Correlation analysis of the physiological factors controlling fundamental voice frequency

James E. Atkinson

New London Laboratory, Naval Underwater Systems Center, New London, Connecticut 06320
(Received 7 June 1977; revised 10 August 1977)

A technique has been developed to obtain a quantitative measure of correlation between electromyographic (EMG) activity of various laryngeal muscles, subglottal air pressure, and the fundamental frequency of vibration of the vocal folds (F_o). Data were collected and analyzed on one subject, a native speaker of American English. The results show that an analysis of this type can provide a useful measure of correlation between the physiological and acoustical events in speech and, furthermore, can yield detailed insights into the organization and nature of the speech production process. In particular, based on these results, a model is suggested of F_o control involving laryngeal state functions that seems to agree with present knowledge of laryngeal control and experimental evidence.

PACS numbers: 43.70.Bk

INTRODUCTION

This paper reports on one aspect of a continuing study to determine the physiological correlates of the changes in fundamental voice frequency (F_0) that carry the prosodic or intonational aspects of language. In particular, a technique is described that provides a quantitative measure of correlation between the physiological factors and the acoustic signal. It will be shown that such an analysis can yield detailed insights into the organization and nature of the speech production process.

Many experimental studies have investigated the physiological factors controlling F_0. These have included electromyographic (EMG) studies of the laryngeal muscles by Faaborg-Andersen (1965), Lieberman *et al.* (1970), and Ohala (1970), as well as aerodynamic studies of subglottal, transglottal, and oral air pressure by Ladefoged (1962), Bouhuys, Proctor, and Mead (1966), and Lieberman (1967). It is known that both laryngeal and respiratory functions have an effect on F_0, but the exact role of each is still not clear. With the exception of some carefully conducted quantitative studies involving the effects of subglottal pressure on F_0 (e.g., Lieberman, 1967; Ohala, 1970) and a recent study by Flanagan *et al.* (1976), which involved digital analysis of laryngeal control functions[1] most studies in this area have been analyzed subjectively as a result of data acquisition and processing limitations. That is, the investigator looks for peaks in the physiological measure that correspond to peaks in the acoustical data, in essence performing a visual correlation. There are two limitations to such an approach: First, no quantitative measure of degree of correlation or relative ordering of the factors is obtained; and, second, different observers may interpret the same results differently. This paper addresses two basic questions: (1) Can a meaningful quantitative measure of correlation between F_0 and the various physiological factors be obtained? (2) Does such a measure shed any new light on the speech production process?

I. EXPERIMENTAL PROCEDURE

A. Speech material and data collection

Twenty repetitions of each of the 12 utterances in Table I were obtained from a single native speaker of American English. These utterances were chosen carefully to include various stress placements in both statement and question frames. (Throughout this article, underscoring indicates emphasis.)

For each of these utterances, data were obtained for seven variables as functions of time. These particular physiological factors were chosen because they have been implicated repeatedly in F_0 control. The variables are

F_0, fundamental frequency;

P_s, subglottal pressure;

SH, averaged EMG data from the sternohyoid muscle;

LCA, averaged EMG data from the lateral cricoarytenoid muscle;

V, averaged EMG data from the vocalis muscle;

TABLE I. List of test utterances.

Bev loves Bob.
Bev loves Bob.
Bev *loves* Bob.
Bev loves *Bob*.
Bev loves Bob?
Bev loves Bob?
Bev *loves* Bob?
Bev loves *Bob*?
He had plans to leave[1]. (Blueprints to drop off.)
He had plans to leave[1]. (To depart.)
He had plans to *leave*.
He had *plans* to leave.

Note: Superscript 1 = main sentence stress.

 0001-4966/78/6301-0211$00.80 ©1978 Acoustical Society of America

CT, averaged EMG data from the cricothyroid muscle; and

ST, averaged EMG data from the sternothyroid muscle.

All these data were obtained at Haskins Laboratories, New Haven, Connecticut, using the Haskins Pitch Extraction (PEX) program (Luketala, 1973). The EMG data were obtained using hooked-wire electrodes (Hirose, 1971) and processed on the Haskins EMG data processing system (Port, 1971). At least 18 tokens of each utterance were obtained and averaged by the processing system. Subglottal pressure P_s was measured directly using a catheter inserted between the cricoid and thyroid cartilages (for details see Atkinson, 1973). Mean P_s was obtained by averaging. Figure 1 presents an example of typical raw data. Figure 2 presents an example of typical averaged data for one utterance. The fundamental frequency contour used represents the average of several repetitions of each utterance.

B. Correlation analysis

All these data were sampled at a 60-Hz rate and stored for processing by a Univac 1108 computer. Correlation coefficients were calculated between all possible pairs of these variables (e.g., F_0–SH, F_0–LCA, ..., ST–P_s) using the Pearson-Product-Moment formula,

$$r_{x,y} = \frac{\sum_{j=1}^{N}(x_j - \overline{X})(y_j - \overline{Y})}{N\sigma_x\sigma_y} , \qquad (1)$$

where $r_{x,y}$ is the correlation coefficient between variables x and y, x_j and y_j are the jth sample of variables x and y, $\overline{X}$ and $\overline{Y}$ are the mean of variables x and y, σ_x and σ_y are the standard deviation of x and y, and N is the number of samples.

The sampled and digitized data were stored in the Univac 1108 computer as a sequence of successive samples representing each variable as a function of time. This is shown schematically in Fig. 3. When the correlation coefficient is calculated, the effective sample size (N) can assume different values, depending on the method of subdividing the data. The subdivisions were suggested by known or suspected physiological, linguistic, or acoustic evidence.

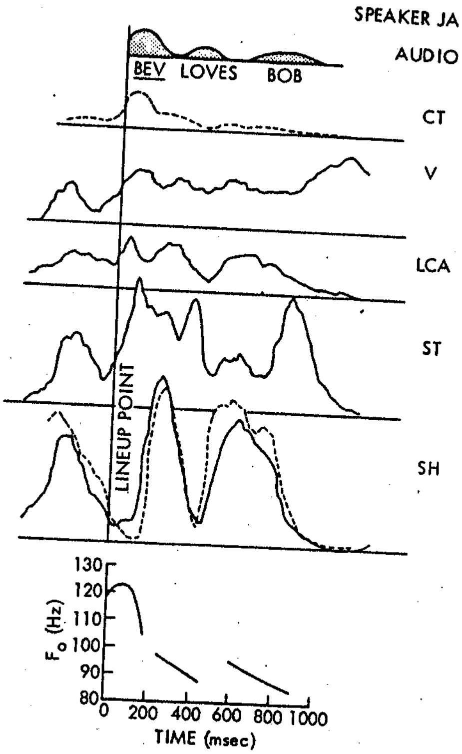

FIG. 2. Typical averaged data from speaker JA for "Bev loves Bob."

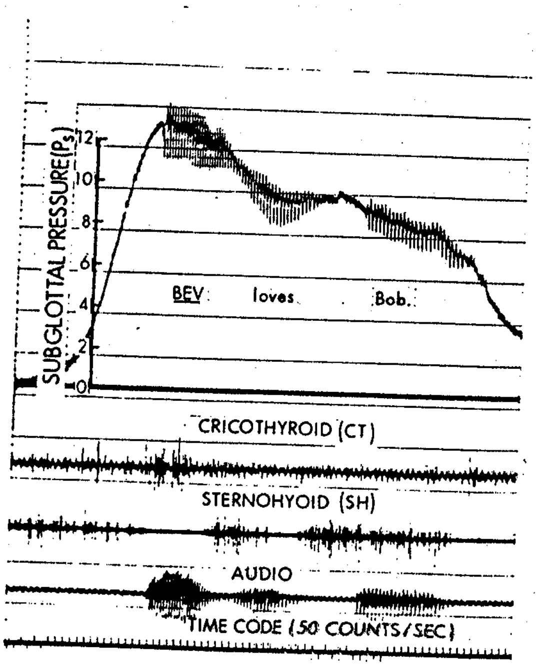

FIG. 1. Typical raw data from speaker JA for "Bev loves Bob."

