

Parametric Analysis of Musical Vibrato in Voice and Instrument Performance

Mingfeng Zhang (Univ. of Rochester)

ABSTRACT

Vibrato is an important music performance technique for both voice and various music instruments. In this paper, a signal processing framework for vibrato analysis is presented. In this framework, music vibrato is treated as a generalized descriptor of music timbre and its parameterized feature dimensions are proposed. The authors implemented signal analysis algorithms for segmenting sonic partials from musical sound and performing magnitude-frequency tracking of sonic components. Based on these analysis algorithms, various vibrato parameters are extracted from the magnitude/frequency tracking results. The authors also implemented extensive visualization functionalities that allows in-depth interaction with vibrato feature tracks and their parameterizations..

1. INTRODUCTION

The music vibrato is termed a “pulsation in pitch, intensity and timbre” [1] because of its effectiveness in artistic rendering. However, this sonic expression is still largely treated as a mythology in music conservatories. In music pedagogy, music teachers use demonstration, body gestures, and metaphors to convey their artistic intensions. Modern computer sound analysis tools are seldom employed. In sound design, audio engineers manually inspect the audio spectrogram and manipulate a large amount of vibrato notes to condition them to content context. However, only a blurred magnitude/frequency trajectory can be observed from the audio spectrogram and these interactions are severely limited. Thus the goal of this paper is to demystify the music vibrato by providing extensive signal analysis, pattern recognition and interactive tools and to help the musicians and sonic artist to better explore musical vibrato in an interactive manner.

Our proposed sound analysis methods are based on high-precision magnitude-frequency tracking. We aim at achieving higher analytic resolution compared to the manual inspection of audio spectrogram. Our proposed analysis framework is based on magnitude-frequency tracking of sonic partials. We segment the music tones into harmonic partials using band-passed filters. We then perform magnitude-frequency tracking algorithm from these segmented harmonic partials. In our implementation, each separated partial overtones are modeled as a quasi-monochromatic component, which is a sinusoidal signal with narrow-band magnitude modulation and frequency modulation. Based on this signal model, the magnitude track and the frequency track are extracted from each harmonic partial. The magnitude track and the frequency track are further parameterized to extract descriptive features for visualization.

2. MAGNITUDE-FREQUENCY TRACKING ALGORITHMS

In this section, we introduce algorithms that obtains the magnitude track and the frequency track from each sonic partials of music tones.

2.1. Signal Model

Each segmented sonic partial is modeled as quasi-monochromatic signal, which can be represented as:

$$S(t) = a(t) \cos \phi(t)$$

$$\phi(t) = 2\pi f_c t + 2\pi \int_{-\infty}^t b(t) dt + \phi_0$$

where $a(t)$ and $b(t)$ are two slow-varying random processes with their bandwidth much smaller than the "monochromatic" frequency f_c , ϕ_0 denotes the initial phase.

For convenience, the quasi-monochromatic signal is usually denoted as:

$$S(t) = a(t)e^{j\phi(t)}$$

In the following theory part, we will stick to this definition because using exponential function makes the analysis part more succinct. This definition can be chanced to cos based definition by choosing the real part.

In this signal model, $a(t)$ is called magnitude because it is the envelope of the signal. $b(t)$ is called instantaneous frequency deviation because

$$\frac{d\phi(t)}{dt} = f + b(t)$$

The bandwidth of quasi-monochromatic signal can be calculated as [2]

$$B = \sqrt{\int a'^2(t)dt + \int (2\pi f_c + 2\pi b(t) - \langle \omega \rangle)^2 a^2(t)dt}$$

where $\langle \omega \rangle$ denotes the mean frequency

$$\begin{aligned} \langle \omega \rangle &= \int [2\pi f_c + 2\pi b(t)] |s(t)|^2 dt \\ &= \int [2\pi f_c + 2\pi b(t)] a^2(t) dt \end{aligned}$$

From this representation, we see the bandwidth of $s(f)$ is determined by $a(t)$ and $b(t)$. Because $a(t)$ is a slow-varying random process, and $b(t)$ is small-value the bandwidth B very small, so the spectra content is concentrated in the vicinity of f_c . Thus $s(t)$ acts approximately like a pure sinusoidal signal (which is called monochromatic ,or " one color"). For this reason $s(t)$ is called quasi-monochromatic signal.

