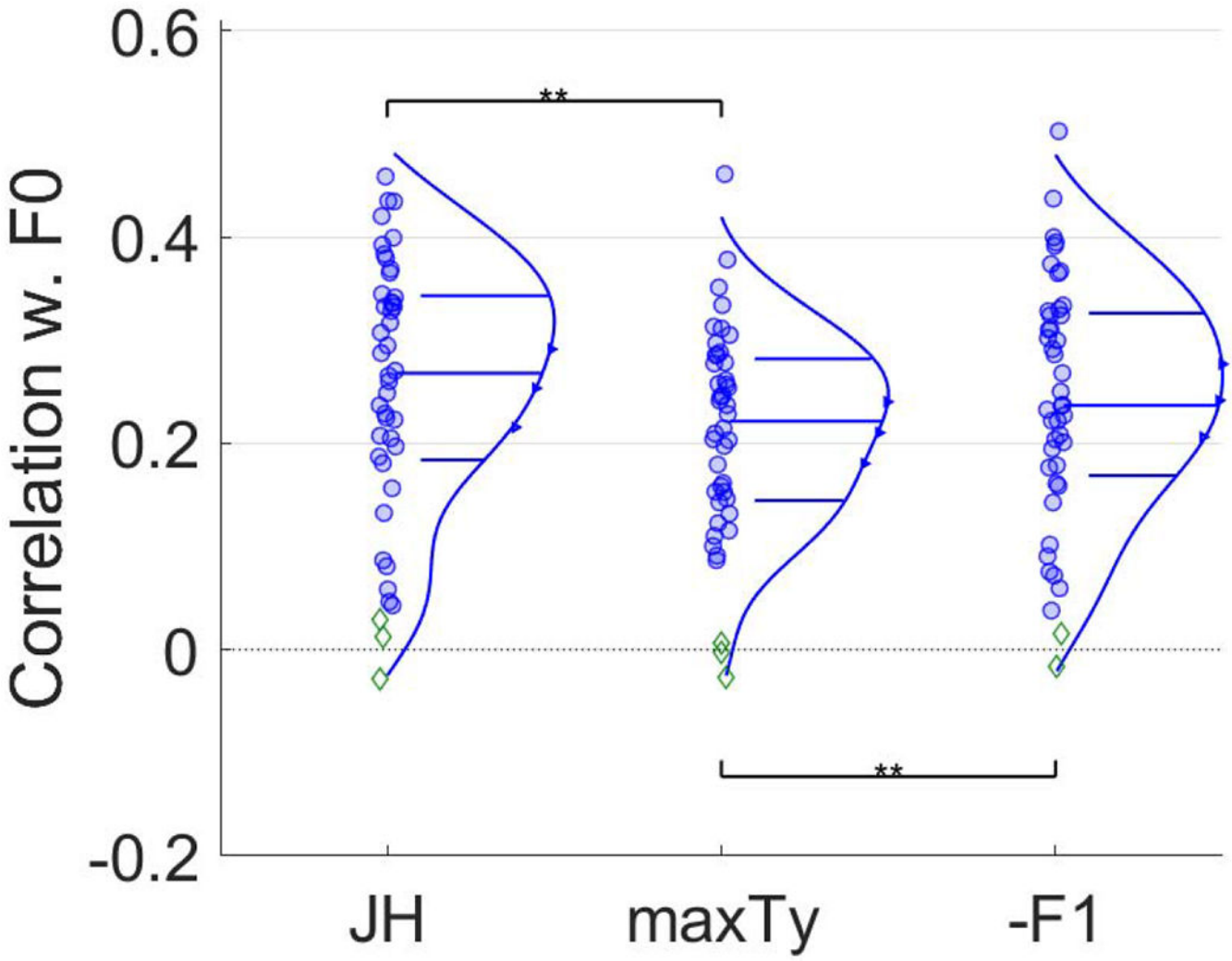

Figure 6.

Articulatory and acoustic correlates of IF0. Each datapoint (circle or diamond) indicates the correlation off0 with each articulatory or acoustic parameter (JH, maxTy, and–F1)for one speaker. Filled circles indicate the individual correlation coefficient for one speaker was significantly different from zero; unfilled diamond symbols indicate otherwise.