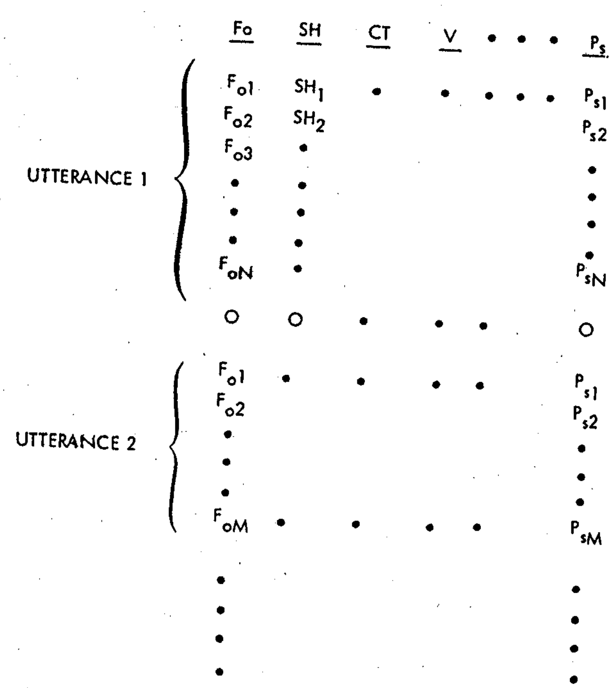

FIG. 3. Data-storage array for correlation analysis.

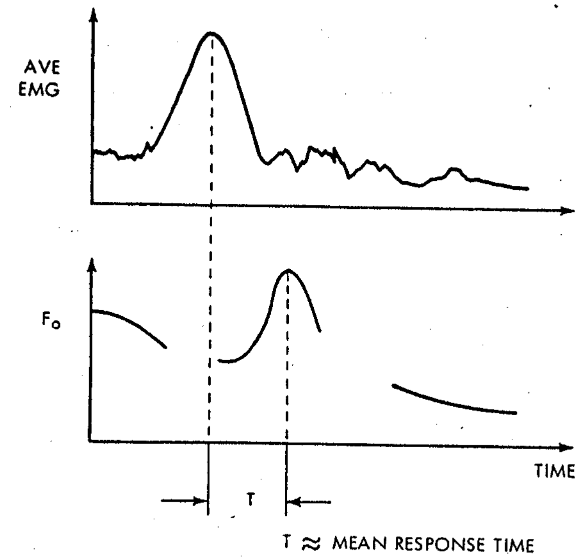

FIG. 4. Schematic representation of mean response time.

There are two points that deserve special mention concerning the correlation analysis. First, it can be seen in Fig. 2 that the physiological factors are all continuous functions of time, beginning before the onset of phonation and continuing after the end of phonation. The F_0 contour, on the other hand, is a discontinuous function; it is nonexistent during both the voiceless intervals and silent periods when there is no periodic vibration of the vocal folds. During these periods, many of the laryngeal muscles are involved with adducting and abducting the vocal folds to assume the proper voice state and bear no direct relation to F_0 *per se*. In view of this, only the voiced portions of the utterances were used in calculating correlations. Second, when the correlation between the EMG data and F_0 is performed, a correction must be made to account for the fact that it takes a finite time for the muscle to contract fully after being innervated. This is because of the electrochemical transfer of energy, the stretching of tendons and other connective tissue, and the inertia of the muscle mass. Therefore, in order to correlate a muscle action with the resultant acoustic effect, the muscle activity must be shifted forward in time relative to the acoustic event. Figure 4 shows this for a hypothetical example.

In general, the amount of shift required to compensate for muscle contraction time depends on the type of muscle and its size (Sawashima, 1973). Although estimates are available for contraction time of various laryngeal muscles from other species (e.g., cat, dog), no actual data exist on human laryngeal muscles. To obtain a measure of this time lag, a discrete form of cross correlation was performed in which the averaged EMG data for each muscle was shifted relative to the F_0 data in incremental steps, and the correlation coefficient was calculated for each successive shift.

TABLE II. Comparison of MRT with contraction time.

Muscle	Atkinson MRT (ms)	Other studies CT (ms)	Other studies Species
V	15	14–21	Cat, Dog
LCA	15	14–16	Dog
CT	40	30–44	Cat, Dog
SH	120	50	Cat
ST	70	No data	
TH	No data	52	Dog

Assuming a causal relationship between F_0 and each of the muscles, we would expect such an analysis to yield correlation coefficients that increased to the point at which the time shift equals the contraction time and decreased with further time shifts. Figure 5 shows that this is precisely what happened. In the figure the different muscles have different response times. The smaller intrinsic muscles [cricothyroid (CT), lateral cricoarytenoid (LCA), and vocalis (V)] are quite fast, whereas the larger extrinsic muscles [sternohyoid (SH) and sternothyroid (ST)] are considerably slower. It should be noted that the accuracy of these values must be limited to ± 10 ms because of the time-shift-increment size used.

We have termed the time lag measured in this way as the mean response time (MRT), which is defined as the amount of time shift to yield maximum correlation between averaged EMG activity and F_0. This is not truly a measure of muscle contraction time as that term is usually defined. Generally, muscle contraction time refers to the impulse response, or twitch contraction time of a single fiber to an impulse excitation (Sawashima, 1973). Nevertheless, good agreement can be seen between the MRT measured in this study and the muscle contraction times measured in other studies and in other species. Table II shows this correspondence. These data on contraction times are cited from Sawashima (1973), as gathered from various sources.

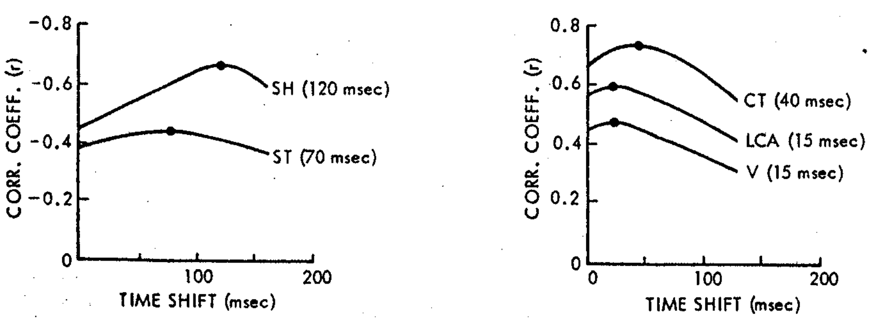

FIG. 5. Results of cross-correlation analysis.

In particular, the intrinsic muscles (V, LCA, and CT) show excellent agreement. Note that these muscles are quite similar as to size and function in cat, dog, and human. On the other hand, the extrinsic muscles show considerable differences in size between these species. The SH, especially for cat, is much smaller than in human and, thus, we would expect it to have a much faster contraction time, as Table II indicates.

The results of this analysis are consistent with what is already known from other studies. This procedure supports the claim that a cross-correlation analysis of this sort does, in fact, offer a method of measuring indirectly contraction times in human muscles. Such a measure is not otherwise available and could be an important tool for future research in this area.

In the following paragraphs, all the EMG data have been shifted by the MRT for each muscle before calculating correlation coefficients between muscle activity and F_0. However, before proceeding to the results, it is important to note one final point. The laryngeal muscles are involved in many articulatory gestures not specifically related to regulating F_0. Among these are abduction and adduction of the vocal folds for the voiced–voiceless distinction and for periods of silence within an utterance. Even within the voiced sounds, there is evidence that various degrees of glottal opening may be involved between different classes of sounds (e.g., see Atkinson, 1973; Lisker *et al.*, 1969; Sawashima, 1973; Hirose and Gay, 1972). All these gestures, as well as general head, neck, and jaw movements involving the extrinsic laryngeal muscles, which are not related to F_0 *per se*, can lower the degrees of correlation. This should be kept in mind in the following discussion when interpreting the correlation coefficients obtained.

