Ann. Bull. RILP, No. 14, 101-111 (1980)

# ANALYSIS OF PITCH CONTROL IN SINGING

Hiroya Fujisaki*, Mariko Tatsumi* and
Norio Higuchi*

## 1. Introduction

Needless to say, changes in voice fundamental frequency play important roles both in speech and in singing. In many spoken languages, temporal patterns of the fundamental frequency are major manifestations of the suprasegmental structure of a spoken message and are used to convey both lexical and syntactic information, while in some languages they carry even segmental information. Their role in singing is also crucial since they carry the melodic information. Studies of fundamental frequency control (henceforth pitch control) in speech and singing are thus quite important in elucidating the underlying mechanism by which such information is encoded and transmitted to a listener.

The dynamic process of pitch control in speech has been studied rather extensively. For example, the fundamental frequency contours (henceforth pitch contours) of spoken words and sentences were analyzed and modeled by Fujisaki and others, separating features that are related to the linguistic content of an utterance from those related to the voice control mechanism.[1-2] This approach has been applied to analysis of the word accent both in Japanese and in English,[3-4] as well as to analysis of the sentence intonation of Japanese.[5] The dynamic characteristics of the larynx in rapid pitch changes were also studied by Ohala and Ewan, indicating a tendency for a downward pitch change to be more rapid than an upward pitch change, as well as a tendency for the transition time to be unaffected by the magnitude of the pitch change.[6] On the other hand, studies of pitch control in singing have been rather scarce. The speed of rapid pitch changes in singing, however, was investigated recently by Sundberg in an experiment in which the subjects were asked to alternate repeatedly between two given pitches in a *legato*-like performance.[7] His results show mean response times of about 50 to 80 msec with systematic changes due to differences in sex and amount of professional training.

It is to be expected, however, that the speed (hence the response time) of pitch changes will certainly depend on expression, manner of performance, and possibly on tempo. The purpose of the present study is to analyze the characteristics of pitch transitions from one note to another under various singing conditions, using techniques already developed for the analysis of pitch contours of speech.

---

* Department of Electrical Engineering, Faculty of Engineering, University of Tokyo.

## 2. Materials and Subjects

The material for the present study consisted of a number of two-note sequences sung with the vowel [a] in various manners of performance (i. e., in various degrees of articulation of the two notes). They were: a) a note with an *appoggiatura*, b) two notes sung in *staccato*, c) two notes sung in *non legato*, d) two notes sung in *legato*, and e) two notes sung in *portamento*. Since, however, the two notes sung in *staccato* are generally separated by a silent gap of considerable duration, they were left out of the subsequent analysis. The subjects were two females with a similar range and quality of voice. One (subject MT) was a voice trainer with 10 years of professional training in singing, and the other (subject YS) was a student at a music conservatory with three years of training in singing. The pitches of the two notes were selected to suit their voice range, and the intervals were either a musical fourth ($A_4$ - $D_5$) or an octave ($D_4$ - $D_5$). The sequences were produced in both directions, upward and downward. Each sequence was repeated several times in 3/4 time with an M. M. setting of 100, except in the case of *portamento* where the beat was approximately M. M. 80. These sequences were sung at three levels of volume: *forte* (*f*), *mezzopiano* (*mp*), and *pianissimo* (*pp*). A minimum of five samples were collected for each of the conditions and subjects. For the sake of comparison, speech materials were also recorded. These were isolated utterances of two words in the *Tokyo* dialect of Japanese: "ame" [a͡me] (candy) and "āme" [āme] (rain). These two words possess an identical phonemic structure but differ in the accent type manifested mainly in their pitch contours, the former being the "low-high" type and the latter being the "high-low" type.

## 3. Method of Analysis

The analysis of pitch transitions in singing involves two stages of processing, i. e., (1) extraction of the fundamental frequency trajectory, and (2) extraction of the parameters of the trajectory.

