Analysis and Modeling of Fundamental Frequency Contours of Hindi Utterances

Hiroya Fujisaki[1] *and Sumio Ohno*[2]

[1]Professor Emeritus, University of Tokyo, Tokyo, Japan

fujisaki@alum.mit.edu

[2]School of Computer Science, Tokyo University of Technology, Hachioji, Japan

ohno@cc.teu.ac.jp

10.21437/Interspeech.2005-505

Abstract

This paper describes the results of a preliminary study on the applicability of the command-response model to F_0 contours of spoken Hindi, an official language of India with almost 400 million native speakers in the world. Analysis of observed F_0 contours of a number of utterances by two native speakers indicated that the model with provisions for positive and negative accent commands applies quite well to all the utterances analyzed, and the estimated commands are found to be closely related to the linguistic contents of the utterances. One of the peculiar features of F_0 contours of Hindi is the occurrence of a negative accent command at most phrase-initial positions, often followed by a positive accent command.

1. Introduction

Hindi is one of the major languages belonging to the Indo-European language family. As of 1999, about 366 million people use it as the first language, and 121 million people use it as the second language in the world. It ranks fifth in the number of native speakers after Mandarin, Spanish, English, and Bengali. Being a direct descendant of Sanscrit, it started to emerge as Apabhramsha in the 7th century A.D. Among several dialects of Apabhramsha, Khali Boli was established as the official language by the Mughal empire, and by the end of the 17th century it was often called "Hindi". It remained so during the British rule after the fall of the empire. According to the constitution of India established in 1949, Hindi was accepted as the official language of the Union, while English was regarded as the subsidiary official language. In addition to Hindi and English,, each state can have an official language of its own.

It is thus quite important from the viewpoint not only of basic speech science but also of information technology to elucidate the acoustic-phonetic characteristics of Hindi, and to utilize the findings to establish effective means for synthesis and recognition of speech of Hindi. While most textbooks and papers on phonetics give descriptions of its segmental characteristics, relatively little has been published on the quantitative characteristics of its prosody, especially of its intonation, of which the primary acoustic correlate is the contour of the fundamental frequency of voice (henceforth the F_0 contour). It has been shown by Fujisaki and his coworkers that the command-response model, originally developed for the process of generation of F_0 contours of Common Japanese [1, 2], applies quite well, after certain language-specific modifications, to F_0 contours of many other languages including Chinese, English, Estonian, German, Greek, Korean, Portuguese, Spanish, Swedish, and Thai [3]. The model has also been shown to apply to F_0 contours of Basque [4], French [5], and Italian [6].

The present paper describes the results of a preliminary study conducted by the authors to test the applicability of the command-response model to F_0 contours of utterances of Hindi, as well as to find specific features that will have to be added to the original model, with an aim to utilize it in a TTS system of Hindi.

2. A model for the generation process of F_0 contours of Hindi utterances

Careful observation of F_0 contours of Hindi utterances suggests that the mechanism of laryngeal control for Hindi intonation is essentially the same, at least qualitatively, as that for other languages investigated so far, but requires accent commands of both positive and negative polarities in order to be able to generate the fall-rise patterns found in F_0 contours of many utterance samples [7].

Figure 1 shows the model for the process of generation of F_0 contours of Hindi utterances we propose on the basis of these considerations. The phrase commands are impulses while the accent commands are pedestal functions. These commands are applied to the respective control mechanisms which are assumed to be critically-damped, and which produce phrase and accent components. These components are then added to a constant component ln F_b to produce the final ln $F_0(t)$. For the rest of the paper, we shall use the word 'F_0-contour' to indicate ln $F_0(t)$. Physiological and physical evidences supporting the model were presented elsewhere [3].

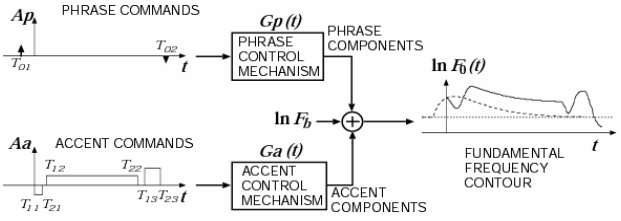

Figure 1: *The command-response model for the process of generating F_0 contours of Hindi utterances.*

Thus the F_0 contour as a function of time can be expressed by the following equations:

$$\ln F_0(t) = \ln F_b + \sum_{i=1}^{I} Ap_i Gp(t - T_{0i}) + \sum_{j=1}^{J} Aa_j \{Ga(t - T_{1j}) - Ga(t - T_{2j})\}, \quad (1)$$

