1146

Journal of Phonetics (1998) **27**, 125–142
Article No: jpho1999.0091
Available online at http://www.idealibrary.com on IDEAL®

Cricothyroid activity in high and low vowels: exploring the automaticity of intrinsic F0

D. H. Whalen

Haskins Laboratories, 270 Crown St., New Haven, CT 06511, U.S.A.

Bryan Gick

Haskins Laboratories, 270 Crown St., New Haven, CT 06511, and Department of Linguistics, Yale University, New Haven, CT 06520, U.S.A.

Masanobu Kumada

Haskins Laboratories, 270 Crown St., New Haven, CT 06511, and Department of Surgery, Yale University School of Medicine, New Haven, CT 06520, U.S.A. and Department of Physiology I, National Defense Medical College, Tokorozawa, Japan Saitama 359

and

Kiyoshi Honda

ATR Human Information Processing Research Laboratories, 2-2 Hikaridai Seika-cho Soraku-gun, Kyoto, Japan and Waisman Center, University of Wisconsin, Madison, WI 53705, U.S.A.

Received 12th May 1998, and accepted 23rd June 1999

Although vowels can be produced with any F0 in a speaker's range, the high vowels tend to be produced with a higher F0 than low vowels. This "intrinsic F0" (IF0) has been found for every language that has been examined for it (Whalen & Levitt, 1995) and has also been found in the babbling of prelinguistic infants (Whalen, Levitt, Hsiao & Smorodinsky, 1995), suggesting that it is an automatic consequence of articulation. Nonetheless, some researchers have suggested that IF0 is a deliberate enhancement of the perception of vowel height (Diehl & Kluender, 1989; Kingston, 1992; Fahey & Diehl, 1996). The only positive evidence in favor of this view is that EMG activity for the cricothyroid (CT) muscle has been reported to be higher for high vowels than for low, suggesting active control of F0. The present experiment examines CT activity in four English-speaking subjects saying isolated vowels. In one condition, target tones that differed by the same amount as the IF0 magnitude itself were presented for the subjects to match; in the other condition, there were no targets. CT activity for the condition with targets was higher for the high vowels for only one of the four subjects; the patterns for the other three were negative, neutral or mixed. Since only two subjects from the previous literature are comparable to the present work, the basic assumption of higher CT activity for high vowels must be called into

Correspondence to D. H. Whalen Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695, U.S.A. E-mail: whalen@lenny.haskins.yale.edu.

0095–4470/99/020125 +18 $30.00/0

question. Further, when F0 was shifted by an amount equivalent to that seen in IF0, it was found that the high vowels needed more CT activity to effect a change than the low vowels did. This indicates that it is impossible to compare absolute values of CT activity as indications of direct F0 control. In the condition without targets, CT activity for three subjects followed the pattern of the target condition, being neutral, negative or mixed. Thus, the EMG evidence cited in favor of IF0 being deliberate is contradicted, leaving a preponderance of evidence that IF0 is an automatic consequence of successful vowel articulation.

1. Introduction

We have known for a century now that the fundamental frequency (F0) of high vowels is typically higher than that of low vowels (Meyer, 1896–1897). In the ensuing years, many other studies have found this "intrinsic F0" (IF0) for many languages. Whalen & Levitt (1995) found 58 studies covering 31 languages representing 11 of the world's 29 major language families. All of them reported IF0, and no one language stood out as having atypical values. There have been many attempts at explaining the effect (Honda, 1983; Steele, 1986; Ohala & Eukel, 1987; Sapir, 1989; Fischer-Jørgensen, 1990; Honda & Fujimura, 1991), but there has been no definitive explanation as yet. The present paper examines an aspect of IF0 that is relatively independent of the exact cause of IF0, except for the issue of whether it is deliberate or automatic.

One typical feature of proposed explanations is that IF0 is an automatic aspect of vowel articulation, whether it is due to the acoustic coupling of F1 and F0, the pull of the tongue on the hyoid bone, or some other, as yet unknown cause. Although the earlier explanations were incomplete in various ways, there was no theoretical reason for thinking that IF0 was anything other than an automatic consequence. Indeed, there is strong circumstantial evidence that this view is correct. The universality of the effect is the first indication, but it is equally important to note the kinds of languages that have IF0. Tone languages, which use F0 to signal lexical distinctions, might be presumed to want to control F0 more tightly than other languages, and yet they too show IF0 in addition to the tone differences. Further, the size of the language's vowel inventory does not affect IF0 (Whalen & Levitt, 1995), indicating that irrespective of how crowded the vowel space is, the size of the effect remains the same. Finally, even prelinguistic babbling shows the effect (Whalen et al., 1995), a result that would be difficult to account for if IF0 were anything other than automatic.

One early suggestion that IF0 might, nonetheless, be deliberate came from Gandour & Weinberg (1980), who discovered that even esophageal speakers exhibited F0 changes in the direction of IF0. Since the proposed links between the tongue and the larynx were physically missing in these speakers, it seemed likely that the differences were introduced deliberately. (Other influences of articulation on the tension of the esophageal flap might come into play, though this has not been examined directly.) It was then inferred that this meant that the differences were deliberate in normal speakers as well. However, such a conclusion is unwarranted: if esophageal speakers reintroduce deliberately an effect that was formerly automatic, it would help in some small way in making their speech sound more natural. Such a choice of action would tell us nothing about what typical speakers do.

More recently, several researchers have claimed that IF0 is a deliberate enhancement of the vowel height dimension (Diehl & Kluender, 1989; Kingston, 1992; Fahey & Diehl, 1996), which is presumed to depend not on the frequency of F1 but on the differences between F1 and F0 (Traunmüller, 1981, 1994). This latter, "distance" theory is based on experiments with synthetic vowels that cover a wide range of F0s, which are heard by subjects as representing a single vowel if the acoustic distance between F1 and F0 is the same in all of them. If the distance is small, a high vowel is heard, while a large difference is perceived as a low vowel. According to the "enhancement" theory, IF0 makes this distance even more robust, since raising the F0 of a high vowel will make the distance between it and F1 smaller, while the lower F0 of low vowels will make that distance larger. Because this presumed perceptual strategy does not depend on articulation, the assumption is that IF0 would be deliberately introduced into the system to enhance the perception of vowel height. The assumptions of the two theories are not completely compatible, since the distance theory claims that the differences should be constant, while enhancement theory assumes that exaggerating the differences will be perceptually useful. Presumably, enhancement theory would posit a boundary value between the two categories, so that being further from that value would be less likely to be misheard. The lack of a clear explanation of the articulatory mechanism responsible for IF0 is also mentioned as a reason for its being deliberate. Kingston (1992) also argues that the commonly found lack of IF0 in the lower portion of a speaker's F0 range (Hombert, 1977; Zee, 1980; Ladd & Silverman, 1984; Whalen & Levitt, 1995) is evidence for the deliberate nature of IF0.