2.2. Hilbert Transform Based Magnitude-Frequency Tracking

The detection method based on Hilbert transform, or analytic signal, is a standard method for magnitude-frequency tracking in time-frequency analysis. Suppose that we have the source signal:

$$S(t) = a(t) \cos \phi(t)$$

$$\phi(t) = 2\pi f_c t + 2\pi \int_{-\infty}^t b(t) dt + \phi_0$$

In our implementation, one sonic partial is obtained using bandpass filtering. Figure 1 shows an oboe note before (a) and after filtering (b). The fundamental sonic partial is picked out in this process and serves as $S(t)$. Its Hilbert transform is:

$$\begin{aligned} S_h(t) &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(x)}{x-t} dx \\ &= \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{s(t+x)}{x} dx \end{aligned}$$

Because $S(t)$ is a monochromatic signal, which means the bandwidth of $a(t)$ and $b(t)$ are much smaller than f_c , the Hilbert transform of $S(t)$ can be easily calculated using Bedrossian's theorem as:

$$S_h(t) = a(t) \sin \phi(t)$$

From here, we can easily perform magnitude tracking as:

$$a(t) = \sqrt{S^2(t) + S_h^2(t)}$$

The frequency tracking can be achieved by calculating the phase variations $\phi(t)$ as

$$\phi(t) = \arctan \frac{S_h(t)}{S(t)}$$

from(3) we have the frequency deviation as:

$$\begin{aligned} b(t) &= \frac{1}{2\pi} \frac{d\phi(t)}{dt} - f_c \\ &= \frac{1}{2\pi} \left[\frac{S(t)\dot{S}_h(t) - \dot{S}(t)S_h(t)}{S^2(t) + S_h^2(t)} \right] - f_c \end{aligned}$$

where $\dot{S}(t)$ denote the derivative of $S(t)$.

A related method is called heterodyne filter method. From the representation of quasi-monochromatic signal, we observe that it can be viewed as a baseband signal be modulated to a carrier frequency of f_c . Thus we can "demodulate" it back to baseband to perform detection by multiplying it with the carrier as:

$$\begin{aligned} S_{hs1}(t) &= \cos 2\pi f_c t S(t) \\ &= \frac{1}{2} a(t) \cos \tilde{\phi}(t) + \frac{1}{2} a(t) \cos(4\pi f_c t + \tilde{\phi}(t)) \end{aligned}$$