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

$$Ga(t) = \begin{cases} \min[1 - (1+\beta t)\exp(-\beta t), \gamma], & \text{for } t \geq 0, \\ 0, & \text{for } t < 0, \end{cases} \quad (3)$$

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

F_b : baseline value of fundamental frequency,
I : number of phrase commands within the utterance,
J : number of accent commands within the utterance,
Ap_i : magnitude of the *i*th phrase command,
Aa_j : amplitude of the *j*th accent command,
T_{0i} : timing of the *i*th phrase command,
T_{1j} : onset of the *j*th accent command,
T_{2j} : offset of the *j*th accent command,
α : natural angular frequency of the phrase control mechanism, empirically set equal to 3/s,
β : natural angular frequency of the accent control mechanism, empirically set equal to 20/s,
γ : relative ceiling level of accent components, empirically set equal to 0.9.

3. Speech material and method of analysis

3.1. Speech material

The text for the speech material was taken from the book "Practical Conversations in Hindi" by H. Ishida, Daigaku Shorin Publishing Co., Tokyo (1993). It is a conversation on the events taking place on the birthday of Mahatma Gandhi, and consists of 16 sentences.

A part of the speech material was recorded at the Tata Infotech, Limited, in Mumbai, India, while another part was taken from the pre-recorded cassette tape appended to the book. Table 1 lists the texts and their English translations for two sentences, while Figure 2 shows the syntactic structures of the two Hindi sentences (along with their rough phonetic transcriptions). The speakers are one male (AS) and one female (nk) native speakers of Hindi. The average speech rates for the two speakers are 4.0 and 3.6 syllables per second, respectively.

3.2. Analysis procedure

The speech signal was digitized at 10kHz with 16bit precision. The fundamental frequency was extracted at 10 ms intervals by the modified autocorrelation analysis of the LPC residual. The measured F_0 contour was aligned with the speech waveform by visual inspection of the waveform and its sound spectrogram. Using the model's formulation described in Section 2, it is possible to extract the model parameters from an observed F_0 contour by the method of Analysis-by-Synthesis.

Table 1: *Two examples of Hindi sentences used for the current study – excerpts from a conversation on Mahatma Gandhi's birthday.*

	Hindi sentences	English translation
A	गाँधी जी की समाधि पर फूल चढ़ाने के लिए लोग राजघाट जाते हैं।	People go to Rajghat to offer flowers at Sir Gandhi's tomb.
B	लोग यहाँ आते हैं और महात्मा जी की सिखाई बातों पर अमल करने की कसम दोहराते हैं।	People come here and renew their pledge to put Sir Mahatma's teachings into practice.

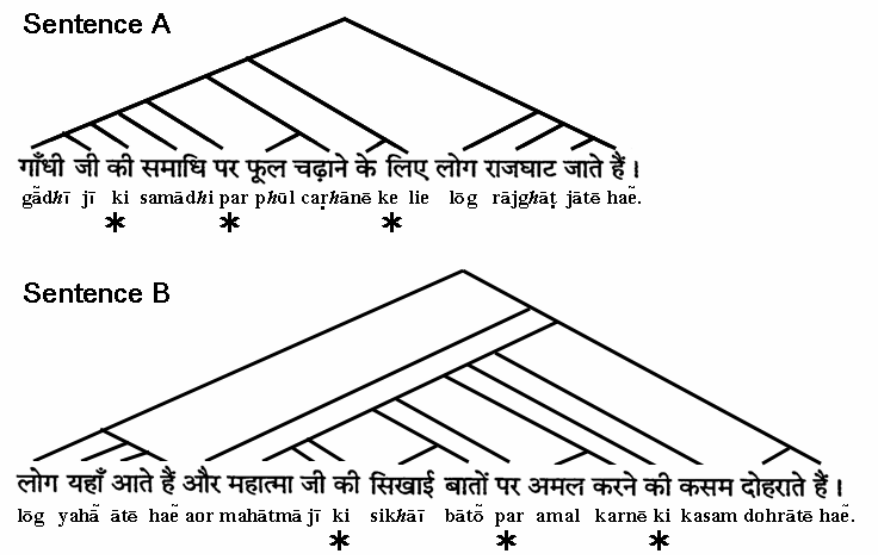

Figure 2: *Syntactic structures of the two Hindi sentences shown in Table 1.* * *indicates a postposition.*

4. Experimental results

Panels (a) and (b) in Figure 3 show the results of analysis of one sample each of the utterances of Sentence A by the male (AS) and the female (nk) speaker, while panels (c) and (d) show those of Sentence B. Each panel shows, from top to bottom, the speech waveform, measured F_0 values (+symbols), the model-generated best approximation (solid line), the baseline frequency F_b (dotted line), the phrase commands (impulses), and the accent 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 accent components.