### 3. 1 Extraction of the fundamental frequency trajectory

The fundamental periods are detected pitch synchronously using both short-term autocorrelation analysis and waveform peak detection. These fundamental periods are converted to fundamental frequency values, which are further smoothed and interpolated to produce a trajectory (henceforth $F_0$-trajectory) uniformly sampled at intervals of 10 msec.

In the sung materials, a trajectory thus extracted usually consists of an initial segment which is quasi-stationary except for the case of *appoggiatura*, and a transitional segment characterized by a smooth movement toward a stationary final segment. The initial and final segments are usually accompanied by a *vibrato*, i. e., an almost periodic modulation of the fundamental frequency, whose amplitude is considerably diminished in the transitional segment. As pointed out by Vennard and Sundberg, the *vibrato* generally aligns in phase with the transition [7-8]. Although the analysis of interaction between these two components is certainly worthwhile, we

are concerned here only with the transient component whose parameters reflect the characteristics of the pitch control mechanism.

3.2 Extraction of parameters of an $F_0$-trajectory

The extraction of parameters of an observed $F_0$-trajectory is based on an approximate functional formulation (model) of the process of its production. Our previous studies on pitch contours of speech suggest that a proper formulation of the $F_0$-trajectory in singing should be based on the logarithmic scale of the fundamental frequency. They also suggest that the transient component can be regarded as the response of the pitch control mechanism to a hypothetical command to switch notes. If we assume the command to be a step function, the pitch control mechanism can be functionally approximated by a second-order linear system, and the $F_0$-trajectory of a two-note sequence may be represented by

$$\ln [F_0(t) / F_i] = \ln (F_f / F_i)\ f(\beta, \gamma, t)\ u(t), \quad (1)$$

where

$$f(\beta, \gamma, t) = 1 - \left[\cos \beta\sqrt{1-\gamma^2}\, t + \frac{\gamma}{\sqrt{1-\gamma^2}} \sin \beta\sqrt{1-\gamma^2}\, t\right] \exp(-\beta\gamma t), \quad \text{for } \gamma<1,$$

$$= 1 - (1+\beta t) \exp(-\beta t), \quad \text{for } \gamma=1,$$

$$= 1 - \left[\cosh \beta\sqrt{\gamma^2-1}\, t + \frac{\gamma}{\sqrt{\gamma^2-1}} \sinh \beta\sqrt{\gamma^2-1}\, t\right] \exp(-\beta\gamma t), \quad \text{for } \gamma>1,$$

u(t) denotes the unit step function, $F_i$ and $F_f$ respectively denote the initial and final values of the transition, and $\beta$ and $\gamma$ are parameters characterizing the second-order linear system. In particular, $\gamma$ is the damping factor, and the three conditions for $\gamma$ in the above equations correspond to the under-damped, the critically-damped, and the over-damped cases, respectively. The origin of the time axis is selected at the onset of transition.

Although it is possible to measure such characteristics as the rise/fall times directly from an $F_0$-trajectory, the above formulation gives us more insight into the underlying mechanism of pitch control. Parameters such as $\beta$ and $\gamma$ can be obtained from a measured $F_0$-trajectory by finding its best approximation given by the above equations, which can then be used to determine the rise/fall times. For the sake of comparison with the results obtained by Ohala and Sundberg, we adopt here the same definition of rise/fall time as in their studies. Namely, the rise/fall time is defined as the time required for the pitch to change from 1/8 to 7/8 of the total range of transition.

The above formulation for the $F_0$-trajectory of sung notes follows the same line of thought as that for the $F_0$-trajectory of a spoken word, except that in the latter we assume an additional "voicing component" of the following shape

$$g(t) = h(t - t_i) - h(t - t_f), \quad (2)$$

where

$$h(t) = A_V \alpha t \exp(-\alpha t)\ u(t),$$

which serves as the baseline upon which the accent component is superposed on the logarithmic scale of the fundamental frequency. In Eq. (2), $A_V$ and $\alpha$ respectively denote the amplitude and the rate of change of the voicing component, while $t_i$ and $t_f$ denote the onset and the offset of the command for voicing. Thus the method of analysis for the speech material is essentially the same as that for the sung material except for the addition of parameters characterizing the voicing component.

