Title	歌声におけるF0動的変動成分の抽出とF0制御モデル
Author(s)	齋藤，毅
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1570
Rights	
Description	Supervisor:赤木 正人，情報科学研究科，修士

EXTRACTION OF F0 DYNAMIC CHARACTERISTICS AND DEVELOPMENT OF F0 CONTROL MODEL IN SINGING VOICE

Takeshi Saitou (010046)

School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 2002

Keywords: singing voice perception, singing voice synthesis, fundamental frequency (F0), F0 dynamic characteristics, F0 control model.

1 INTRODUCTION

Singing voice has more dynamic and complicated characteristics than speaking voice. The characteristics are contained in the fundamental frequency (F0), especially. Therefore, F0 contours are important factors in order to synthesize high-quality singing voices. It becomes necessary to clarify F0 characteristics and to develop a method to control F0 contours.

In the previous studies, it is reported that some F0 dynamic characteristics are peculiar to singing voice. However, the reports did not mention the effects of F0 dynamic characteristics on singing voice perception. Although F0 control models dealing with speaking voices have been proposed, they can not control F0s of singing voices.

In this paper, we extract F0 dynamic characteristics in singing voices, and demonstrated how much the F0 dynamic characteristics influence on singing voice perception, through psychoacoustic experiments. In addition, we propose a F0 control model that can control F0 dynamic characteristics in singing voice.

2 ANALYSIS OF F0 DYNAMIC CHARACTERISTICS OF SINGING VOICE

Previous studies reported that there are three characteristics in F0 contours of singing voices [1].

1. Melody and note.
2. Stability of F0 in one note.
3. Dynamic characteristics which are observed in singing voice only.

Numbers 1 and 2 are characteristics related to melody. In this paper, we examined F0 dynamic characteristics mentioned in No.3.

Singing voice data

The singing voice data for our studies were obtained from recordings of three adults singing a Japanese children's song "Nanatsunoko". The singers were asked to sing with Japanese vowel /a/ only, to simplify the experimental conditions. The songs were recorded on a DAT with 48-kHz sampling and 16-bit accuracy.

F0 estimation method

F0 contours with dynamic characteristics were estimated using an F0 extraction method, TEMPO in STRAIGHT[2]. The TEMPO can accurately extract dynamic characteristics in F0 contours.

2.1 Extraction of F0 dynamic characteristics

Figure 1 shows an estimated F0 along the logarithmic axis. We extracted three F0 dynamic characteristics as follows.

Overshoot,Undershoot : The larger deflection than the value of the target note at the time of pitch change.

Vibrato • Fine-fluctuation : Vibrato is a periodic frequency modulation (4∼7 Hz) observed when a singer sings one note. Fine-fluctuation is irregularly rapid fluctuation over 10Hz frequency modulation, which remains after removing Vibrato.

Preparation : Sudden vibration of the reverse direction of pitch change. It is observed just before pitch change.

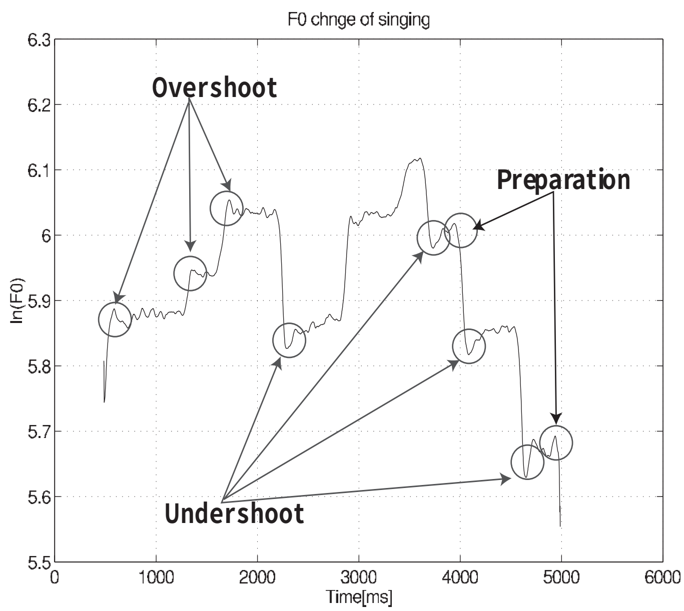

Figure 1: The F0 extracted by TEMPO

It is reported that Overshoot, Undershoot, Vibrato and Fine-fluctuation are the characteristics peculiar to singing voices [1, 3, 4]. In this paper, we extracted Preparation as a new characteristic. We carried out psychoacoustic experiments to demonstrate how much the F0 dynamic characteristics influence on singing voice perception.

2.2 Effect of F0 dynamic characteristics on singing voice perception

In order to investigate effects of F0 dynamic characteristics on singing voice perception, we performed the following psychoacoustic experiment.

