Analysis for Vibrato with Arbitrary Shape and its Applications to Music

S. W. Lee, Minghui Dong and Paul Yaozhu Chan
Human Language Technology Department, Institute for Infocomm Research, A*STAR, Singapore 138632
E-mail: {swylee, mhdong, ychan}@i2r.a-star.edu.sg Tel: +65-64082796

***Abstract*— Vibrato in singing voice and instrument performance is essential for perceived quality. It also creates individual performer styles. Over the years, vocal teachers and musicians often examine this musical characteristic by hand. This results in limited scope of music pieces being analyzed. This paper presents a vibrato analysis method, which is suitable for vibrato tones with arbitrary shape and quantitative analysis, enabling various music applications. To evaluate the performance of the proposed analysis method, we have conducted an experiment on estimation accuracy. It is shown that the error of estimating rate and extent are less than or equal to 6% and 9%, respectively. Some applications of the analysis results, namely, vibrato analysis and singing synthesis, are presented at the end.**

I. INTRODUCTION

Singing voice synthesis has been one of the emerging and popular research topics recently [1]-[3]. Nevertheless, the analysis of singing voice and music is often manually done by professional musicians [4], [5]. As technologies advance, more and more performance on singing, musical instruments and music applications are available. For example, the growing amount of music performance clips from professionals and amateurs in the internet, VOCALOID singing synthesizer [6], LADIDA entertainment applications for mobile phones [7]. Automatic analysis methods are, hence, beneficial to the music industry, leading to various kinds of applications. These include singing voice synthesis and karaoke applications, understanding and characterization of individual singing voices, music information retrieval, computer-assisted vocal training etc.

Pleasant singing voice and musical instrument performance are often accompanied by vibrato [8], [9]. Vibrato is a periodic fluctuation in pitch of a musical tone [4], [10]. Fig. 1 shows two examples of vibrato extracted from a singing voice and an instrument sound. These plots are the fundamental frequency (F0) contours over time. The left one is from a pop song singing produced by a female singer, whereas the right one is from a C4 note produced by a violin. Note that vibrato is not necessarily sinusoidal, as shown in the right figure. Vibrato is essential for pleasant performance [8], [9]. Singers and instrument players use it to color a tone, personalize the performance and express their emotions for a given piece of music. Furthermore, vibrato has been found to contribute most to perceived voice quality, compared to other singing features [11].

Despite of these important roles, there are not many quantitative nor statistical studies on vibrato characteristics [4], [5], [12]-[15]. A singing skill evaluation was proposed in [16], which focuses on the singing pitch accuracy and vibrato, in particular, aiming at detecting the presence of vibrato. Traditionally, vibrato study has relied on manual inspection by musicians [4], [5], [8], [10]. Several recent analysis methods ([12] and the spectrum modeling in [17]) are restricted to sinusoidal vibrato only. Likewise, most of the vibrato studies are on Western operatic singing [13] or instrumental sounds in classical music [14]. The abundant singing and music data in the pop song domain is left unexplored.

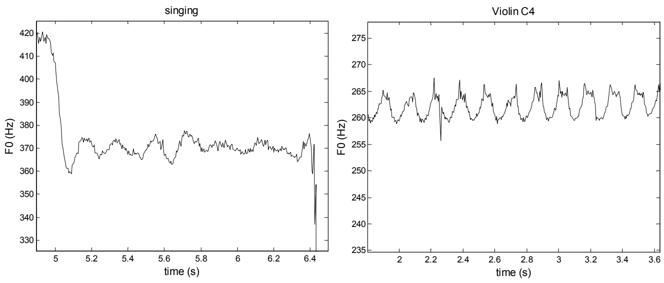

Fig. 1 Examples of vibrato. (left) A vibrato segment extracted from pop song singing. (right) A vibrato segment extracted from a C4 violin note.

What are the vibrato characteristics in pop songs? Is it possible to incorporate these characteristics in synthetic singing, if any? To answer the above questions, an automatic vibrato analysis method is necessary. This paper presents an analysis method, estimating the two parameters *rate* and *extent*. This method is suitable for vibrato with arbitrary shape. Evaluation of the estimation performance (this is seldom reported in previous publications).