The evidence presented earlier contradicts these assumptions. If the perceptual preservation of vowel height were the crucial factor in the deliberate use of IF0, then one would expect that languages that used vowel height relatively little would be more inclined to reduce or eliminate IF0. Whalen & Levitt (1995) found no evidence of any language lacking IF0, and no effect of the size of the vowel inventory, even though one would expect that large inventories would make enhancing a height distinction more necessary. Moreover, the lack of a clear explanation of the source of the effect is a very weak argument against its being automatic. Indeed, an immediate difficulty for enhancement accounts is the fact that vowels can be produced—and perceived—at almost any pitch in a speaker's range. If the F1/F0 difference were truly the perceptual determinant of vowel height, then we would either constantly be mistaking the vowels of stressed syllables because of their higher F0 (Fry, 1958) or changing the F1 to accommodate the new F0. Listeners are apparently sensitive to the effect of vowel height on F0, and thus perceptually "parse" this information so that high and low vowels on the same F0 sound as if the high vowels have the lower pitch (Silverman, 1987; Fowler & Brown, 1997). Finally, the absence of IF0 at the lowest part of the range is merely another indication of how complex the control of F0 is. The strap muscles become involved in the lowering of pitch (Erickson, Baer & Harris, 1983; Hallé, 1994), directly affecting the relationship of the thyroid cartilage with the hyoid bone and thus, in all likelihood, with the changes in the oral cavity as well (see also Honda, Hirai, Masaki & Shimada, in press). That such an interaction could obscure or eliminate an otherwise automatic effect on F0 is not surprising, but it would not be expected to suppress an intentional effect.

The only remaining positive evidence in favor of IF0 being a deliberate enhancement, then, is the activity of the cricothyroid (CT) muscle. CT narrows the angle between the cricoid and thyroid cartilages, increasing tension on the vocal folds which, all else being

equal, raises F0. This muscle has been shown to be active when F0 is raised (Hirose & Gay, 1972; Atkinson, 1978; Roubeau, Chevrie-Muller & Saint Guily, 1997), and so an increase in activity could easily be an indication of planning for a raised F0. Indeed, reports from several languages have found that there is higher CT activity for higher vowels (Autesserre, Roubeau, Di Cristo, Chevrie-Muller, Hirst, Lacau *et al.*, 1987; Dyhr, 1990; Honda & Fujimura, 1991; Vilkman, Aaltonen, Laine & Raimo, 1991). These authors have, in general, concluded that the higher CT activity indicates a deliberate raising of F0, although Vilkman *et al.* suggest that the activity takes place "in order to avoid opening of the cricothyroid visor during increased vertical pull in the laryngeal region" (1989, p. 202).

A note of caution against treating CT as the main determinant of F0 can be found in Honda & Baer (1981) and Honda (1983), who also found a correlation between activity of the genioglossus muscle and IF0. The genioglossus attaches to the mandible anteriorally and radiates posteriorally into the tongue and to the hyoid bone, so that it advances the tongue when it contracts. It may be that the fronting of the tongue pulls the hyoid bone forward, which would tend to pull the thyroid cartilage forward through the connection at the lateral thyroid ligament (Honda, 1983). The action of the CT, then, may be synergistically involved in many laryngeal adjustments, so that the interpretation of any one muscle must be done within limitations. Honda also found paradoxical CT activity in the lower F0 region of a speaker's range (in this case, at the end of sentences), in which increases in CT were correlated with decreases of F0.

The present study was designed to determine whether the increased CT activity found in previous research replicates and is indicative of deliberate planning on the part of the speaker. The total number of subjects tested in previous studies is rather small (seven, from three different language backgrounds), and only one study looked at four subjects together. For the further analysis of the results, our approach was to test an implicit assumption in the interpretation of the EMG measurements: A particular change in CT activity should result in a particular change in F0 regardless of the vowel being produced. If this assumption is incorrect, then a simple interpretation of CT values across different vowels seems impossible. If it takes, say, one unit of CT activity to raise F0 by 1 Hz for /ɑ/ but two units of CT activity to effect the same F0 change for /i/, then it is impossible to compare the baseline values of CT activity, and if we cannot make any claims about the baseline activity, then the only positive evidence for the deliberateness of IF0 would disappear as well. Even if we cannot, at that point, explain the automatic mechanism in full, this in addition to the evidence from the universal distribution and babbling would point to an automatic mechanism.

2. Experimental method

We used electromyography (EMG) to measure muscle activity in the CT. Our goal was to examine CT activity associated with changes in F0 when these F0 changes were of the same magnitude as those found in IF0. In order to induce these changes in F0, we had subjects match a target tone. We also took steps to avoid having the subjects begin singing, since the muscles of the larynx are recruited in unusual ways during singing (Sundberg, 1987).

2.1. *Subjects*

The subjects were four colleagues from Haskins Laboratories and the Linguistics Department of Yale University. Two were male (M1 and M2) and two were female (W1 and W2). None reported any speech pathology.

2.2. *Stimuli*

Two conditions were run for all subjects except M1, for whom only one condition was collected. This condition was a set of productions made in response to an auditorily presented target tone, which the subjects were asked to match at the beginning of the vowel. (The two female subjects performed this condition twice, the males only once.) The other condition was a set of vowels produced in isolation and without any target tone to match.

To generate the target tones, we needed to know the subject's typical F0 for his/her productions of isolated vowels. This was accomplished in a pre-test condition. Before the EMG session, subjects recorded ten repetitions of each of the vowels /ɑ i u/ in random order. The F0 was measured at the beginning of each of these vowels, and these numbers were used as the basis for the tones to be matched. The IF0 difference between the high and low vowels in these natural productions was used as the interval separating the tones. Stimuli were created with one, two and three times this interval, added to the base value of the vowel. Additionally, the average F0 value minus the interval was used to create one lower tone. Separate lists were made for /ɑ/ and for /i u/ so that the lowest tone occurred only with the /ɑ/ and the highest only with the /i/ and /u/ (see Table I).

The tones, 300 ms in duration, were created in a synthesis package (SWS, P. E. Rubin, Haskins Laboratories) that allowed the specification of sine waves representing the first five harmonics of the fundamental. There was a linear intensity ramp over the first 20 ms, after which the the intensity remained steady for 80 ms. There was a linear decline in amplitude over the last 200 ms of the tone. Amplitudes of the first three harmonics were equal, while the fourth and fifth were at one-half the amplitude of the others. These tones were easy to hear as the intended pitch (and were still completely non-speech-like) but not quite as irritating over the course of many repetitions as pure sine waves would have been.

TABLE I. Mean F0 values produced by the speakers for the vowels in isolation, and the generated F0 values for the target tones for each speaker

	Pre-test production			Target values					
Subject	ɑ	i/u	IF0	ɑ − 1	ɑ i/u − 1	ɑ + 1 i/u	ɑ + 2 i/u + 1	ɑ + 3 i/u + 2	i/u + 3
M1	117	123	6	111	117	123	129	135	141
M2	100	104	4	96	100	104	108	112	116
W1	155	160	5	150	155	160	165	170	175
W2	192	204	12	180	192	204	216	228	240

"IF0" is the difference between the high vowels and the low vowel. Targets were the original mean; original plus 1, 2 or 3 IF0 intervals; and original minus 1 IF0 interval. The lowest value was used only for /ɑ/ and the highest only for /i/ and /u/.