Psychoacoustic experiment 1

A. Stimuli

We eliminated each F0 dynamic characteristic from F0 contours and re-synthesized singing voices using the modified F0s by the Klatt formant synthesizer as follows.

- The singing voice synthesized using the extracted F0 from a real song. (NORMAL)
- The singing voice removed Overshoot and Undershoot components. (NO-OUS)
- The singing voice removed Vibrato and Fine fluctuation components. (NO-VIB)
- The singing voice removed Preparation components. (NO-PRE)
- The singing voice whose F0 is smoothed by an FIR low-pass filter (cut-off frequency is 5Hz). Dynamic characteristics components in the song are removed from the F0, then melody is distorted. (SMS)

B. Procedure

The experiment adopted the Scheffe's method of paired comparison (Evaluation measures are seven-grade: -3 to 3). It was carried out in the sound-proof chamber. Subjects are six graduate students with normal hearing ability.

C. Result and Discussion

Figure 2 shows the result of the experiment. The numbers below the horizontal axis express naturalness as singing voice. The result indicates that the effects of three F0 dynamic characteristics on singing voice perception are large. Moreover, the effect of Overshoot and Undershoot is the largest.

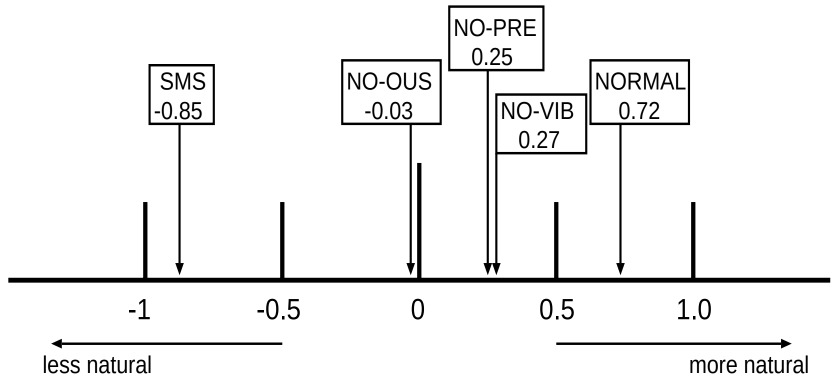

Figure 2: The relation of the naturalness of stimulus 1

3 THE F0 CONTROL MODEL FOR SINGING VOICE

For F0 control models [5, 6] for speaking voices, it is difficult to control and generate F0 contours including dynamic characteristics of singing voices. In this paper, we develop a new method which can control F0 contours of singing voices.

3.1 Schematic graph of F0 control model

Figure 3 shows a schematic graph of an F0 control model. This model can control F0 dynamic characteristics and generate F0 contours using five components, and can generate F0 contours.

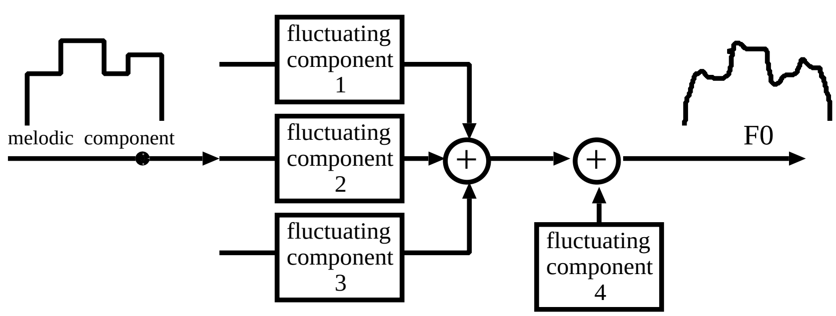

Figure 3: The schematic graph of F0 control model

melodic component : The external component of melody, described by sum of step functions. (Input component of the system)

fluctuating component 1 : The component controlling Overshoot and Undershoot, expressed with a 2nd order damping model.

fluctuating component 2 : The component controlling Vibrato, expressed with a 2nd order oscillation model (no-loss).

fluctuating component 3 : The component controlling Preparation, expressed with a 2nd order damping model.

fluctuating component 4 : The components controlling Fine-fluctuation. This is expressed with irregularity rapid oscillation which the amplitude of about 5Hz at the maximum.

The transfer function of 2nd order system is given by the following formula.

$$H(s) = \frac{\Omega}{s^2 + 2\zeta\Omega s + \Omega^2}$$

This F0 control model can control and generate F0 contours which includes each dynamic characteristic, by determining the optimal control parameters for ζ and Ω. An F0 contours generated by the model is shown in Fig. 4.