## 4. Results

Figure 1 illustrates one example each of pitch transitions across the interval of a fourth ($A_4$ - $D_5$) in the sung material of subject MT produced at *mezzopiano* under eight different conditions, i. e., upward and downward transitions sung at four different degrees of articulation (*appoggiatura*, *non legato*, *legato*, and *portamento*). The symbol (+) indicates a value of the fundamental frequency sampled at 10 msec intervals on the measured $F_0$-trajectory, while the curve in each panel indicates the best approximation based on Eq. (1) of the preceding section. These and other results of analysis indicate that the model adopted in the present study can provide very good approximations to almost all the $F_0$-trajectories observed, and hence its parameters can be regarded as good indices for the dynamic properties of the pitch control mechanism.

Quite naturally, the rate of pitch change is seen to vary to a large extent with the degree of articulation, and also to vary with the direction, i. e., a downward transition is faster than an upward transition especially in *appoggiatura* and in *non legato*. The $F_0$-trajectories are clearly underdamped in these fast transitions, while they are almost critically damped in normal and slower transitions (*legato* and *portamento*). The *vibrato* component, which appears as deviations of the measured points from the smooth trajectory given by the model, is seen to be suppressed during the transition.

Figure 2 shows one example each of pitch contours of the speech material uttered by the same subject, together with the best approximations produced by the model of the word pitch contour under the assumption of critical damping.

Table 1 lists the mean values of $\beta$ and $\gamma$ averaged over several samples each of the eight conditions, obtained from the analysis of the material sung by MT, together with the mean rise/fall times $\tau^*$ defined in the preceding section and calculated from $\beta$ and $\gamma$. For the sake of comparison, parameter values obtained from the speech material are also listed. The rise/fall times for the speech material were also calculated from $\beta$ and $\gamma$ disregarding the existence of the voicing component. The rate of transition in the spoken material is considerably slower than that of the pitch transition in *legato* singing.