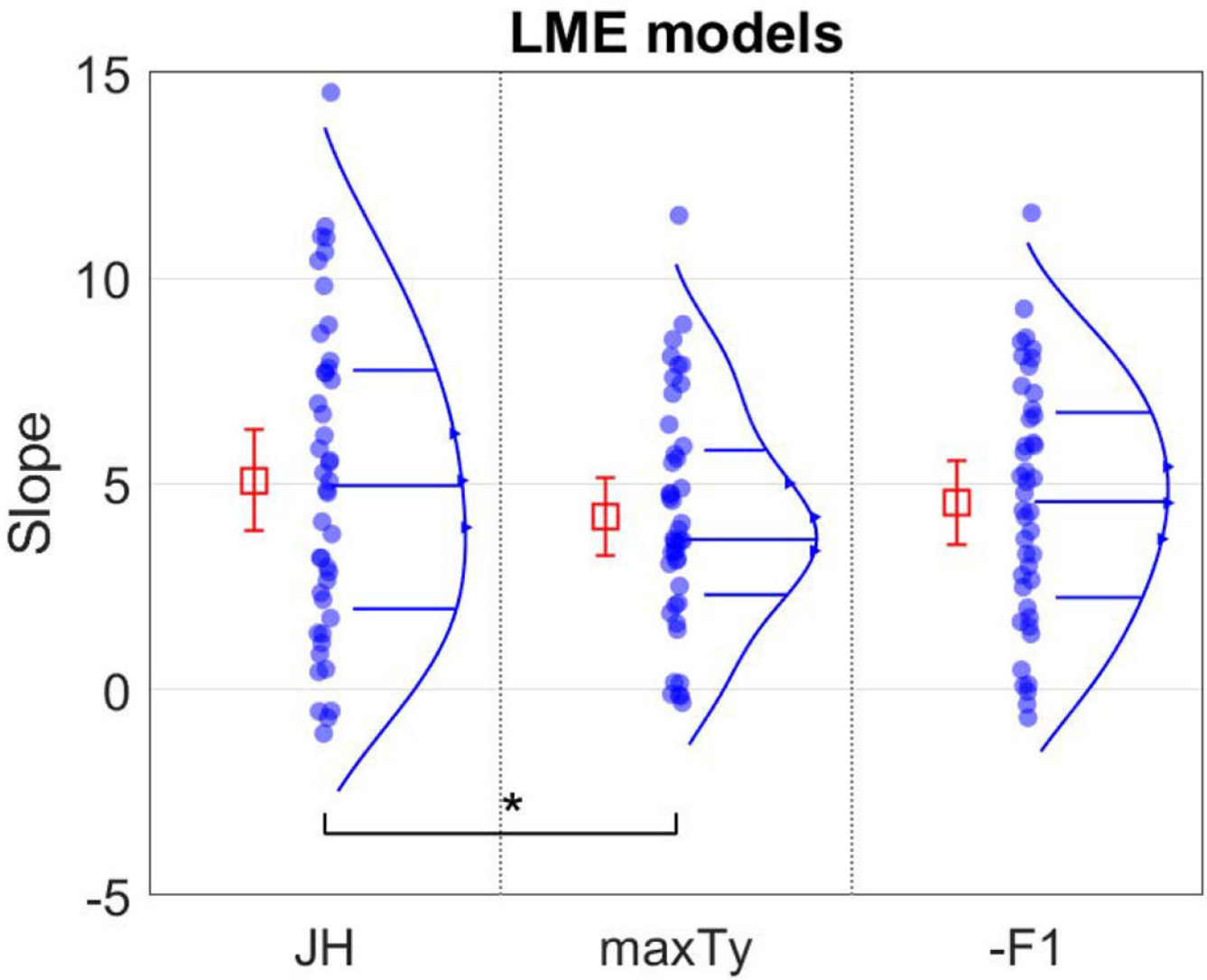

Figure 7.

Coefficients of LME models predicting f0 from one of the three variables JH, maxTy and −F1 as the main effect with by-speaker random intercept and random slope. Each circle indicates the random slope of a speaker superimposed on the main effect. Squares indicate the coefficient of the main effect and error bar the 95% confidence interval.