With this vibrato analysis method, quantitative study of enormous audio data from various domains, in addition to Western operatic singing, will be available in the near future. Singing voice synthesis and singer characterization are other applications for the estimated rate and extent. Previous singing and music studies [4], [10] showed that, the vibrato characteristics are generally considered to be constant within a singer, i.e. the singer is often unable to change it or it is difficult, though not impossible to change by intensive vocal training. Therefore, the analysis results act as singer identification or discrimination features. Similar ideas have been adopted in [15].

II. Review on Vibrato and Past Analysis Methods

Being a periodic fluctuation in pitch, vibrato is characterized by two parameters: rate and extent. The rate represents the number of vibrato cycles per second. The extent describes how far the F0 rises and falls from the average pitch of a vibrato segment concerned (as illustrated in Fig. 2). By vibrato analysis here, it refers to the estimation of these two parameters.

In Western operatic singing, the rate is generally between 5-8 Hz, whereas the extent ranges from 30 to 150 cents [10], [18].

Vibrato analysis methods, in particular, those automatic ones, are rare. In the following, some representative analysis methods are briefly reviewed before introducing our method.

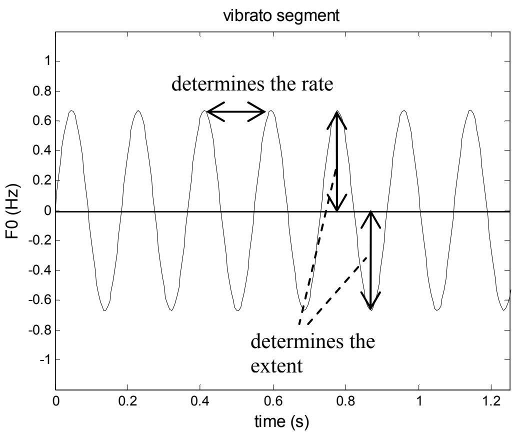

Fig. 2 Vibrato parameters: rate and extent.

A. *Prame's Method*

To analyze the vibrato characteristics of vocal singing, Prame measured vibrato rate with a KAY DSP Sonagraph machine [5]. Given a singing record, the spectrogram was visually inspected to locate the wave troughs. The time between consecutive wave troughs was recorded by using the cursors on the screen and the rate was calculated as the inverse of the time difference.

Prame worked on vibrato rate only. No measurement method was suggested for extent, whereas the method for vibrato rate is done manually. It is also stated in [5] that as the singing intensity varies over time, it is difficult to locate wave troughs on the screen. This is often common at the end of vibrato tones where the intensity is typically low. Consequently, the estimates from this manual measurement are erroneous and the scope of measurements is limited too.

B. *Pang's Maximum Likelihood Estimation*

Given a vibrato segment, this method [12] first extracts the F0 contour. It is assumed that this F0 contour is sinusoidal. Maximum likelihood (ML) is then applied with the representation below.

$$x(n) = a(n) + b(n)\sin\left(2\pi\frac{v(n)}{f_{frame}}n + \theta\right) + w(n), \quad (1)$$

where $x(n)$ is the estimated F0 contour. $v(n)$ and $b(n)$ are the rate and extent parameters respectively. $a(n)$ represents the average pitch. θ and $w(n)$ are the phase shift and noise term respectively. f_{frame} is calculated as the ratio between the sampling frequency and the non-overlapping number of samples between two consecutive frames. ML estimation is used to find $a(n)$, $b(n)$ and $v(n)$ by maximizing a cost function with respect to a vibrato rate hypothesis [12]. By searching a set of possible rate hypotheses, the one that optimizes the cost function is selected as $v(n)$. With this rate estimate, the average pitch $a(n)$ and the extent $b(n)$ are further estimated.

The Pang's ML method serves as automatic vibrato feature estimation. However, this method intrinsically suffers from non-sinusoidal vibrato segments, where (1) is not valid. Vibrato tones are not necessarily sinusoidal. There are vibrato tones, e.g. those from some musical instruments (as shown in the right hand side of Fig. 1) are non-sinusoidal.

C. *Vibrato Detection Method from Nakano et al.*

A singing evaluation system has been proposed by Nakano *et al.* [16]. Pitch accuracy and vibrato are used as the features to judge the quality of singing and classify the singing as 'good' or 'poor' by Support Vector Machine (SVM). The system focuses on the detection of the presence of vibrato. Specifically, based on the knowledge from Western operatic singing, restrictions on accepted rate and extent are imposed. Parameters regarding the length of detected vibrato section, frequency spectrum of the F0 segment and number of times of the F0 segment crossing its average value are used [16]. Their system focuses on the presence of vibrato and compares the segment under consideration with those labeled as good singing and those labeled as poor singing, rather than explicitly learning the vibrato characteristics in different types of music.

