ANALYSIS AND INTERPRETATION OF FUNDAMENTAL FREQUENCY CONTOURS OF BRITISH ENGLISH IN TERMS OF A COMMAND-RESPONSE MODEL

Hiroya Fujisaki   *Sumio Ohno*   *Takashi Yagi*   *Takeshi Ono*

Science University of Tokyo
2641 Yamazaki, Noda, 278-8510, Japan

ABSTRACT

In order to test the validity of authors' command-response model for the generation of F_0 contours in the analysis and interpretation of F_0 contours of British English, F_0 contours of utterances containing sections that are not usually found in those of the common Japanese are analyzed. The results indicate that these F_0 contours can actually be generated by selecting appropriate patterns of commands for phrase and accent in a way that is specific to British English, while using the same model for the mechanism of F_0 control as for the common Japanese.

1. INTRODUCTION

In many languages of the world, the contour of the fundamental frequency (henceforth F_0 contour) of speech play an important role in conveying information on lexical accents of words, syntactic structures of sentences, and foci of discourses. The relationship between the F_0 contour and the underlying information can be elucidated if one has an objective, quantitative model of the process whereby the F_0 contour is generated. Such a model has been shown by Fujisaki and his coworkers, initially for F_0 contours of the common Japanese [1,2]. It models both the input commands that provide linguistic information and the mechanism which generates the F_0 contour as the response. The model has since been extended to apply to F_0 contours of a number of languages including American English, Chinese, German, Korean, Spanish and Swedish [3–8], indicating that the model of the mechanism is quite similar across speakers of various languages tested, but the patterns

of input commands are naturally language-specific. It was argued, however, that the model would fail to generate some F_0 contour sections commonly found in utterances of British English [9]. The purpose of the present study is to look into the applicability of the model to these cases and to demonstrate the generality of our approach.

2. THE BASIC MODEL AND LANGUAGE-SPECIFIC MODIFICATIONS

It is widely accepted that F_0 contours of many languages are characterized by relatively slow undulations (henceforth global components) which roughly correspond to larger phrases, clauses, and sentences, and by relatively fast rise/fall patterns (henceforth local components) which correspond to either lexical tones of syllables (in tone languages such as Chinese) or lexical accent of words (in non-tone languages such as Japanese and English).

A close examination of the F_0 contour of the common Japanese, plotted on the logarithmic scale of frequency as a function of time, suggests that the global component has a shape similar to the impulse response of a second-order linear system (i.e., a mass-viscosity-stiffness system), and the local component has a shape similar to the step response of another second-order linear system with a shorter time constant than the former one. These considerations have led to a model for the generation mechanism of the F_0 contour shown in Fig. 1.

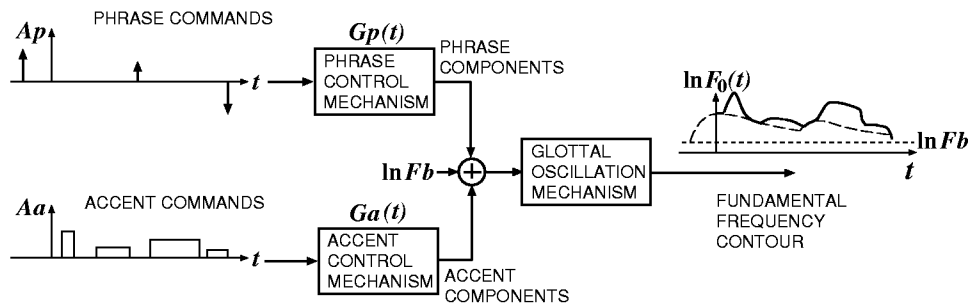In this model, the phrase commands are assumed to be impulses applied to the phrase control mechanism to gen-

Figure 1: A command-response model for F_0 contour generation for the common Japanese.

erate the phrase components, while the accent commands are assumed to be positive stepwise functions applied to the accent control mechanism to generate the accent components. Both mechanisms are assumed to be critically damped second-order linear systems, and the sum of their outputs, i.e., the phrase components and accent components, is superposed on a baseline value ($\ln Fb$) to form an F_0 contour, as given by the following equation:

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

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

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

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

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 within an utterance, while the parameter γ is set equal to 0.9.

