Physiological and Physical Mechanisms for Fundamental Frequency Control in Some Tone Languages and a Command-Response Model for Generation of Their F_0 Contours

Hiroya Fujisaki[1], *Sumio Ohno*[2] *and Wentao Gu*[3, 1]

[1]University of Tokyo, [2]Tokyo University of Technology, [3]Shanghai Jiaotong University
fujisaki@alum.mit.edu, ohno@cc.teu.ac.jp, wtgu@gavo.t.u-tokyo.ac.jp

Abstract

This paper first presents the physiological and physical properties of the vocal fold and the laryngeal structure involving intrinsic laryngeal muscles that are mainly responsible for the generation of F_0 contours with global components and positive local components. It then takes up the mechanism involving extrinsic laryngeal muscles for generating negative local components. Based on these evidences, a command-response model is derived for generating F_0 contours of tone languages with an extremely high accuracy. Experimental results support the validity of the model for Mandarin, Thai, and Cantonese.

1. Introduction

In many languages of the world, the contour of the voice fundamental frequency (F_0 contour) plays an important role in conveying linguistic, para-linguistic and non-linguistic information. This is accomplished by controlling the frequency of vibration of the vocal cords mainly through various intrinsic and extrinsic laryngeal muscles. As far as the linguistic information is concerned, information on the syntactic structure is mainly expressed by relatively slow changes (global components), while information on the word accent/syllable tone is expressed by relatively rapid changes (local components) of the F_0 contour. Although the basic mechanism is the same in most languages, certain differences may exist among languages. While the mechanism for F_0 control is fairly clear for languages whose F_0 contours have only positive local components, it is not so for languages that use also negative local components.

In the present paper, we will first present the physiological and physical properties of the vocal fold and the laryngeal structure that supports a model for the generation process of F_0 contour with global components and positive local components. It then takes up Mandarin and Thai as examples of languages that use both positive and negative local components to express tones, and explains the mechanism involving extrinsic laryngeal muscles that is responsible for the generation of negative local components.

2. Vocal cord length and voice fundamental frequency [1]

2.1. Stress-strain relationship of skeletal muscles

The stress-strain relationship of skeletal muscles including the human vocalis muscle has been widely studied [2, 3]. Figure 1 shows the earliest published data on the relationship between tension and stiffness [2].

The data shown in Figure 1 indicate the existence of a very good linear relationship between tension and stiffness over a wide range of values, and can be approximated quite well by the following equation:

$$dT \ / dl \ = \ a \ + \ bT \ , \qquad (1)$$

where T indicates the tension, l indicates the length of the muscle, and a indicates the stiffness at $T = 0$. This leads to the stress-strain relationship

$$T = (T_0 + a / b) \exp\{ b(l - l_0)\} - a / b, \qquad (2)$$

where T indicates the static tension applied to the vocal cord, and l_0 indicates its length at $T = T_0$. When $T_0 >> a / b$, Equation (2) can be approximated by

$$T = T_0 \exp \ (bx), \qquad (3)$$

where x indicates the change in vocal cord length when T is changed from T_0.

On the other hand, the fundamental frequency F_0 of vibration of an elastic membrane is given by

$$F_0 = c_0 \sqrt{T / \sigma}, \qquad (4)$$

where σ is the density per unit area of the membrane and c_0 is a constant inversely proportional to the size of the membrane. From Equations (3) and (4) we obtain

$$\log_e F_0 = \log_e \{ c_0 \sqrt{T_0 / \sigma} \} + (b / 2)x. \qquad (5)$$

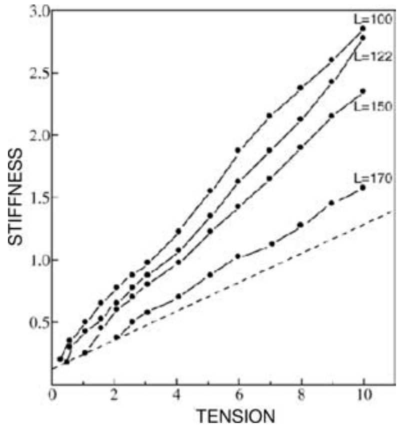

Figure 1*: Stiffness as function of tension at rest (- - - -) and during isometric tetanic contraction initiated at different original length. In the top curve contraction is initiated at a length below 100 (equilibrium length = 100). Ordinate: stiffness in arbitrary units. Abscissa: tension in arbitrary units. (From [2])*