III. Proposed Analysis Method for Vibrato with Arbitrary Shape

As vibrato tones may be in any arbitrary shape, not necessarily sinusoidal, the periodic regularity becomes the major characteristic. Any parametric representation regarding the shape is inapplicable. Our vibrato analysis aims to provide reliable estimation of both rate and extent for a given vibrato segment, such that analysis of vibrato observation and generation of synthetic vibrato are facilitated.

At the beginning, Tandem-STRAIGHT [19], a unified estimation for spectrum, F0 and aperiodicity is applied for pitch tracking. Tandem-STRAIGHT provides accurate estimation of spectrum and F0 and has been widely used in speech and singing synthesis [20], [21]. With the output F0 contour, the following describes the proposed estimation method for the rate and extent.

Let $x(n)$ denotes the F0 contour. $x(n)$ is first normalized with its average value. This normalization is necessary for finding the relative rise and fall for the extent and the underlying rate remains unchanged. The mean of the normalized F0 contour is then removed, so as to better capture the periodicity in subsequent stage. Let $\hat{x}(n)$ and $\tilde{x}(n)$ be the normalized F0 and the zero-mean, normalized F0 respectively. The autocorrelation function $R_{\tilde{x}}(m)$ is calculated on $\tilde{x}(n)$.

$$R_{\tilde{x}}(m) = \begin{cases} \sum_{n=0}^{N-m-1} \tilde{x}(n+m)\tilde{x}(n), & m \geq 0 \\ R_{\tilde{x}}(-m), & m < 0 \end{cases}, \quad (2)$$

where N is the data length of $\tilde{x}(n)$. $R_{\tilde{x}}(m)$ is smoothed with a moving average window. A seven-tap window is used in the current implementation. Finally, the non-zero lag (m^*, where $m^* > 0$) maximizing $R_{\tilde{x}}(m)$ is selected and

$$rate = \frac{1}{m^* \cdot \Delta F0}, \quad (3)$$

where $\Delta F0$ denotes the sampling period of F0 estimates in Tandem-STRAIGHT. $\Delta F0$ is 0.05 s in current implementation. The autocorrelation function and m^* selection rely on the periodic nature of vibrato only. Any vibrato shape, whether it is sinusoidal vocal singing or non-sinusoidal instrument sound, etc., is applicable.

The extent is estimated as the mean of the maximum displacements and minimum displacements from the average pitch value observed in each vibrato period. No vibrato characteristic is assumed, except its periodic nature. Fig. 3 illustrates the calculation of extent. A normalized vibrato segment $\hat{x}(n)$ taken from a F#4 cello note is shown.

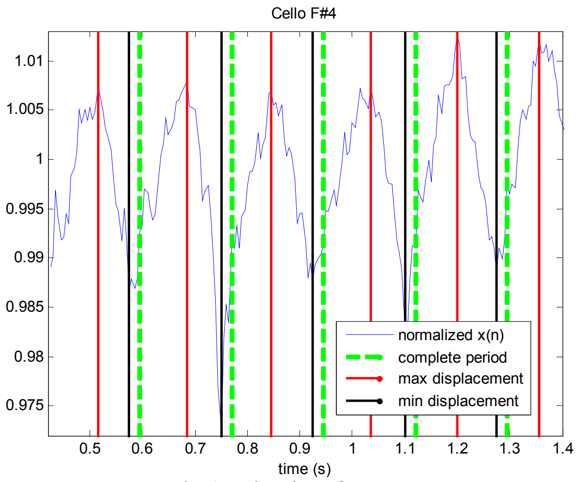

Fig. 3 Estimation of extent.

