A research on the discrimination of fine fluctuation in fundamental frequency contained in the singing voice

Hironori Kitakaze

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

February 15, 2000

Keywords: fundamental frequency, fine fluctuation, modulation frequency, deviation, singing voice.

1 Introduction

Synthesis of high quality singing voice is aimed by many researchers. Parameter controls to the most suitable values become necessary to synthesize the quality singing voice. Dynamics of fundamental frequency (f_0) is one of important parameters. Dynamics of f_0 is divided into three components. One is a component of a melody and another is a component of regular fluctuation like vibrato. The third component is a fine fluctuation component, that other two components were removed from. The components of a melody and a regular fluctuation have been examined by many researchers [1]. However, fine fluctuation components were not studied, because fine fluctuation components were believed not to be important. Recently, it was found that fine fluctuation components influence on naturalness of singing voices. Then, fine fluctuation components came to be taken seriously [2].

This paper examines the fine fluctuation component with one for the main subject. One hypothesis is made from results that became distinct here. "Fine fluctuation components of f_0 contained in singing voices influence on perception of singing voices, and magnitude of influences depends on modulation frequency and deviation width of f_0." The hypothesis will be verified by auditory experiments.

2 Analysis of fine fluctuation component of f_0

Fine fluctuation components of f_0 were extracted from singing voices. Modulation frequencies and deviation widths of fine fluctuation components are analyzed. Furthermore, some factors involved in perception are examined. The analyzed results, lead to hypothesis about perception of fine fluctuation component.

2.1 Singing voice data

Singing voices are collected from students (Two males, three females) who are going to a vocal school or an actor training place. The song is a Japanese children's song "NANATUNOKO" of the public vocal music tune with Japanese vowel /a/ to simplify an experiment. Recorded singing voices were partitioned into 4 bars. These singing voices were used for analysis of fine fluctuation components.

2.2 Extraction of f_0

$F_0 s$ must be extracted with high precision to analyse fine fluctuation components accurately. $TEMPO2$ proposed by Kawahara [3] is used for f_0 extraction. The modulation frequencies can be extracted by TEMPO2 with high precision as up to 1/5 f_0. An example of an analyzed wave is shown in Figure 1.

2.3 Extraction of fine fluctuation component

A FIR low-pass filter with 201 taps is applied to f_0s to estimate melody components. The cut-off frequency is 5Hz. An example of an f_0 and a melody component is shown in Figure 2. The fine fluctuation component can be extracted by subtracting the melody component from the f_0 (Figure 3). Cent is used as unit.

2.4 Analysis of fine fluctuation component

Modulation frequencies and deviation widths of fine fluctuation components are analyzed. The modulation frequencies are analyzed using running spectra (4), and the deviation widths are analyzed using anplitudee histograms of fine fluctuation conponents (Fig 5, sampling period:1ms, bin width:1cent).

2.5 Result and discussion (establishment of the hypothesis)

The analyzed results of the modulation frequencies are as follows.

- When f_0 fluctuates lergely, it became clear that frequency element of $20Hz$ was contained in the modulation frequency. When fluctuation was small in the f_0, it became clear that the frequency component of $15Hz$ was contained in the modulation frequency.