II. RESULTS

A fundamental assumption in this study is that F_0 is a dependent variable and all the other factors measured are independent variables that, singly or in combination, are capable of causing changes directly in F_0. There seems to be ample evidence to support this assumption, namely, a direct causal relationship between CT activity and F_0, for example, has been shown in many studies. This includes work with excised larynxes, both human (Van den Berg and Tan, 1959) and animal (Baer, 1975), as well as *in vivo* EMG studies on humans (Faaborg-Andersen, 1965; Lieberman *et al.*, 1970; Ohala, 1970; Collier, 1975).

Similarly, several studies have shown a general correlation between V activity and F_0. The most consistent and convincing studies have involved singing or isolated words (Gay and Harris, 1971; Hirano, Ohala, and Vennard, 1969; Gay *et al.*, 1972). Studies involving speech in longer utterances have shown a somewhat lower correlation and greater variability (Gårding, Fujimara, and Hirose, 1970; Hirose, Simada, and Fujimara, 1970; Simada and Hirose, 1970). Although it is clear that V is capable of causing changes in F_0, exactly how and to what extent remains less clear.

Primarily LCA is an adductor, but some studies have found a relationship between LCA activity and F_0, especially at high F_0 (Van den Berg and Tan, 1959; Hirano, Ohala, and Vennard, 1969; Ohala, 1970; Gay *et al.*, 1972). This has not been as consistent as that of V or CT, yet is clear that the LCA is capable of regulating F_0.

The extrinsic laryngeal muscles (SH and ST in particular) have been found to be involved with lowering F_0. These muscles also show increased activity at both very low and very high F_0 (Faaborg-Andersen, 1965; Ohala, 1970; Ohala and Hirose, 1970; Collier, 1975; Erickson, 1975). The exact nature of the involvement is uncertain owing to the complex interrelationship between these muscles, the intrinsic laryngeal muscles, and the suprahyoid muscles. It appears that extrinsic muscles are capable of causing F_0 changes (Sonninen, 1968; Kotby and Haugen, 1970; Ohala, 1972), but they are also involved with other gestures not directly related to F_0.

Finally, several recent studies have shown, all else being equal, that F_0 is related directly to P_s (Van den Berg, 1960; Ladefoged, 1962; Lieberman, 1967; Ohala, 1970; Atkinson, 1973; Collier, 1975). There is some disagreement as to the sensitivity of F_0 to pressure changes (values from 2 to 20 Hz/cm H_2O have been obtained), but there is no dispute that P_s is capable of causing directly F_0 changes.

These studies support the claim that the various muscles and P_s are capable of causing F_0 changes. Nevertheless, one must be careful in interpreting the results of any correlation analysis. A high correlation coefficient between variables does not necessarily mean that there is a causal relationship between them. They may be only incidentally or stochastically related or there may be a third unsampled variable to account for the observed variations. Properly interpreted, however, a correlation analysis offers strong supporting evidence that can be very useful in quantifying and ranking the variables involved.

A. Averaging over all utterances

The first step in the analysis was to calculate the correlation coefficient between all factors averaged over all 12 utterances. The results given in Table III show many highly significant correlations with F_0. These results are consistent with previous studies. The CT (the muscle that has most consistently shown a direct relation with F_0 control) shows the highest correlation. The SH shows the next highest correlation and the fact that it is negative agrees with the studies that indicate it is in-

TABLE III. Correlation matrix including all utterances.

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.65	0.59	0.45	0.72	−0.19	−0.42	F_0
	1.00	−0.24	−0.24	−0.40	−0.26	0.25	SH
		1.00	0.36	0.85	−0.19	−0.24	LCA
$N=568$			1.00	0.44	0.01	−0.37	V
$\lvert r \rvert > 0.11$ for $P < 0.01$				1.00	−0.08	−0.37	CT
					1.00	0.15	P_s
						1.00	ST

volved with F_0 lowering. The LCA correlation is next highest. The very high correlation between LCA and CT (0.85), however, suggests that this may be due to synergistic activity of those two muscles rather than a direct causal relationship with F_0. A moderate correlation with F_0 and also rather high correlations with both the LCA and CT are shown by V. This suggests a general synergy between all three of these intrinsic muscles and that they may act in combination to control F_0.

The moderately high negative correlation of the ST with F_0 also is to be expected. Like the SH, it has often been suggested as a pitch-lowering mechanism. The low correlation between P_s and F_0 initially was quite surprising. However, this may be accounted for by the fact that this analysis included both statements and questions that have markedly different F_0 contours. The more detailed analyses to follow indicate that in some cases there are very significant correlations between F_0 and P_s.

The relative magnitude and rank order of correlations found in this overall analysis are consistent with expectations based on related studies.

B. Rising versus falling F_0

Many investigators have made a distinction (at least implicitly) between mechanisms for raising and lowering F_0 (e.g., Zemlin, 1968; Collier, 1975). Tensing the intrinsic laryngeal muscles (particularly the CT) generally has been considered the normal pitch-raising mechanism. Several possible mechanisms have been suggested for pitch lowering. These most often involve relaxing the CT (Simada and Hirose, 1970; Lieberman, 1970), tensing the SH (Fromkin and Ohala, 1968; Ohala, 1970), or combinations of these and other factors.

A logical extension of our analysis was to subdivide the data in terms of rising and falling F_0, calculate correlation coefficients for each, and determine if different mechanisms seemed to be involved. To do this, the slope of F_0 with respect to time was first calculated. This variable F_0' was defined as

$$F_0'(i) = F_0(i) - F_0(i-1) \ .$$

Two correlation matrices then were computed. The first included only data points corresponding to rising F_0 (e.g., $F_0' > 0$); the second, only the falling F_0 data points (e.g., $F_0' < 0$). The results are given in Tables IV and V. The factors with the highest correlation, CT

TABLE IV. Correlation matrix for rising F_0 including all utterances.

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.59	0.63	0.43	0.71	−0.23	−0.37	F_0
	1.00	−0.24	−0.19	−0.37	0.07	0.08	SH
		1.00	0.40	0.84	−0.44	−0.32	LCA
			1.00	0.38	−0.18	−0.40	V
				1.00	−0.38	−0.36	CT
					1.00	0.45	P_s
						1.00	ST

$N = 180$

$|r| > 0.19$ for $P < 0.01$

TABLE V. Correlation matrix for falling F_0 including all utterances.

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.65	0.36	0.39	0.68	0.46	−0.37	F_0
	1.00	−0.07	−0.21	−0.39	−0.42	0.23	SH
		1.00	0.27	0.61	−0.02	−0.03	LCA
			1.00	0.54	0.08	−0.29	V
				1.00	0.29	−0.38	CT
					1.00	0.03	P_s
						1.00	ST

$N = 370$

$|r| > 0.14$ for $P < 0.01$

and SH, show no significant differences between rising and falling F_0. This tends to refute any hypothesis that calls for functional specialization of these two muscles into specific pitch-raising and pitch-lowering mechanisms. The high correlations of these muscles suggests they are actively involved in both rising and falling F_0 patterns. The V and ST have somewhat lower correlations and, like the CT and SH, show no significant differences between rising and falling F_0. There is a significant difference between rising and falling F_0, however, for P_s and the LCA. P_s has a significantly higher correlation for falling F_0 (0.46 versus −0.23). This will be discussed more fully later. Conversely, the LCA shows a significantly higher correlation for rising F_0 (0.63 versus 0.36), although this may be a result of the higher synergistic correlation between LCA and CT for rising F_0 (0.84 versus 0.61).