Based on m^*, the normalized vibrato segment $\hat{x}(n)$ is partitioned into a finite number of complete periods. Let K denotes the number of periods in total. In each period, the maximum displacement ($d_{max,k}$) and minimum displacement ($d_{min,k}$) are calculated by

$$d_{max,k} = 1200 \times |\log_2 \widehat{x_k}(n^*_{max})| \quad (4)$$
$$d_{min,k} = 1200 \times |\log_2 \widehat{x_k}(n^*_{min})| \quad (5)$$

where $\widehat{x_k}(n^*_{max})$ and $\widehat{x_k}(n^*_{min})$ are the normalized F0 values where the maximum displacement and minimum displacement in k-th period occur, respectively. As vibrato segments generally contain both positive and negative displacements from the average pitch value, the location of maximum displacement and minimum displacement are often the highest peak and the lowest valley respectively. The base 2 logarithm operation and scaling by 1200 convert the displacement into cents. Consequently,

$$extent = \frac{1}{2K} \sum_{k=1}^{K} \left(d_{max,k} + d_{min,k}\right). \quad (6)$$

Referring to Fig. 3, $\hat{x}(n)$ is segmented into complete periods, as indicated by the thick dotted line boundaries. Within each period, the maximum displacement and minimum displacement are located, which are indicated by a red solid line and a black solid line, respectively.

IV. Experiments on Vibrato Analysis and Applications to Music

The proposed vibrato analysis method is applied to music recordings, specifically, vocal singing and instrument sounds. The vibrato tones are analyzed accordingly. The following gives the details of the experiments. Several examples of potential applications are presented at the end.

A. Recording Setup and Data Sets

Two data sets are used in our experiments. The first data set D1 contains singing voice recordings made from a female professional singer. Ten Mandarin Chinese pop songs are recorded. Each song lasts about four mins. These pop songs are selected based on the singing skills of the singer, rhythms and the pitch ranges, such that she sings comfortably. The corresponding MIDI files are prepared in advance. During the recording, the music score and lyrics are given to the singer. While the MIDI file is played on the headphones, she sings and her singing voice is recorded synchronously. To ensure pleasant and natural singing, the singer is allowed to practise as many times as she wants.

As a result, for each pop song, the data contain: singing voice, music score, MIDI file and the lyrics. The singing voice is sampled in 48 kHz and stored in 24-bit .WAV format. The music score and the lyrics are used during data collection only.

The second data set D2 contains sole instrument sounds from flute, cello, violin and bass (vibrato samples). They are originated from the McGill classic collection [22]. Please refer to Table I for details. These samples are in 44.1 kHz, stereo channel and stored in 24-bit .WAV format. The duration in total in the two data sets is 45 mins 53 s.

Table I
Instrument Sounds in Data Set D2

instrument	Note	No. of files
Flute	A4-A6, A#4-A#6, B4-B6, C4-C7, C#4-C#6, D4-D6, D#4-D#6, E4-E6, F4-F6, F#4-F#6, G4-G6 and G#4-G#6	37
Cello	A2, A3 (stop & open), A4, A#2-A#4, B2-B4, C2-C5, C#2-C#5, D2, D3	47

	(stop & open), D4-D5, D#2-D#5, E2-E5, F2-F5, F#2-F#5, G2 (stop & open), G3-G5, G#2-G#4	
Violin	A#3-A#4, A3, A4 (stop), A5, B3-B4, C#4-C#5, C4-C5, D#4-D#5, D4 (stop), E4, F#4-F#5, F4-F5, G#4, G4	21
Bass	A#3, A1 (stop & open), B2, C#1, C#3-C#4, C1-C3, D#1, D#3-D#4, D1, D2 (stop), D3, E4, F#1-F#2, F1-F3, G2 (stop)	23

B. Vibrato Analysis

Pitch tracking is done as the first step with Tandem-STRAIGHT [19]. The pitch search ranges for D1 and D2 are set to 80-1100 Hz and 50-2100 Hz, respectively. Vibrato segments are then manually located by inspecting the output F0 contour. There are 234 segments from D1 and 112 segments from D2. These vibrato segments are input to the proposed analysis method.

Fig. 4 shows the scatter plot of rate and extent observed in singing voice (from D1). A two-dimensional (2-D) normal distribution is fit by minimum variance unbiased estimator (MVUE). The mean rate and extent are 5.07 Hz and 27.97 cents respectively. By using the proposed vibrato analysis method, it is now possible to analyze vibrato tones in a statistical manner. In the current implementation, vibrato segments are located manually. For enormous amount of data, some detection scheme, e.g. imposing restrictions of accepted rate and extent as in [16], can be incorporated together with the proposed analysis method. With these vibrato features (rate and extent), analysis, modeling and subsequent applications are facilitated, as shown below.