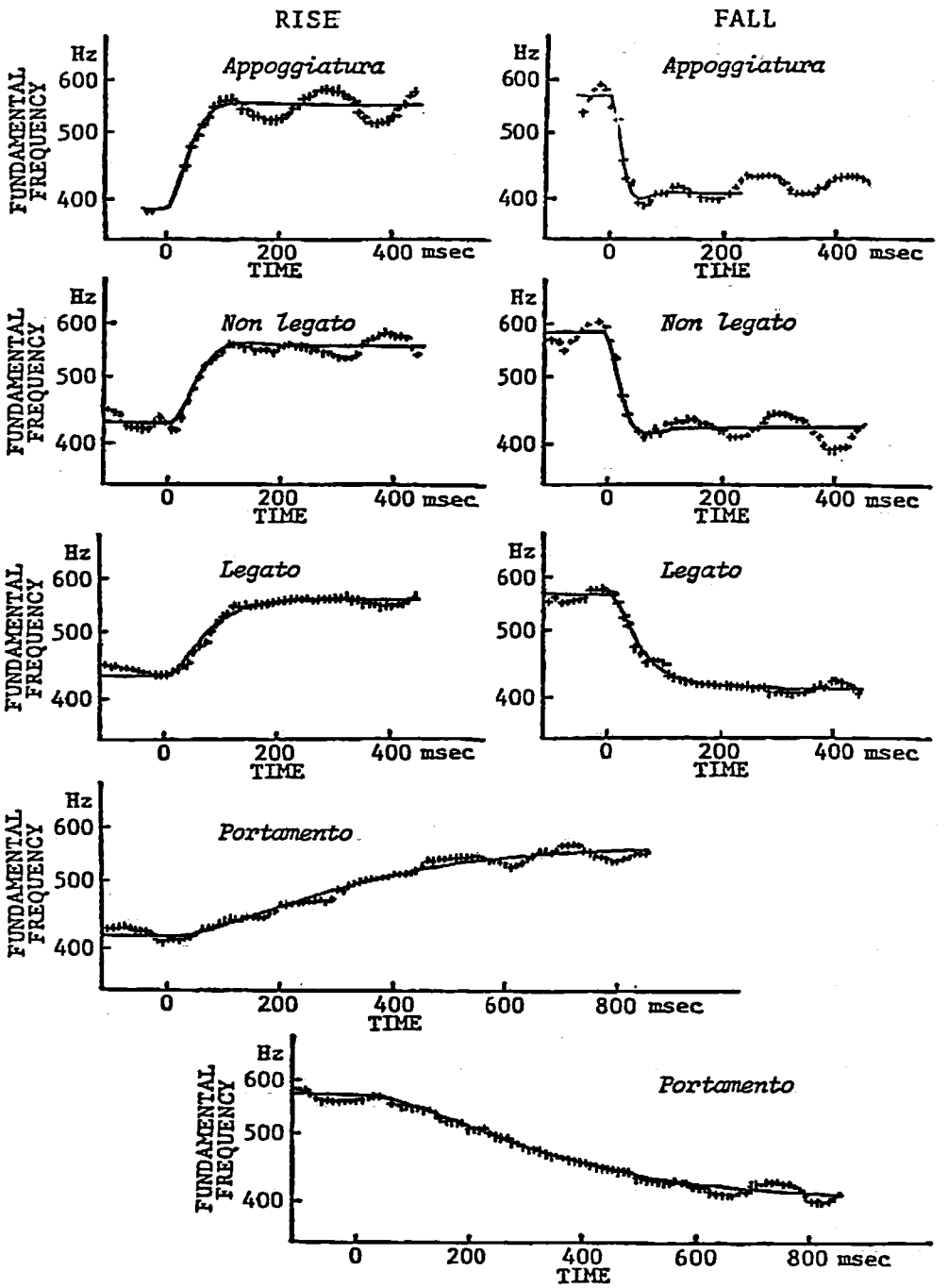


Fig. 1 Analysis of $F_0$-trajectories in singing. Typical results obtained from samples of two-note sequences separated by an interval of a fourth (subject MT).

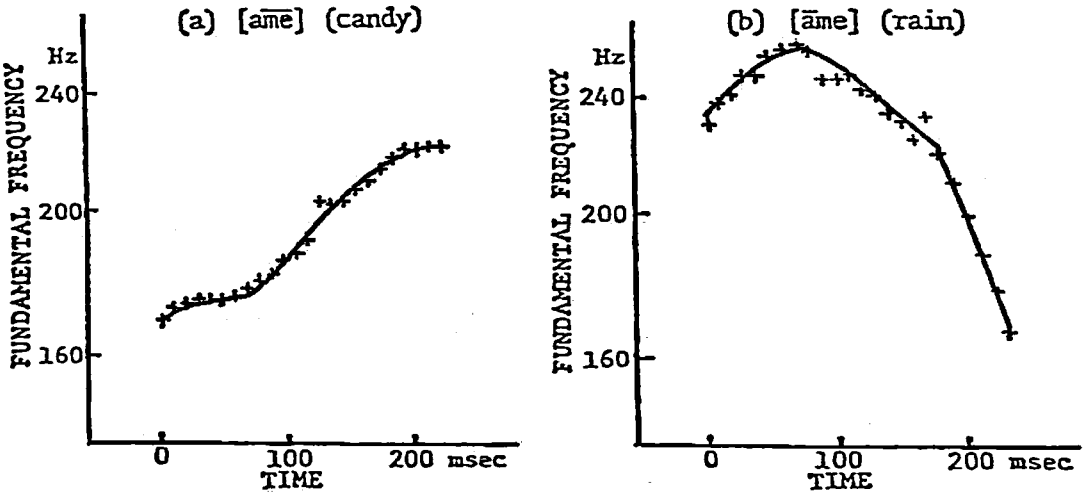


Fig. 2 Analysis of $F_0$-trajectories in speech. Results obtained from (a) [ām̄e] (candy) and (b) [āme] (rain) uttered by subject MT.