Thus, although there are some differences between rising and falling F_0, the muscles most often implicated by other studies as mechanisms for pitch raising and pitch lowering (CT and SH) show no such differences based on the data of this study.

C. Temporal and spectral partitioning

To this point, the results obtained have included all the test utterances and, thus, represent an average over many different F_0 contours. Since these different F_0 contours may reflect unique physiological control patterns, a more detailed analysis was necessary to refine the results.

There are two convenient and logical dimensions in which the data may be segmented to obtain finer resolutions: the time domain and the frequency domain. The units of division in the time domain correspond to sentences, words, or syllables, whereas in the frequency domain the data are divided into "bins" of F_0.

There is a fundamental difference in the underlying assumption between time and frequency partitioning. Segmenting by time inherently involves categorization into linguistic units (e.g., sentences, words, syllables). This implicitly assumes that the physiological correlates of F_0 are organized or programmed at some higher and, presumably, more abstract linguistic level. Segmenting by frequency involves an acoustical categorization in terms of F_0 intervals. This implicitly assumes that the physiological control of F_0 is organized or programmed at this more concrete level.

216

TABLE VI. Analysis by time partitioning into sentences.

Test utterances	F_0 SH	F_0 LCA	F_0 V	F_0 CT	F_0 P_s	F_0 ST
Bev loves Bob.	−0.83	−0.31	0.09	0.58	0.97	−0.29
Bev loves Bob.	−0.48	0.51	0.69	0.93	0.95	0.16
Bev *loves* Bob.	−0.81	−0.37	0.45	0.72	0.89	0.10
Bev loves *Bob*.	−0.63	0.09	0.43	0.34	0.74	−0.09
Bev loves Bob?	−0.70	0.54	0.42	0.65	−0.35	0.24
Bev loves Bob?	−0.31	0.70	0.72	0.65	−0.76	−0.77
Bev *loves* Bob?	−0.89	0.73	0.73	0.74	−0.24	−0.81
Bev loves *Bob*?	−0.75	0.84	0.59	0.94	−0.52	−0.13

Since there is no clear justification *a priori* for choosing one approach over the other, both time and frequency partitioning were investigated. We shall discuss more fully the notion of linguistic versus acoustical programming after those results are presented.

Regardless of whether the analysis involves time or frequency partitioning, the proper choice of partition size can be crucial and often is difficult to determine. Essentially we are filtering the data into several contiguous bins of time or frequency. Such an operation is optimized (i.e., the signal-to-noise ratio is maximized) if the filter width is matched to the width of the signals to be analyzed. The ideal filter, of course, can be determined only if the signal is known exactly. In this study the signals to be analyzed are the prosodic features of speech and, in particular, the changes in F_0 (in both time and frequency) that signify these features. The temporal and spectral characteristics of these features are not known exactly, but sufficient data exist to allow reasonable approximations to the ideal bin size to be determined. As we shall see, statistical sampling limits ultimately impose a far more stringent restriction on the choice of bin size (particularly in the time domain).

D. Linguistic analysis (time partitioning)

The logical (and, in fact, the only feasible) method of segmenting in time involves breaking the data into linguistic units for finer analysis. Thus, the first analysis might partition the data into individual sentences. In terms of the matched-filter concept, this would correspond to matching the time bin to a phonetic feature like [± breath-group] (Lieberman, 1967, 1970) that encompasses an entire phrase or sentence. The next (finer) analysis might subdivide each sentence into individual words or even syllables. This corresponds to matching the time bin to a phonetic feature like [± prominence] (Lieberman, 1967, 1970) whose domain is the syllable. An analysis using this resolution could be extremely interesting. Unfortunately, in analyzing each syllable individually the sample size within each time bin is reduced drastically, causing the variability to increase. It, thereby, becomes impossible to interpret with any confidence the significance of a particular correlation coefficient. For this reason the finest time partitioning possible in this study involved entire sentences.

For this analysis, only the "Bev loves Bob" utterances from Table I were used. In this way, there was an equal number of statements and questions, and all the utterances were composed of the same segmental elements so that the only difference was in intonation, which is the factor of immediate interest. Furthermore, all the segmental phonemes in these utterances are voiced, thus minimizing the perturbations in F_0 caused by going from the voiced to voiceless, or silent, state and vice versa.

Correlation matrices were computed for each of these utterances and the results are tabulated in the Appendix, Tables A-I–A-VIII. These matrices encompass a great deal of information. For convenience, the most pertinent correlations (those involving F_0) have been summarized in Table VI.

Looking down each column in Table VI, we see considerable variability from sentence to sentence. This suggests that all the factors interact in a complicated fashion to produce the desired F_0 pattern, but the exact nature of the interaction is not clear from these data. When these results are divided into two sentence types (statements and questions), however, a pattern begins to emerge. From the data in Table VI, the mean correlation coefficients for the four statements and the four questions were calculated. These are presented in Table VII. There are only two variables that show striking differences between statements and questions. The LCA shows a very high correlation with F_0 in questions but not in statements. P_s shows just the reverse, a very high correlation with F_0 in statements but not in questions.[2] As might be expected, these results are very similar to those of the preceding section where the data were subdivided into rising versus falling F_0. Recalling that the questions all involve a sharply rising F_0 contour and the statements a generally falling F_0 contour, we see that the same two variables (LCA and P_s) were the only ones to show significant differences in each analysis, and the results of the two analyses are consistent with each other.

Apparently, there are significant differences between the statement and question forms. The most striking difference is seen in the correlation of F_0 with P_s. In questions, there is no significant correlation between F_0 and P_s, and it is clear that the intrinsic laryngeal muscles (CT, LCA, and V) are controlling F_0. However, in statements, P_s has the highest correlation with F_0, while CT and SH show lesser but still significant convolutions.

It is quite clear from this that under different conditions different physiological factors may play the dominant role in determining F_0, presumably depending on the nature of the F_0 change involved.

TABLE VII. Correlation by sentence type (statements versus questions).

Parameters	Statements	Questions
F_0 SH	−0.69	−0.66
F_0 LCA	−0.02	0.70
F_0 V	0.42	0.62
F_0 CT	0.64	0.74
F_0 P_s	0.89	−0.46
F_0 ST	−0.03	−0.37

TABLE VIII. Analysis by frequency partitioning.

Parameters	Fundamental frequency (F_0) bins (Hz) 80–100	100–120	120–140	140–160
F_0 SH	0.19	−0.60	−0.18	−0.10
F_0 LCA	0.30	−0.09	0.49	0.41
F_0 V	−0.10	0.22	0.29	0.22
F_0 CT	0.15	0.23	0.51	0.71
F_0 P_s	0.59	0.22	−0.53	−0.51
F_0 ST	−0.13	−0.02	−0.30	0.33
$\|r\|$ for $P<0.01$	0.23	0.15	0.35	0.30
	Low F_0	Mid F_0	High F_0	

In statements in which there is normally a level or slowly falling F_0 contour, it is, in a sense, easier simply to let F_0 follow the normal P_s pressure contour. However, statement contours also can be modified by moderate F_0 rises (as on syllables with emphasis) and by sharp falls immediately following emphatic syllables. These changes must involve laryngeal adjustments. The single most important factor in statements, however, seems to be P_s. In questions, on the other hand, sharp rises in F_0 are required that can be caused only by the tensor and adductor muscles of the larynx (V, CT, and LCA). For these utterances, P_s plays a minor role in controlling F_0 and the laryngeal muscles predominate.

E. Acoustic analysis (frequency partitioning)

The frequency partitioning analysis subdivides the data in terms of the frequency rather than the time domain. That is, bins of F_0 are established and correlation coefficients are calculated within each frequency bin. The choice of bin size represented a compromise between the desire for fine resolution and the sample size required in each bin to yield statistically valid results. A value of 20 Hz was chosen as the smallest bin size consistent with these bounds. In terms of the matched-filter concept, this also represents a reasonable approximation to the frequency excursions required by this speaker to signal the prosodic features (Atkinson, 1973, 1976).

