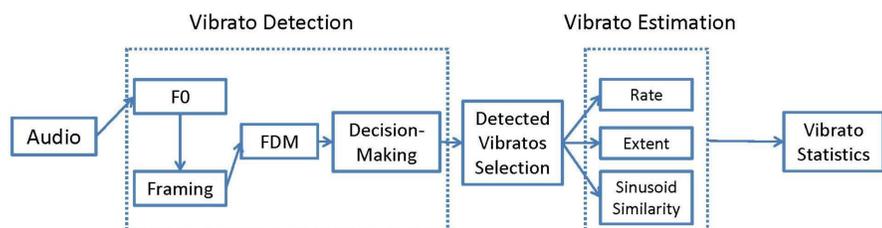


1. Introduction

- Performers display different interpretations of the same piece of music. For scripted music, these differences lie mainly in the expressive dimensions, such as tempo, dynamics, vibrato and portamento. The mathematics underlying these expressive nuances can further performer-style and cross-cultural analyses.
- In current vibrato research, music audio must first be manually segmented into notes, which is a time consuming pre-processing step that does not scale well to large datasets.
- We present an automatic performance analysis system focused on vibrato that can greatly speed systematic research and advance performance pedagogy.

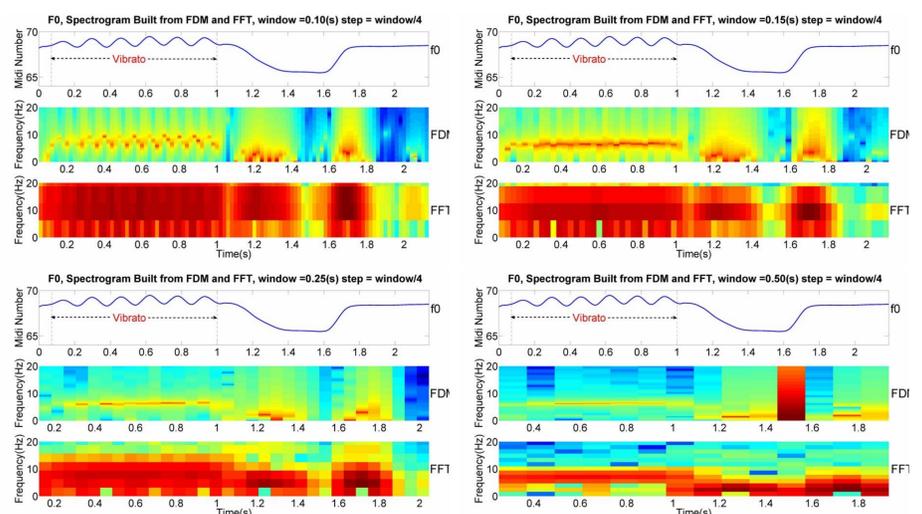
2. Vibrato Detection and Estimation



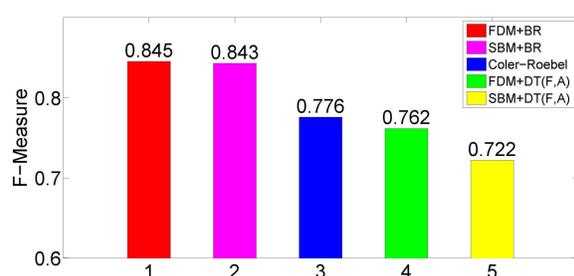
- Apply the Filter Diagonalization Method (FDM) to the f_0 of each frame for sinusoidal decomposition.
- Apply a decision-making mechanism to the frequency and amplitude information of the sinusoid with the largest amplitude in each frame.
 - **Decision Tree:** uses the frequency and amplitude of the sinusoid with the largest amplitude. Requires frequency and amplitude range thresholds.
 - **Bayes' Rule:** assigns a probability to vibrato existence in each frame based on the probability density distribution of the frequency and the amplitude of the sinusoid with the largest amplitude.
- Extract vibrato parameters (rate, extent, sinusoid similarity) from the detected vibratos.

3. Advantages of FDM

- Models time signals as parametric exponentially decayed sinusoids.
- Obtains the frequencies and amplitudes of a given number of sinusoids in a selected frequency range without the limitation of the uncertainty principle in FFT, i.e., it is able to obtain the sinusoids from short frames.
- Extracts the sinusoids directly from the original signal without deriving them from the spectral information, i.e., avoiding an error-prone peak-picking process in the spectrum.

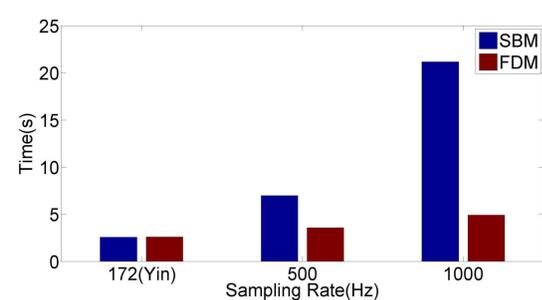


4. Evaluation



• Vibrato Detection

- Dataset has 64 audio passages ranging from 2-12s, including string, woodwind, brass and voice.
- FDM+BR has the best performance.
- FDM based method has better efficiency than widely used parametric damped sinusoid modeling subspace-based method.



• Vibrato Estimation

- Dataset has 64 audio passages ranging from 2-12s, including string, woodwind, brass and voice.
- Synthesized data: rate, 99.28%; extent, 98.32%; sinusoid similarity, 96.81%.
- Manually annotated real data: rate, 96.00%; extent, 90.20%; sinusoid similarity, 81.30%.

5. Conclusions

- Proposed an automatic performance analysis system focused on vibrato using Filter Diagonalization Method and statistics.
- Performs better at vibrato detection and vibrato estimation than state-of-the-art methods. Greatly speeds systematic expressive research.

Acknowledgement

This work is supported in part by the China Scholarship Council.