Fifteen blocks of productions with target tones were collected. Each block contained three repetitions of the five target tones for one of the three vowels. The subject was told which vowel would be produced in the upcoming block. The vowels were alternated in a pseudo-random fashion across the experiment. This procedure resulted in the collection of 15 repetitions of each vowel for each target tone. Since no firm criteria exist for the acoustic difference between speaking and singing, we relied on our perception to determine whether the subjects remained in speaking mode.

For the condition containing utterances without tonal targets, the subjects produced three blocks of the three vowels, repeating each of the vowels five times in succession, for a total of 15 repetitions of each vowel. One such condition was produced before the target condition, and one after, for a total of 30 utterances per vowel.

2.3. *Procedure*

Hooked-wire electrodes were inserted into the anterior portion (*pars recta*) of the CT (Hirose, 1971). An electrode was inserted at a point above the anterior cricoid arch and approximately 5 mm lateral to the midline. It was directed posterolaterally and slightly upwards toward the inferior thyroid tubercle. The placement was checked by four tests of the accuracy of insertion. First, we found increased activity for raising the F0 in a frequency glissando, in both chest register and head register (falsetto). Such correlations indicate that CT and/or thyroarytenoid (TA) are being recorded. The second task was opening the jaw, during which the CT should be inactive and nearby strap muscles active. We had no activity during this task. Similarly, the third task of raising the head showed no activity, further verifying that the strap muscles had been avoided. For the fourth and final task, subjects swallowed so that we could discriminate between CT and TA activation. CT is suppressed during swallowing but TA is quite active; we had no activation evident during swallowing, indicating that our signals were primarily from CT. Bilateral insertions were attempted for all subjects, but only M1 had successful signals on both sides.

EMG signals were filtered at 2 kHz and digitized at 5 kHz, while the speech signal was filtered at 10 kHz and sampled at 20 kHz. EMG signals were further processed by a triangular window 12 ms in duration.

The conditions were run in the following sequence: first, a block of targetless vowels (this was not done for M1): next, a block of vowels produced in response to target tones: then, another block of targetless vowels. For the two female subjects, there was an additional pair of blocks, targeted and then targetless, which were collected opportunistically given the continued strength of the EMG signal.

3. Results—target condition

The measurement techniques were similar for all of the following analyses. F0 was measured near the beginning of the vowel, before the F0 fall began. They were measured with an autocorrelation function in the HADES program (Rubin, 1995). The subjects were successful in following the instructions to begin with the intended pitch and then allow it to lower (as in statement intonation). The average durations of these isolated vowel utterances were approximately 475, 350, 400 and 375 ms for M1, M2, W1 and W2, respectively. Although vocal intensity was not a measure of interest, it remained fairly

consistent across the productions. All utterances were produced with an effort level typical of normal conversation.

EMG was measured as the average activation (from the rectified and smoothed signal) in the 150 ms prior to vowel onset. Previous research has found that the CT muscle activation anticipates the changes in F0 by approximately 100 ms (Atkinson, 1978; Sapir, McClean & Luschei, 1984), but the shape of the signal for the female talkers indicated that important activity was taking place 150 ms before vowel onset. While inclusion of the entire region of activation was important for characterizing the female patterns, the inclusion of low levels of activation in the average lowered the apparent strength of the signal for the male talkers. This averaging technique is similar to that of Honda (Honda, 1983, 1985; Honda & Fujimura, 1991), but different from that of other researchers (Autesserre *et al.*, 1987; Dyhr, 1990; Vilkman *et al.*, 1991), who measured peak CT activity instead.

The absolute levels of the EMG recordings are somewhat smaller than those in other published studies (Honda, 1983; Löfqvist, McGarr & Honda, 1984; Löfqvist, Bear, McGarr & Story, 1989), but the nature of the signal itself indicates that the muscle was accurately recorded. The inflections in the signal were sharp, and one to three phases of each spike were clearly indicated. When the muscle is only weakly recorded, an intrinsic averaging takes place and the peaks in the signal are smooth. The phases of such signals are also not clearly articulated. As mentioned earlier, the window used was large enough to include all the activation for the female speakers, and so there was some averaging of inactivity for the male speakers. Although the measurements can thus be attributed to CT, the signals from M1's left side were near the lower limit of resolution and are not included in group statistics.

The results for the vowel productions with targets are presented in several ways. First, the accuracy of the F0 matching is assessed. Then, three ways of assessing CT activity are shown.

Subjects were fairly successful in producing a range of F0s in response to the target tone. Table II presents the correlation of the target tone and the F0 attained for all four subjects. (Recall that four of the five targets were common to all vowels, with an additional low target for /ɑ/ and high for /i/ and /u/.) All correlations of target and attained F0 were significant at the 0.01 level. The slopes are less than one, indicating that there was not as much F0 change as was intended, but there was still a broad range of F0s that were achieved. The actual F0 attained was analyzed, without further regard to the target tone that elicited it. Table II presents the means of these F0s. Each subject showed a difference among the vowels by an ANOVA ($F(2, 117) = 4.16, p < 0.05$ for M1; $F(2, 222) = 39.77, p < 0.0001$ for M2; $F(2, 447) = 25.56, p < 0.0001$ for W1; $F(2, 389) = 30.83, p < 0.0001$ for W2). The difference appears between the high vowels and the low.

TABLE II. Column 1: Correlation values between the target F0 and the actual F0. Column 2: Slope of the fitted function. Columns 3–5: Average F0 values by vowel, collapsing across target F0

Subj:	*r*	Slope	ɑ	i	u
M1	0.63	0.37	125.8	128.4	129.8
M2	0.69	0.45	103.0	110.8	114.3
W1	0.69	0.87	172.5	176.3	176.7
W2	0.88	0.84	206.4	218.4	218.4

TABLE III. Unadjusted and adjusted mean CT activation levels for the analysis of all productions, in microvolts. Adjusted means use actual F0 as a covariate. Significant differences among the values are indicated by an asterisk before the subject initials. The numbers in the last row represent the average percent change from /ɑ/ to /i/ or /u/, based on the percent change for each subject

	ɑ	i	u
Unadjusted			
*M1 (left)	6.68	7.88	10.60
*M1 (right)	25.62	25.72	27.03
*M2	26.15	25.59	25.84
W1	20.95	20.76	20.24
*W2	14.30	15.78	14.09
% change†		1.92	− 0.13
Adjusted			
*M1 (left)	7.46	7.98	9.72
*M1 (right)	26.40	25.82	26.14
M2	26.34	25.55	25.69
*W1	21.05	20.56	20.35
*W2	15.92	15.13	13.14
% change†		− 3.12	− 6.06

† Mean percent change from /ɑ/ (excluding M1 left).