$$\tilde{\phi}(t) = 2\pi \int_{-\infty}^t b(t) dt + \phi_0$$

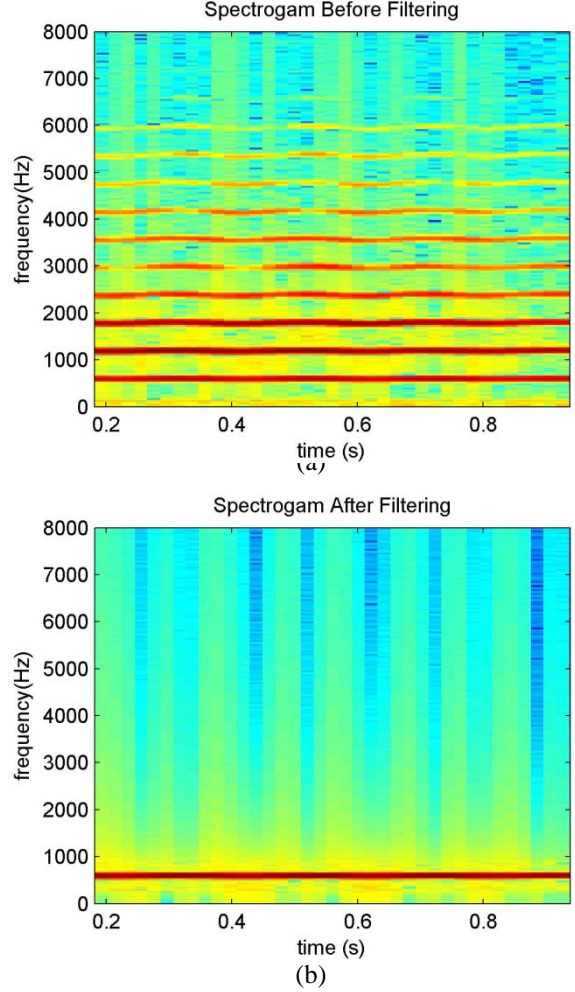


Figure 1. An oboe note before (a) and after filtering (b). The fundamental sonic partial is picked out in this process.

We can then apply a low-pass filter to $S_{hs1}(t)$ to eliminate the $2f_c$ component and result in:

$$S_{h1}(t) = \frac{1}{2} a(t) \cos \tilde{\phi}(t)$$

we similarly form a quadrature component as:

$$S_{hs2}(t) = \sin 2\pi f_c t S(t)$$

Similarly, after low-pass filtering, we have:

$$S_{h2}(t) = -\frac{1}{2} a(t) \sin \tilde{\phi}(t)$$

Using the same procedure as Hilbert transform method, we get the detection results as:

$$\begin{aligned} a(t) &= \sqrt{2} \sqrt{S_{h1}^2(t) + S_{h2}^2(t)} \\ b(t) &= \frac{1}{2\pi} \frac{d\tilde{\phi}(t)}{dt} \\ &= \frac{1}{2\pi} \left[\frac{S_{h1}'(t)S_{h2}(t) - S_{h1}(t)S_{h2}'(t)}{S_{h1}^2(t) + S_{h2}^2(t)} \right] \end{aligned}$$

From these derivations, we see that the heterodyne filter method is equivalent to Hilbert transform method. The only different is that the heterodyne filter method conduct all the processing in the baseband, in the vicinity of zero frequency; while Hilbert transform method performed all processing in the pass-band around f_c .

3. TRACKING RESULTS

3.1 Magnitude-Frequency Tracking

The detection results of analytical signal method are illustrated in Figure 2. Figure 2 (a) is the time domain audio waveform. Figure 2 (b) is the instantaneous amplitude extracted from the fundamental sonic partial, which is similar to the signal envelop in Figure 2 (a). Figure 2 (c) is the instantaneous frequency extracted from the fundamental sonic partial.

In Figure 3 we illustrate the AM/FM detection results using alternative methods. Figure 3 (a) is the signal amplitude obtained by block processing. We split the signal of the fundamental component to 20ms analysis frame. Then we calculate signal energy in every 20ms frames. The analysis resolution here is low than analytical signal method because we have to use a signal frame of enough length to calculate its energy. Figure 3 (b) is the instantaneous frequency detected by counting zero-crossing rate. In this method we first normalize the magnitude of the signal component. Then we calculate the rate the signal value oscillate across zero as instantaneous frequency. From Figure 3 we observe that we achieved similar results as in Figure 2, however, the result using analytical signal method (Figure 2) have better resolution.

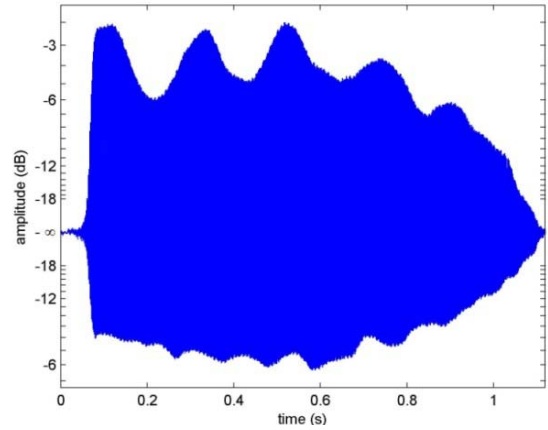
From the audio waveform and the extracted feature tracks in Figure2. We can observe that the magnitude tracking result faithfully retained the shape of the audio magnitude. In the future, We also plan to apply auditory models for human response. The frequency track can also be observed in audio spectrogram (Figure 1) as the oscillation curve in sonic partial but that trajectory is blurred. Our extracted frequency track conforms the shape of sonic partials, while provides much higher time resolution.

