Proceedings of 20th International Congress on Acoustics, ICA 2010

23-27 August 2010, Sydney, Australia

A Study of Vibrato Features to Control Singing Voices

Naoto Migita (1), Masanori Morise (1) and Takanobu Nishiura (1)

(1) Graduate School of Science and Engineering, Ritsumeikan University, Japan

PACS: 43.72.Ar, 43.75.Rs

ABSTRACT

Singing synthesis applications have recently been developed to create natural singing voices. Applications such as VOCALOID can synthesize singing voices by inputting lyrics and scores, and musical expressions such as vibrato and portamento are added to the singing voices by controlling pitches. This control is crucial to synthesize natural singing voices, but users are required to edit the pitch contour precisely, which means that the sound quality depends on the skill of the user. If the singing voices were controlled by using professional singers' musical expressions, users could synthesize natural singing voices more easily. We created a singing database composed of several singers' vibrato and portamento and then extracted their musical expressions and analyzed the vibrato to synthesize natural singing voices. The database consists of four professional singers (two females and two males) who sang the five Japanese vowels. The vibrato pitches were set to thirteen pitches of an octave that each singer could sing. This database contains not only the singers' own singing voices but also imitations of other singers. We used the database to analyze the differences of vibrato between professional singers and to determine how to control vibrato by impersonation. In a previous study, a vibrato model in the form of a fundamental frequency (F0) contour was proposed to synthesize natural singing voices. This model can control the vibrato rate (speed of pitch fluctuation) and vibrato extent (width of pitch fluctuation). Vibrato sections are extracted and vibrato features are calculated from the F0 contour of singing voices and defined as average in the vibrato section. It was reported that these vibrato features varied according to the type of song. Song type was the only focus of this study, whereas in our study, we focus on the difference of vibrato based on the singer. We propose using vibrato duration (the ratio between vibrato length and voice length) and time fluctuations of the vibrato rate and vibrato extent to analyze professional singers' vibrato. The time fluctuation of the vibrato rate is calculated from period of F0 contour, and the time fluctuation of the vibrato extent is calculated from the instantaneous amplitude of the F0 contour. Results showed the different vibrato features of the professional singers and also indicated that when a singer imitates another singer, the vibrato features are consciously controlled.

INTRODUCTION

Singing synthesis applications have been developed to synthesize singing voices by inputting lyrics and scores. The applications can add musical expressions such as vibrato and portamento to the singing voices by controlling pitches. Although users are required to edit the pitch contour manually, adding natural musical expressions to singing voices is difficult. If the singing voices are controlled by using professional singers' musical expressions, users will be able to synthesize natural singing voices easily.

Previously, musical expressions of singing voices have been analyzed, and then the method to control fundamental frequency (F0) which means the pitch of the voice has been proposed [1]. Especially, vibrato was focused as one of the musical expressions and vibrato features such as vibrato rate and vibrato extent were proposed [2]. It was reported that the vibrato features varied according to the type of song. However the differences of the vibrato features among the singers were not focused. This is because recording of a lot of vibrato singing voices is indispensable and the amature singers can't express the vibrato clearly.

We designed the vibrato database by recording several professeional singers' singing voices to analyze the difference of the vibrato features among the singers. Conventional features and proposed features were extracted from the database and analyzed by histograms of each vibrato feature. In this paper, the vibrato features to control singing voices and results of the analysis are reported.

CONVENTIONAL VIBRATO FEATURES

Vibrato is expressed by vibrating the pitch of the singing voice and defined as the periodic pitch fluctuation. Thus F0 contours of vibrato singing voices were analyzed and two vibrato features were proposed [2]. The vibrato features consist of vibrato rate and vibrato extent which are calculated from the F0 contour in the vibrato section. Vibrato rate is the speed of the pitch fluctuation and vibrato extent is the depth of the pitch fluctuation. The vibrato features are calculated from Eqs. (1) and (2). The parameters (R_n and E_n) for calclulating the vibrato features are extracted from the F0 contour in the vibrato section as shown in Fig. 1. The F0 contour of a vibrato singing voice is calculated by the conventional method [3]. N in Eqs. (1) and (2) means the number of parameters which are extracted from the F0 contour in the vibrato section. Therefore, vibrato rate and vibrato extent are defined as the average of parameters (R_n and E_n) in the vibrato section. In conventional study, the method to control the vibrato features for synthesizing the natural vibrato was proposed [4]. It was reported that the most natural vibrato rate was about 6.3[Hz] and vibrato extent was about 68~84[cent].