The first approach to analyzing the EMG results was to run an analysis of variance for the target conditions, with and without F0 as a covariate. If CT activity increases with the higher F0, then there should be a significant difference in CT activity across the vowels in the ANOVA, as had been found previously. A difference in the ANCOVA, on the other hand, would indicate that something besides F0 is contributing significantly to the CT activity. In Table III, we see that the unadjusted results fail to replicate previous studies of IF0. An ANOVA that treats each F0/EMG pair as a case, with the grouping factors of Speaker and Vowel (which allows for the difference in number of repetitions for the male and female speakers) shows no overall effect of Vowel on EMG value ($F(2, 1338) = 1.76$, n.s.), although there is an interaction of Speaker and Vowel $F(6, 1338) = 16.18$, $p < 0.0001$). Table III shows which speakers, analyzed separately, show a significant effect. While three of the four subjects show significant effects, they are mixed in direction: one had greater activation for high vowels (M1), one lower (M2) and one was higher for one vowel and lower for the other (W2). The overall percentage change from /ɑ/ to /i/ was 1.9% and to /u/, − 0.1%. With the present four subjects, then, we fail to replicate the previously found higher activation for CT with high vowels.

If the CT levels are comparable across vowels, there should be no residual effect left over after F0 has been partialled out. While this assumption is not made explicit in any previous publications, it must at least be largely true for there to be any sense in comparing the levels across the different vowels. In Table III, it becomes clear that the vowels do not achieve these F0 differences in the way: when F0 is partialled out (fitting CT activation with a single regression line for F0, and analyzing the residuals), both /i/ and /u/ are significantly lower in activity and /ɑ/, for all four speakers. If F0 had been changed in predictable steps regardless of which vowel was being articulated, then there should not have been any significant differences.

TABLE IV. Mean F0 values for only the target tones that were common to all three vowels

	a	i	u	i minus a	u minus a
M1	135	134	139	− 1	4
M2	105	109	112	4	7
W1	176	173	174	− 3	− 2
W2	207	215	217	8	10

If subjects are controlling IF0 deliberately, then they should have sufficient control over F0 in these vowels to be able to match the target F0 regardless of vowel when the tone target was the same. There were four such tones for each speaker (Table I). For each subject, an ANOVA was conducted that used target F0 (TF0, four levels) and vowel (V, three levels) as grouping factors for the obtained F0. Individual productions were entered as cases. With the 15 repetitions, this resulted in 180 cases for the male speakers and 360 for the females (who had two repetitions of the target condition). All four speakers achieved different F0s with different targets, ($F(3, 168) = 19.86, 35.28$, $p < 0.001$ for M1 and M2, respectively; $F(3, 348) = 16.08. 270.31$, $p < 0.001$ for W1 and W2, respectively). This result matches what we have already seen in the analysis of the complete data set. As Table IV shows, however, all the speakers differed by vowel ($F(2, 168) = 3.86$, $p < 0.05$, 24.59, $p < 0.001$, for M1 and M2; $F(2, 348) = 7.37, 61.77$, $p < 0.001$ for W1 and W2). The two factors did not interact ($F(6, 168) = 1.10, 0.87$, n.s., for M1 and M2; $F(6, 348) = 1.42$, 1.79, ns. for W1 and W2). For three of the four subjects, there was a difference in the direction of IF0, and one that was of the same magnitude as the IF0 in the isolated productions of the pre-test. For W1, the significant difference indicated that *lower* values were produced for the high vowels. For all four subjects, though, the intention to control F0 directly by matching target tones did not result in the absence of vowel effects.

The analysis of covariance indicates that there is something different in the way that F0 changes are effected for the different vowels. The second analysis for the vowels with targets tests this notion by examining the slopes of the function relating F0 and CT. If the CT values are to be comparable between vowels, then the slopes of the regression lines should be the same across vowels. If not, the levels of CT activity are not comparable. The slopes for our subjects show divergence between the low vowel and the high vowels (see Fig. 1). Since the slopes were calculated on the specific CT activity collected by the electrodes as they happened to be inserted into the muscle, the magnitudes of the signal will differ for every subject (or even for the two sides of the muscle for M1), even if the muscle activity were very similar. Therefore, the slopes can only be compared across the three vowels within one subject's results. For three of the four subjects, the slope for /ɑ/ is considerably lower than that for the high vowels (see Table V). The exception is W1, who had an atypical relationship between CT and F0 in other respects. These differences were tested statistically by running an ANCOVA (with Actual F0 as the covariate) on the variables of Vowel and CT activity. A further analysis which treats Actual F0 as another independent variable will reveal that the slopes of the regression lines between Actual F0 and CT activity are significantly different if there is an interaction between Vowel and Actual F0 (StatView, 1998, pp. 99–102). These interactions were significant for both males ($F(2,219) = 12.51$ and 3.66 for M1 and M2, respectively, $p < 0.0001$ and 0.05, respectively), and for both females ($F(2, 444) = 10.85$ and 36.12 for W1 and W2,

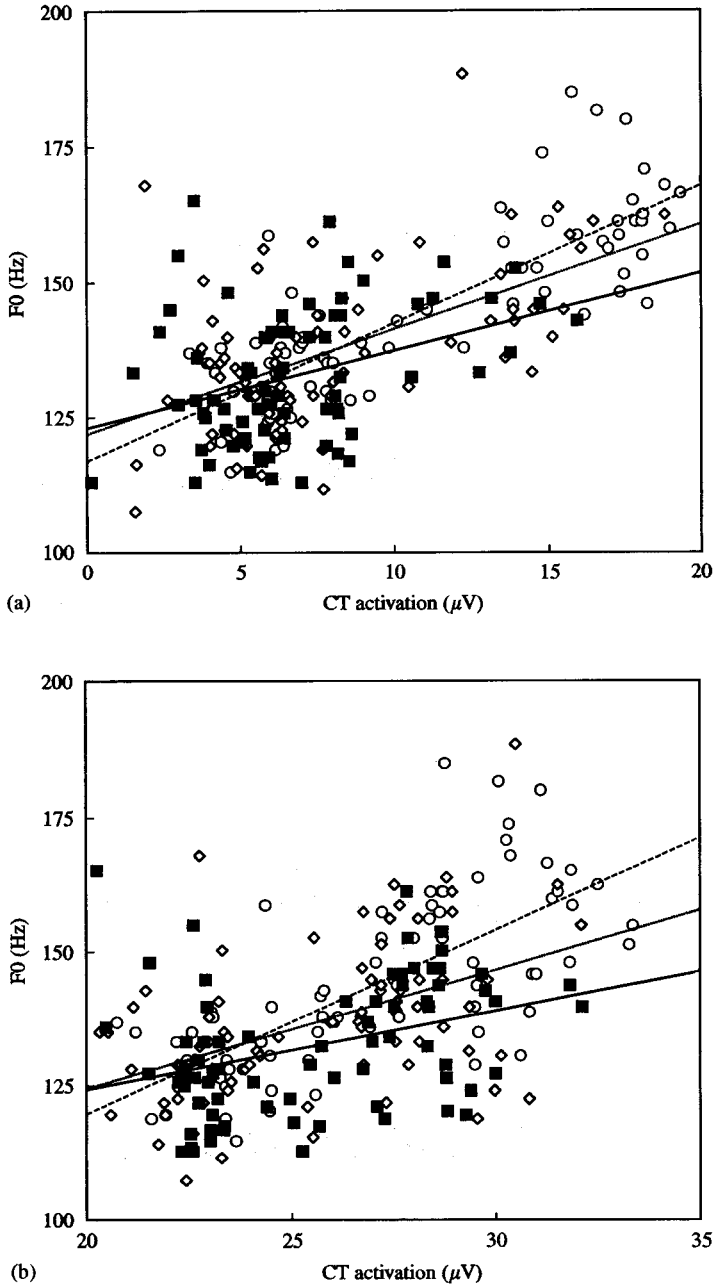

Figure 1. Plots of CT activation by F0 for the three vowels, plotted separately for each subjects: (a) M1, left. (b) M1, right. (c) M2. (d) W1. (e) W2: —, ■ F0-a; ·····, ◇ F0-i; - - - -, ○ F0-u.