3.2 Correlation between Magnitude and Frequency

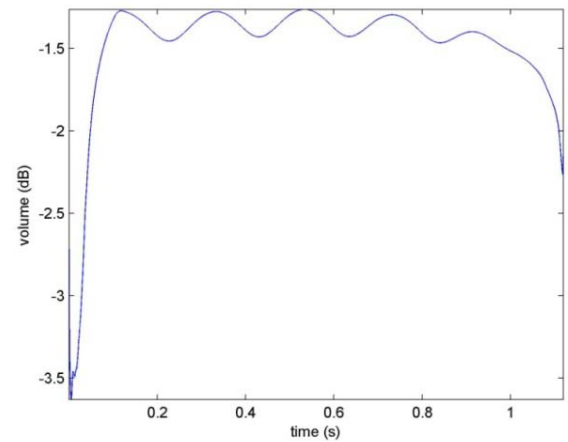
Figure 4 illustrates the correlation trajectory between instantaneous amplitude and instantaneous frequency, by plotting a point at time at the coordinate location of the amplitude-frequency space. From Figure 4 we can observe that the volume variation and the pitch variation are well synchronized.

4. VISUALIZATIONS

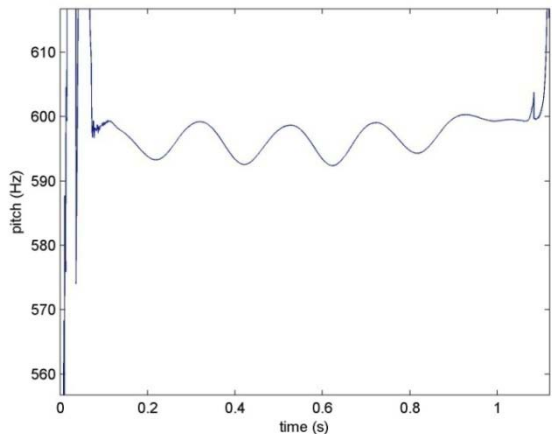
In our implementation, we provided visualization options perform in-depth analysis of the extracted magnitude-frequency tracks and compare vibrato parameters. These visualization functions are nice for interactions.



(a)

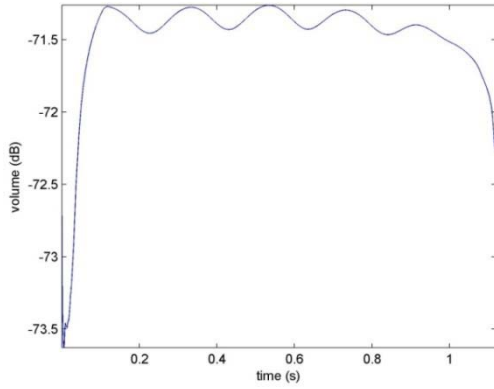


(b)

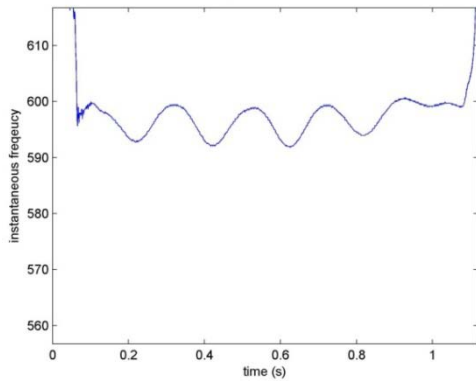


(c)

Figure 2: A visualization example of musical vibrato features. (a) is the the time domain audio waveform. (b) is the the volume variation with time extracted from the fundamental sonic partial, which is similar to the signal envelop in (a). (c) is the pitch variation with time extracted from the fundamental sonic partial. We can observe that the magnitude tracking result faithfully retained the shape of the audio magnitude. The frequency track can also be observed in audio spectrogram (Figure 1), but we provide better time resolution.