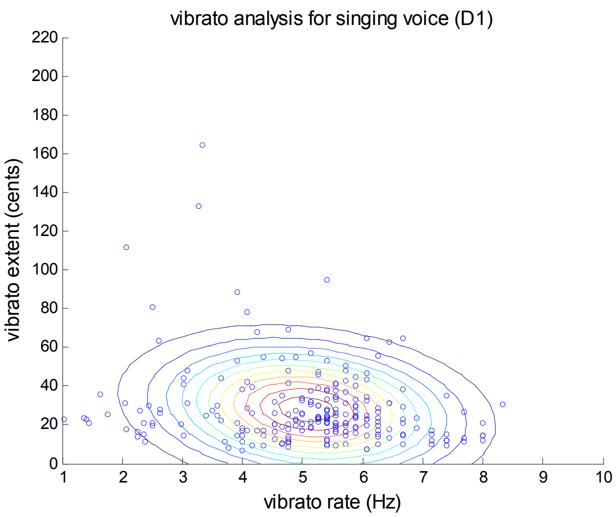

Fig. 4 Scatter plot of rate and extent observed in singing voice.

Comparing with the typical vibrato values ($5 \leq rate \leq 8$ Hz, $30 \leq extent \leq 150$ cents) in Western operatic singing, this singer exhibits a slower and smaller vibrato. Constructing a random variable by rate and extent as the first and second elements respectively), the covariance matrix is

$$\begin{bmatrix} 2.01 & -5.31 \\ -5.31 & 385.5 \end{bmatrix}.$$

It is found that this singer has a smaller range of extent than typical Western operatic singing (the standard deviation of extent is 19.63 cents). This experiment demonstrates an example of vibrato analysis for a single singer in an automatic, statistical manner. Vibrato in pop songs can be analyzed similarly using data of different collections of singing voice.

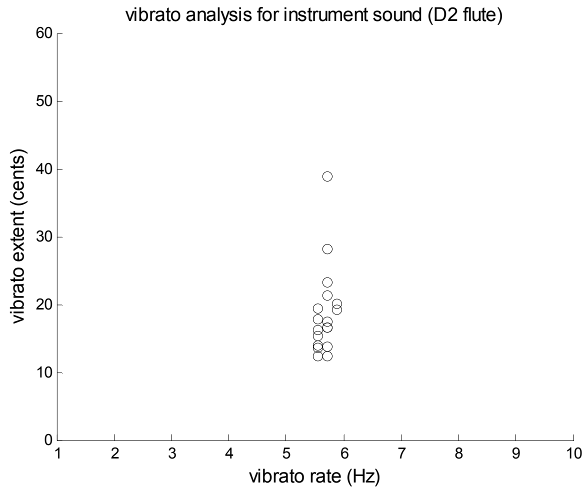

Fig. 5 Scatter plot of rate and extent observed in flute.

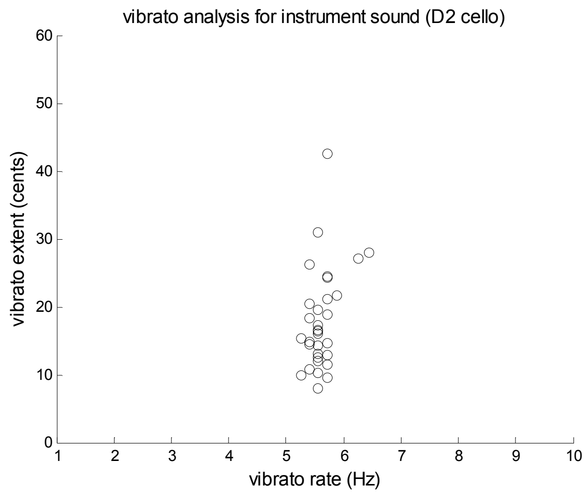

Fig. 6 Scatter plot of rate and extent observed in cello.