Table 1. Characteristic parameters and rise/fall times of $F_0$-trajectories in singing and in speech. The results for singing are from two-note sequences separated by an interval of a fourth (subject MT).

| | Rise | | | Fall | | |
|---|---|---|---|---|---|---|
| | $\beta$ ($sec^{-1}$) | $\gamma$ | $\tau^*$ (msec) | $\beta$ ($sec^{-1}$) | $\gamma$ | $\tau^*$ (msec) |
| *Appoggiatura* | 40 | 0.76 | 54 | 75 | 0.50 | 20 |
| *Non legato* | 38 | 0.85 | 67 | 57 | 0.77 | 38 |
| *Legato* | 27 | 0.85 | 91 | 33 | 0.94 | 87 |
| *Portamento* | 6.8 | 0.95 | 419 | 6.9 | 0.93 | 409 |
| Speech | 18 | 1.00 | 167 | 18 | 1.00 | 171 |

The mean values of $\beta$ and $\gamma$ are plotted in Fig. 3 on the $\beta$ - $\gamma$ plane to illustrate the differences in characteristics of pitch transition in the eight conditions of singing. There exists a high negative correlation between the two parameters. The mean rise/fall times for the upward and downward transitions over the interval of a fourth are compared in Fig. 4 for each of the four manners of performance, while those for transitions over the interval of an octave are compared in Fig. 5 for *appoggiatura*, *non legato*, and *legato*. From Figs. 4 and 5 it can be seen that mean rise/fall times are not greatly affected by the pitch interval between the two notes, as already pointed out both by Ohala and Ewan[6] and by Sundberg[7].

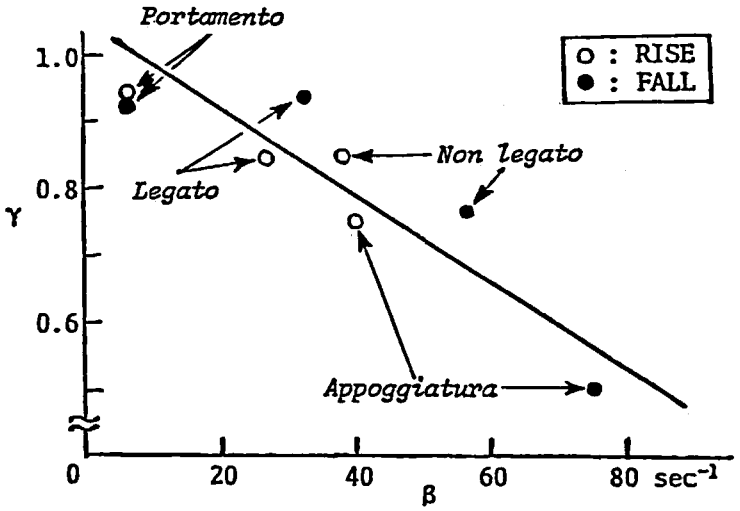


Fig. 3 Parameters of various $F_0$-trajectories in singing. The results are from two-note sequences separated by an interval of a fourth (subject MT).

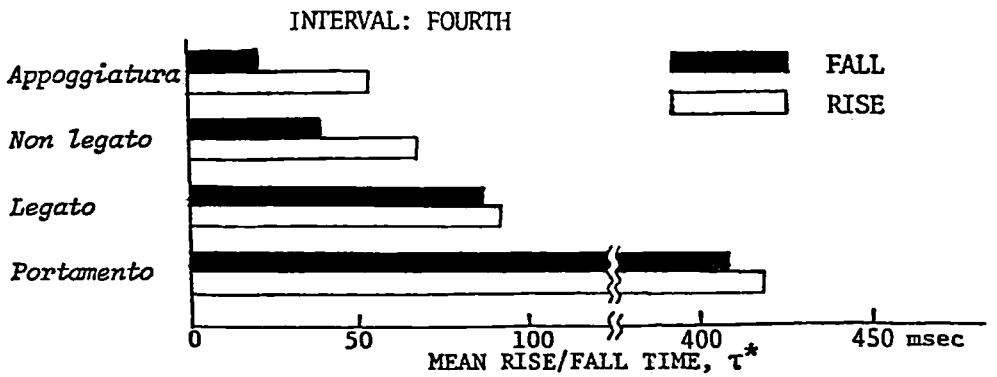


Fig. 4 Rise/fall times of $F_0$-trajectories in singing. The results are from two-note sequences separated by an interval of a fourth (subject MT).