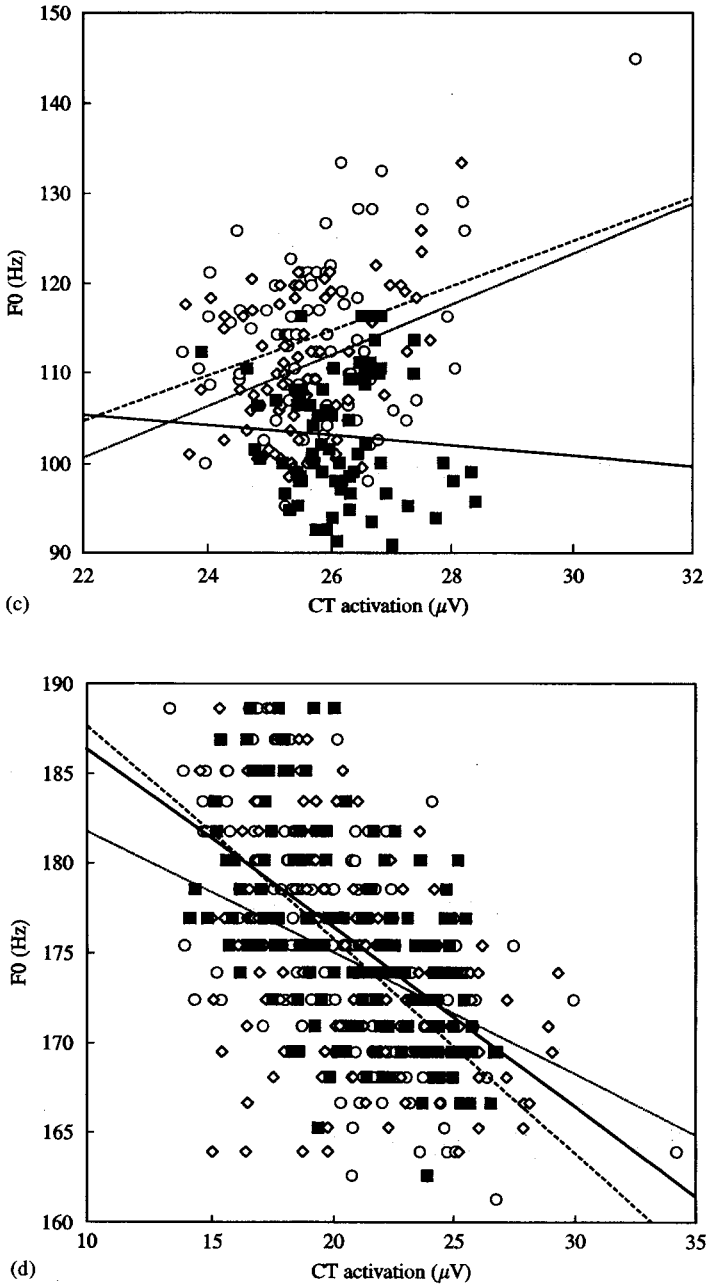

Figure 1. (*Continued*).

respectively, $p < 0.0001$ for both). Thus, the difference apparent in the figure is statistically reliable.

The analysis of the slopes of the correlations between CT and F0 can be extended to three of the other studies as well. Studies by Dyhr (1990) and Autesserre *et al.* (1987) do

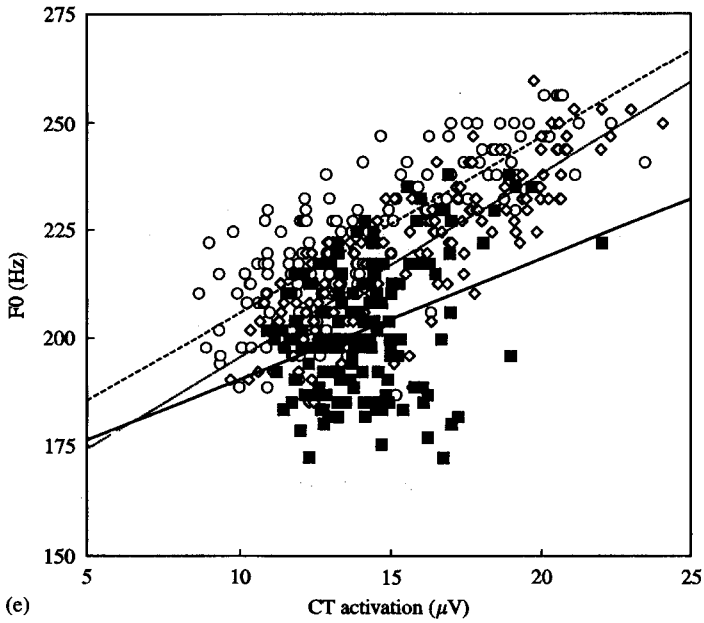

Figure 1. (*Continued*).

TABLE V. Slopes for the regression line of F0 and CT: M1, M2, W1 and W2 represent the current subjects. The other data is computed from the sources cited. Each of these represents one subject, though the two Honda papers are for the same subject at different times

Subject/ study	ɑ		i		u		æ	
	Slope	*r*	Slope	*r*	Slope	*r*	Slope	*r*
M1	1.47	0.35**	2.22	0.46**	3.43	0.70**		
M2	− 0.57	− 0.07	2.83	0.35**	2.49	0.33**		
W1	− 0.99	− 0.58**	− 0.68	− 0.33**	− 1.19	− 0.60**		
W2	2.82	0.35**	4.27	0.83**	4.09	0.79**		
Honda & Baer 1981	10.26	0.67*	12.92	0.93**	17.53	0.96**	9.97	0.67*
Honda 1985	− 0.03	− 0.17	0.12	0.61*				
Vilkman *et al.* 1991	1.35	0.30*	2.18	0.66**	1.71	0.46**	1.52	0.29*

* $p < .05$; ** $p < 0.01$.

not contain enough information to determine the relationship. But the results from the study by Honda (1985) and from Vilkman *et al.* (1991) can be derived from the figures, and values from the study by Honda & Baer (1981) were still available. The calculated slopes are given in Table V. For the Honda subject (in both reports), there is a small difference in slope between the high and low vowels, and the pattern is consistent with the present results. The Vilkman *et al.* subject also exhibited the pattern that was found in the present subjects. Studies by Vilkman *et al.* and Honda & Baer include the

TABLE VI. Average CT activity (in microvolts) for those productions that fell within the range of F0s for isolated vowels

	ɑ	i	u
M1 (left)	5.92 (10)	5.67 (9)	5.86 (7)
M1 (right)	24.55 (10)	25.26 (9)	23.76 (7)
*M2a	26.10 (23)	25.32 (10)	26.15 (10)
M2b	26.23 (24)	25.94 (20)	25.75 (16)
W1a	– (3)	– (0)	– (0)
W1b	20.36 (40)	19.30 (18)	18.96 (34)
*W2	13.66 (39)	12.51 (28)	11.92 (19)

(Number of tokens contributing to the mean is given in parentheses.) Subjects showing a significant difference among the three vowels are indicated with an asterisk. The "a" after a subject's initials indicates that the range from the pre-test was used, while the "b" indicates that a range based on the isolated wovel productions from the EMG experiments was used.

additional low vowel /æ/, which has a slope similar to that of /ɑ/. Thus, with the exception of one subject (W1) who used CT in an atypical way, there is a different relationship between the activation needed to change F0 for high vowels than for low vowels. This relationship makes the overall appearance of a difference in CT activity difficult to interpret as an indication of intention.

