

REAL-TIME VOCAL VIBRATO TRAINING SYSTEM

David Heid, Lei Fan

Univeristy of Rochester
Department of Electrical and Computer Engineering

ABSTRACT

A vocal vibrato training system providing real-time visual feedback to the user has been created for the purpose of helping singers achieve healthy and consistent vibrato. This feedback is based on quantitative measurements of the singers' vibrato rate and extent and compared to acceptable values as defined by Ingo Titze[1]. Additionally, a preliminary questionnaire was given to trained and untrained singers to gauge their understanding of vocal vibrato. Recordings of several students were made after taking this survey. Results show an inconsistent understanding of vibrato and its underlying cause, with recordings that reflect this ambiguity.

Index Terms— vocal, vibrato, real-time, rate, extent

1. INTRODUCTION

Vocal vibrato has proven to be difficult to quantify through objective measurement. In a personal meeting with otolaryngologist Dr. John Ingle, we learned that current stroboscopic and endoscopic imaging methods of the voice do not function properly with the use of vibrato. This requires patients to sing with straight-tone (no vibrato), a practice some voice teachers view as being unhealthy for the voice. Vibrato measurement through sub-glottal pressure is highly invasive, requiring singers to have a tube inserted into their vocal tract or a bag placed over their mouth. This makes audio signal processing techniques ideal for measuring vocal vibrato.

The purpose of a vocal vibrato training system is to help singers adjust their technique in response to visual feedback. Feedback displaying quantitative measurements of the vibrato rate and extent are meant to visually verify when appropriate placement and technique adjustments have been made. This system is not meant to replace the vocal coach and students should not become dependent on the system's feedback. With proper kinesthetic awareness, this system can improve rehearsal efficiency during lessons and individual practice.

The vibrato can be determined by four parameters, which vibrato rate and extent, regularity and the waveform of the undulations, see figure1.

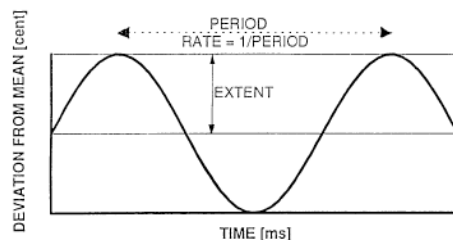


Figure 1 Definition of vibrato parameters

The vibrato rate defines the number of undulations per second. The extent defines how far phonation frequency departs up and down from its average during a vibrato cycle. The regularity shows how similar the frequency excursions are to one another. The waveform is generally more or less similar to a sine wave. Since regularity and waveform are of lesser perceptual relevance than rate and extent. We use rate and extent as the key parameters to judge the quality of vibrato.

Based on a survey conducted by Isidoro Ferrante[2] on vibrato rate and extent with recordings collected from 1900 to 2010, the distributions of vibrato rate (Hz) and extent (cent) are shown in the figure2 and figure3 below.

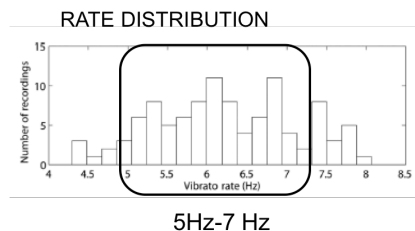


Figure 2 Distribution of vibrato rate

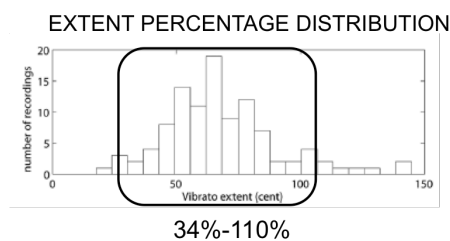


Figure 3 Distribution of vibrato extent percentage

The rate in and extent percentage distribution are reflected in the real-time implementation to determine what is a ‘good vibrato’.

2. VIBRATO IMPLEMENTATION IN MATLAB

The Matlab implementation of vibrato is based on the block diagram, see figure4.