By the method of Analysis-by-Synthesis, it is possible to determine the number, magnitude and timing of the phrase and accent commands that will generate an F_0 contour which is the closest approximation to that of an actual utterance, say in the sense of the least mean squared error in the logarithmic scale of F_0.

There are languages, however, which call for negative local components. For example, the four tones in the Standard Chinese are conventionally classified as High (Tone 1, T1), Rising (Tone 2, T2), Low (Tone 3, T3), and Falling (Tone 4, T4). A close examination of the F_0 contours of these tone types indicates that the local component corresponding to T1 is always positive, while those corresponding to the three other tones are partially (T2 and T4) or entirely (T3) negative. This suggests that, instead of word accent commands of only positive polarity used for F_0 contours of the common Japanese, we should assume tone commands that can have both positive and negative polarities. In fact, the best approximation can be obtained by assuming a positive 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 [4].

The example of tone commands in Chinese illustrates that what constitutes language specificity of intonation is mainly the constraints on the patterning (including polarity) of the input commands, while the model of the control mechanism is more or less similar across speakers of different languages, as it was already shown by our studies of intonation of languages other than Japanese.

However, it was argued that the model lacks the capability to represent certain F_0 contour characteristics found in British English [9]. Namely, it was pointed out that the current model has two limitations in that (1) it uses a fixed time constant for the accent component and thus cannot account for the variations in the gradient of rises and falls, and (2) it cannot model long continuously rising sections of F_0 contour. Whereas the former point (1) is not true since out model actually can deal with variations in the gradient of rises and falls, there are certain features in F_0 contours of British English intonation, including the latter point (2), which, at first sight, may look to be impossible to be represented by our model. We will therefore show that these features can actually be realized within the framework of our original model, simply by modifying the patterning of the input commands for phrase and accent.

3. SPEECH MATERIAL AND METHOD OF ANALYSIS

The speech material adopted for the current study consisted of 529 conversational utterances of British English uttered by 31 native speakers (16 males and 15 females), taken from the recorded tapes of Linguaphone's Intermediate (British) English Course. The sppech signal was digitized at 10 kHz with 16 bit precision, and the fundamental frequency was extracted at 10 ms intervals by a modified autocorrelation analysis of the LPC residual.

Preliminary analysis and inspection of F_0 contours of all the utterances indicated that about 87 % of the total F_0 contours showed patterns that are quite similar to those of Japanese and could be generated by the same model as for Japanese, but the remaining 13 % contained sections with features that are not common to F_0 contours of Japanese.