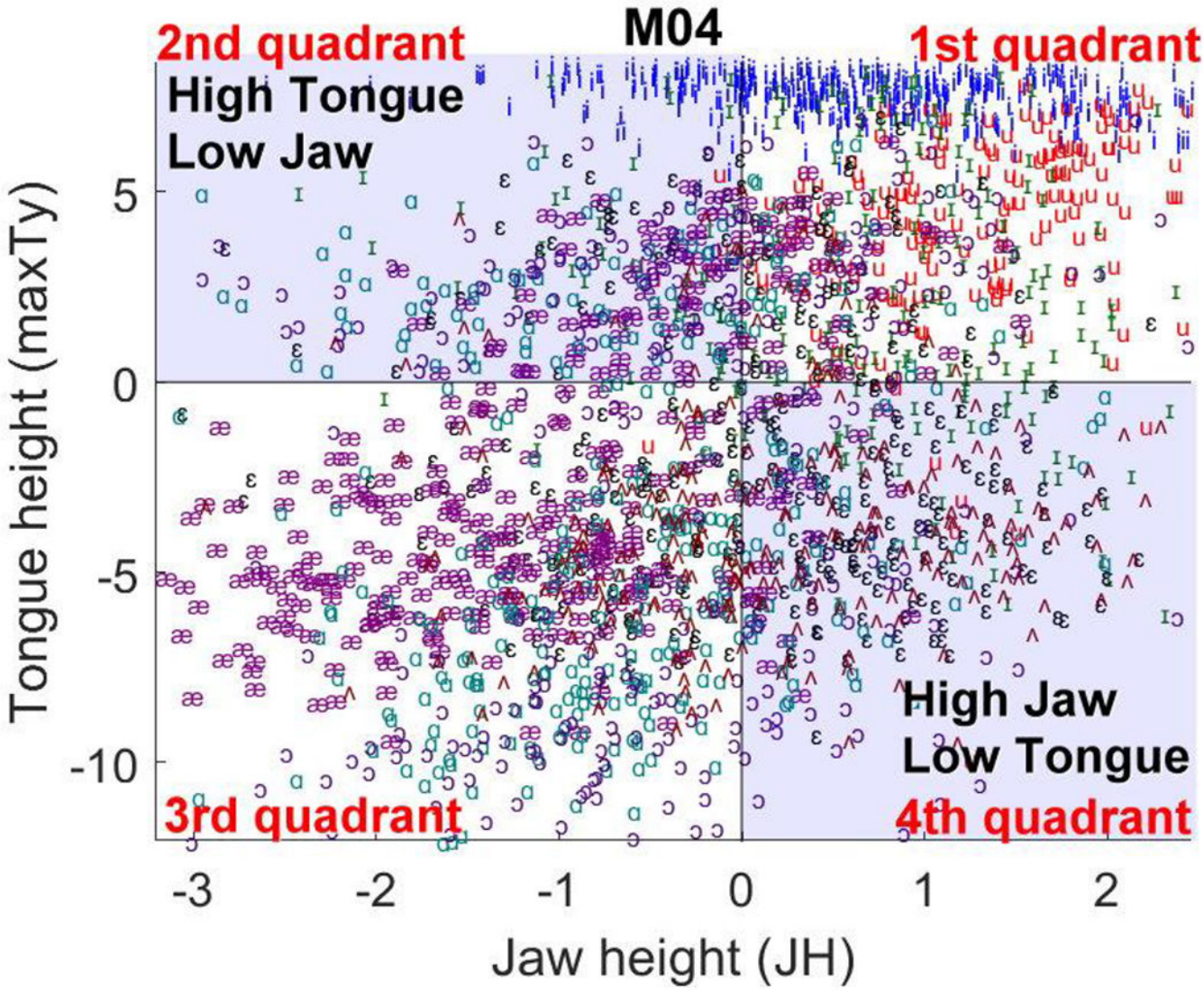

Figure 8.

A scatterplot of tongue height (maxTy) against jaw height (JH) for the speaker M04, centered at the medians. The origin was set at the medians of JH and maxTy. The 2nd quadrant contains vowels produced with the tongue position higher than the median and the jaw position lower than the median and the 4th quadrant shows the reverse relationship. Because these do not follow the expected positive correlation shown by the first and third quadrants, these samples are relatively uncorrelated for tongue versus jaw height effects.

