ISCA Archive
http://www.isca-speech.org/archive

6th International Conference on Spoken
Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

PHYSIOLOGICAL MECHANISMS FOR FUNDAMENTAL FREQUENCY CONTROL IN STANDARD CHINESE

Hiroya Fujisaki [1], *Ryou Tomana* [1], *Shuichi Narusawa* [1], *Sumio Ohno* [2] *and Changfu Wang* [3]

[1] Science University of Tokyo [2] Tokyo University of Technology
[3] University of Science and Technology of China

ABSTRACT

This paper first presents the physiological and physical properties of the vocal fold and the laryngeal structure that support a model for the generation process of F_0 contour with global components and positive local components. It then takes up Standard Chinese as an example 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.

1. INTRODUCTION

In may languages of the world, the contour of 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 folds 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 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, such as Standard Chinese.

In the present paper, we will first present the physiological and physical properties of the vocal fold and the laryngeal structure that support a model for the generation process of F_0 contour with global components and positive local components. It then takes up Standard Chinese as an example 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 Fig. 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:

$$dT/dl = a + bT, \tag{1}$$

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

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

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

$$T = T_0\ \exp(bx), \tag{3}$$

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

$$F_0 = c_0\sqrt{T/\sigma}, \tag{4}$$

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 Eqs. (3) and (4) we obtain

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

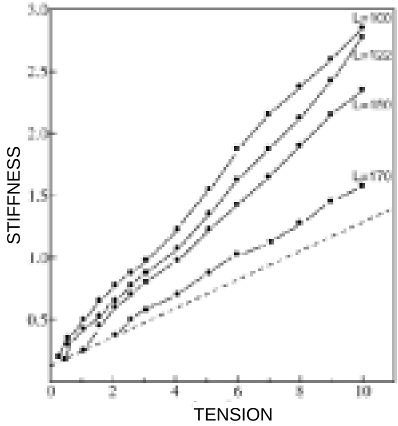


Figure 1: Stiffness as function of tension at rest (- - - -) and during isometric tetanic contraction initiated at different original length. In the top curve contraction is initiated at a length below 100 (equilibrium length = 100).
Ordinate: stiffness in arbitrary units.
Abscissa: tension in arbitrary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

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 [4], 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

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

where the constant $c_0\sqrt{T_0/\sigma}$ in Eq. (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 $F_0(t)$.

2.2. Role of Cricothyroid Muscle

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 Fig. 2. The translation and the rotation of the thyroid can be represented by separate second-order systems as shown in Fig. 3, and both cause small changes in vocal cord length.

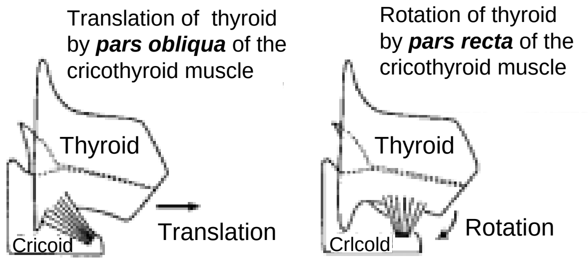


Figure 2: The roles of *pars obliqua* and *pars recta* of the cricothyroid muscle in translating and rotating the thyroid cartilage.

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, Eq. (6) can be rewritten as

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

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. A Functional Model for the Process of F_0 Contour Generation

The foregoing analysis of the physiological and physical mechanisms for controlling F_0 supports the functional model for the process of F_0 contour generation shown in Fig. 4 [7-9]. In this model, the activities of CT *pars obliqua* over a short time interval as compared to the time constant of the translational mechanism are named 'phrase commands' and are represented by impulses, while activities of CT *pars recta* for a certain duration, with rapid onset and offset, are named 'accent commands' and are represented by square waves.

The mechanism that produces changes in $\log_e F_0(t)$ from the phrase commands is named 'phrase control mechanism' and its outputs are named 'phrase components.' Likewise, the mechanism that produces changes in $\log_e F_0(t)$ from the accent commands is named 'accent control mechanism' and its outputs are named 'accent components.' The outputs of these two mechanisms are added to a constant component $\log_e F_b$ to produce the final $\log_e F_0(t)$. Although a further mechanism ('glottal oscillation mechanism') is required to obtain the glottal source waveform, this final stage can be disregarded in the discussion of $\log_e F_0(t)$. For the rest of the paper, we shall use the word 'F_0 contour' to indicate $\log_e F_0(t)$.