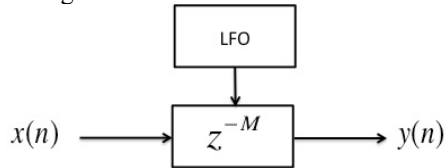


Figure 4 Vibrato Implementation Block Diagram

The primary idea for implementation is the drive the delay line with a low frequency oscillator. In this way the delay samples would vary periodically, which is precisely a vibrato effect. To test the model, we use an 800Hz sine wave as the input with a 6Hz low frequency oscillator produce the vibrato effect. The spectrogram of the input sine wave and the output signal with vibrato effect are plotted, see figure5 and figure6.

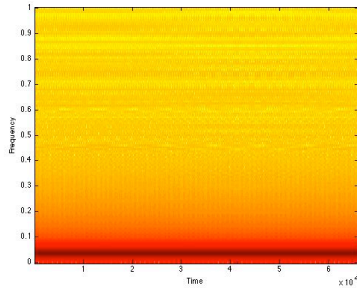


Figure 5 Spectrogram of 800Hz sine wave

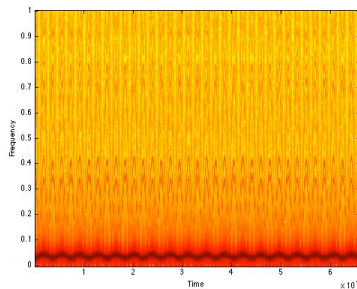


Figure 6 Spectrogram of 800Hz sine wave with vibrato

3. REAL-TIME VOCAL VIBRATO TRAINING SYSTEM IMPLEMENTATION

3.1. Tracking of fundamental frequency

Cepstral method is used to track the fundamental frequency. Assume that a vocal signal is the result of convolution the excitation signal $e[n]$ with the filter. From this perspective, a cepstrum is the result of taking the Inverse Fourier Transform (IFT) of the logarithm of the estimated spectrum of a signal. If the log amplitude spectrum contains regularly spaced harmonics, then the Fourier analysis will show a peak corresponding to the spacing between the harmonics, which corresponds to the fundamental frequency. The value of the fundamental frequency will be used for real-time calculation of rate and extent.

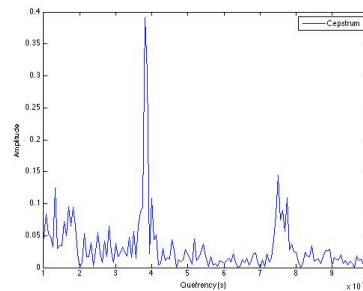


Figure 7 Cepstrum of a vocal signal

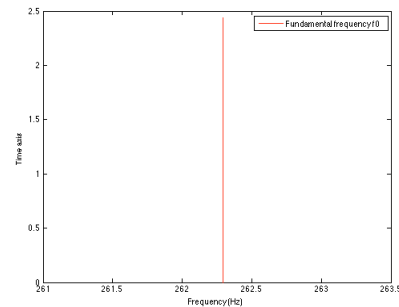


Figure 8 Plot of fundamental frequency

3.2. Rate and extent calculation in real-time

Fourier Transform is taken on the vocal signal first for frequency domain analysis. The peaks are then detected by applying findpeaks function. As defined above, rate is the frequency variation between two local maxima. The average value of the rate is dynamically calculated. The extent is calculated by referring to the extent of the fundamental frequency. The percentage value of extent is calculated block by block to give real-time feedback. Based on the rate and extent distribution of a survey, the rate range of ‘good vibrato’ lies between 5 and 7 Hz, the extent percentage of ‘good vibrato’ lies between 34%-110%. The real-time value has to be in both ranges to get the feedback as a ‘good vibrato’.

3.3. Real-time visual feedback design

The real-time visual feedback is designed based on the shape changing of an ellipse and the position varying of rate and extent sliders. For the ellipse, the horizontal axes correspond to the rate and the vertical axes correspond to the extent. The shape of ellipse will change dynamically due to the variations of rate and extent. The two sliders at the bottom take in the dynamic value of rate and extent. The position of each slider will change correspondingly. When you reach the two ranges set as 'good vibrato', the color of the ellipse will turn green, which indicates a successful vocal training.