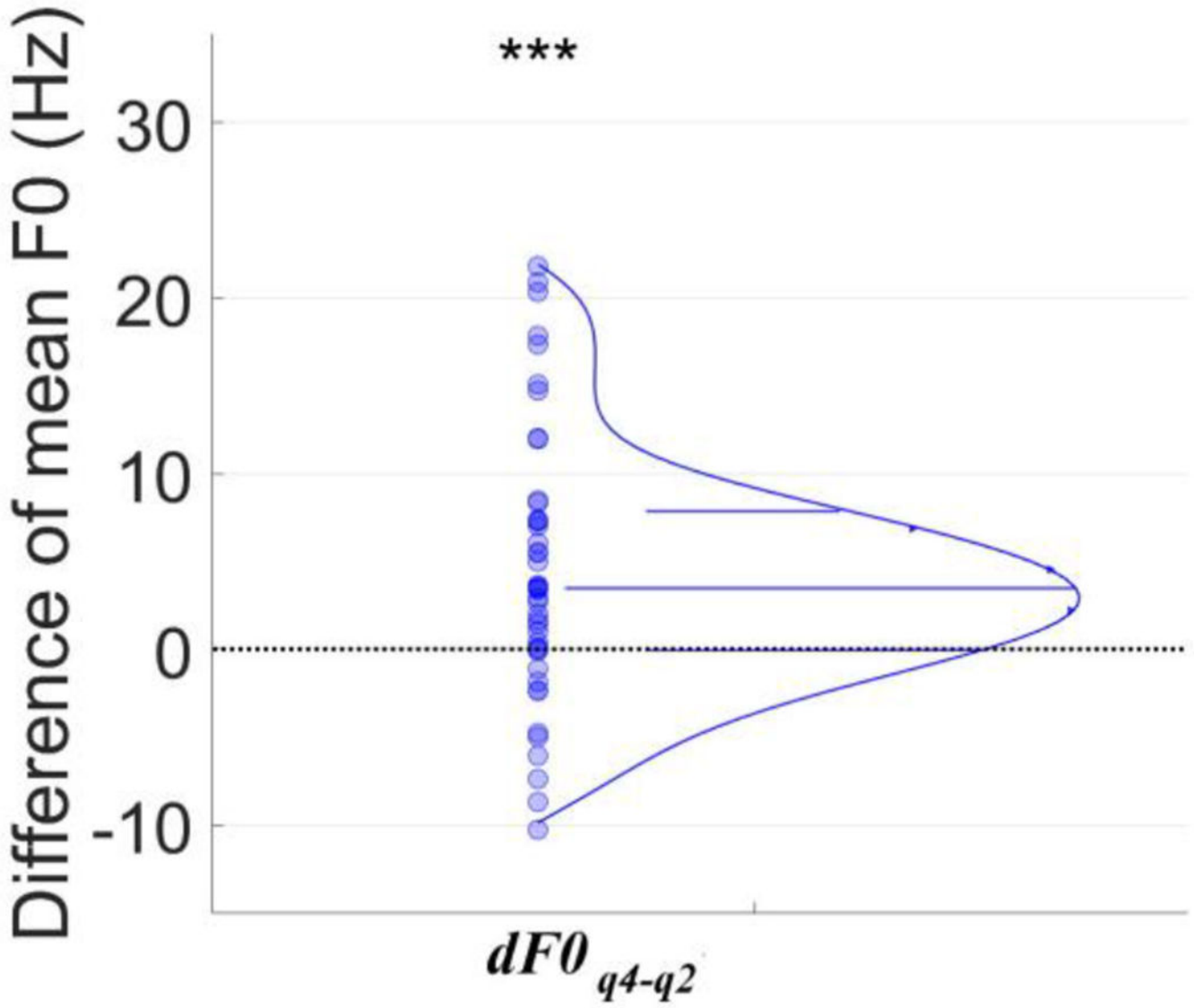

Figure 9.

Distributions of quadrant 4 - quadrant 2 f0 difference ($dF0_{q4-q2}$) across 44 speakers. Positive value $dF0_{q4-q2}$ indicates the contribution of jaw height to f0 is greater than that of tongue height; negative value, the opposite. Each filled circle represents one speaker. One-sample t-*test revealed that the mean of '$dF0_{q4-q2}$ 'across 44 speakers was significantly greater than zero ($p < 0.001$ ***).*