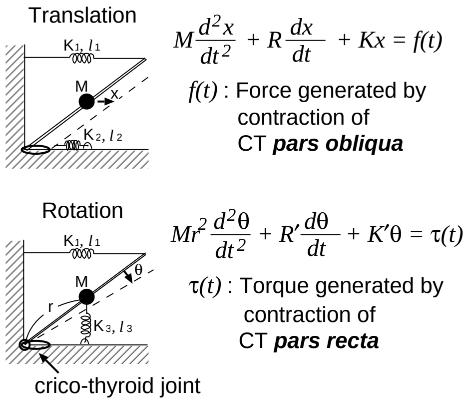


Figure 3: Equations of translation and rotation of the thyroid cartilage.

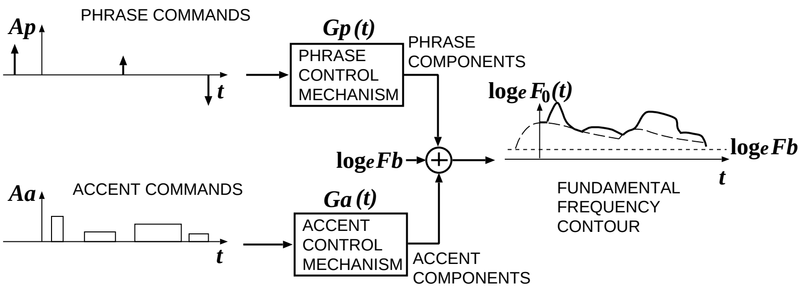


Figure 4: A functional model for the process of generating F_0 contours.

In this model, the F_0 contour is expressed by

$$\begin{aligned}\log_e F_0(t) &= \log_e Fb + \sum_{i=1}^{I} Ap_i Gp(t - T_{0i}) \\ &+ \sum_{j=1}^{J} Aa_j \{Ga(t - T_{1j}) - Ga(t - T_{2j})\},\end{aligned} \tag{8}$$

$$Gp(t) = \begin{cases} \alpha^2 t \exp(-\alpha t), & \text{for } t \geq 0, \\ 0, & \text{for } t < 0, \end{cases} \tag{9}$$

$$Ga(t) = \begin{cases} \min[1 - (1 + \beta t)\exp(-\beta t), \gamma], & \text{for } t \geq 0, \\ 0, & \text{for } t < 0, \end{cases} \tag{10}$$

where $Gp(t)$ represents the impulse response function of the phrase control mechanism and $Ga(t)$ represents the step response function of the accent control mechanism. The symbols in these equations indicate

Fb : baseline value of fundamental frequency,
I : number of phrase commands,
J : number of accent commands,
Ap_i : magnitude of the ith phrase command,
Aa_j : amplitude of the jth accent command,
T_{0i} : timing of the ith phrase command,
T_{1j} : onset of the jth accent command,
T_{2j} : end of the jth accent command,
α : natural angular frequency of the phrase control mechanism,
β : natural angular frequency of the accent control mechanism,
γ : relative ceiling level of accent components.

Parameters α and β are assumed to be constant at least within an utterance, while the parameter γ is set equal to 0.9. A rapid downfall of F_0, often observed at the end of a sentence and occasionally at a clause boundary, can be regarded as the response of the phrase control mechanism to a negative impulse for resetting the phrase component.

It has been found that the model can generate very close approximations to observed F_0 contours of speech of many languages including English [10], German [11], Greek [12], Japanese [8, 9], Korean [13] and Spanish [14], by assuming only positive commands for the accent components. These findings suggest that the on/off pattern of activeties of CT *pars recta* is solely responsible for the realization of word accent in these languages.

3. F_0 CONTROL IN STANDARD CHINESE

3.1. Languages with Positive and Negative Commands for the Local Components

Analysis of F_0 contours of several languages including Standard Chinese and Swedish, however, indicates that the local components (associated with tones in the case of Standard Chinese) 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 to obtain good approximations to the observed F_0 contours. For example, the F_0 contours of the four tones in Standard Chinese, conventionally classified as High (Tone 1, T1), Rising (Tone 2, T2), Low (Tone 3, T3), and Falling (Tone 4, T4), can be approximated quite well by assuming a positive tone command for T1, a negative one followed by a positive one for T2, a negative one for T3, and a positive one followed by a negative one for T4 within a syllable [15]. Similar switching of polarity is found to occur also in Swedish [16], in which a disyllabic word of Accent 1 is characterized by a positive F_0 excursion for the first syllable followed by a negative excursion for the second syllable. The order of polarity is reversed in a disyllabic word of Accent 2.