The data from all the utterances were divided into four frequency bins, each 20 Hz wide, and the various correlation coefficients were calculated within each bin. The results are presented in Table VIII.

Scanning any one of the frequency bins reveals significant differences between the physiological variables, depending on the particular F_0 bin. Similarly, scanning across the frequency regions for any one of the physiological factors reveals that the degree of correlation with F_0 differs significantly from bin to bin. Different combinations of physiological factors appear to be involved for different ranges of F_0. For low F_0 (80–100 Hz), by far the highest correlation with F_0 is obtained for P_s. In the midrange (100–120 Hz), the SH shows a much higher correlation with F_0 than any other factor. At high F_0 (above 120 Hz), there is little difference between the two F_0 bins.[3] CT and LCA both show high correlations in this region.

These interesting results suggest that different laryngeal modes (i.e., different physiological control mechanisms and/or different modes of vibration of the vocal folds) may be involved in different F_0 regimes. We shall return to this point later.

The next logical extension of this analysis by frequency partitioning would be to look within each F_0 bin to see if there is further differentiation in terms of rising versus falling F_0. However, this further partitioning reduces the sample size in some bins to the point where meaningful correlation coefficients cannot be calculated. In fact, the only bin having sufficient data points for both rises and falls is the midrange (100–120 Hz). Correlation coefficients for this range are presented in Table IX, which shows no major qualitative differences between rising and falling F_0. The SH has the highest correlation with F_0 for both cases. It does not appear, then, that there are specific mechanisms exclusively for pitch raising or pitch lowering, at least in this F_0 range.

The implications of the results from this and the preceding sections will be discussed in the next section, and an attempt will be made to unify them into a cohesive hypothesis concerning the nature of F_0 control.

III. DISCUSSION

Before we discuss the theoretical implications of this study, a summary of the major results will be given. For this purpose it is assumed that the degree of correlation between a laryngeal muscle and F_0 is, in some way, a measure of the relative importance of that muscle in controlling F_0. This will allow us to rank order the various physiological factors. Using this metric, we can summarize the role of each factor in rank order as follows:

(1) Cricothyroid Muscle (CT). The CT shows the highest and most consistent overall correlation with F_0 (Table III). The correlation coefficient (r) is typically around 0.7 and is positive. The CT seems to be equally important in both raising and lowering F_0 (Tables IV and V) and shows no significant difference between statements and questions (Table VII). However, it does behave differently in different F_0 ranges (Table VIII). At high F_0 (120–160 Hz for this speaker), it is the single most important factor; while in the low and mid F_0 ranges, it plays a much lesser role, and other factors predominate.

TABLE IX. Correlation coefficients for rising and falling F_0 within the mid F_0 range (100–120 Hz).

Parameters	Rising F_0	Falling F_0
F_0 SH	−0.78	−0.49
F_0 LCA	0.00	−0.24
F_0 V	−0.01	0.34
F_0 CT	0.24	0.20
F_0 P_s	0.25	0.23
F_0 ST	0.28	−0.13
$\|r\|$ for $P<0.01$	0.28	0.18

(2) Sternohyoid Muscle (SH). The SH shows a high, consistent, negative correlation with F_0 (r is typically around −0.6). No significant difference is seen between rising and falling F_0 (Tables IV and V) or between statements and questions (Table VIII). Like the CT, the SH shows significant differences as a function of F_0 range (Table VIII). In the mid F_0 range (100–120 Hz for this speaker), it is the single most important factor in F_0 control; while in the other F_0 ranges, it has a low correlation with F_0. This does not mean that it plays no role in those ranges, only that it does not seem to regulate F_0 directly.

(3) Lateral Cricoarytenoid Muscle (LCA). In not one of the analyses is the LCA the dominant factor. However, in general, LCA shows a high overall correlation ($r \simeq 0.6$). It also shows significant differences between rising and falling F_0 and between statements and questions because the LCA is involved quite heavily with rising F_0 and with questions and very little with falling F_0 and statements (Tables IV, V, and VII). As a function of F_0 range, it is moderately involved at high F_0 (120–160 Hz), somewhat less at low F_0 (80–100 Hz), and not at all for mid F_0 (100–120 Hz). However, in all cases where LCA correlates highly with F_0, it also correlates highly with CT. This suggests that in these cases the LCA may actually be compensating for CT effect, perhaps to maintain medial compression, rather than in controlling F_0 *per se*. This interpretation requires further investigation but appears to be consistent with other findings.

(4) Vocalis Muscle (V). The V has a moderate correlation with F_0 overall ($r \simeq 0.4$), shows no significant difference between rising and falling F_0, or between statements and questions, and slight differences as a function of F_0 range. Aside from some synergistic activity with the LCA and CT, it appears to play a minor role in controlling F_0 in speech.

(5) Sternothyroid Muscle (ST). Similarly, the ST shows a moderate correlation with F_0 ($r \simeq -0.4$), is never the dominant factor under any analysis condition, and appears to play a minor role in F_0 control.

(6) Subglottal Pressure (P_s). P_s has the most varied relationship with F_0. Overall P_s shows the lowest correlation with F_0 and would seem to be the least important factor. P_s shows a significant difference between rising and falling F_0, being moderately high for falling F_0 and very low for rising F_0, and a dramatic difference between statements and questions. In statements, P_s is easily the dominant factor and, in fact, shows the highest correlation obtained under any analysis condition; in questions it plays no apparent role. In terms of F_0 ranges, P_s is the highest correlate in the low F_0 range (80–100 Hz) and elsewhere is insignificant.

With regard to this summary, let us reiterate that a high correlation does not necessarily prove a causal relationship between two variables. However, in cases where a high correlation exists and where it is known from independent studies that the particular factor is capable of causing the observed changes in F_0, it seems reasonable to assume (as we have in this discussion) that this reflects more than a coincidental unrelated occurrence.

There are two particularly interesting aspects to these results that will be discussed. The first is the fact that different physiological factors seem to be in control in different F_0 ranges. This leads to the formulation of a model of F_0 control involving "laryngeal state functions." The second is the fact that when the data are split in terms of the acoustic variable F_0 or its derivative, lower correlations are obtained than when the data are split by linguistic category (statement versus question). This leads to a consideration of the notion of linguistic versus acoustical organization in programming the F_0 contour.

A. Laryngeal state functions in the control of F_0

It seems clear from Table VIII that different physiological factors serve to control F_0 over different F_0 regions. In each of the F_0 ranges, there is a different factor that dominates by a clear margin. At low F_0 (below 100 Hz), P_s is the major factor. At mid F_0 (100–120 Hz), it is the SH; and at high F_0 (above 120 Hz), the CT, aided by the LCA, predominated. This suggests that there may be distinct regions of F_0 within the normal speech range that correspond to different laryngeal "states" (i.e., different modes of vibration of the vocal folds and/or different physiological control mechanisms). Within each of these states a particular set of physiological factors predominates.

These controlling state functions clearly are different in the different regions. This reflects the fact that the vocal folds must undergo dramatic changes in length and tension to achieve the range of F_0 required in normal speech. Thus, to produce low F_0, the vocal folds must be short, thick, and relatively slack. On the other hand, to produce high F_0, the vocal folds must be long, thin, and taut. To achieve this wide range of conditions requires the coordinated action of the extrinsic, as well as the intrinsic, laryngeal muscles. The complex interconnected set of muscles that can, directly or indirectly, affect the vocal folds must act as a system. The intrinsic muscles, of course, have the most direct effect on vocal fold tension. However, they, in turn, are affected and constrained by the extrinsic muscles, particularly at the extremes of this range. The results of this study and the data from other studies, when interpreted in this light, suggest that the extrinsic muscles play a crucial role in determining the gross laryngeal configuration within which the intrinsics can function. This concept is at the heart of the "external frame function" (Sonninen, 1968).