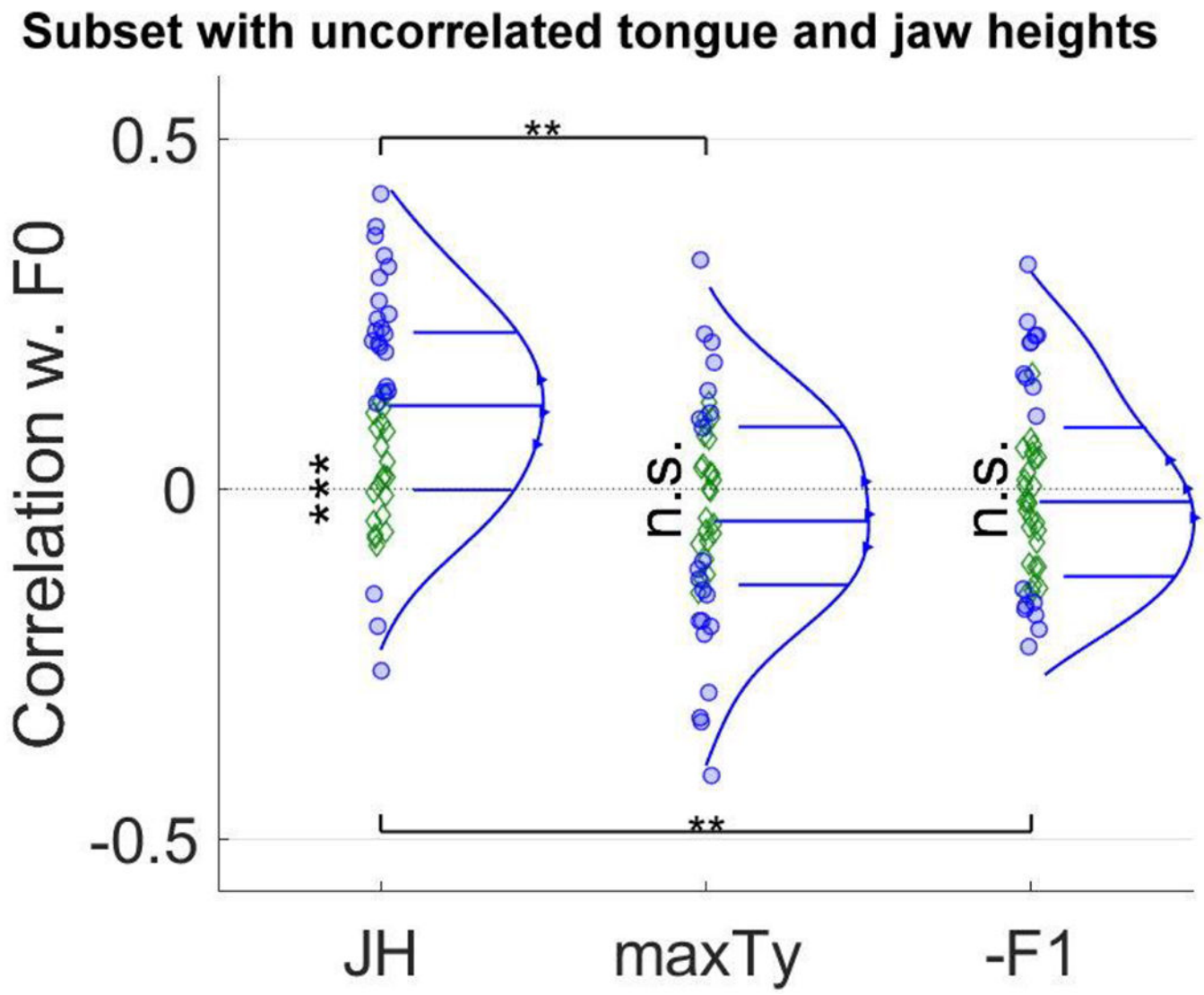

Figure 10.

Correlations of jaw height (JH), tongue height (maxTy), and negated F1 (−F1) with f0, calculated with the subset in which tongue height and jaw height are relatively uncorrelated. Significance symbols displayed vertically indicate whether the mean of each variable is significantly different from zero. Significance symbols displayed horizontally indicate whether the two variables are significantly different from each other.