Figure 5 illustrates an example of approximation of the F_0 contour of the utterance of Standard Chinese:

> *Mu4 ni2 hei1 buo2 lan3 hui4 bu2 kui4 shi4 dian4 zi3 wan4 hua1 tong3.* (The Munich exposition is really an electronic kaleidoscope.)

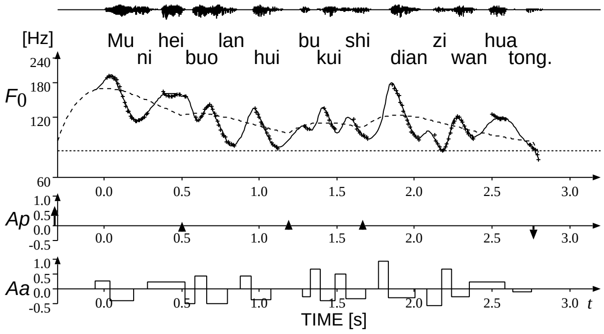


Figure 5: An example of Analysis-by-Synthesis of the F_0 contour of an utterance of Standard Chinese.

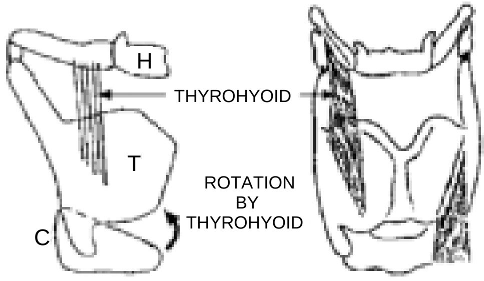
Analysis-by-Synthesis of F_0 contours was conducted on a number of utterances of Standard Chinese, and the results indicate that the mean value of β for negative tone commands is significantly smaller than that for positive tone commands, suggesting that the lowering of F_0 in T2, T3, and T4 is caused by an additional mechanism involving muscles other than CT.

3.2. Roles of Extrinsic Laryngeal Muscles

Although several hypotheses have already been presented on the possible mechanisms for the active lowering of F_0 [17], none seems to be satisfactory since these hypotheses do not take into account the activities of muscles that are directly connected to the thyroid catilage 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 Standard Chinese [18], but the mechanism itself is not 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 [19, 20], one of the present authors suggested the active role of the thyrohyoid (henceforth TH) muscle in F_0 lowering in these languages. Figure 6 shows the relationship between the hyoid bone, thyoid and cricoid cartilages, and TH in their lateral and frontal views, and Fig. 7 shows thier relationships with three other muscles: VOC (thyrovocalis muscle), CT, and SH.



C: cricoid cartilage. T: thyroid cartilage. H: hyoid bone

Figure 6: Role of thyrohyoid in laryngeal control.

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 shown in Fig. 7.

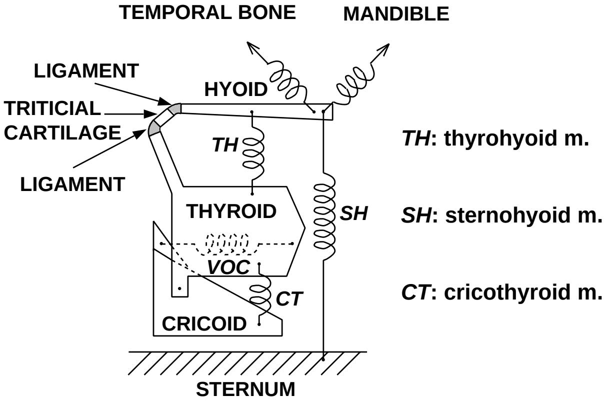


Figure 7: Mechanism of F_0 lowering by activities of TH and SH.