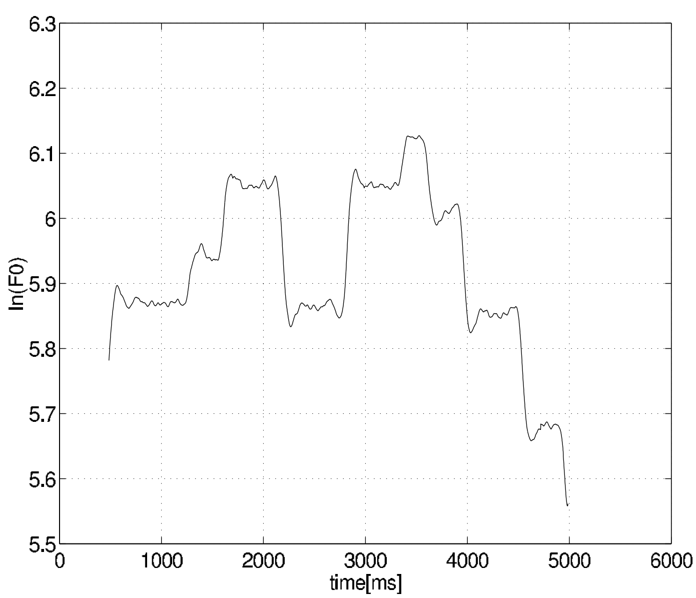

Figure 4: The F0 generated by the F0 cotrol model

3.2 Singing voice synthesis

We synthesized singing voices by STRAIGHT using generated F0 contours and performed the following psychoacoustic experiment in order to evaluate availability of the F0 control model and the generated F0 contours.

Psychoacoustic experiment 2

A. Stimuli

The six stimuli were used in the experiment as follows.

NORMAL : the synthesized song using a real F0.

SYN-All : the synthesized song using all F0 dynamic characteristics.

SYN-OUS : the synthesized song using Overshoot and Undershoot.

SYN-PRE : the synthesized song using Preparation.

SYN-VB : the synthesized song using Vibrato and Fine fluctuation.

SYN-BASE : the synthesized song using a melodic component only.

The spectrum of synthesized singing voices was the same in all stimuli. The psychoacoustic experiment was carried out on the same procedure and conditions as those of the last experiment.

B. Results and Discussion

The result is shown in Fig. 5. The result show that F0 control model can produce as natural singing voices as NORMAL. This indicats that all of F0 dynammic characteristics are important for F0 contour generation.

4 CONCLUSION

In this paper, we extracted some F0 dynamic characteristics in singing voice and demonstrated how much the F0 dynamic characteristics influence on singing voice perception, through psychoacoustic experiments. The results

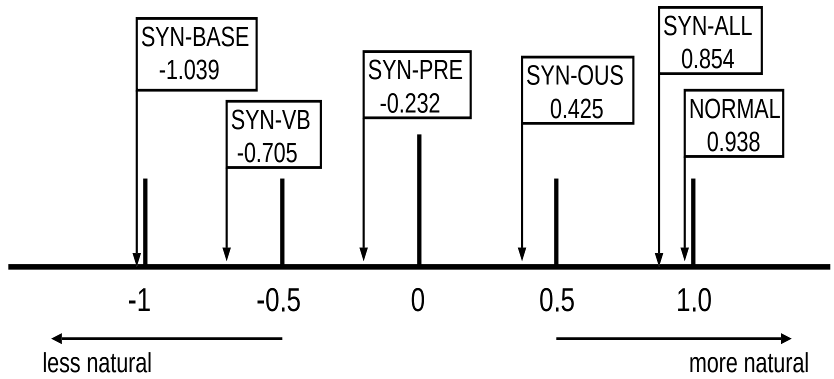

Figure 5: The relation of the naturalness of stimulus 2

show that F0 dynamic characteristics, especially Overshoot, Undershoot, Vibrato, Fine-fluctuation, and Preparation affect singing voice perception.

We also proposed an F0 control model that can control F0 dynamic characteristics and can generate F0 contours. The F0 control model can produce natural singing voices by controlling the F0 dynamic characteristics.

References

[1] M. Yatabe and H. Kasuya, "Dynamic characteristics of fundamental frequency in singing" Meeting of the acoustical society of Japan, 3-8-6, 1998.

[2] H. Kawahara *et al.*, "Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity" Proc. Eurospeech99, pp.2781-2784, 1999.

[3] W. Odagiri and H. Kasuya, "Study of analysis, synthesis and perception of vocal vibrato" Meeting of the acoustical society of Japan, 1-7-5, 1999.

[4] H. Kitakaze, "A research on the discrimination of fine fluctuation in fundamental frequency contained in the singing voice" JAIST master thesis, 2000.

[5] H. Fujisaki and M. Tatsumi, "Analysis control in singing" "Vocal fold physiology", UNIVERSITY OF TOKYO PRESS, pp.347-363, 1981.

[6] T. Moriyama, H.Ogawa and S.Tenpaku, "A new control model based on rising and falling fundamental frequency" Proc. of ASA and ASJ Third Joint Meeting, pp.1171-1176, 1996.