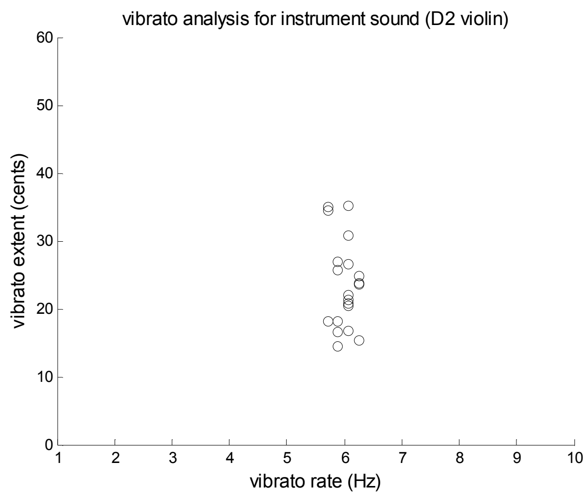

Fig. 7 Scatter plot of rate and extent observed in violin.

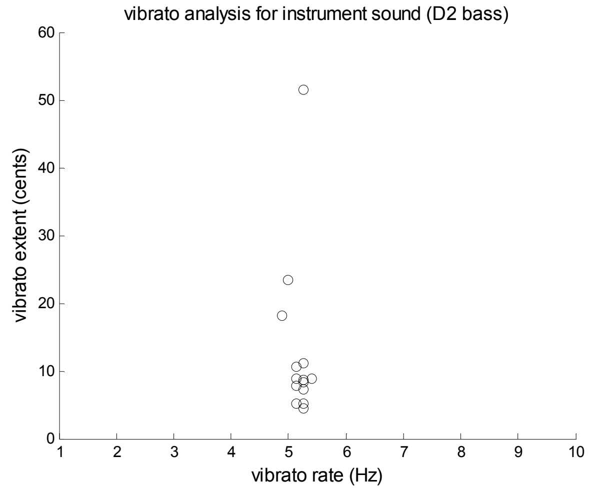

Fig. 8 Scatter plot of rate and extent observed in bass.

Fig. 5-8 show the distributions of rate and extent observed in instrument sounds. These figures are in the order of flute, cello, violin and bass, from top to bottom, respectively. The vibrato characteristics of these four instruments are similar. The rate is always around 5 to 6 Hz. The ranges of extent in flute, cello and violin are highly similar. Bass has a wider range of extent. As these vibrato rates are generally constant, normal distribution fitting is not used. Compared with the vibrato characteristics in singing voice, the rate exhibited in D2 is fairly unchanged.

C. *Estimations of Rate and Extent*

The estimation of rate and extent are evaluated in this section. Regular vibrato segments are selected from the two data sets and the evaluation is done on this subset. 93 vibrato segments are selected. The periods, peaks and troughs in F0 contours are located manually. The rate and extent estimated from the proposed analysis method are then compared to those found by the manual labels (named as references).

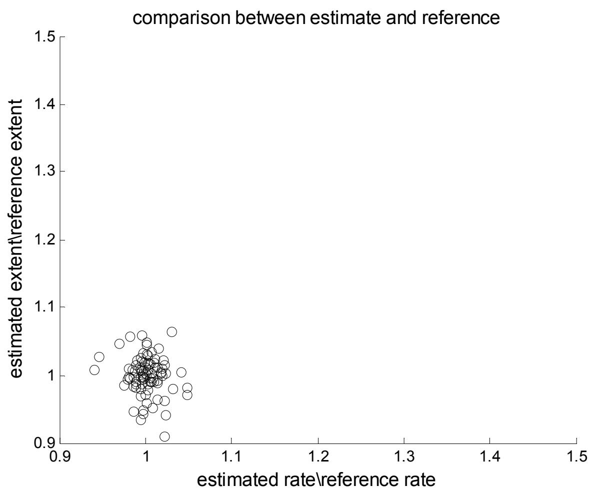

Fig. 9 Evaluation of the estimation of rate and extent. The estimates are compared with the references obtained from manual labels.

Referring to Fig. 9, the comparison is done by calculating the ratio between the estimate and the reference. The x-axis is for the rate and the y-axis is for the extent. Both ratios are found to be close to one, showing that our estimates are close to the references. The error of rate and extent are within 6% and 9% respectively (This is found by computing the distance between one and the corresponding farthest estimate). This indicates the rate and extent are accurately estimated.

D. *Application to Singing Voice Synthesis*

One of the potential applications of our vibrato analysis method is singing voice synthesis. There are two major categories of voice synthesis systems. The first category generates singing voice directly from lyrics, similar to conventional text-to-speech systems [3], [6]. The vocal characteristics are totally determined by pre-defined singer recordings.