$$\frac{1}{vibrato\ rate} = \frac{1}{N} \cdot \sum_{n=1}^{N} R_n, \tag{1}$$

$$vibrato\ extent = \frac{1}{N} \cdot \sum_{n=1}^{N} E_n. \tag{2}$$

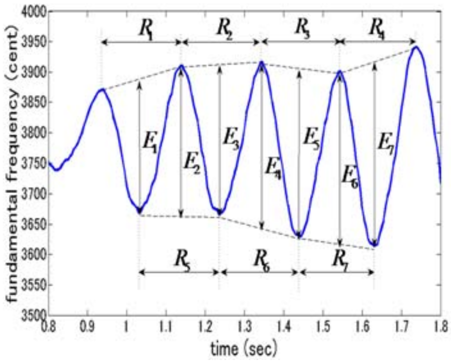

Figure 1. F0 contour of a vibrato singing voice

PROPOSED VIBRATO FEATURES

Conventional study defined the averages of the parameters (R_n and E_n) in the vibrato section as vibrato rate and vibrato extent [2]. The parameters fluctuate with time as shown in Fig. 1. In this paper, the time fluctuations of vibrato rate and vibrato extent are focused. The time length of the vibrato section is also focused.

Time Fluctuation of Vibrato Rate

Solid line in Fig. 2 shows the time fluctuation of the vibrato rate extracted from the F0 contour of a vibrato singing voice by the conventional method [5]. Vibrato rate fluctuates and rises gradually in case of this singing voice. The approximate line of the time fluctuation of the vibrato rate is calculated for analyzing the time fluctuation of vibrato rate and shown by dotted line in Fig. 2. In this paper, the gradient of the dotted line calculated by least square of the time fluctuation of vibrato rate is defined as the time fluctuation of vibrato rate.

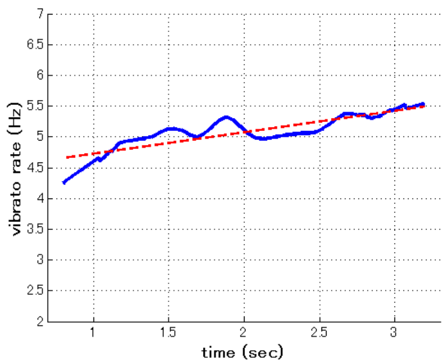

Figure 2. Time fluctuation of vibrato rate (solid line) and the approximate line of the time fluctuation (dotted line)

Time Fluctuation of Vibrato Extent

Solid line in Fig. 3 shows the F0 contour of a singing voice. The instantaneous amplitude of the F0 contour is calculated for analyzing the time fluctuation of vibrato extent and shown by dotted line in Fig. 3. The instantaneous amplitude is represented by the absolute value of $f_a(t)$ calculated by Hilbert transform of the F0 contour $f(t)$ and given by the Eqs. (3), (4) and (5). IDFT denotes the inverse discrete Fourier transform in the Eq. (4). j represents imaginary, t represents time, $f_h(t)$ represents the imaginary part of $f_a(t)$, $F(\omega)$ represents the spectrum of $f(t)$ and ω represents angular frequency. In this paper, the standard deviation of the instantaneous amplitude is defined as the time fluctuation of vibrato extent.

$$f_a(t) = f(t) + jf_h(t), \tag{3}$$

$$f_h(t) = IDFT(F_h(\omega)), \tag{4}$$

$$F_h(\omega) = \begin{cases} -jF(\omega), & \omega > 0 \\ jF(\omega), & \omega < 0. \end{cases} \tag{5}$$

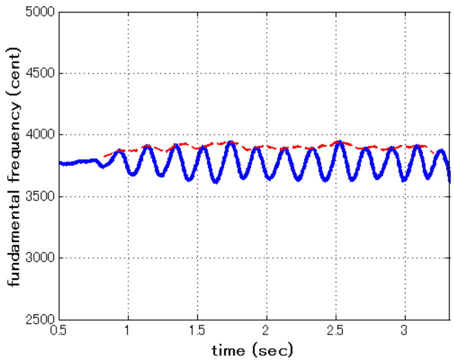

Figure 3. F0 contour of a vibrato singing voice (solid line) and the instantaneous amplitude of the F0 contour (dotted line)