4. SUMMARY

Physiological mechanisms and biomechanical modeling have been described to explain both the origin of the relationship between a change in the vocal cord length and the corresponding change in the logarithm of F_0, and the origin of global and local components of the F_0 contour, firstly for languages such as Common Japanese in which local components are always positive. The roles of extrinsic laryngeal muscles, especially of the thyrohyoid and the sternohyoid muscles, in producing negative local components in languages such as Standard Chinese, have then been explained, on the basis of the detailed physiological structures involving the sternum, the hyoid bone and its connection to the thyriod cartilage.

REFERENCES

[1] Fujisaki, H.: "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.), Raven Press, pp. 347–355 (1988).

[2] Buchthal, F. and Kaiser, E.: "Factors determining tension development in skeletal muscles," *Acta Physiol. Scand.*, vol. 8, pp. 38–74 (1944).

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

[4] Honda, K., Hibi, S., Kiritani, S., Niimi, S. and Hirose, H.: "Measurement of the laryngeal structure during phonation by use of a stereoendoscope," *Ann. Bull. Res. Inst. Logoped. Phoniatr. Univ. Tokyo*, No. 14, pp. 73–78 (1980).

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

[6] Fink, B.R. and Demarest, R.J.: *Laryngeal Biomechanics*, Harvard Univ. Press (1978).

[7] Öhman, S.: "Word and sentence intonation: A quantitative model," *Speech Transmission Laboratory Quarterly Progress and Status Report, KTH*, 2–3/1967, pp. 20–54 (1967).

[8] Fujisaki, H. and Nagashima, S.: "A model for the synthesis of pitch contours of connected speech," *Annual Report of Engineering Research Institute, University of Tokyo*, vol. 28, pp. 53–60 (1969).

[9] Fujisaki, H. and Hirose, K.: "Analysis of voice fundamental frequency contours for declarative sentences of Japanese," *J. Acoust. Soc. Jpn (E)*, vol. 5, no. 4, pp. 233–242 (1984).

[10] Fujisaki, H. and Ohno, S.: "Analysis and modeling of fundamental frequency contours of English utterances," *Proc. EUROSPEECH'95*, vol. 2, pp. 985–988 (1995).

[11] Mixdorff, H. and Fujisaki, H.: "Analysis of voice fundamental frequency contours of German utterances using a quantitative model," *Proc. 1994 Int'l Conf. Spoken Language Processing*, vol. 4, pp. 2231–2234 (1994).

[12] Fujisaki, H., Ohno, S. and Yagi, T.: "Analysis and modeling of fundamental frequency contours of Greek utterances," *Proc. EUROSPEECH'97*, vol. 1, pp. 465–468 (1997).

[13] Fujisaki, H.: "Analysis and modeling of fundamental frequency contours of Korean utterances — A preliminary study —, in *Phonetics and Linguistics — in honour of Prof. H. B. Lee*, pp. 640–657 (1996).

[14] Fujisaki, H., Ohno, S., Nakamura, K., Guirao, M. and Gurlekian, J.: "Analysis of accent and intonation in Spanish based on a quantitative model," *Proc. 1994 Int'l Conf. Spoken Language Processing*, vol. 1, pp. 355–358 (1994).

[15] Fujisaki, H., Hallé P. and Lei, H.: "Application of F_0 contour command-response model to Chinese tones," *Reports of Autumn Meeting, Acoust. Soc. Jpn.*, vol. 1, pp. 197–198 (1987).

[16] Fujisaki, H., Ljungqvist, M. and Murata, H.: "Analysis and modeling of word accent and sentence intonation in Swedish," *Proc. 1993 Int'l Conf. Acoust., Speech, & Signal Processing*, vol. 2, pp. 211–214 (1993).

[17] Honda, K. and Fujimura, O.: "Intrinsic Vowel F_0 and Phrase-Final F_0 Lowering: Phonological vs. Biological Explanations," *Vocal Fold Physiology – Acoustic, Perceptual, and Physiological Aspects of Voice Mechanisms*, Singular Patlishing Group, San Diego (1991).

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

[19] Erickson, D.: *A Physiological Analysis of the Tones of Thai*, Ph. D. Dissertation, University of Connecticut (1976).

[20] Erickson, D.: "Laryngeal muscle activity in connection with Thai tones," *Ann. Bull. Res. Inst. Logoped. Phoniatr. Univ. Tokyo.*, vol. 27, pp. 135–149 (1993).