Strictly speaking, the first term varies slightly with x, but the overall dependency of $\log_e F_0$ on x is primarily determined by the second term on the right hand side. This linear relationship was confirmed for sustained phonation by an experiment in which a stereoendoscope was used to measure the length of the vibrating part of the vocal cord [7], and will hold also when x is time-varying. Thus we can represent $\log_e F_0(t)$ as the sum of a constant term and a time-varying term, such that

$$\log_e F_0(t) = \log_e F_b + (b/2)x(t), \qquad (6)$$

where the constant $c_0\sqrt{T_0/\sigma}$ in Equation (5) is rewritten as F_b to indicate the existence of a baseline value of F_0 to which the time-varying term is added when the logarithmic scale is adopted for $F_0(t)$. It is to be noted, however, that the first term ($\log_e F_b$) can be regarded to be approximately constant only as long as the speaker maintains the same speaking style and emotional state. For example, F_b is found to be appreciably higher when the speaker is angry than when he/she is not.

2.2. Role of cricothyroid muscles

Analysis of the laryngeal structure suggests that the movement of the thyroid cartilage relative to the cricoid cartilage has two degrees of freedom [5, 6]. One is horizontal translation due presumably to the activity of *pars obliqua* of the cricothyroid muscle (henceforth CT); the other is rotation around the cricothyroid joint due to the activity of pars recta of the cricothyroid muscle, as illustrated by Figure 2. The translation and the rotation of the thyroid can be represented by separate second-order systems as shown in Figure 3, and both cause small changes in vocal cord length.

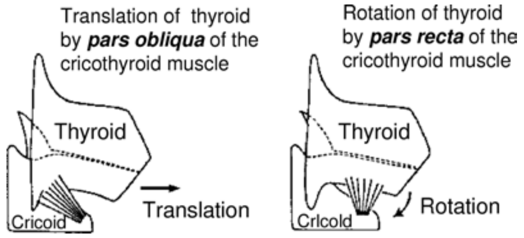
Translation of thyroid by **pars obliqua** of the cricothyroid muscle
Thyroid
Translation
Cricoid
Rotation of thyroid by **pars recta** of the cricothyroid muscle
Thyroid
Crlcold
Rotation

Figure 2: *The roles of pars obliqua and pars recta of the cricothyroid muscle in translating and rotating the thyroid cartilage.*

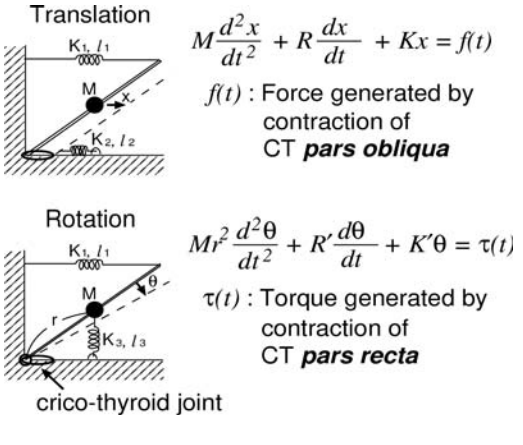
Translation
K1, l1
M
x
K2, l2
$M\frac{d^2x}{dt^2} + R\frac{dx}{dt} + Kx = f(t)$
f(t) : Force generated by contraction of CT **pars obliqua**
Rotation
K1, l1
M
θ
r
K3, l3
crico-thyroid joint
$Mr^2\frac{d^2\theta}{dt^2} + R'\frac{d\theta}{dt} + K'\theta = \tau(t)$
τ(t) : Torque generated by contraction of CT **pars recta**

Figure 3: *Equations of translation and rotation of the thyroid cartilage.*

An instantaneous activity of *pars obliqua* of the CT, contributing to thyroid translation, causes an incremental change x_1(t), while a sudden increase or decrease in the activity of *pars recta* of CT, contributing to thyroid rotation, causes an incremental change x_2(t) in vocal cord length. The resultant change is obviously the sum of these two changes, as long as the two movements are small and can be considered independent from each other. In this case, Equation (6) can be rewritten as

$$\log_e F_0(t) = \log_e F_b + (b/2)\{x_1(t) + x_2(t)\}, \qquad (7)$$

which means that the time-varying component of $\log_e F_0$(t) can be represented by the sum of two time-varying components. Since the translational movement of the thyroid cartilage has a much larger time constant than the rotational movement, the former is used to indicate global phenomena such as phrasing, while the latter is used to indicate local phenomena such as word accent.