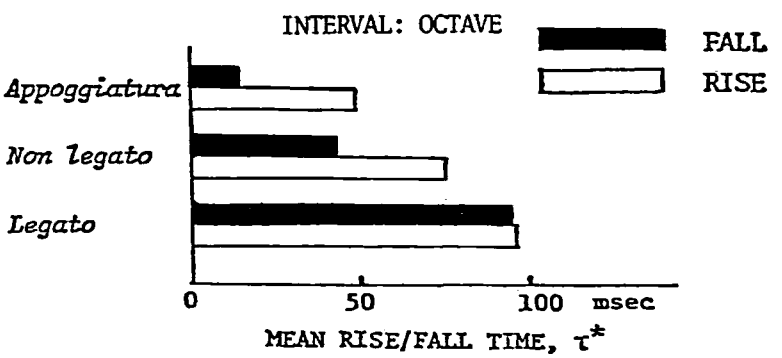


Fig. 5 Rise/fall times of $F_0$-trajectories in singing. The results are from two-note sequences separated by an interval of an octave (subject MT).

Analysis of materials sung at different levels indicated that changes in the volume do not appreciably affect the rate of transition in most cases. Exceptions are the downward pitch transitions in the case of *appoggiatura* and *non legato*, where a marked increase was observed in the rate of transition by going from *mezzopiano* to *forte*. In fact, a fall time of as short as 17 msec was observed in one sample of *appoggiatura* sung by MT, though with a considerable amount of overshoot due to insufficient damping.

All the foregoing results were obtained from the analysis of materials sung by MT, but the results for subject YS also showed similar tendencies. Comparison of results for the two subjects indicates that the differences in parameters of pitch transition are rather small in slower transitions, but become more marked in rapid transitions. For example, the minimum rise/fall time observed in YS was 36 msec, which was twice as long as the minimum value for MT. The variability of the final value of pitch transition was also found to be somewhat greater in YS. These differences may be ascribed to differences in the period and amount of training between the two subjects.

## 5. Discussion

Although the present study is concerned mainly with the analysis of the dynamic characteristics of pitch control in singing, the results can be interpreted to shed light on the underlying mechanism of pitch control both in speech and in singing.

The essential features of the present analysis are the success in:
(1) the functional formulation of the dynamic characteristics of pitch control on the logarithmic scale of the fundamental frequency, and
(2) the approximation of the $F_0$-trajectory in terms of the response characteristics of a second-order linear system. Although these results were obtained empirically, we will not present our interpretations based on some theoretical considerations and published data on the physical properties of skeletal muscles.

The closeness of approximation of (logarithmic) $F_0$-trajectories by our model strongly suggests that the logarithmic fundamental frequency may actually reflect the mechanical motion of an element in the laryngeal mechanism which, from the point of view of pitch control, can be approximated by a second-order linear system. More specifically, we present the following hypotheses and show evidences supporting these hypotheses.
Hypothesis (1). The logarithmic fundamental frequency varies linearly with the displacement of a point in the laryngeal structure.
Hypothesis (2). The displacement of the point reflects the mechanical motion of a mass element connected with stiffness and viscous resistance elements.