The only consistent result from the various studies on the SH (and for the most part the other strap muscles) has been the finding that there is high activity at low F_0 and little or no activity at high F_0. This refers only to the normal speech range, not to very high F_0 (as in singing) where other adjustments may be required. It is particularly interesting in this study that the SH shows a high correlation with F_0 only in the mid F_0 range (100–120 Hz). At low F_0, it is very active but there is no correlation with F_0 *per se*. At high F_0, it is inactive and there again is no correlation with F_0. Thus, it does not control F_0 directly (except possibly in the mid F_0 range, which we shall discuss in more detail later), but it does seem to be setting up the gross configuration, the laryngeal state required to allow F_0 to be controlled within each of these ranges.

At low F_0, then, the SH (and possibly other strap muscles) is active and generally remains active for as long as low F_0 is required. The exact mechanism is not yet clear, but several studies have suggested that this may lower the larynx as a whole and may affect the horizontal and/or vertical dimensions of the vocal folds as well as their tension. The result is short, thick, and rather slack vocal folds. Given this configuration, the tensor muscles can cause changes in F_0 by changing the length and tension of the vocal folds, but always within the constraints imposed by the extrinsics. It also happens, however, that given this configuration, the sensitivity to P_s is maximized owing to the aerodynamic forces acting when the vocal folds are thick and slack. Van den Berg and Tan (1959) in their study on excised human larynxes, indicate that in the low pitches of the chest voice the Bernoulli forces (and P_s) predominate. Similarly, Stevens (1963) in discussing laryngeal states, suggests that in the low-frequency region the vocal folds are slack and F_0 is more sensitive to P_s than to vocal fold tension. The data of the present study agree very nicely with both of these statements. Table VIII shows that P_s at low F_0 is the major factor in controlling F_0 by a clear margin. In statements involving generally low and falling F_0, Table VII shows that P_s again is the major factor.

On the other hand, in order to achieve high F_0, it appears that the SH must be inactive in order to allow the tensor muscles to make the vocal folds long, thin, and taut. It also happens that, given this laryngeal configuration, the sensitivity to aerodynamic effects (P_s) is minimized. This was seen in the present study by the fact that at high F_0 only the CT and LCA showed significant correlations with F_0. Van den Berg and Tan (1959) also support this result. They indicate that at high F_0 the deformation forces (caused by the V, LCA, and CT) acting on the vocal folds and adjacent tissue predominate in the control of F_0. Likewise, Stevens (1963) suggests that in this region F_0 is controlled primarily by vocal fold tension and is insensitive to P_s changes.

It is not clear from these data whether the mid F_0 range (100–120 Hz in this speaker) should be treated as a separate state or merely as a transition region between the high F_0 and low F_0 states. Arguments can be made for either approach. On the one hand, it has been suggested (Atkinson, 1973, 1973) that in English there may be only two states and that the midrange is simply a transitional region, with the controlling mechanism being the SH muscle. In this model, the SH is seen as a binary state function (tense or lax) that sets the general configuration of the larynx for either low F_0 or high F_0. Thus, when tense, SH does not control F_0 directly (and hence has a low correlation with F_0), but it adjusts the larynx to allow the low F_0 state. When lax, SH again

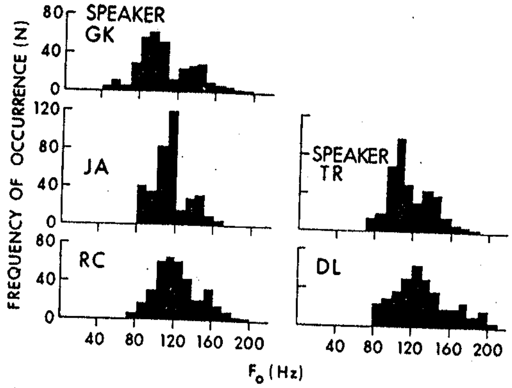

FIG. 6. Histogram of F_0 distributions for various speakers.

shows no direct correlation with F_0, but sets the larynx to allow the high F_0 state. The strong negative correlation in the midrange indicates just such a state change or transition. When going from the low F_0 to the high F_0 state, the SH is becoming lax and F_0 is rising rapidly as the state changes. When going from the high F_0 to the low F_0 state, the SH is becoming tense and F_0 is falling. Thus, the strongest correlation is seen just in these transitions between laryngeal states. The data of this study and that of Van den Berg (1960) seem to support a two-state model of this type.

On the other hand, data from tone languages, like Thai (Abramson, 1962 and Erickson, 1975), which employs three distinct F_0 ranges, might lead one to argue for a three-state system. At this point we favor the two-state model, primarily because of the Van den Berg (1960) data, but feel that we simply do not have enough data to decide conclusively.[4]

Clearly, the notion of laryngeal state functions suggested here is still in its formulative stages and much research remains to be done. However, there are several points that seem to support such a model. First, data from another study involving several speakers (Atkinson, 1976) suggest that the occurrence of laryngeal states within the normal voice range is a general phenomenon and not idiosyncratic of the present subject. Figure 6 presents histograms of F_0 distribution for several speakers, including the subject in the present study (JA). Notice that every speaker has a bimodal or even trimodal distribution involving fairly distinct regions of F_0. This is consistent with the notion of different laryngeal modes (i.e., different modes of vibration of the vocal folds and/or different physiological control mechanisms).

Perhaps the strongest supporting evidence comes from the careful experiments on excised human larynxes by Van den Berg (1960). In that study, two distinct modes of operation were found involving different physiological control functions within what is generally considered the normal chest register. These were called the chest and midregisters. The physiological factors involved and the laryngeal characteristics seen in these registers are in excellent agreement with the present findings.

In fact, on a larger scale, the whole notion of voice registers (e.g., fry, chest, falsetto) supports the pres-

TABLE X. Laryngeal state function characteristics.

State	Vocal folds	Strap muscles	Primary F_0 control
Low F_0	Short, thick, slack	Tense	P_s
High F_0	Long, thin, taut	Lax	CT

ent model. It is well known and has been verified experimentally that each of these different registers corresponds to different F_0 regions, is characterized by a different laryngeal configuration, and is controlled by different physiological factors. We are suggesting that the same thing happens on a more limited scale within at least one of these registers (chest). Within the normal register there are different laryngeal states corresponding to different F_0 regions. Each state is characterized by a different laryngeal configuration and is controlled by different physiological factors. To avoid confusion with the term "register," we have chosen to call these laryngeal states and refer to the controlling factors as state functions. Table X summarizes the major characteristics of this model of F_0 control.

What we have proposed is based on limited data from a single speaker. There surely are individual differences in terms of absolute F_0 ranges, particular muscles employed, and amount of overlap between states, just as there are such individual differences regarding the major voice registers. We have implicated the SH rather heavily, based on this particular speaker. This is not to say that all speakers would use the SH. The ST or other extrinsic muscles may well be implicated in other speakers. We also do not mean to imply that there are sharp absolute F_0 values delineating the states. As with the major voice registers, we would expect rather loosely defined boundaries between states and some overlap in terms of F_0 range.

Finally, when we indicate a specific factor as controlling one state, this does not mean that other factors cannot control F_0 in that range. For example, P_s is the major factor in low F_0, but surely the CT and other factors can and do affect F_0. The single highest correlate, however, is P_s. This presumably reflects the fact that at low F_0 (or in statements) this is, in some sense, the simplest, easiest way to produce the required F_0 contour. This, of course, holds true for the factors involved in the other F_0 states as well.