2.3. Polarity of Local Components

The foregoing analysis of physiological and physical mechanisms for controlling $F_0(t)$ provides a basis for the command-response model, proposed by the present author, for languages with only positive local components [7, 8]. In this case, a rapid increase in the activity of CT *pars recta* for a certain time interval is represented by a positive pedestal function and named 'accent command,' while a sudden activity of CT *pars obliqua* over a shorter time interval as compared to the time constant of the translational mechanism is represented by an impulse function and named 'phrase command.' The resulting changes in $\log_e F_0(t)$ caused by these commands are called 'accent component' and 'phrase component,' respectively. It is to be noted that the lowering of $\log_e F_0$ in this case occurs due to the sudden decrease in the activity of CT *pars recta*, and does not require an increase in the activity of other muscles. For the rest of the paper, we shall use the word 'F_0 contour' to indicate $\log_e F_0(t)$.

Analysis of F_0 contours of several languages including Mandarin, Thai and Swedish, however, indicates that the local components (associated with tones in the case of Mandarin and Thai) are not always positive but can be both positive and negative. In other words, it is necessary in these languages to posit commands of both positive and negative polarities for the local components, the latter causing active lowering of $\log_e F_0$ below the phrase component.

2.4. Role of extrinsic laryngeal muscles

Although several hypotheses have already been presented on the possible mechanisms for the active lowering of F_0, none seems to be satisfactory since these hypotheses do not take into account the activities of muscles that are directly connected to the thyroid cartilage and are antagonistic to CT *pars recta* in rotating the thyroid cartilage in the opposite direction.

Several EMG studies have shown that the sternohyoid (henceforth SH) muscle is active when the F_0 is lowered in Mandarin [9, 10], the five tones of Thai [11] as well as of the grave accent of Swedish [12], but the mechanism itself has not been made clear since SH is not directly attached to the thyroid cartilage, whose movement is essential in changing the length and hence the tension of the vocal cord.

On the basis of an earlier study on the production of tones of Thai, the present author suggested the active role of the thyrohyoid (henceforth TH) muscle in F_0 lowering in these languages [13]. Figure 4 shows the relationship between the hyoid bone, thyoid and cricoid cartilages, and TH in their lateral and frontal views, and Figure 5 shows their relationships with three other muscles: VOC (thyrovocalis muscle), CT, and SH.

The activity of SH stabilizes the position of the hyoid bone, while the activity (hence contraction) of TH causes rotation of the thyroid cartilage around the crico-thyroid joint, in a direction that is opposite to the direction of rotation when CT is active, thus reducing the length of the vocal cord and thereby reducing its tension, and eventually lowering F_0. This is made possible by the flexibility of ligamentous connections between the upper ends of the thyroid cartilage and the two small cartilages (triticial cartilages) and also between these cartilages and the two ends of the hyoid bone, as in Figure 5.

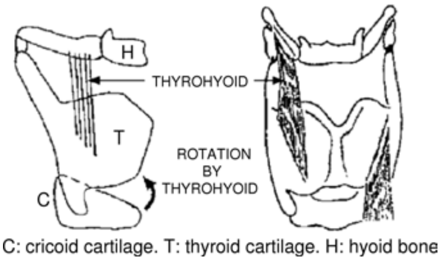

Figure 4: *Role of thyrohyoid in laryngeal control*

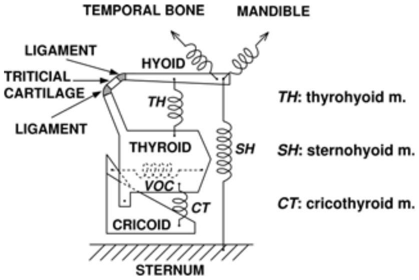

Figure 5: *Mechanism of F_0 lowering by activities of TH and SH.*

3. Mathematical representation of F_0 contours of tone languages with positive and negative local components

The foregoing analysis leads to a model for the generation process of F_0 contours of tone languages from phrase commands and tone commands of positive and negative polarities, as shown in Figure 6.

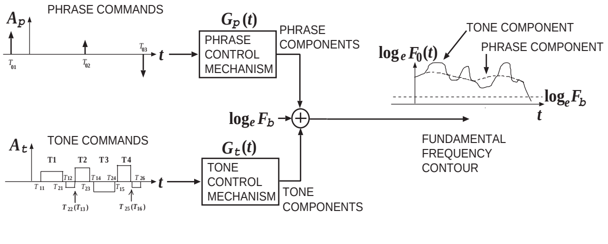

Figure 6: *The F_0 contour generation model for Mandarin.*

In this model, the F_0 contour can be given by the following mathematical formulation:

$$\log_e F_0(t) = \log_e F_b + \sum_{i=1}^{I} A_{pi} G_p(t - T_{0i})$$
$$+ \sum_{j=1}^{J} A_{tj} \{G_t(t - T_{1j}) - G_t(t - T_{2j})\}, \qquad (8)$$