(a)



(b)

Figure 3: AM/FM detection Results Using Alternative Methods. (a) is the signal amplitude obtained by calculating signal energy in every 20ms frames. (b) is the instantaneous frequency detected by counting zero-crossing rate.

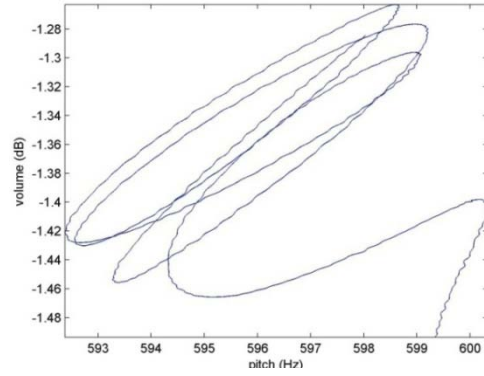


Figure 4: A visualization the correlation trajectory between instantaneous amplitude and instantaneous frequency.

In time-domain, our toolbox provide three-dimension visualization functions that enable us to compare multiple sonic partials. In Figure 5 we provide a visualization example that compares the magnitude track and the frequency track extracted from the fundamental partial of an oboe note. The instantaneous frequency, magnitude and the time location decide the spatial coordinates of a visualization element.

In Figure 6 we provided a frequency domain comparison of the magnitude track and the frequency track extracted from the same sonic partial. The magnitude, delay and the frequency decide the spatial locations. This visualization provides a complete signal profile in frequency domain. This figure get all these features in 3d, we call it roller-coaster graph, the phase delay is the difference in fft phase between AM and FM,

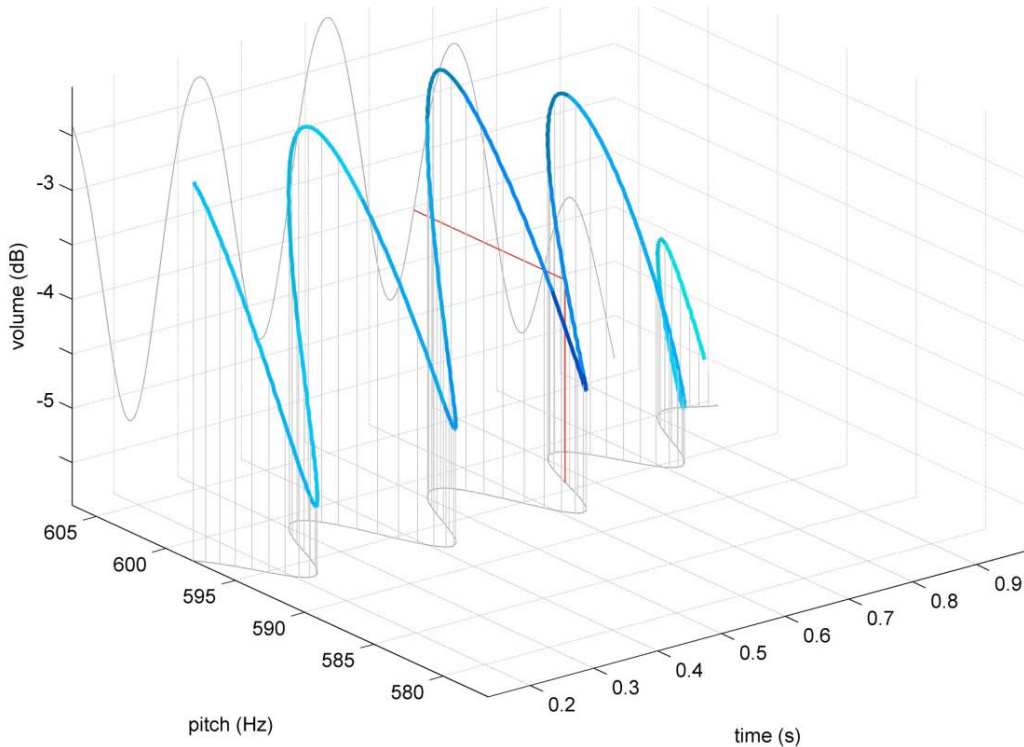


Figure 5: A visualization example of musical vibrato features that compares the variational pattern of multiple feature dimensions. in a three dimensional space. The pitch deviation, loudness and the feature time location decide the spatial coordinates of a visualization element. $f_0 = 596.75\text{Hz}$, Vibrato frequency = 5.02 Hz.

a " π " delay means inverse phase, a "0" delay mean synchronized here, the 5 HZ component have a phase delay of around "0", "0" here is no delay, means that the AM, FM is very synchronized.