The final way of looking at the CT activity for the productions with targets was to analyze only those productions, regardless of the target, that happened to be near each speaker's typical value for the vowel in isolation without targets. This, to some degree, prefigures the no-target condition, to be discussed next. The range of F0 values considered to match was the mean value of the vowel plus or minus one-half of the IF0 difference (i.e., the high vowel means minus the low vowel mean). Thus, for speaker W2, the range for /ɑ/ was 186–198 while the range for /i/ and /u/ was 199–210. This analysis is complicated by the fact that two subjects (M2 and W1) changed their F0 for the isolated vowels between the pre-test and the experiment (as will be seen in Table VII); for these subjects, the ranges were computed both for the original values, which would have been matched if they successfully matched the stimulus tones, and for the isolated values obtained the same day in the other, no-target condition. Subject W1 had virtually no productions in her original range, and so no analysis was possible. As seen in Table VI, the CT activity was not higher for high vowels than the low ones. The two subjects who showed significant differences among the vowels had lower values for /i/ (and, for W2, /u/) than for /ɑ/.

4. Results—no target condition

The mean F0s and EMG activations for the two or three repetitions of the condition of vowel productions without tone targets are presented in Table VII. (This is the condition that was not collected for M1.) Unlike the previous reports mentioned in the introduction, there was no tendency for the overall activation to be higher for the high vowels than for the low vowels. A separate ANOVA was run for each subject's EMG activation levels, with the factors of Vowel (3 levels) and the Block (2 levels for M2 and W2, 3 levels

TABLE VII. F0 values (in Hz) and CT activity (in μV) for the isolated productions of vowels without target tones. Each line for a speaker represents the average over one repetition of the condition. There were two for M2 and W2, three for W1

	a		i		u	
Subject:	F0	CT	F0	CT	F0	CT
M2	109	26.1	124	26.3	128	26.0
	113	26.6	119	26.1	122	25.8
W1	178	16.4	183	15.0	183	16.5
	177	18.9	179	16.2	181	17.6
	178	21.4	183	16.5	181	19.3
W2	193	13.3	201	14.9	194	10.2
	192	13.2	204	15.2	200	11.3

for W1). The results for Vowel were significant for two speakers (for M2: $F(2,84) < 1$, n.s.; W1: $F(2,126) = 21.55$, $p < 0.0001$; W2: $F(2,84) = 79.25$, $p < 0.0001$). The repetitions differed for one subjects (for M2: $F(1,84) = 1.47$, n.s.; W1: $F(2,126) = 23.31$, $p < 0.0001$; W2: $F(1,84) = 2.04$, n.s.). The interaction was significant for one subject (for M2: $F(2,84) = 1.04$, n.s.; W1: $F(4,126) = 2.61$, $p < 0.05$; W2: $F(2,84) = 1.66$, n.s.).

The same factors were used to analyze the F0 values (also in Table VII). The results for Vowel were highly significant for all speakers (for M2: $F(2,84) = 46.57$, $p < 0.0001$; W1: $F(2,126) = 24.97$, $p < 0.0001$; W2: $F(2,84) = 16.50$, $p < 0.0001$). The repetitions differed for one subject (for M2: $F(1,84) = 2.95$, $p < 0.10$; W1: $F(2,126) = 5.84$, $p < 0.01$; W2: $F(1,84) = 3.21$, $p < 0.10$). The interaction was significant for one subject (for M2: $F(2,84) = 6.37$, $p < 0.01$; W1: $F(4,126) = 2.26$, $p < 0.10$; W2: $F(2,84) = 1.66$, $p < 0.10$).

The EMG differences among the three vowels yielded a significant ANOVA for W1 and W2. For W1, the low vowel had the largest activation, despite having the lowest F0. For W2, the two high vowels went in opposite directions and averaged almost exactly the activation of the low vowel (13.23 *vs.* 12.98). These results are at odds with previous results, since only the consistent subject shows the paradoxical lower CT value for the high vowels. These mixed effects can be compared with the results obtained from a sample of the same size for the F0s themselves. Here, all three subjects have a highly significant effect of the IF0 itself. One subject (M2) showed a somewhat reduced IF0 in the second repetition of this no-target condition, but the pattern remained. (W1 had a main effect of repetition, but the overall difference was only 2 Hz; she was very consistent in each repetition.) In this condition, there does not seem to be a direct contribution of CT to the changes made in F0.

The lack of replication of previous results must be viewed in terms of the relative amount of data available in the various studies, and the methods used to analyze the EMG data. With the heavy experimental demands imposed by the EMG techniques, there have been fewer subjects run than would be ideal. Indeed, given the range of behavior available just in our four subjects, it is clear that the contribution of CT to the control of F0 within the speaking range is quite complicated and needs the study both of more subjects and more muscles within subjects. However, since the previous arguments about the deliberateness of IF0 have depended on just the CT muscle, it is worth comparing the previous results with the current ones to see where the differences lie.

The most difficult study to assess is the one by Dyhr (1990), who also had the largest number of subjects (four). He does not report any means of activation nor any statistics of any sort; so it is difficult to evaluate his descriptions of the results. He claims that the high vowels have earlier onsets of activation and higher peaks. All the eight CT plots that are presented in the paper (representing 16 of the 306 vowels analyzed) show the earlier onset of CT activity for the high vowel, but only five show larger peaks. Without a statistical test, it is impossible to know whether this amplitude difference is reliable. The difference in timing is probably due to the greater distance that the root of the tongue must travel for high vowels (with the wide pharynx) compared with the low vowels (with constricted pharynx). However, it is hard to know if these are the correct timing relationships, since the plots of the EMG signals have been shifted in his figures, based on the location of peaks in CT and F0. The alignment of peak CT activation to peak F0 may not be the most appropriate way to align the signals. As for the amplitude difference, it is not clear even if it is present; if it is, it might not remain if an average EMG activation over a fixed window (as in the current experiment) were used rather than peak intensity. Since Dyhr's high vowels had steeper onset slopes, it is quite possible that the difference in activation would disappear across a window rather than looking solely at the peaks. In any case, it may be that there is a difference in the shape of the CT activation rather than a simple linear increase in amplitude. This study, then, must be treated with some caution, and probably should not serve as a foundation for other theories except for further refinements of experimental technique.