Supporting evidence for Hypothesis (1) can be found in the stress-strain relationship of muscles. Although we do not have data from the vocalis muscle, the following experimental relationship has been known to apply between the tension T and the elongation $x$ of skeletal muscles in general[9, 10]:

$$T = a\,(e^{bx} - 1). \qquad (3)$$

In the present study, we regard $x$ as the elongation of the vocalis muscle due mainly to the displacement of its anterior end. If $e^{bx} \gg 1$, the above equation can be approximated by

$$T = a e^{bx} . \tag{4}$$

On the other hand, the frequency of vibration of strings as well as membranes with simple structures varies generally in proportion to the square root of their tension[11]. This relationship will hold even for the vibration of the vocal fold, which can be regarded as an elastic membrane to a first-order approximation. Thus

$$f_0 = c_0 \sqrt{T} . \tag{5}$$

From Eqs. (3) and (4) we obtain

$$ln\, f_0 = \frac{b}{2} x + ln\, (\sqrt{a} \cdot c_0) , \tag{6}$$

where, strictly speaking, $c_0$ also varies slightly with $x$, but the overall dependency of $ln\, f_0$ is primarily determined by the first term on the right-hand side of Eq. (6)

Hypothesis (2) can be supported by the analysis of the mechanical properties of the laryngeal structure whose major elements are shown in Fig. 6. If we adopt a coordinate fixed with the cricoid cartilage (and the trachea which is connected more or less tightly with the cricoid), the thyroid cartilage can be regarded as one major mass element supported by two stiffness elements (the cricothyroid and the vocalis muscles) and rotating around the cricothyroid joint with a viscous resistance. The two stiffness elements are also accompanied by viscous resistances which represent their internal losses. If we denote the angular displacement of the thyroid by $\theta$, its rotation can be described by the following equation of motion:

$$I\ddot{\theta} + R\dot{\theta} + (c_1 K_1 + c_2 K_2)\theta = \tau(t) , \tag{7}$$

where I represents the moment of inertia, R represents the combined viscous loss, $K_1$ and $K_2$ represent the stiffness of the cricothyroid and vocalis muscles, and $\tau(t)$ represents the torque caused by the contraction of the cricothyroid muscle. Since the anterior end of the vocalis muscle is fixed at a point on the thyroid cartilage, the following relationship holds approximately between $x$ and $\theta$ for small angular displacement of the thyroid cartilage:

$$x = c_3 \theta . \tag{8}$$

For a unit-step forcing function u(t), $x$ can then be given by

$$x = c_4 f(\beta, \gamma, t)\, u(t) , \tag{9}$$

where $f(\beta, \gamma, t)$ is the same as in Eq. (1), and

$$\beta = \sqrt{\frac{c_1K_1 + c_2K_2}{I}}, \qquad \gamma = \frac{R}{2\sqrt{I\,(c_1K_1 + c_2K_2)}}.$$

Equations (6) and (9) with initial and final conditions will lead to Eq. (1).

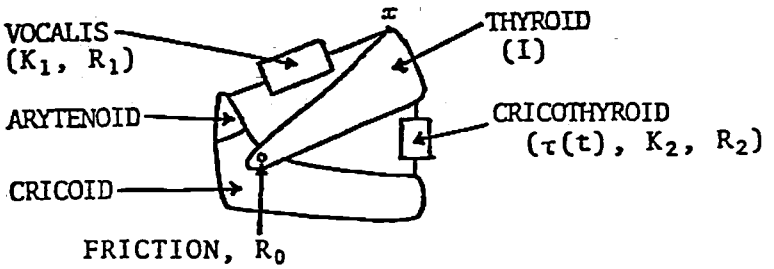


Fig. 6 Simplified laryngeal structure showing only those elements that exert dominant influences on the dynamics of the voice fundamental frequency.

As one of the possible ways to control the rate of pitch transition, we may assume that only the stiffness of the related laryngeal muscles is changed. In this case, the following hyperbolic relationship is expected to hold between $\beta$ and $\gamma$:

$$\beta \cdot \gamma = R / 2I. \qquad (10)$$

While the results of Fig. 3 indicate a definite negative correlation between $\beta$ and $\gamma$, the relationship is not strictly hyperbolic. The relatively large value of $\gamma$ for larger $\beta$ (*appoggiatura*) can be ascribed to the increase of viscous resistance within the muscle itself in the case of stronger contraction[12, 13], while the tendency of $\gamma$ to approach 1 for smaller values of $\beta$ (*portamento*) may be the consequence of better coordination between antagonistic muscles to accomplish optimal control.