This analysis/visualization framework can also be utilized to analyze feature tracks extracted from different harmonic partials in both the time domain and the frequency domain. A group of ensemble visualization functions are also implemented to analyze a group of music files, or comparing parameters extracted from multiple harmonic partials from the same music tone. For example, we can also analyze the magnitude-frequency tracks from multiple harmonic partials from a music note. In Figure 7, we analyze the sonic partials from a Saxophone tone. Here we plot the vibrato parameter of modulation index. The term modulation index describes how deep the modulation is. The signal with high peak-valley distance will have a higher modulation index, it is defined as:

$$M = \frac{\frac{1}{T} \sum_{i=1}^T P_i}{V}$$

Here P_i is several peak-valley distance obtained from the vibrato waveform. Here we use MIR toolbox for peak-picking. V is the complete dynamic range. Here we take the mean value for peak-valley distance to make this measurement more robust. The modulation index for magnitude track is called "AM Index", The modulation index for frequency track is called "FM Index". These two terms conform to related terms in amplitude/frequency modulation in communication engineering.

Figure 7 (a) is the magnitude tracking results for this music note. Here we only show part of the results. (b) is AM index for 20 harmonic partials. (c) is their FM index, in cent, a cent is 1/100 (one-hundredth) of a semitone. Musicians like this unit because it conveys the deviation from normal pitch grid and thus intuitive.

We also implemented batch processing functions to compares multiple music notes. Using this visualization framework, we can also analyze other sound features from a group of audio files. In Figure 8 as an example, we analyses a group of trumpet tones. We have 35 trumpet tone here, their fundamental frequency from E3 to D6. Here we analyze the modulation indices for the 1st harmonic partials of these 35 trumpet tones. In Figure 8 (a) we get AM indices for each note, each note is placed at the frequency location of their fundamental frequency. In Figure 8 (b) we get FM index in percentage for each note. This "ensemble analysis" let us effective navigate a sound database.

5. CONCLUSIONS

We achieve high analysis precision for both magnitude tracking tasks and frequency tracking tasks using the detection method based on analytic signals. Based on the analysis methods of individual sonic partials, we generalize the analysis to complex harmonic structures and multiple music notes. We also implemented visualization system for analysis. The proposed analysis tools serves as a musical communication to convey musical concepts in an intuitive manner, while maintaining solid signal processing background, thus is suitable for various applications that require high analytic precision such as physical modeling based re-synthesis, electronic instrument timbre design, music performance analysis, and music cognition experimentations.

6. REFERENCES

- [1] C. E. Seashore, *The Vibrato*. University of Iowa studies, New series, no. 225. 1932.
- [2] L. Cohen, *Time-Frequency Analysis*, Prentice Hall, 1995.