Another category produces singing voice by converting a lyrics-reading speech into singing voice [20], [21]. The vocal characteristics of the input speech are hence preserved. In particular, good singing voice can be generated from any individual; even he or she is not good at singing. Given a speaking voice reading the lyrics of a song and the MIDI file, the system first decomposes the speech into F0, spectrum and aperiodicity. Modifications on these attributes are then made and the resultant attributes are combined finally to produce the singing voice. F0 modification includes the additions of vibrato fluctuation, overshoot and preparation, etc. [21]. Ideally, the modified F0, spectrum and aperiodicity should be close to those observed in real singing voice.

In the following illustration, we incorporate the above analysis findings into architecture similar to the second category. The singing voice in D1 is first decomposed into the three attributes using Tandem-STRAIGHT [19]. Next, the F0 is manipulated to add the vibrato analysis results, whereas the singing spectrum and aperiodicity are adopted directly.

Based on the estimated rate and extent, artificial vibrato segments are generated in the F0 contours. These F0 contour may or may not contain some other minor fluctuations, such as overshot and preparation. First, all constant MIDI F0 segments last for at least 0.5 s are identified. These will be the locations of the artificial vibrato. With the normal distribution results for rate and extent, artificial vibrato $v(n)$, in the form of sinusoid, is generated by

$$\hat{e} = 2^{\frac{\bar{e}}{1200}} \quad (7)$$

$$v(n) = m(n)\left[(\hat{e}-1)\sin\left(\frac{2\pi\bar{r}n}{f_{ps}}\right)+1\right] \quad (8)$$

where $\bar{r}$ and $\bar{e}$ are the distribution means of rate and extent respectively. $m(n)$ and f_{ps} are the constant pitch value and the F0 sampling frequency respectively. The corresponding singing F0 segment is replaced by $v(n)$ and the singing voice is generated accordingly.

The informal listening test shows that the singing voice with artificial vibrato is pleasant. In the following, the contribution of artificial vibrato on perceived voice quality on singing synthesis is verified.

A subjective listening test is conducted. It consists of 30 questions. Each question compares stimuli generated from three methods: (A) recorded singing voice with Tandem-STRAIGHT analysis and synthesis (no modification on singing F0 contour); (B) singing synthesis with artificial vibrato generated by (8); and (C) singing synthesis with constant F0 during the instant when artificial vibrato takes place in method (B). All stimuli contain one line of lyrics.

Listeners are asked to compare and rate the quality of the three methods by mean opinion score (MOS). Possible MOS ranges from 1 (bad) to 5 (excellent). During the test, listeners can play the stimuli as many times as they wish. There are 21 subjects participated, constituting $30 \times 21 = 630$ responses.

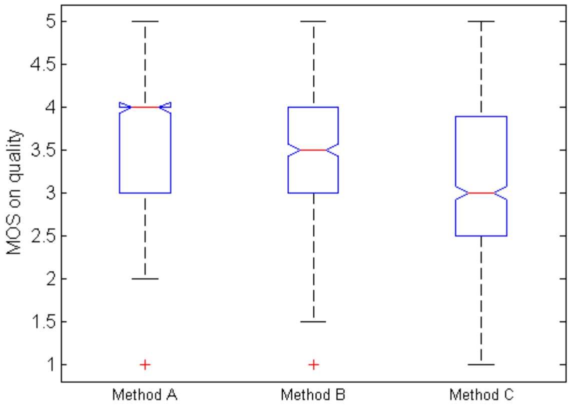

Fig. 10 Box plot of the MOS result on quality.

Fig. 10 shows the box plot of the MOS result on quality. Compared with the recorded singing voice (Method A), the quality achieved by synthesizing singing voice with artificial vibrato (Method B) is not as high as the recorded singing. Comparing Method B and C, it is shown that the quality of synthetic singing with artificial vibrato (Method B) is significantly better than the one with constant F0 (Method C) with 95% confidence. This indicates that the proposed artificial vibrato is useful for promoting the output quality of speech-to-singing synthesis (In speech-to-singing synthesis system, if MIDI is used alone, the pitch values during a long note are modified to constant values accordingly).