B. Linguistic versus acoustic analysis

It is interesting to compare the results of the acoustical analysis in Tables IV, V, and VIII, in which the data were parsed in terms of F_0 or its derivative, with the results in Table VII, in which the data were parsed in terms of the linguistic categories, statement or question. Table XI consolidates these data, presenting the highest correlation between each physiological factor and F_0 under the two criteria (acoustical and linguistic categorization) and indicates under what condition that value was obtained. In every instance, the highest correlation was obtained when the data were separated in terms of the linguistic variable (statement versus question).

Although we cannot explain this conclusively, it seems to suggest that the F_0 contour for an utterance is preprogrammed as a holistic event in terms of a larger linguistic unit, such as the breath group. It is not produced, nor (presumably) can it be perceived, as a sequence of units (of word or syllable length) to be processed in a left-to-right sequence like so many beads on a string. That is, the intonational features of speech, like the segmental, are heavily encoded at a fairly abstract linguistic level and are carried on the entire intonation pattern. This idea has been discussed by Lieberman (1967), based on other evidence, in terms of the ±breath-group feature and is mentioned here only as another interesting insight suggested by a correlation analysis.

IV. CONCLUSIONS

In this paper a technique has been presented that allows a quantitative measure of the relationship between F_0 and the various physiological factors that can control F_0. Although the results are based on fairly limited data, the insights gained do emphasize the potential importance and range of implications that may result from the use of such a technique.

It is obvious from the results presented that there are many ways of controlling F_0 and that many different muscles (both laryngeal and respiratory) may be involved at any given instant. This is strong evidence against any one-to-one mapping between phonetic features, articulatory implementation, and acoustic signal. The correlations obtained indicate only a certain propensity toward one particular mechanism, which in some sense is simplest and, thus, used most often to achieve the desired result.

The major results of this study are as follows:

(1) A correlation analysis can provide a quantitative measure of the relationship between F_0 and the various physiological factors that may control F_0. The results agree with the general findings from other studies and seem plausible in terms of our knowledge of laryngeal mechanisms. That is, no obviously erroneous results were obtained.

(2) The technique employed produced an indirect measurement of contraction time for the human laryngeal muscles that agrees well with contraction times of the same muscles in other species.

TABLE XI. Comparison of acoustic versus linguistic partitioning.

Muscle	Ling. (Sent)	Acoustic F_0 bins	Acoustic F_0 slope
SH	−0.69 (statements)	−0.60 (mid F_0)	−0.65 (fall)
LCA	0.70 (questions)	0.49 (hi F_0)	0.63 (rise)
V	0.62 (questions)	0.29 (hi F_0)	0.43 (rise)
CT	0.74 (questions)	0.71 (hi F_0)	0.71 (rise)
P_s	0.89 (statements)	0.59 (low F_0)	0.46 (fall)
ST	−0.37 (questions)	−0.33 (hi F_0)	−0.37 (rise)

(3) The results indicate that P_s seems to be the dominant factor in controlling F_0 at low values of F_0 and in statements. Laryngeal tension, on the other hand, predominates at high F_0 and in questions involving the marked breath group.

(4) The findings in this study led to speculation about a possible model of F_0 control in terms of laryngeal state functions corresponding to different ranges of F_0, analogous to the differences between voice registers (like falsetto, chest), but occurring as substates within what is considered as the chest register. Basically, these laryngeal state functions would adjust the larynx to achieve the desired F_0 range and, depending on the resulting laryngeal configuration, would tend to be most sensitive to particular physiological factors, namely, P_s at low F_0 and laryngeal tensors at high F_0. The strap muscles (in particular the SH in this speaker) seem to play an important role in producing the different laryngeal states.

ACKNOWLEDGMENTS

The author gratefully acknowledges the guidance of Philip Lieberman, the helpful comments of Donna Erickson, Tom Baer, and Tom Gay at the Haskins Laboratories, and David Kennedy at the Naval Underwater Systems Center, as well as the two anonymous technical reviewers, and the financial support of the Long-Term Training Program of the Naval Underwater Systems Center. This paper is a revised section of a doctoral dissertation in linguistics at the University of Connecticut and an expanded version of an oral paper presented to the 85th Meeting of the Acoustical Society of America in Boston, Massachusetts, April 1973. All the correlation coefficients have been recalculated for this paper using an improved algorithm. This has resulted in some quantitative differences between this study and the basic dissertation. However, the two studies are in qualitative agreement.

APPENDIX

TABLE A-I. Correlation matrix for "Bev loves Bob."

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.83	−0.31	0.09	0.58	0.97	−0.29	F_0
	1.00	0.62	−0.30	−0.17	−0.87	0.45	SH
		1.00	0.10	0.03	−0.46	−0.04	LCA
			1.00	−0.24	0.09	0.05	V
				1.00	0.56	0.19	CT
					1.00	−0.29	P_s
						1.00	ST

$N = 35-40$
$|r| > 0.40$ for $P < 0.01$

TABLE A-II. Correlation matrix for "Bev loves Bob."

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.48	0.51	0.69	0.92	0.95	0.16	F_0
	1.00	−0.14	−0.61	−0.58	−0.59	−0.64	SH
		1.00	0.55	0.57	0.52	−0.33	LCA
			1.00	0.86	0.63	0.32	V
				1.00	0.87	0.29	CT
					1.00	0.25	P_s
						1.00	ST

$N = 37-43$
$|r| > 0.39$ for $P < 0.01$

TABLE A-III. Correlation matrix for "Bev loves Bob."

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.81	−0.37	0.45	0.72	0.89	0.10	F_0
	1.00	0.69	−0.40	−0.55	−0.74	−0.12	SH
		1.00	−0.25	−0.17	−0.15	−0.05	LCA
			1.00	0.84	0.27	−0.24	V
				1.00	0.55	−0.22	CT
					1.00	−0.03	P_s
						1.00	ST

$N = 38-45$
$|r| > 0.39$ for $P < 0.01$

TABLE A-IV. Correlation matrix for "Bev loves Bob."

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.63	0.09	0.43	0.34	0.74	−0.09	F_0
	1.00	0.07	−0.32	0.05	−0.75	0.06	SH
		1.00	0.30	0.58	−0.44	−0.33	LCA
			1.00	0.06	0.22	0.07	V
				1.00	0.03	−0.63	CT
					1.00	−0.11	P_s
						1.00	ST

$N = 38-45$
$|r| > 0.39$ for $P < 0.01$

TABLE A-V. Correlation matrix for "Bev loves Bob?"

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.70	0.54	0.42	0.65	−0.35	0.24	F_0
	1.00	−0.56	−0.64	−0.60	0.39	−0.19	SH
		1.00	0.84	0.98	−0.87	−0.46	LCA
			1.00	0.77	−0.80	−0.38	V
				1.00	−0.86	−0.39	CT
					1.00	0.61	P_s
						1.00	ST

$N = 39-46$
$|r| > 0.38$ for $P < 0.01$

TABLE A-VI. Correlation matrix for "Bev loves Bob?"

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.31	0.70	0.72	0.65	−0.76	−0.77	F_0
	1.00	−0.15	−0.26	−0.34	0.33	0.51	SH
		1.00	0.37	0.79	−0.81	−0.69	LCA
			1.00	0.46	−0.73	−0.38	V
				1.00	−0.71	−0.61	CT
					1.00	0.62	P_s
						1.00	ST

$N = 44-51$
$|r| > 0.36$ for $P < 0.01$

TABLE A-VII. Correlation matrix for "Bev loves Bob?"

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.89	0.73	0.73	0.74	−0.24	−0.81	F_0
	1.00	−0.58	−0.58	−0.70	0.03	0.47	SH
		1.00	0.97	0.67	0.12	−0.71	LCA
			1.00	0.66	0.15	−0.70	V
				1.00	−0.16	−0.61	CT
					1.00	0.08	P_s
						1.00	ST

$N = 41-48$
$|r| > 0.37$ for $P < 0.01$

TABLE A-VIII. Correlation matrix for "Bev loves Bob?"