where

$$G_p(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & t \geq 0, \\ 0, & t < 0, \end{cases} \qquad (9)$$

$$G_t(t) = \begin{cases} \min[1 - (1 + \beta_1 t) \exp(-\beta_1 t), \gamma_1], & t \geq 0, \\ 0, & t < 0, \end{cases}$$
$$\text{(for positive tone commands),} \qquad (10)$$

$$G_t(t) = \begin{cases} \min[1 - (1 + \beta_2 t) \exp(-\beta_2 t), \gamma_2], & t \geq 0, \\ 0, & t < 0, \end{cases}$$
$$\text{(for negative tone commands),}$$

where $G_p(t)$ represents the impulse response function of the phrase control mechanism and $G_t(t)$ represents the step response function of the tone control mechanism. The symbols in these equations indicate

F_b : baseline value of fundamental frequency,
I : number of phrase commands,
J : number of tone commands,
A_{pi} : magnitude of the *i*th phrase command,
A_{tj} : amplitude of the *j*th tone command,
T_{0i} : timing of the *i*th phrase command,
T_{1j} : onset of the *j*th tone command,
T_{2j} : end of the *j*th tone command,
α : natural angular frequency of the phrase control mechanism,
β_1: natural angular frequency of the tone control mechanism to positive tone commands,
β_2: natural angular frequency of the tone control mechanism to negative tone commands,
γ_1: relative ceiling level of positive tone components,
γ_2: relative ceiling level of negative tone components.

Although both β and γ should take different values depending on the polarity of commands as in Equation (10), the use of a common value for both β and γ irrespective of command polarity was found to be acceptable in almost all cases.

4. Application of the model to analysis of F_0 contours of Mandarin, Thai, and Cantonese

It is possible to use the above-mentioned model to analyze an observed F_0 contour and estimate the underling commands by the procedure known as Analysis-by-Synthesis. We have already applied it to the analysis of F_0 contours of several tone languages including Mandarin, Thai and Cantonese. Because of space limitations, however, only one example for each language is shown here to illustrate the model's ability of generating very close approximations to observed F_0 contours.

Figure 7 shows the results of analysis of an utterance of the following sentence in Mandarin:

Mu4 ni2 heil buo2 lan3 hui4 bu2 kui4 shi4 dian4 zi3 wan4 hual tong3.
(The Munich exposition is really an electronic kaleidoscope.)

The figure shows, from top to bottom, the speech waveform, the measured F_0 values (+symbols), the model-generated best

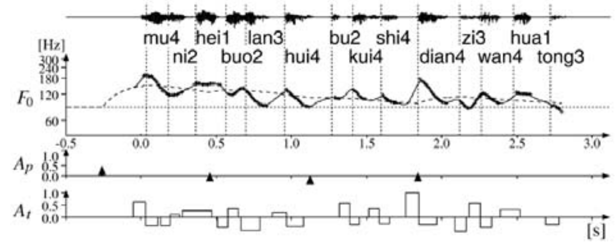

Figure 7: *Analysis of an F_0 contour of a Mandarin utterance.*

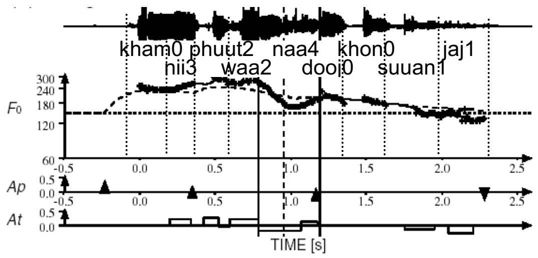

Figure 8: *Analysis of an F_0 contour of a Thai utterance.*

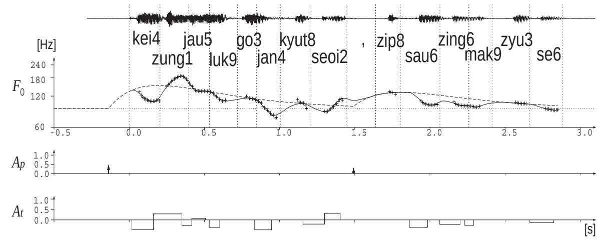

Figure 9: *Analysis of an F_0 contour of a Cantonese utterance.*

approximation (solid line), the baseline frequency (dotted line), the phrase commands (impulses), and the tone commands (pedestal functions). The dashed lines indicate the contributions of phrase components, and the differences between the F_0 contour and the phrase components correspond to the tone components. As seen from the figure, the model is capable of generating an extremely close approximation to the observed contour, and the patterns of estimated commands show the prosodic structure of the utterance quite well. In particular, the figure shows that the patterns of commands for Mandarin tones are: positive for Tone 1 (H), initially negative and then switched to positive for Tone 2 (R), negative for Tone 3 (L), and initially positive and then negative for Tone 4 (F).