V. CONCLUSIONS

Although vibrato takes an important role in music quality, musicians conventionally resort to study vibrato characteristics by hand. This paper presents a vibrato analysis method, which is suitable for vibrato tones with arbitrary shape. With the proposed analysis method, extraction of major vibrato characteristics, namely rate and extent, is facilitated. The experiment on estimation performance shows that rate and extent are accurately estimated with error less than or equal to 6 % and 9 % respectively. This vibrato analysis method not only acts as a mean to analyze vibrato characteristics, but also leads to music applications, such as singing voice synthesis.

REFERENCES

[1] *Synthesis of Singing Challenge (Special Session), Proc. Interspeech*, Aug. 2007.

[2] M. Akagi, "Rule-based voice conversion derived from expressive speech perception model: How do computers sing a song joyfully?" in *Proc. ISCSLP*, Tutorial 01, Nov. 2010.

[3] K. Saino, H. Zen, Y. Nankaku, A. Lee, and K. Tokuda, "HMM-based singing voice synthesis system," in *Proc. Interspeech*, Sep. 2006.

[4] R. Timmers and P. Desain, "Vibrato: Questions and answers from musicians and science," in *Proc. ICMPC*, Aug. 2000.

[5] E. Prame, "Measurements of the vibrato rate of ten singers," *J. Acoust. Soc. Amer.*, vol. 96, pp. 1979-1984, Oct. 1994.

[6] H. Kenmochi, and H. Ohshita, "VOCALOID – Commercial singing synthesizer based on sample concatenation," in *Proc. Interspeech*, Aug. 2007.

[7] P. Kirn (2009, Oct. 6). iPhone Day: LaDiDa's Reverse Karaoke Composes Accompaniment to Singing [Online]. Available: http://createdigitalmusic.com/2009/10/iphone-day-ladidas-reverse-karaoke-composes-accompaniment-to-singing/

[8] W. T. Bartholomew, "A physical definition of "Good voice-quality" in the male voice," *J. Acoust. Soc. Amer.*, vol. VI, pp. 25-33, Jul. 1934.

[9] Y. Meron and K. Hirose, "Synthesis of vibrato singing," in *Proc. ICASSP*, Jun. 2000.

[10] J. Sundberg, *The Science of the Singing Voice*, Northern Illinois University Press, 1987.

[11] T. Saitou and M. Goto, "Acoustic and perceptual effects of vocal training in amateur male singing," in *Proc. Interspeech*, Sep. 2009.

[12] H.-S. Pang, "On the use of the maximum likelihood estimation for analysis of vibrato tones," *Applied Acoustics*, vol. 65, pp. 101-107, Jan. 2004.

[13] J. A. Diaz and H. B. Rothman, "Acoustical comparison between samples of good and poor vibrato in singers," *Journal of Voice*, vol. 17, pp. 179-184, Jun. 2003.

[14] P. Herrera and J. Bonada, "Vibrato extraction and parameterization in the spectral modeling synthesis framework," in *Proc. Digital Audio Effects Workshop*, Nov. 1998.

[15] T. L. Nwe and H. Li, "Exploring vibrato-motivated acoustic features for singer identification," *Trans. Acoust., Speech, Language Processing*, vol. 15, pp. 519-530, Feb. 2007.

[16] 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*, Sep. 2006.

[17] S. Rossignol, P. Depalle, J. Soumagne, X. Rodet, and J.-L. Collette, "Vibrato: detection, estimation, extraction, modification," in *Proc. Digital Audio Effects Workshop*, Nov. 1999.

[18] C. E. Seashore, "A musical ornament, the vibrato," in *Psychology of Music*, McGraw-Hill Company, 1938, ch. 4, pp. 33-52.

[19] H. Kawahara, M. Morise, T. Takahashi, R. Nisimura, T. Irino, and H. Banno, "TANDEM-STRAGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation," in *Proc. ICASSP*, Mar. 2008.

[20] T. L. Nwe, M. Dong, P. Chan, X. Wang, B. Ma, and H. Li, "Voice conversion: From spoken vowels to singing vowels," in *Proc. ICME AdMIRe Workshop*, Jul. 2010.

[21] T. Saitou, M. Goto, M. Unoki, and M. Akagi, "Speech-to-singing synthesis: Converting speaking voices to singing voices by controlling acoustic features unique to singing voices," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, Oct. 2007.

[22] F. Opolko and J. Wapnick, *McGill University Master Samples: A 3-DVD Set*, 2006.