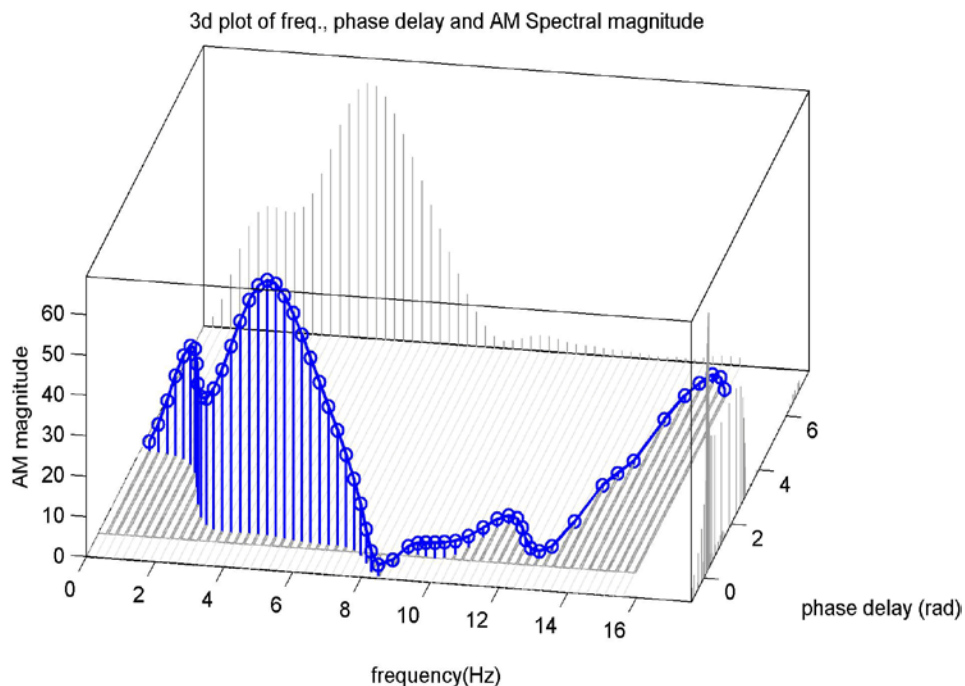
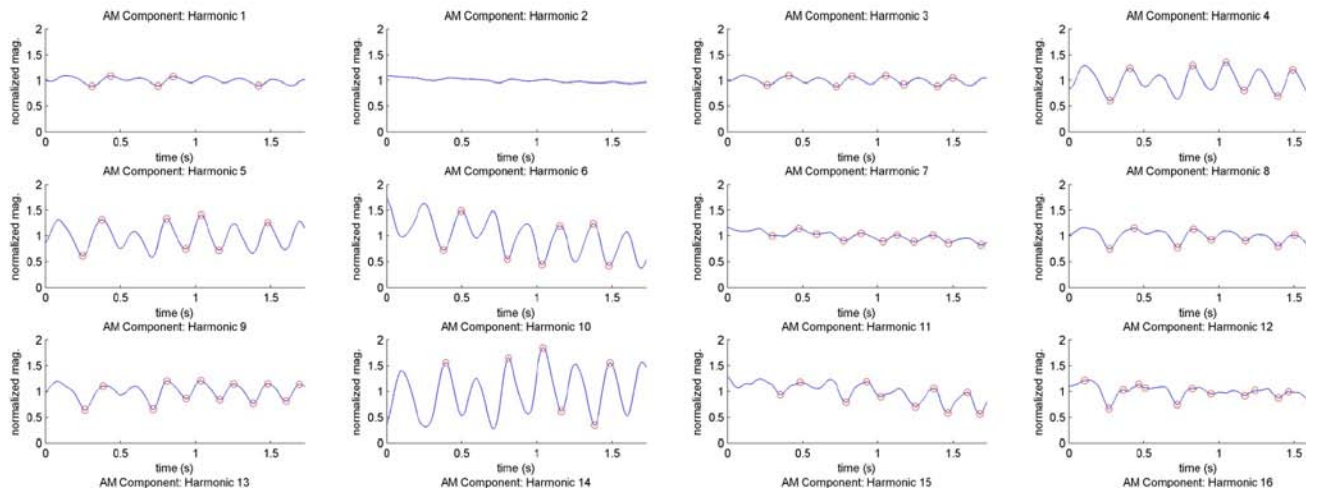
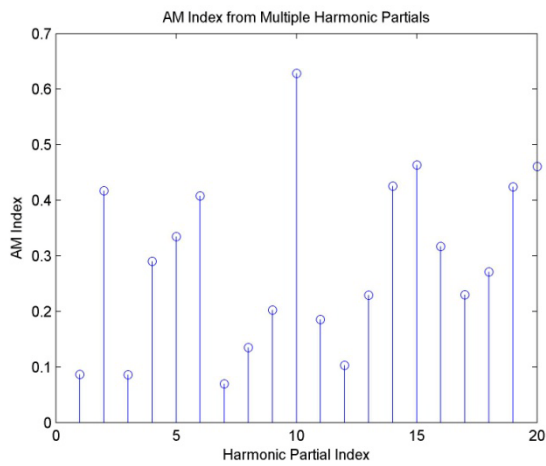