The results shown in these panels as well as the results of analysis of all other utterances in the current speech material indicate that

(1) The F_0 contours of Hindi utterances can be approximated quite well by the command-response model, if we allow both positive and negative polarities for the accent commands, and also assume the same set of values for the parameters α, β, and γ as we adopted for speakers of other languages.

Furthermore, comparison of the results of Fig. 3 with the text of the sentences and their syntactic structures indicates that

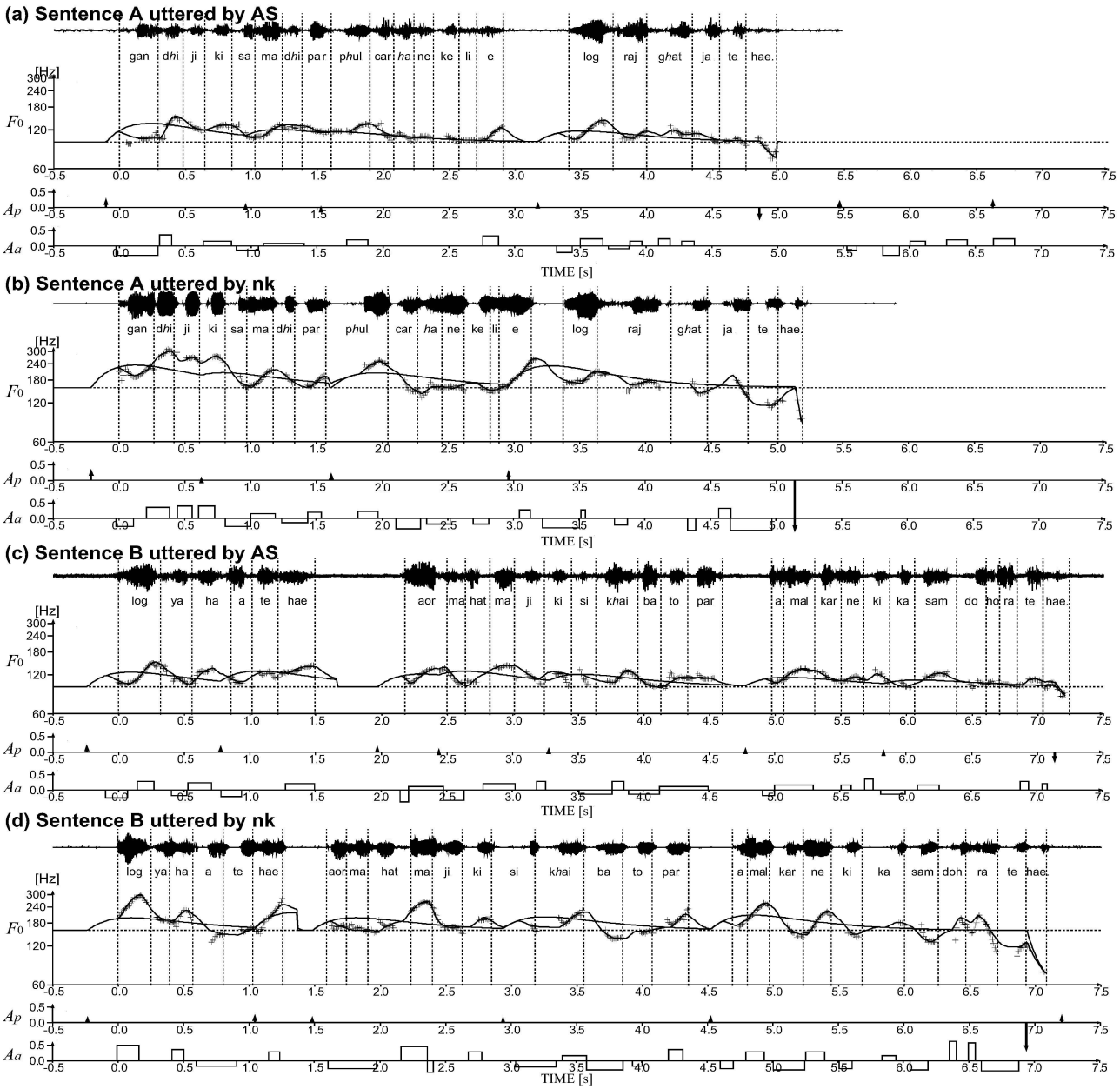

Figure 3: *Results of Analysis-by-Synthesis of F_0 contours of the two Hindi utterances by two native speakers. (a) An utterance of sentence A by the male speaker (AS), (b) An utterance of sentence A by the female speaker (nk), (c) An utterance of sentence B by the male speaker (AS), (d) An utterance of sentence B by the female speaker (nk).*

(2) The phrase commands tend to occur at syntactic boundaries such as at or near the ends of postpositions, and at or near the beginning of conjunctions. Their occurrence, however, is not deterministic. Certain variability exists between speakers.