Figure 8 shows the results of analysis of an utterance of the following sentence in Thai:

Kham0 nii3 phuut2 waa2 naa4 dooj0 khon0 suuan1 jaj1.
(This word is pronounced as naa4 by most speakers.)

The results show that the patterns of commands for Thai tones are: negative for Tone 1 (L), initially positive and then switched to negative for Tone 2 (F), initially zero and then positive for Tone 3 (H), initially negative and then positive for Tone 4 (R), and zero for Tone 0 (M).

Figure 9 shows the results of analysis of an utterance of the following sentence in Cantonese:

Kei4 zung1 jau5 luk6 go3 jan4 kyut3 seoi2, zip3 sau6 zing6 mak6 zyu3 se6.
(Among them, six persons were de-hydrated, so they received intravenous injection.)

The results of this and other utterances of Cantonese indicate that the patterns of commands for Cantonese tones are: positive for Tone 1, initially negative and then switched to positive for Tone 2, zero for Tone 3, quite negative for Tone 4, initially negative and then zero for Tone 5, and negative for Tone 6. As for the three entering tones, the initial command is positive for Tone 7, zero for Tone 8, and negative for Tone 9, while the later command is always extremely negative so that voicing is interrupted.

5. Conclusions

We have explained the physiological and physical mechanisms for controlling the fundamental frequency of speech for languages having both positive and negative local components in the F_0 contour, and described a model for generating F_0 contours from a set of phrase and tone commands. Analysis-by-Synthesis of observed F_0 contours of tone languages including Mandarin, Thai, and Cantonese has shown that the model can generate very good approximations to observed contours, and allows one to estimate the underlying commands that are closely related to the lexical, syntactic and prosodic information of the utterance.

6. References

[1] Fujisaki, H., 1988. A note on the physiological and physical basis for the phrase and accent components in the voice fundamental frequency contour. In *Vocal Physiology, Voice Production, Mechanisms and Functions*, O. Fujimura (ed.). New York: Raven Press, 347–355.

[2] Buchthal, F.; Kaiser, E., 1944. Factors determining tension development in skeletal muscles. *Acta Physiol. Scand.* 8, 38–74.

[3] Sandow, W., 1958. A theory of active state mechanisms in isometric muscular contraction. *Science* 127, 760–762.

[4] Honda, K.; Hibi, S.; Kiritani, S.; Niimi, S.; Hirose, H., 1980. Measurement of the laryngeal structure during phonation by use of a stereoendoscope. *Ann. Bull. Res. Inst. Logoped. Phoniatr. Univ. Tokyo* 14, 73–78.

[5] Zemlin, W. R., 1968. *Speech and Hearing Science, Anatomy and Physiology*. New Jersey: Prentice Hall, Inc.

[6] Fink, B. R.; Demarest, R. J., 1978. *Laryngeal Biomechanics*. Harvard Univ. Press.

[7] Fujisaki, H.; Nagashima, S., 1969. A model for the synthesis of pitch contours of connected speech. *Annual Report of Engineering Research Institute, University of Tokyo* 28, 53–60.

[8] Fujisaki, H.; Hirose, K., 1984. Analysis of voice fundamental frequency contours for declarative sentences of Japanese. *J. Acoust. Soc. Jpn (E)* 5(4), 233–242.

[9] Sagart, L.; Hallé, P.; De Boysson-Bardies, B.; Arabia-Guidet, C., 1986. Tone production in modern Standard Chinese: an electromyographic investigation. *Cahiers de Linguistique Asie Orientale.* 15, 205–211.

[10] Hallé, P.; Niimi, S.; Imaizumi, S.; Hirose, H., 1990. Modern Standard Chinese four tones: electromyographic and acoustic patterns revisited. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo* 24, 41-58.

[11] Erickson, D., 1976. A Physiological Analysis of the Tones of Thai, *PhD. Dissertation*, University of Connecticut.

[12] Gårding, E., 1970. Word tones and larynx muscles. *Working Papers, Dept. of Linguistics, Lund University* 3, 20-46.

[13] Fujisaki, H., 1995. Physiological and physical mechanisms for tone, accent and intonation. *Proc. the XXIII World Congress of the International Association of Logopedics and Phoniatrics*. Cairo, 156-159.