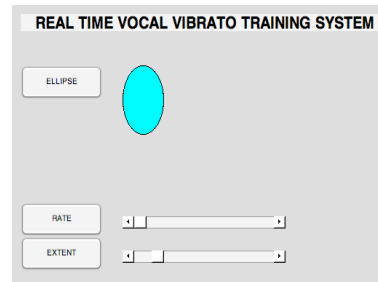


Figure 12 Extent slider position changes in real-time

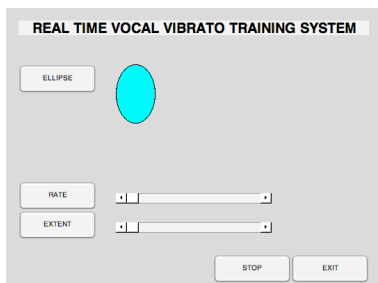


Figure 9 Default settings of real-time vocal vibrato training system

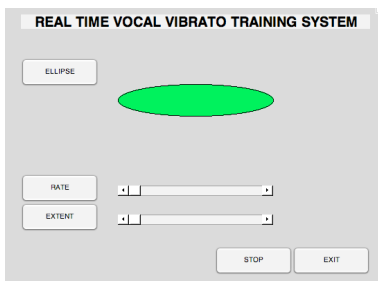


Figure 10 'Good vibrato' indication

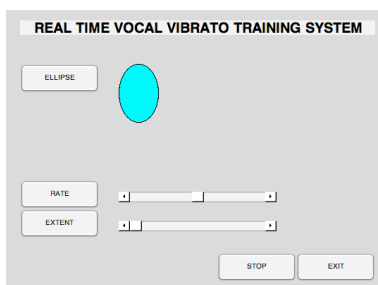


Figure 11 Rate slider position changes in real-time

3.4. Real-time vocal vibrato training system

The real-time training system is implemented by following the flow chart below, see figure 13.

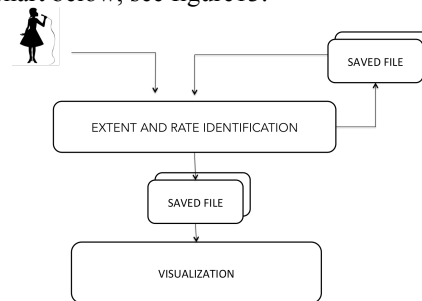


Figure 13 Real-time training system implementation flow chart

Vocal signal from the singer is recorded in real-time. The instant percentage extent and average rate values are calculated in real-time. The instant values further determine the visual feedback, which are the shape of the ellipse and the positions of the sliders.

4. CONCLUSIONS

Initial results show that there is still much to be learned about the cause of vibrato and what is considered to be healthy vibrato usage. Our current training system shows promise based on synthetic vibrato and limited user testing, but technical adjustments still need to be made to more accurately measure the rate and extent. It would be useful to record expert vocal coaches and compare their vibrato parameters to the ideal parameters given by Titze which are implemented in our system.

Future work includes improving measurement accuracy, displaying numeric values for the rate and extent, and creating an interface that allows the vocal coach to set the range of acceptable vibrato rate and extent. Additionally, it would be useful to measure the consistency of vibrato. Inconsistent vibrato or a late onset of vibrato is not desirable from a classical voice perspective, and is often associated with vocal tension. Additional study is also needed to

understand what vocal coaches judge to be healthy vibrato. Long term, the desire is to explore correlations between subjective expert opinions of what is good vibrato and objective measurement of these parameters.

5. REFERENCES

[1] Titze, Ingo, Brad Story, Marshall Smith and Russel Long, "A reflex resonance model of vocal vibrato," Acoustical Society of America, 2002

[2] Ferrante, Isidoro, "Vibrato rate and extent in soprano voice: A survey on one century of singing," University degli studi di Pisa, Pisa, Italy, July 4, 2011.