Vibrato Duration

Figure 4 also shows the F0 contour of a singing voice. In this F0 contour, the time length of the vibrato section is shorter than that of the voice section. A singer with this singing voice controls the start time of vibrato. Vibrato duration is defined by Eq. (6) for analyzing the relationship between the time length of the vibrato section and that of the voice section.

$$Vibrato\ \ duration = \frac{T_1}{T_2}. \tag{6}$$

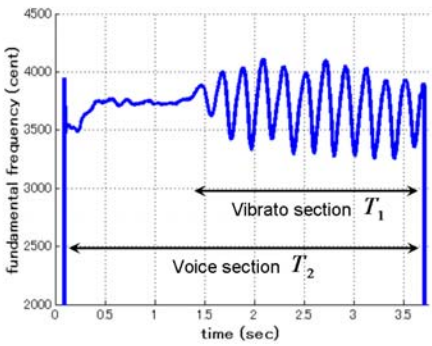

Figure 4. F0 contour of a singing voice

ANALYSIS OF VIBRATO FEATURES

We constructed vibrato database to analyze the vibrato features of several singers. Vibrato features were extracted from singing voices of the database and the differences of the features based on each singer were analyzed. Vibrato rate, vibrato extent and the time fluctuations of the vibrato features were extracted from the F0 contour in the vibrato section. Vibrato duration was also extracted from the time length of the vibrato section. The F0 contour was calculated from the waveform of the vibrato singing voice by the conventional method called “STRAIGHT” [3], and the vibrato section was extracted by using amplitude spectra calculated by short-time Fourier transform of the F0 contour.

Vibrato Database

Vibrato database was designed by recording four professional singers’ (two females and two males) singing voices. The voices used for the singing were the five Japanese vowels (/a/,/i/,/u/,/e/,/o/). The range of the pitche was an octave that each singer could sing. In this database, not only singers’ own singing voices but also the singing voices they imitate other singers were recorded. The imitated singers were selected from famous Japanese pop singers and enka (Japanese ballad) singers. The number of all vibrato data is 520 and the detail is shown in Table 1. Sampling frequency is 96[kHz] and quantization is 24[bit].

Table 1. Conditions of vibrato database

	Singers' own name	Imitated singers' name	Pitches
Female 1	Yoko Aramaki	Hikaru Utada (pop)	B3...B4
Female 2	Kaori Senda	Hibari Misora (enka)	C2...C3
Male 1	Naoto Fuga	GACKT (pop)	D3...D4
Male 2	Kazuo Nishi	Hiroshi Itsuki (enka)	E3...E4

Result of Vibrato Rate

The histograms of Fig.5 show distributions of vibrato rates extracted from 520 singing voices in vibrato database. In singers’ own singing voices, the averages of vibrato rates of female 1, female 2, male 1 and male 2 are 5.2[Hz], 3.8[Hz], 3.8[Hz] and 5.3[Hz] respectively. The averages are different from 6.3[Hz] of the conventional study, whereas the individualies of vibrato rates are confirmed. In case of female 2, vibrato rates of the singer’s own singing voices are also different from the imitation singing voices. In other words, female1 expressed vibrato for Hikaru Utada by controlling vibrato rate.

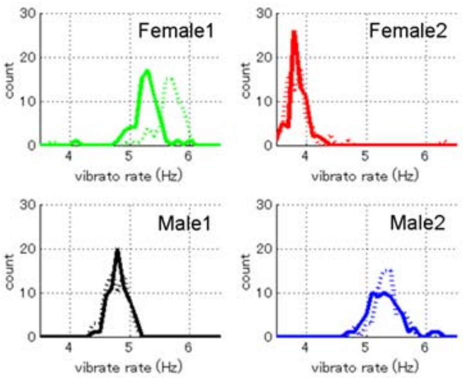

Figure 5. Histograms of vibrato rates
(Solid lines: singers’ own voices, dotted line: imitation voices.)

Result of Time Fluctuation of Vibrato Rate

Figure 6 represents the histograms of time fluctuations of vibrato rates extracted from the vibrato database. All vibrato rates fluctuate gradually and the range of the time fluctuations is over -1.0 and under 1.0. In case of female 2, the average of the time fluctuation is negative. On the other hand, the averages of the time fluctuations for male 1 and male 2 are positive. Hence, the time fluctuations of the vibrato rates are different by the singers. However singer’s own singing voices are similar with the imitation singing voices in all singers.