4. RESULTS OF ANALYSIS

Figure 2 shows examples of F_0 contours containing sections that are not usually found in F_0 contours of the common Japanese and their analysis results. Each panel displays, from top to bottom, the speech waveform, approximate positions of some of the word boundaries (vertical dotted lines), measured F_0 values ($+$ symbols), the best approx-imation by the model (solid line), the baseline frequency (horizontal dotted line), the phrase commands (impulses), and the accent commands (stepwise functions). The broken lines indicate the phrase components wherever they differ from the model's approximation, and the differences be-tween the solid line and the broken lines indicate the accent components.
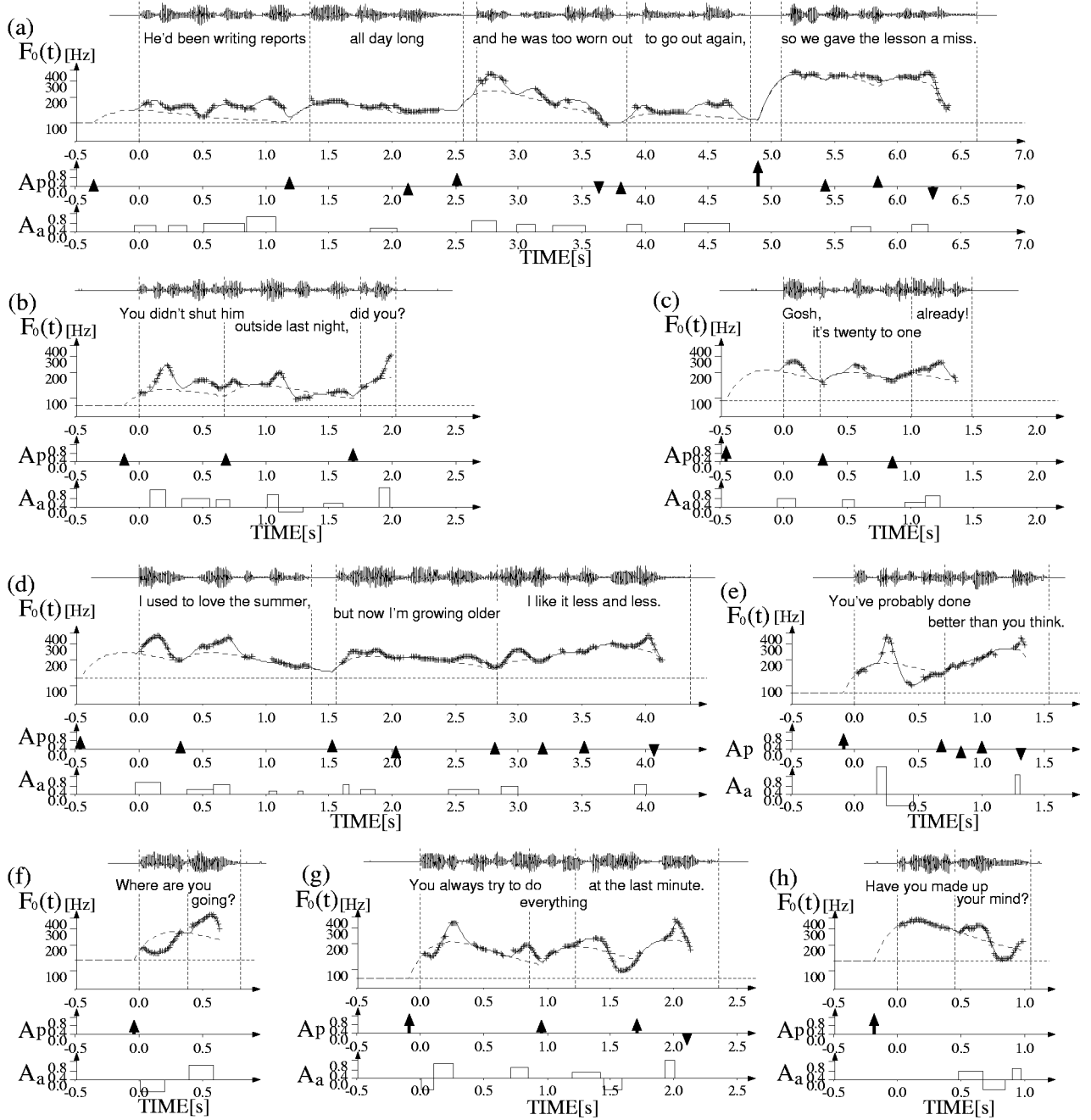

Figure 2: Examples of F_0 contours containing sections that are not usually found in F_0 contours of the common Japanese and their analysis results.

These sections can be classified into the following five types:

Type 1: Nearly flat sections at relatively low F_0 values (*i.e.,* close to the baseline frequency Fb, such as the initial part (up to 2.3 sec) of the F_0 contour of panel (a).

Type 2: Nearly flat sections at relatively high F_0 values, such as the final part (beyond 5.0 sec) of the F_0 contour of panel (a).

Type 3: Relatively short rising sections, usually at the end of an F_0 contour, such as the final positions of the F_0 contours shown in panels (b), (c), and (d).

Type 4: Relatively long, continuously rising sections, such as the middle and final parts of the F_0 contour of panel (e).

Type 5: Local dips, such as those found in F_0 contours of panels (f), (g), (h), as well as in (b) and (e).