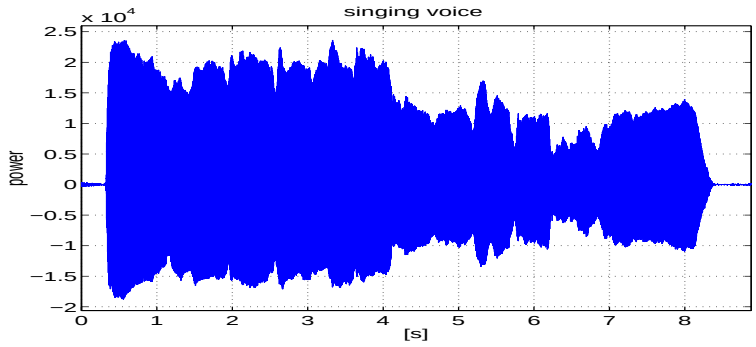

Figure 1: The example of the analysis wave shape

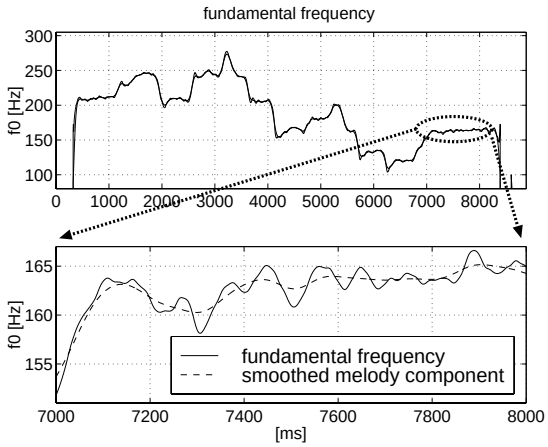

Figure 2: f_0 and The contour component of melody

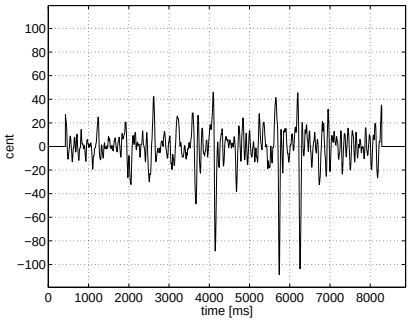

Figure 3: fine fluctuation component of the f_0 that it was extracted

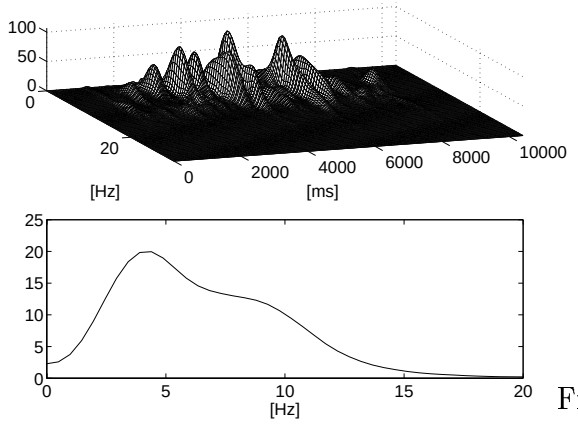

Figure 4: spectrum of fine fluctuation component

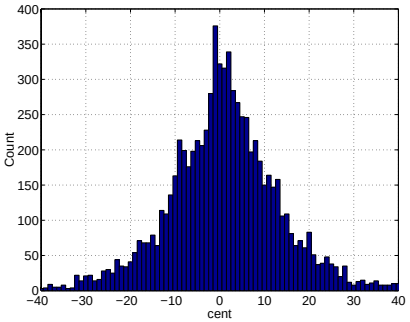

Figure 5: histogram of fine fluctuation component(sampling period:1ms, bin width:1cent)

- As for the f_0, it was found that a rapid fluctuation was not caused.

The analyzed results of the deviation widths are as follows.

- The average of standard deviations of the deviation widths is $\pm 20 cent$.
  Since a chromatic scale of is $100 cent$, this fluctuates as much as 1/5 of a half tone musical scale.

- The deviation width goes up to $\pm 60 cent$ when f_0 changes rapidly.

From these findinfs, humans may have a potential for perceiving "modulation frequency and deviation width" of fine fluctuation components in the singing voices. Thus, we set up the following hypothesis. "Fine fluctuation components of f_0 contained in singing voices influence on perception of singing voices, and magnitude of influences depend on modulation frequency and deviation width of f_0."

3 Influence on perception of fine fluctuation component

3.1 Preliminary Experiment

3.1.1 Method

A problem that " humans can perceive fine fluctuation components" is confirmed. This experiment is to examine whether difference in "the voice which a fine fluctuation component exists in", and "the voice which a fine fluctuation component doesn't exist in" can be perceived. If it can be confirmed, it is understood of fine fluctuation component influences on perception of singing voices.

3.1.2 Result

Cognition rate was 78.8% as results of the experiment. Therefore, the fine fluctuation components influence on perception. Then the following experiments are meaningful.

3.2 Experiment1

3.2.1 Method

Detection threshold for existence of f_0 modulation is confirmed in Experiment 1. This experiment is basis to verify the hypothesis. Therefore, detection threshold is examined using some synthesized voices. The dependence of modulation frequency and the deviation width are examined too. If it can be confirmed, influence of fine fluctuation components on perception of the singing voice can be shown quantitatively.