On the other hand, differences in the rate for upward and downward transitions can be explained by referring to the stress-strain relationship of Eq. (3). The incremental stiffness, as given by $\partial T/\partial x$, is obviously greater at larger values of $x$. Since the initial value of $x$ is greater in the downward transitions, the stiffness is greater and hence produces a larger value of $\beta$ than in the upward transitions.

## 6. Conclusions

Dynamic characteristics of pitch control in singing have been investigated using techniques already developed for the analysis of pitch control in speech. It was found that the trajectory of the voice fundamental frequency in transitions from one note to another, when expressed in logarithmic units, can be approximated very well by the step response function of a second-order linear system. The parameters of the model were determined by the method of successive approximation to the measured $F_0$-trajectory, and were used to represent the dynamic characteristics of the vocal pitch

control. In general, our results on *legato* singing were in agreement with the findings of Ohala and Ewan as well as with those of Sundberg. On the other hand, it was found that under certain conditions the rise/fall times of pitch transition in singing could be much shorter than 50 msec, which is the smallest value reported by these investigators. The speed of pitch transition in speech was shown to be considerably lower than that found in *legato* singing. Furthermore, an interpretation of the observed dynamic characteristics in terms of a possible underlying mechanism of pitch control has been presented on the basis of two hypotheses, supported by some physiological data and theoretical considerations.

Acknowledgment

The authors are grateful to Dr. K. Akazawa for his helpful comments on the preliminary version of the paper.

References

1. Fujisaki, H. and S. Nagashima (1969); "A Model for Synthesis of Pitch Contours of Connected Speech, "Annual Report, Engineering Research Institute, Faculty of Engineering, University of Tokyo 28, 53-60.
2. Fujisaki, H. and H. Sudo (1971); "A Model for the Generation of Fundamental Frequency Contours of Japanese Word Accent, " J. Acoust. Soc. Japan 27, 445-453.
3. Fujisaki, H, and M. Sugito (1978); "Analysis and Perception of Two-Mora Word Accent Types in the *Kinki* Dialect, " J. Acoust. Soc. Japan 34, 167-176.
4. Hirose, K., H. Fujisaki and M. Sugito (1978); "Word Accent in Japanese and English: A Comparative Study of Acoustic Characteristics in Disyllabic Words, " J. Acoust. Soc. Am. 64, Suppl. 1, S114.
5. Fujisaki, H., K. Hirose and K. Ohta (1979); "Acoustic Features of the Fundamental Frequency Contours of Declarative Sentences in Japanese, " Annual Bulletin, Research Institute of Logopedics and Phoniatrics, Faculty of Medicine, University of Tokyo, No. 13, 163-173.
6. Ohala, J. and W. Ewan (1973); "Speed of Pitch Change, " J. Acoust. Soc. Am. 53, 345(A).
7. Sundberg, J. (1979); "Maximum Speed of Pitch Changes in Singers and Untrained Subjects, " J. Phonetics 7, 71-79.
8. Vennard, W. (1967); Singing - the Mechanism and the Technic (revised edition), Carl Fischer, Inc., New York.
9. Buchthal, F. and E. Kaiser (1944); "Factors determining Tension Development in Skeletal Muscle, " Acta Physiol. Scand. 8, 38-74.
10. Sandow, A. (1958); "A Theory of Active State Mechanisms in Isometric Muscular Contraction, " Science 127, 760-762.
11. Slater, J. C. and N. H. Frank (1933); Introduction to Theoretical Physics, McGraw-Hill Book Co., New York.
12. Mashima, H., K. Akazawa, H. Kushima and K. Fujii (1973); "Graphical Analysis and Experimental Determination of the Active State in Frog Skeletal Muscle, " Jap. J. Physiol. 23, 217-240.
13. Akazawa, K. (1980); Personal communication.