As shown by the model-generated F_0 contours of Fig. 2, however, F_0 contours containing these sections can be very closely approximated by selecting appropriate timing, magnitude and polarity for the phrase or accent commands and applying them to the same mechanism as that of the common Japanese. Namely, Type 1 section is generated when the phrase command is very small, Type 2 section is generated when a large phrase command is followed by a succession of relatively small phrase commands, often followed by an utterance-final accent commands, Type 3 section is generated by an utterance-final phrase component and accent component, Type 4 section is generated by a succession of phrase commands whose magnitudes are not very small (often followed by an utterance-final accent command), and Type 5 section is generated by a negative accent command occurring at an accentable syllable.

5. DISCUSSION AND CONCLUSION

Our analysis has shown that all the F_0 contours of 70 British English utterances with features that are not usually found in F_0 contours of the common Japanese can actually be very well approximated by the appropriate selection of timing and magnitude of the phrase commands and the polarity of the accent commands, without modifying the model of the mechanisms of F_0 control. Namely, sections of Type 1, 2, 3 and 4 are generated by using specific patterns for the timing and magnitude for the phrase and the accent commands.

For instance, the interrogative intonation for the English expression "did you?" in panel (b) is generated by an utterance-final phrase command accompanied by an accent command, while the interrogative intonation in the common Japanese requires only an utterance-final accent command. However, the gradually rising intonation such as in panel (e) can also be found in certain dialects of Japanese, though it is not found in the common Japanese. Likewise, the low flat intonation can be found also in Japanese in the case

of monologue. The use of negative accent commands is found in some dialects of Japanese as well as in Swedish. The laryngeal mechanism for such an active lowering of F_0 may involve muscles other than the cricothyroid such as the sternohyoid and the thyrohyoid [10], which are usually not used in the common Japanese.

These results and considerations lead us to believe that the dynamic characteristics of the laryngeal mechanism, as expressed by Eqs. (1), (2) and (3) are essentially similar in speakers of various languages, but the patterns of control commands are more or less specific to individual languages and dialects.

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. 52–60, 1969.

2. Fujisaki, H. and Sudo, H. "A model for the generation of fundamental frequency contours of Japanese word accent," *Journal of the Acoustical Society of Japan*, Vol. 27, pp. 445–453, 1971.

3. Fujisaki, H., Sugito, M. and Hirose, K. "Word accent and sentence intonation in foreign language learning," *Preprints of Papers, Working Group on Intonation, The XIIIth International Congress of Linguists, Tokyo*, pp. 109–119, 1982.

4. Fujisaki, H., Hirose, K., Hallé, P. and Lei, H. "Analysis and modeling of tonal features in polysyllabic words and sentences of the Standard Chinese," *Proceedings of 1990 International Conference on Spoken Language Processing*, Vol. 2, pp. 841–844, 1990.

5. Mixdorff, H. and Fujisaki, H. "Analysis of voice fundamental frequency contours of German utterances using a quantitative model," *Proceedings of 1994 IntInternational Conference on Spoken Language Processing*, Vol. 4, pp. 2231–2234, 1994.

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

7. Fujisaki, H., Ohno, S., Nakamura, K., Guirao, M. and Gurlekian, J. "Analysis of accent and intonation in Spanish based on a quantitative model," *Proceedings of 1994 IntInternational Conference on Spoken Language Processing*, Vol. 1, pp. 355–358, 1994.

8. Fujisaki, H., Ljungqvist, M. and Murata, H. "Analysis and modeling of word accent and sentence intonation in Swedish," *Proceedings of 1993 International Conference on Acoustics, Speech, and Signal Processing*, Vol. 2, pp. 211–214, 1993.

9. Taylor, P. "The rise/fall/connection model of intonation," *Speech Communication*, Vol. 15, pp. 169–186, 1994.

10. Fujisaki, H. "Physiological and physical mechanisms for tone, accent and intonation," *Proceedings of the XXIII World Congress of the International Association of Logopedics and Phoniatrics*, pp. 156–159, 1995.