3.2.2 Result

Synthesized voices were used for the experiment. Results of Experiment 1 are the following. The stimuli whose $f_0 is 125Hz$ and deviation width is 1% can be perceived when modulation frequency is up to $20Hz$. The stimuli whose $f_0 is 250Hz$ and deviation width is 1% can be perceived when modulation frequency is up to $60Hz$. When the deviation width increases, detection threshold increase.

These results and the analysis result of the singing voices are examined. It was clear from these results that the fine fluctuation component of f_0 contained in the singing voices was perceived. We examine conversely whether deviation width of what percent is necessary to perceive difference of the singing voice whose modulation frequency exists at $15Hz$. It was found that the deviation width is 1.1% from the result of the experiment 1. It was cear from deviation widths having existed on the average in the 1.2% in the analysis of the singing voice to the fluctuation component was perceived. Actually, we examine in the prelimimary experiment to the fine fluctuation component was perceived. If modulation frequency changed, it becames clear that detection threshold of deviation width in the f_0 changes. It was understood indicate that there were dependence in the modulation frequency and deviation width. These result.

3.3 Experiment2

3.3.1 Method

Detection threshold of the syntheszed voices which modulation of the f_0 exists together in is examined in experiment 2. Experiment 1 shows that fine fluctuation can be perceived. Perhaps, the difference between smooth melody component and f_0 is not perceive, however , it is possible that change component of modulation may influence on perception. Therefore, detection threshold when modulation exists together in voice pair is examined in this experiment. A problem how much influence on perception in the difference in modulation frequency is examined as the results. Additionally, "magnitude of influence" of the contents in the hypothesis is examined.

3.3.2 Result

Detection threshold was $9Hz$, when the synthesizd voices whose f_0 $125Hz$ and the difference in modulation frequency $5Hz$ were presented. Detection threshold was $12Hz$, when the synthesizd voices whose f_0 was $125Hz$ and the difference in modulation frequency was $10Hz$ were presented. Therefore, it was shown that had the possibility which could be cognized sensitively when the difference in modulation frequency is large. In analysis of singing voice, the result is exposed change of modulation frequency by pitch was obteined. Possibility of the hypothesis was more strongly confirmed by these results.

3.4 Experiment3

3.4.1 Method

Detection threshold of the synthesized voices is confirmed when high with low modulation frequency modulation frequency exists. Detection threshold of melody component such as vibrato is examined in Experiment 3. It is considered whether fine fluctuation components were perceived.

"Influence on perception of the modulation frequency and the deviation width " to verify hypothesis is consider from these stages .

3.4.2 Result

Detection threshold was $20Hz$ when $f_0 = 125Hz$. Detection threshold was $60Hz$ when $f_0 = 250Hz$. This result was the same as that of Experiment 1. Therefore, influence of vibrato component is not taken on this condition. Therefore, it is considered that fine fluctuation components can be perceived without influence of vibrato components.

4 Conclusion

These results verify the hypothesis. Perceptive possibility for fine fluctuation components can be shown in the experiment 1 quantitatively. "Fine fluctuation component influences perception" and "magnitude of perception of singing voices depends on modulation frequency and deviation width" in the hypothesis were confirm.

Not only difference in the due to the modulation but also difference in modulation frequency could show that it influenced perception in the experiment 2 in the actual singing voice perception. Influence vibrato component gave to the perception of the fine fluctuation component was examined in the experiment 3. It was found that it wasn't involved in the perception very much as this result. Therefor,the hypothesis "Fine fluctuation components of f_0 contained in singing voices influence on perception of singing voices, and magnitude of influences depend on modulation frequency and deviation width of f_0." could be proved.

References

[1] Manabu Yatabe, Yasuo Endo, Hideki Kasuya, Takao Koube, 'The movement of the basic frequency of the singing voice ', Meeting of the acoustical society of Japan,3-8-6, (1998).

[2] Wakana Odagiri, Hideki Kawahara, 'An examination about the analysis, composition, perception of vibrato of the singing voice', Meeting of the acoustical society of Japan,1-4-6, (1999).

[3] Hideki Kawahara, Parham Zolfaghari, Alain de Cheveigné, Roy D.Patterson, 'Source Informatiuon Extraction using Fixed Points of a Frequency to Instantaneous Frequency Map', The Acoustical Society of Japan H-99-46, July 8 ,(1999).