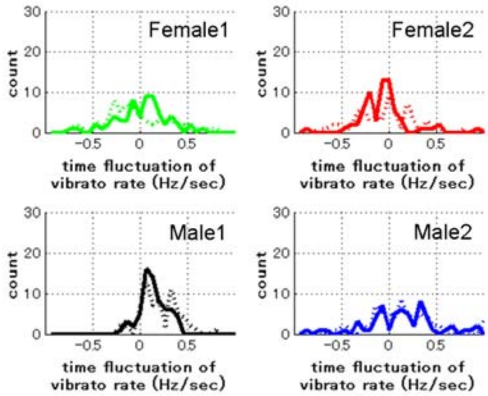

Figure 6. Histograms of time fluctuations of vibrato rates
(Solid lines: singers’ own voices, dotted line: imitation voices.)

Result of Vibrato Extent

The histograms of vibrato extents were calculated as well as the analysis of vibrato rate and given by Fig. 7. In singers’ own singing voices, the averages of vibrato extents of female 1, female 2, male 1 and male 2 are 35[cent], 120[cent], 86[cent] and 52[cent] respectively. Therefore, vibrato extents are different by the singers as well as vibrato rate. The averages of male1 and male 2 are near 68…84[cent] of the conventional study. Vibrato extents of the imitation singing voices for male 1 are higher than the singer’s own singing voices, provided that the imitation singing voices are compared with singer’s own singing voices. This means male 1 expressed vibrato for GACKT by controlling vibrato extent. However, the variance of the distribution of imitation singing voices for GACKT is larger than that of male1’s own singing voices.

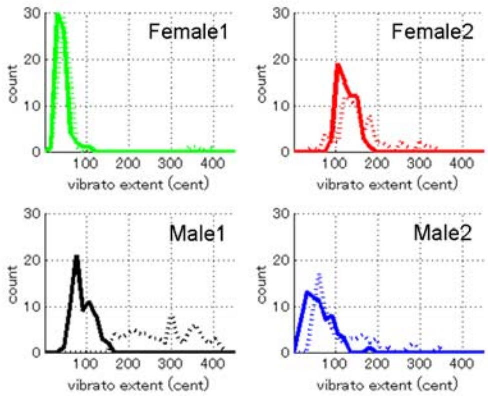

Figure 7. Histograms of vibrato extents
(Solid lines: singers’ own voices, dotted line: imitation voices.)

Result of Time Fluctuation of Vibrato Extent

Figure 8 represents the histograms of time fluctuations of vibrato extents extracted from the vibrato database. The relationships of histograms among singers are similar, provided that Fig. 8 is compared with Fig. 7. The relationships between singers’ own singing voice and imitation singing voice are also similar. Especially, the time fluctuations of male 1’s imitation singing voice are higher than those of male 1’s own singing voices. The time fluctuation of vibrato extent varies according to the vibrato extent.

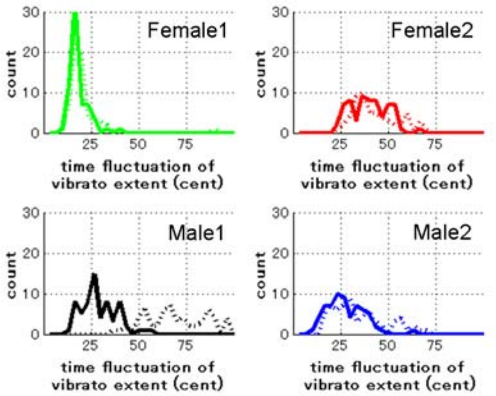

Figure 8. Histograms of time fluctuations of vibrato extents (Solid lines: singers' own voices, dotted line: imitation voices.)

Result of Vibrato Duration

The time length of the vibrato section is also focused. The histograms of vibrato durations extracted from vibrato database are shown in Fig. 9. In case of male 1, vibrato durations of the imitation singing voices are higher than those of the singer's own singing voices. Meanwhile, in case of male 2, vibrato durations of singer's own voices are lower than those of imitation voices. These results show singers control the vibrato start time in vibrato duration by the imitation and vibrato start time is important to express vibrato for GACKT and Hiroshi Itsuki.