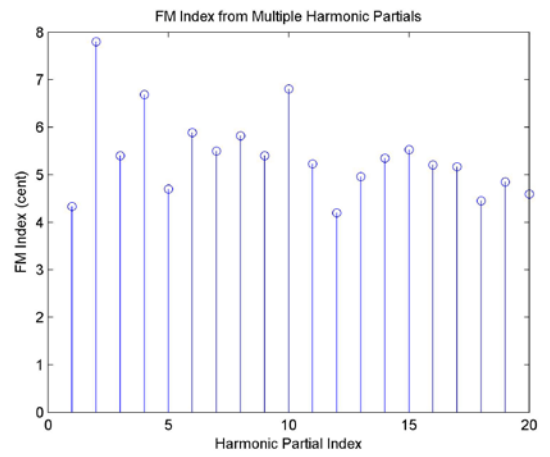
Figure 6: A visualization example of musical vibrato features that compares the variation pattern of multiple feature dimensions in frequency domain. in a three dimensional space. The spectra of pitch deviation, and magnitude decide the spatial coordinates of a visualization element.



(a)

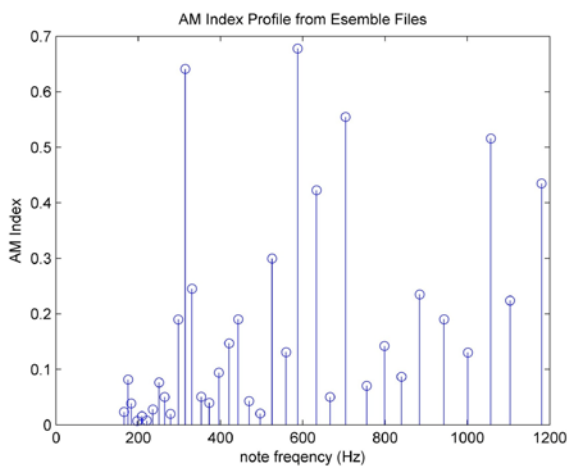


(b)

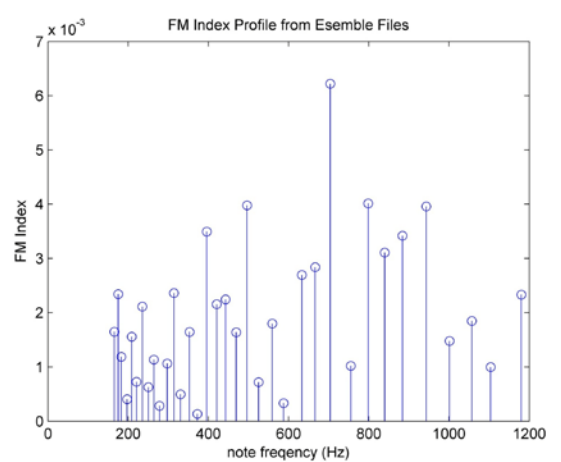


(c)

Figure 7: An ensemble visualization example of musical vibrato features. This is a Alto Saxphone note with a fundamental frequency of 173.9Hz. (a) is the magnitude tracking results for this music note. Here we only show part of the results. (b) is AM index for 20 harmonic partials. (c) is their FM index, in cent, a cent is 1/100 (one-hundredth) of a semitone.



(a)



(b)

Figure 8: Analysis multiple audio files and makes comparisons. This is a a group of 35 trumpet tones with fundamental frequency from E3 to D6. Here we analyze the modulation indices for the 1st harmonic partials In Figure (a) we get AM indices for each note, each note is placed at the frequency location of their fundamental frequency. In (b) we get FM index in percentage for each note.