Each of the other studies (Autesserre *et al.*, 1987; Honda & Fujimura, 1991; Vilkman *et al.*, 1991) investigated a single speaker. Autesserre *et al.* (1987) had the subject perform F0 "melodies" that ranged from 93 to 377 Hz, a far larger range than that used in speech. Additionally, they, as well as Vilkman *et al.* (1991), analyzed the peak intensity rather than averaging over a window as is done here and in Honda & Fujimura (1991). The case for CT being more active for high vowels than for low thus rests on one or two speakers. Of the current four speakers added to the pool, only one showed this pattern, while the others were negative, neutral or mixed. Such individual variability cannot be elucidated without running more subjects, but it is clear that there is no solid basis for claiming that the CT activation levels are indicative of conscious control of IF0.

5. General discussion

The present results show that activation of the cricothyroid (CT) muscle, the one piece of positive evidence that had been adduced for treating intrinsic F0 (IF0) as a deliberate enhancement of the speech signal, does not in fact support an enhancement account. Although there is an overall higher activation for the higher vowels for one subject in the present experiment and two in other studies in the literature, three subjects in this study had other patterns of activation (neutral, negative, and mixed). Further, the activation level cannot be interpreted directly since the amount of activation needed to effect a change in F0 differs for the different vowels. In the no-target condition, the three subjects providing data had the same range of effects as they did in the target condition (neutral, negative, and mixed). Considering the array of arguments against the deliberateness of IF0, the evidence now seems to support only an automatic mechanism.

The CT activity of both the isolated vowel productions without targets and all three analyses of productions with tonal targets produced evidence that any differences

found across vowels are due to the vowel production and not to planned changes in F0. First, in both the target and no-target conditions, the subjects show every possible pattern (larger EMG for high vowels than for low; neutral; negative; and mixed). Although this contradicts what has been claimed in the literature, there are in fact only two other subjects who performed comparable tasks; they both showed higher EMG for high vowels than for low, but there are still only three subjects with this pattern and three without. It is certainly premature to assert that the evidence suggests planning of IF0. Without such evidence, we should assume that an automatic effect is in place.

Another possibility is that the present tone-matching task did not elicit typical IF0 behavior. Our goal in using this task was to effect shifts of F0 comparable to the shifts inherent in IF0. We achieved this goal, as well as the secondary goal of keeping the speakers out of "singing" mode. However, it may still be that any non-linguistic tone matching task results in an unusual implementation of F0. Given the difficulty of determining just which part of the EMG signal attributes to particular vowels in running speech, it may be that this issue cannot be decided with current technology. The present paradigm seems at least as appropriate as those used by previous researchers, especially in light of the similarities reported in Table V. Further, the pattern of EMG activity for the three subjects in this experiment for whom we have data are the same in both conditions: the speakers were either neutral, negative or mixed in both the target and no-target conditions. It seems, then, that the target task was effective in eliciting genuine IF0 behavior. In as much as these results are typical of IF0, then they do not support an argument for deliberate control of IF0.

The present results fail to replicate the reported effects of previous work, but the literature turns out to be less solid than assumed. The study with the most subjects (Dyhr, 1990) was found to have problems of data description that cast doubts on its reliability, and thus undermine any theoretical claims based on it. Another study used F0 sweeps far larger than those found in speech (Autesserre *et al.*, 1987), and so cannot be easily related to the issue of F0 variation in speech. The other studies of this issue (Honda & Fujimura, 1991; Vilkman *et al.*, 1991) report results for two more subjects, both of whom have higher activation levels for high vowels than for low vowels. More subjects are needed to resolve these individual differences, but it is clear that the case for the deliberateness of IF0 cannot rest on the EMG data.

Another way of analyzing the EMG data supported this interpretation: most of the present subjects showed a lower value for CT with high vowels when F0 is factored out. If F0 were increased by a linear function of CT activity, then there would have been no difference among the vowels. With different CT levels needed for effecting a change in F0, the absolute levels are not comparable across the vowels, making it untenable to posit deliberateness even for the three subjects (in this and other studies) with a positive difference in activation levels.

When we look at the productions that have F0s that form the natural range of variation found in IF0, we find that there is no difference in CT activity. This is true of utterances from the target condition (which did have an overall effect of F0 on CT activity) as well as the utterances without target tones. In this last case, there were highly significant differences in F0 itself, so we could expect the power of the analysis to be similar for the CT activity if it were truly under direct control. It was not, again leading to the conclusion that the overall difference in CT activity found between high and low vowels is not an explanation for IF0.

A final possible piece of evidence in favor of deliberateness would be a language that reinterpreted IF0 as a deliberate feature of the language, giving rise to a set of tones historically related to vowel height. The only such report that we are aware of is found in pedagogical material for Passamaquoddy (Nicholas & Francis, 1988; Nicholas, Francis & Nicholas, 1988). As we report elsewhere, the acoustic measurements of their speakers do not bear out the possible tone system (Whalen, Gick & LeSourd, in press). The size of the F0 difference between the vowels is just what we would expect for IF0. Although IF0 has clear effects on perception (Silverman, 1987; Fowler & Brown, 1997), its magnitude is smaller than typically reported for tone differences. Additionally, IF0 occurs in tone languages as well as those without tone (Whalen & Levitt, 1995), so Passamaquoddy would have to be doubly unusual to be considered to have a tone system. Such a system might arise at any time, of course and so null hypothesis claims should be treated with caution. But the lack of such a language is inconsistent with the enhancement proposal.

IF0, then, does not appear to be a deliberate enhancement of the speech signal. Rather, it seems to be a consequence of successful vowel articulation. This allows us to pose the question of why IF0 exists in a different light: given that it should be possible for speakers to make adjustments in their F0 deliberately to overcome IF0, why don't they? We hope to present evidence from later experiments to answer this question.

This research was supported by NIH grant DC-02717 to Haskins Laboratories. We thank Akira Walter Naito for performing an insertion for the experiment. We thank Anders Löfqvist and Donald S. Hailey for their technical assistance. We thank Caral A. Fowler, Harriet S. Magen, Andrea G. Levitt, Thomas Baer, James R. Sawusch, and an anonymous reviewer for helpful comments.

References

Atkinson, J. E. (1978) Correlation analysis of the physiological features controlling fundamental voice frequency, *Journal of the Acoustical Society of America*, **63**, 211–222.

Autesserre, D., Roubeau, B., 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 français: Approche électromyographique. In *Proceedings of the 11th International Congress of Phonetic Sciences*, Vol. 3, pp. 35–38. Tallinn: Academy of Sciences of the Estonian SSR.

Diehl, R. L. & Kluender, K. R. (1989) On the objects of speech perception, *Ecological Psychology*, **1**, 121–144.

Dyhr, N. (1990) The activity of the cricothyroid muscle and the intrinsic fundamental frequency in Danish vowels, *Phonetica*, **47**, 141–154.