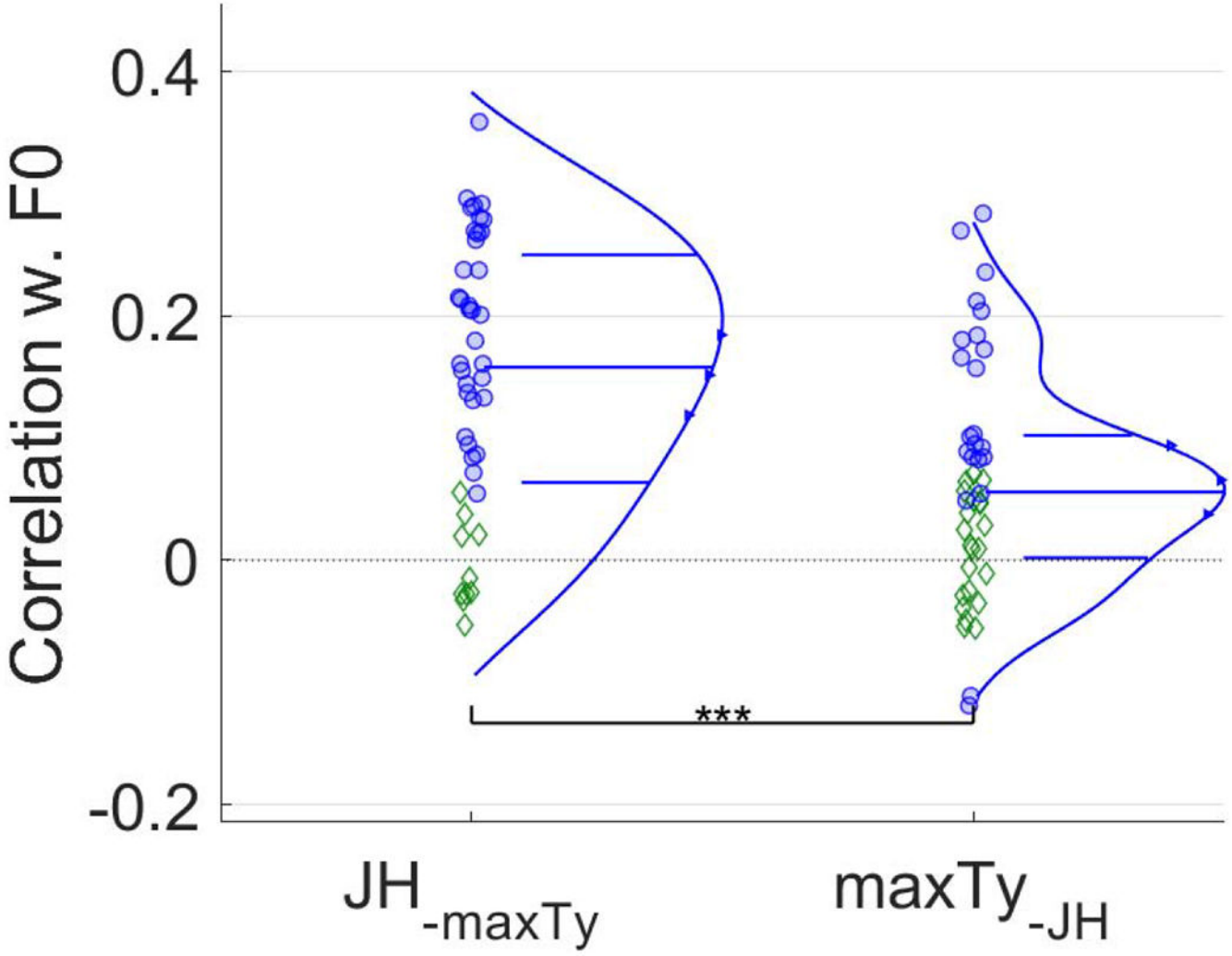

Figure 11.

Left: The remaining jaw height effect on f0 after the removal of tongue height effect. Right: The remaining tongue height effect on f0 after the removal of jaw height effect. Each circle or diamond symbol indicates one speaker.

Table 1.

Mean (across speakers) proportional distributions of vowel tokens in four quadrants. Numbers in parentheses indicate standard deviations (across speakers). Bold indicates the plurality quadrant.

Vowels	1st quadrant	2nd quadrant	3rd quadrant	4th quadrant
i	**80.9** (9.5)%	17.1 (8.8)%	0.1 (0.2)%	0.2 (0.7)%
u	**93.3** (4.1)%	2 (2)%	0.2 (0.6)%	3.7 (3.6)%
I	**57.4** (9.8)%	18.9 (6.7)%	5 (5.3)%	17.5 (7.2)%
ɛ	8.3 (6.3)%	12.2 (7.2)%	**41.9** (11.2)%	34.8 (11.2)%
æ	5.5 (4.1)%	16.6 (9.5)%	**69.6** (11.3)%	7.1 (3.6)%
ɑ	4.2 (3.5)%	13.2 (7.6)%	**72.5** (8.9)%	8.1 (5.5)%
ɔ	11.8 (7.5)%	17.4 (11.1)%	**60.5** (13.5)%	8.5 (5.2)%
ʌ	5.6 (3.8)%	2.2 (2.4)%	**55.6** (11.1)%	35.1 (9.6)%