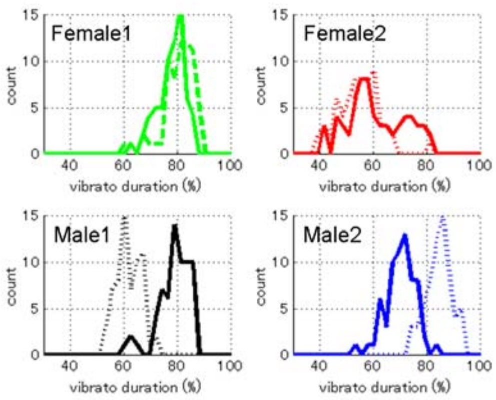

Figure 9. Histograms of vibrato durations (Solid lines: singers' own voices, dotted line: imitation voices.)

DISCUSSIONS

We confirmed that all vibrato features were different by the singers, and the vibrato database was effeicient for analyzing the individualities of the singing voices. Vibrato rates and vibrato extents, the time fluctuations of vibrato extents and vibrato durations are consciously controlled by the imitation, provided that the imitation singing voices of four singers are compared with their own singing voices. On the other hand, the time fluctuations of vibrato rates are not controlled. Figure 10 shows the time flectuations of vibrato rates extracted from singing voices of female 1 and male 1. The time fluctuation of vibrato rate is positive in case of male 1, whereas the time fluctuation is negative in case of female 1. This means the time fluctuation of vibrato rate varies among the singers, but the singers don't control the time fluctuation of vibrato rate when they imitate other singers in the vibrato database. In future work, we will analyse whether the singers control the time fluctuation of vibrato rate.

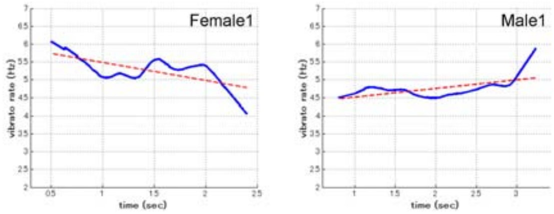

Figure 10. The time fluctuations of vibrato rates (solid lines) and the approximate lines of the time fluctuations (dotted lines)

CONCLUSIONS

Musical expressions of singing voices have been analysed for controlling singing voices. Especially, vibrato was focused as one of the musical expressions and vibrato rate and vibrato extent were proposed as vibrato features in the conventional study. The study showed vibrato features varied according to the type of song. However, the differences of vibrato features among the singers have not been analyzed. We constructed vibrato database consisted of several professional singers' singing voices and analyzed vibrato features to control singing voices. Vibrato rate, vibrato extent, the time fluctuations of the vibrato features and vibrato duration were extracted from the singing voices in the vibrato database. Results of the analysis show the differences of all vibrato features based on each singer. Furthermore, when a singer imitates another singer, the vibrato features are consciously controlled except the time fluctuation of vibrato rate. The result indicates it's difficult for singers to control the time fluctuation of vibrato rate. We confirmed the vibrato features were efficient for expressing the individualities of singing voices. In future works, we will try to extend conventional F0 model by using proposed vibrato features. Conventional F0 model can control vibrato rate and vibrato extent. The proposed features will enable the vibrato model to control the individualities of singers and add the vibrato of professional singers to the synthesized singing voices more naturally.

ACKNOWLEDGMENTS

This work was partly supported by Grants-in-Aid for Scientific Research funded by The Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES

1 T. Saitou, M. Unoki and M. Akagi, "Extraction of F0 dynamic characteristics and development of F0 control model in singing voice," Proc. ICAD2002, pp. 275-278, 2002.
2 T. Nakano, M. Goto and Y. Hiraga, "An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features," In Proc. Interspeech 2006, pp. 1706–1709, 2006.
3 H. Kawahara, "STRAIGHT, Exploration of the other aspect of VOCODER: Perceptually isomorphic decomposition of speech sounds," Acoustic Science and Technology, vol. 27, pp. 349-353, 2006.
4 T. Saitou, M. Unoki and M. Akagi, "A study on a control method of vibrato modulation frequency for synthesizing natural singing-voice," IEICE technical report, vol. 105, pp. 13-18, 2005(in Japanese).
5 M. Morise, H. Kawahara and T. Nishiura, "Rapid F0 estimation for high-SNR speech," In Proc. WESPAC 2009, CD-ROM proceedings, 2009.