(3) Negative accent commands tend to occur mostly at the onsets of prosodic phrases and clauses which always coincide with the beginning of words, but less often also at word-medial syllables. A negative accent command is often immediately followed by a positive accent command, producing a fall-rise pattern of subjective pitch peculiar to Hindi utterances.

(4) As in most other spoken languages, a prosodic phrase that is not final within an utterance is marked by the so-called 'continuation rise' which is caused either by a positive accent command only, or by a combination of a positive accent command and a phrase command for the next part of the utterance.

(5) In addition, the utterances of declarative sentences of the female speaker (nk) possess negative accent commands at utterance-final positions, but it is not so in the utterances of the male speaker (AS). Thus the use of negative accent command at utterance-final positions is found to be more or less speaker-dependent.

To the best of the knowledge of the present authors, there is no mention of this peculiar fall-rise pitch pattern in books and papers on phonetics of Hindi. Informal observations by a few native speakers, however, suggest that it indicates emphasis, often associated with a feeling of respect. The exact factors that contribute to its occurrence, however, certainly require further investigation.

Table 2 shows the percentage of occurrences of negative accent commands at various positions for each of the two speakers.

Table 2: *Percentage of occurrences of negative accent commands at various positions within an utterance*.

Speaker	Position			
	Onset of prosodic clause	Onset of prosodic phrase	Utterance-final	Other
AS	38.9	27.8	5.5	27.8
nk	12.9	19.4	12.9	54.8

5. Discussion and summary

It has already been shown by Fujisaki [3] that spoken languages thus far investigated fall into two groups from the point of view of polarity of the local commands (for tone or accent) in the generation of F_0 contours: (1) those using only positive commands, and (2) those using both positive and negative commands. The latter is further subdivided into two: (2a) those in which the occurrence of a negative command is determined by lexical information, including both tone languages like Standard Chinese and Thai and non-tone languages like Swedish, and (2b) those in which the occurrence of a negative command is determined, not by lexical information, but by other linguistic and paralinguistic factors, like Portuguese.

Due to the preliminary nature of the current study, the speech material we analyzed so far is quite limited in size. Our results, however, clearly indicate that spoken Hindi belongs to the subgroup (2b). Work is already under way to further analyze and elucidate the factors that contribute to the occurrence of negative commands as well as to examine the role of such negative commands in the perception of spoken Hindi.

The current work was partially supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Science and Technology of Japan (No. 12132102, principal investigator: H. Fujisaki). The authors would like to acknowledge Dr. P.V.S. Rao and his colleagues at the Tata Infotech Limited in Mumbai, India for their warm encouragement and kind cooperation.

6. References

[1] Fujisaki, H. and Nagashima, S., "A model for the synthesis of pitch contours of connected speech," Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, Vol. 28, pp. 53-60, 1969.

[2] Fujisaki, H. and Hirose, K., "Modeling the dynamic characteristics of voice fundamental frequency with application to analysis and synthesis of intonation," Preprints of Papers for the Working Group on Intonation, the 13th International Congress of Linguists, H. Fujisaki and E. Gårding, Editors, Tokyo, pp. 57-70, 1982.

[3] Fujisaki, H., "Information, Prosody, and Modeling ― with Emphasis on Tonal Features of Speech ―," Proceedings of Speech Prosody 2004, Nara, pp. 1-10, 2004.

[4] Navas, E. Hernáez, I. and Sánchez, J. M., "Basque intonation modeling for text to speech conversion," Proceedings of International Conference on Spoken Language Processing 2002, Denver, Vol. 4, pp. 2409-2412, 2002.

[5] Bailly, G., "Contribution à la determination automatique de la prosodie du Français parlé à partir d'une analyse syntaxique. Etablissement d'un modèle de génération, PhD thesis, Institut National Polytechnique, Grenoble, 1983.

[6] Rossi, P. S., Palmieri, F. and Cutugno, F., "A method for automatic extraction of Fujisaki-model parameters," Proceedings of Speech Prosody 2002, Aix, pp. 615-618, 2002.

[7] Fujisaki, H. and Ohno, S., "Modeling the generation process of fundamental frequency contours of Hindi utterances," Reports of Fall Meeting, the Acoustical Society of Japan, Vol. 1, pp. 217-218, 2003.