F_0	SH	LCA	V	CT	P_s	ST	
1.00	−0.75	0.84	0.59	0.94	−0.52	−0.13	F_0
	1.00	−0.53	−0.52	−0.59	0.18	−0.20	SH
		1.00	0.87	0.98	−0.77	−0.50	LCA
			1.00	0.78	−0.70	−0.38	V
				1.00	−0.70	−0.43	CT
					1.00	0.68	P_s
						1.00	ST

$N = 40-47$
$|r| > 0.37$ for $P < 0.01$

[1]The excellent paper by Flanagan *et al.* (1976) is in many ways similar to this current study on quantitative measures of laryngeal control. It differs in the muscles studied, however, since they were more interested in the voiced–voiceless distinction and its application in speech synthesis, whereas this study is most interested in F_0 control.

[2]This is an excellent example of why care must be taken in interpreting the results of any correlation analysis. In this case the correlation between F_0 and P_s in questions (−0.46) is certainly significantly large in the statistical sense. If we assumed a causal relationship, however, we would reach the absurd conclusion that, all else being equal, increasing P_s causes F_0 to decrease and decreasing P_s causes F_0 to increase. Clearly this is impossible in the light of all known physiological evidence.

[3]The only exception is the ST, which shows a negative correlation between 120 and 140 Hz and positive between 140 and 160 Hz. It does not seem likely that this could have any major effect on the conclusions and no attempt will be made here to explain this result. The high negative correlation with P_s in this region also is discounted.

[4]Along these lines, however, we are continuing research, using both Thai and English speakers, in an effort to resolve the functions of the strap muscles in particular and the control of F_0 in general (e.g., Erickson and Atkinson, 1975).

Abramson, A. S. (1962). *The Vowels and Tones of Standard Thai: Acoustical Measurements and Experiments* (Indiana University Research Center in Anthropology, Folklore, and Linguistics 20, Bloomington, IN).

Atkinson, J. E. (1973). "Aspects of Intonation in Speech: Implications from an Experimental Study of Fundamental Voice Frequency," Ph.D. dissertation (University of Connecticut, Storrs, CT) (unpublished).

Atkinson, J. E. (1973). "Physiological factors controlling F_0: Results of a correlation analysis," J. Acoust. Soc. Am. 54, 319(A).

Atkinson, J. E. (1976). "Inter- and intra-speaker variability in fundamental voice frequency," J. Acoust. Soc. Am. 60, 440–445.

Baer, T. (1975). "Investigation of Phonation Using Excised Larynxes," Ph.D. thesis (MIT, Cambridge, MA) (unpublished).

Bouhuys, A., Proctor, D. F., and Mead, J. (1966). "Kinetic Aspects of Singing," J. Appl. Physiol. 21, 483–496.

Collier, R. (1975). "Physiological correlates of intonation patterns," J. Acoust. Soc. Am. 58, 249–255.

Erickson, D. (1975). "A Laryngeal Description of Thai Tones," paper presented to the Linguistic Society of America, 28–30 Dec.

Erickson, D., and Atkinson, J. E. (1975). "Function of strap muscles in speech," J. Acoust. Soc. Am. 58, S62(A).

Faaborg-Andersen, K. (1965). "Electromyography of Laryngeal Muscles in Humans-Techniques and Results," *Current Problems in Phoniatrics and Logopedics*, (S. Karger, Basel, Switzerland), Vol. III.

Flanagan, J. L., Rabiner, L. R., Christopher, D., Boch, D. E., and Shipp, T. (1976). "Digital analysis of laryngeal control in speech production," J. Acoust. Soc. Am. 60, 446–455.

Fromkin, V., and Ohala, J. (1968). "Laryngeal Control and a Model of Speech Production," UCLA Working Papers in Phonetics, No. 10, 98–110.

Gay, T. G., and Harris, K. S. (1971). "Some Recent Developments in the Use of Electromyography in Speech Research," J. Speech Hear. Res. 14, 241–246.

Gay, T. G., Strome, M., Hirose, H., and Sawashima, M. (1972). "Electromyography of Intrinsic Laryngeal Muscles During Phonation," Ann. Otol. Rhinol. Laryngol. 81, 401–410.

Gårding, E., Fujimura, O., and Hirose, H. (1970). "Laryngeal Control of Swedish Word Tones," Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo 4, 45–59.

Hirano, J., Ohala, J., and Vennard, W. (1969). "The Function of the Laryngeal Muscles in Regulating Fundamental Frequency and Intensity of Phonation," J. Speech Hear. Res. 12, 616–628.

Hirose, H. (1971). "Electromyography of the Articulatory Muscles: Current Instrumentation and Technique," Haskins Laboratories Status Report on Speech Research SR-25/26, 73–86.

Hirose, H., and Gay, T. (1972). "The Activity of the Intrinsic Laryngeal Muscles in Voice Control: Electromyographic Study," Phonetica 25, 140–164.

Hirose, H., Simada, Z., and Fujimura, O. (1970). "An EMG Study of the Activity of the Laryngeal Muscles During Speech Utterances," Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo 4, 9–25.

Kotby, M., and Haugen, L. (1970). "The Mechanics of Laryngeal Function," Acta-Oto-Laryngol. 70, 74–83.

Ladefoged, P. (1962). "Subglottal Activity During Speech," Proc. IVth Int. Congress Phonetic Sciences, Helsinki (Mouton, The Hague).

Lieberman, P. (1967). *Intonation, Perception, and Language* (MIT, Cambridge, MA).

Lieberman, P. (1970). "Toward a Unified Phonetic Theory," Linguistic Inquiry 1, 307–322.

Lieberman, P., Sawashima, M., Harris, K. S., and Gay, T. (1970). "The Articulatory Implementation of the Breath-Group and Prominence: Cricothyroid Muscular Activity in Intonation," Language 46, 312–327.

Lisker, L., Abramson, A. S., Cooper, F. S., and Schvey, M. H. (1969). "Transillumination of the Larynx in Running Speech," J. Acoust. Soc. Am. 45, 1544–1546.

Luketala, G. (1973). "Pitch Determination by Adaptive Autocorrelation Method," Haskins Laboratories Status Report on Speech Research SR-33, 185–194.

Ohala, J. (1970). "Aspects of the Control and Production of Speech," UCLA Working Papers in Phonetics, No. 15.

Ohala, J. (1972). "How is pitch lowered?" J. Acoust. Soc. Am. 52, 124(A).

Ohala, J., and Hirose, H. (1970). "The Function of the Sternohyoid Muscle in Speech," Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo 4, 41–44.

Port, D. (1971). "The EMG Data System," Haskins Laboratories Status Report on Speech Research SR-25/26, 67–72.

Sawashima, M. (1973). "Laryngeal Research in Experimental Phonetics," in *Current Trends in Linguistics* 12, edited by T. Sebeok (Mouton, The Hague).

Simada, Z. and Hirose, H. (1970). "The Function of the Laryngeal Muscles in Respect to Word Accent Distinction," Annual Bulletin, Research Institute of Logopedics and Phoniatrics, University of Tokyo 4, 27–40.

Sonninen, A. (1968). "The External Frame Function in the Control of Pitch in the Human Voice," Sound Production in Man, Ann. N. Y. Acad. Sci. 1, 55, 68–90.

Stevens, K. N. (1963). "Theory of Vocal-Cord Vibration and its Relation to Laryngeal Features," J. Acoust. Soc. Am. 55, 383(A).

Van den Berg, J. W. (1960). "Vocal Ligaments vs Registers," Current Problems in Phoniatrics and Logopedics 1, 19–34.

Van den Berg, J. W., and Tan, T. S. (1959). "Results of Experiments with Human Larynxes," Pract. Oto-Rhino-Laryngol. 21, 425–450.

Zemlin, W. R. (1968). *Speech and Hearing Science Anatomy and Physiology* (Prentice-Hall, Englewood Cliffs, NJ).