Erickson, D. M., Baer, T. & Harris, K. S. (1983) The role of the strap muscles in pitch lowering. In *Vocal fold physiology: contemporary research and clinical issues* (D. Bless & J. Abbs, editors), pp. 279–285. San Diego: College-Hill Press.

Fahey, R. P. & Diehl, R. L. (1996) The missing fundamental in vowel height perception, *Perception and Psychophysics*, **58**, 725–733.

Fischer-Jørgensen, E. (1990) Intrinsic F0 in tense and lax vowels with special reference to German, *Phonetica*, **47**, 99–140.

Fowler, C. A. & Brown, J. M. (1997) Intrinsic F0 differences in spoken and sung vowels and their perception by listeners, *Perception and Psychophysics*, **59**, 729–738.

Fry, D. B. (1958) Experiments in the perception of stress, *Language and Speech*, **1**, 126–152.

Gandour, J. & Weinberg, B. (1980) On the relationship between vowel height and fundamental frequency: evidence from esophageal speech, *Phonetica*, **37**, 344–354.

Hallé, P. A. (1994) Evidence for tone-specific activity of the sternohyoid muscle in modern standard Chinese, *Language and Speech*, **37**, 103–123.

Hirose, H. (1971) Electromyography of the articulatory muscles: current instrumentation and technique, *Haskins Laboratories Status Report on Speech Research*, **SR26/26**, 73–86.

Hirose, H. & Gay, T. J. (1972) The activity of the intrinsic laryngeal muscles in voicing control: an electromyographic study, *Phonetica*, **25**, 140–164.

Hombert, J. M. (1977) Consonant types, vowel height and tone in Yoruba, *Studies in African Linguistics*, **8**, 173–190.
Honda, K. (1983) Relationship between pitch control and vowel articulation. In *Vocal fold physiology: Contemporary research and clinical issues* (D. Bless & J. Abbs, editors), pp. 127–137. San Diego: College-Hill Press.
Honda, K. (1985) Variability analysis of laryngeal muscle activities. In *Vocal fold physiology: biomechanics, acoustics and phonatory control* (I. R. Titze & R. C. Scherer, editors), pp. 127–137. Denver: Denver Center for the Performing Arts.
Honda, K. & Baer, T. (1981) External frame function, pitch control, and vowel production. In *Transcripts of the 10th Symposium on Care of the Professional Voice* (V. Lawrence, editor), pp. 66–73. New York: The Voice Foundation.
Honda, K. & Fujimura, O. (1991) Intrinsic vowel F0 and phrase-final F0 lowering: phonological *vs.* biological explanations. In *Vocal fold physiology: acoustic, perceptual, and physiological aspects of voice mechanisms* (J. Gauffin & B. Hammarberg, editors), pp. 149–157. San Diego, CA: Singular Publishing Group.
Honda, K., Hirai, H., Masaki, S. & Shimada, Y. (in press) Role of vertical larynx movement and cervical lordosis in F0 control, *Language and Speech*.
Kingston, J. (1992) The phonetics and phonology of perceptually motivated articulatory covariation, *Language and Speech*, **35**, 99–113.
Ladd, D. R. & Silverman, K. E. A. (1984) Vowel intrinsic pitch in connected speech, *Phonetica*, **41**, 31–40.
Löfqvist, A., Baer, T., McGarr, N. S. & Story, R. S. (1989) The cricothyroid muscle in voicing control, *Journal of the Acoustical Society of America*, **85**, 1314–1321.
Löfqvist, A., McGarr, N. S. & Honda, K. (1984) Laryngeal muscles and articulatory control, *Journal of the Acoustical Society of America*, **76**, 951–954.
Meyer, E. A. (1896–1897) Zur Tonbewegung des Vokals im gesprochenen und gesungenen Einzelwort, *Phonetische Studien (Beiblatt zu der Zeitschrift Die Neuren Sprachen)*, **10**, 1–21.
Nicholas, J. A. & Francis, D. A. (1988) *Nihtawi skicinuwatu, Book I: Passamaquoddy/Maliseet*. Guilford, CT: Jeffrey Norton Publisher.
Nicholas, J. A., Francis, D. A. & Nicholas, A. (1988) *Passamaquoddy/Maliseet reference book*. Guilford, CT: Jeffrey Norton Publisher.
Ohala, J. J. & Eukel, B. W. (1987) Explaining the intrinsic pitch of vowels. In *In honor of Ilse Lehiste* (R. Channon & L. Shockey, editors), pp. 207–215. Dordrecht: Foris.
Roubeau, B., Chevrie-Muller, C. & Saint Guily, J. L. (1997) Electromyographic activity of strap and cricothyroid muscles in pitch change, *Acta Otolaryngolical*, **117**, 459–464.
Rubin, P. E. (1995) HADES: a case study of the development of a signal analysis system. In *Applied speech technology* (A. Syrdal, R. Bennett & S. Greenspan, editors), pp. 501–520. Boca Raton, FL: CRC Press.
Sapir, S. (1989) The intrinsic pitch of vowels: Theoretical, physiological and clinical considerations, *Journal of Voice*, **3**, 44–51.
Sapir, S., McClean, M. D. & Luschei, E. S. (1984) Time relations between cricothyroid muscle activity and the voice fundamental frequency (F0) during sinusoidal modulations of F0, *Journal of the Acoustical Society of America*, **75**, 1639–1641.
Silverman, K. E. A. (1987) *The structure and processing of fundamental frequency contours*. Unpublished PhD thesis, University of Cambridge.
StatView (1998) *StatView Reference*. Cary, NC: SAS Institute.
Steele, S. A. (1986) Interaction of vowel F0 and prosody, *Phonetica*, **43**, 92–105.
Sundberg, J. (1987) *The science of the singing voice*. DeKalb, IL: Northern Illinois University Press.
Traunmüller, H. (1981) Perceptual dimension of openness in vowels, *Journal of the Acoustical Society of America*, **69**, 1465–1475.
Traunmüller, H. (1994) Conventional, biological and environmental factors in speech communication: a modulation theory, *Phonetica*, **51**, 170–183.
Vilkman, E., Aaltonen, O., Laine, U. & Raimo, I. (1991) Intrinsic pitch of vowels—a complicated problem with an obvious solution? In *Vocal fold physiology: acoustic, perceptual and physiological aspects of vice mechanisms* (J. Gauffin & B. Hammarberg, editors), pp. 159–166. San Diego, CA: Singular Publishing Group.
Whalen, D. H., Gick, B. & LeSourd, P. S. (in press) Intrinsic F0 in Passamaquoddy vowels. In *Papers from the 30th Algonquina Conference* (D. Pentland, editor), Winnipeg: University of Manitoba.
Whalen, D. H. & Levitt, A. G. (1995) The universality of intrinsic F0 of vowels, *Journal of Phonetics*, **23**, 349–366.
Whalen, D. H., Levitt, A. G., Hsiao, P.-L. & Smorodinsky, I. (1995) Intrinsic F0 of vowels in the babbling of 6-, 9- and 12-month-old French- and English-learning infants, *Journal of the Acoustical Society of America*, **97**, 2533–2539.
Zee, E. (1980) Tone and vowel quality, *Journal of Phonetics*, **8**, 247–258.