AES 143rd Convention Student Technical Papers Award

Audio Engineering Society

Convention Paper 9885

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

Modeling the effects of rooms on frequency modulated tones

Sarah R. Smith[1] and Mark F. Bocko[1]

[1] *University of Rochester, Department of Electrical and Computer Engineering*

Correspondence should be addressed to Sarah Smith (`sarahsmith@rochester.edu`)

ABSTRACT

This paper describes how reverberation impacts the instantaneous frequency tracks of modulated audio signals. Although this effect has been observed in a number of contexts, less work has been done relating these deviations to acoustical parameters of the reverberation. This paper details the instantaneous frequency deviations resulting from a sum of echoes or a set of resonant modes and emphasizes the conditions which maximize the resulting effect. Results of these models are compared with the observed instantaneous frequencies of musical vibrato tones filtered with the corresponding impulse responses. It is demonstrated that these reduced models may adequately reproduce the deviations when the signal is filtered by only the early or low frequency portion of a recorded impulse response.

1 Introduction

Instantaneous frequency tracking is used in a wide range of signal processing applications including audio source separation and other music information retrieval tasks. In particular, many algorithms for source separation use the principle of common modulation to assign frequency components to individual sources. Yet, past work has demonstrated that the presence of reverberation in a signal can alter the observed frequency tracks, reducing the correlation among harmonic partials from the same source [1, 2]. Although these reverberant effects may impede the analysis of the source signals, they also contain information about the associated acoustic system. For example, deviations in the overtone frequency or phase trajectories of instrumental tones can be used in instrument identification [3]. This paper demonstrates how certain acoustic parameters impact the instantaneous frequency of modulated tones, thereby enabling the use of instantaneous frequency analysis to characterize resonant systems including rooms or musical instruments.

Past observations of reduced overtone correlations have appeared in the context of both instrumental and vocal vibrato analysis. Anomalies in the overtone frequencies of violin tones performed with vibrato were noticed as early as the 1990s, when it was observed that higher overtones of these notes often exhibit a greater modulation width than their lower counterparts [4]. More recently, inconsistencies between the instantaneous phases of instrumental overtones have been shown to vary significantly between instruments of different classes, with string instruments exhibiting more deviations than woodwind or brass instruments [3]. The connection between reverberation and these deviations was first made in the context of vocal vibrato with the intention of developing an analysis algorithm that would extract the source frequency modulation and minimize the reverberant deviations [1].

After presenting the sinusoidal model used in this analysis, the paper is organized as follows. Section 2 provides a summary of the relevant time-frequency concepts related to tracking instantaneous frequency and defines an estimate of instantaneous frequency in terms of the phase of the short-time Fourier transform. Sections 3 and 4 model the frequency deviations resulting from a set of discrete reflections or a sum of resonant modes and compare these predictions to to the observed deviations when a tone is filtered by only the early or low frequency portion of a recorded impulse response.

1.1 Sinusoidal signal model

The sustained portion of many audio signals is well modeled as the sum of sinusoidal components, often referred to as partials, each of which may have a time varying amplitude and phase as shown in (1). Each component $s_n(t)$ of the overall signal $s(t)$ is said to have an instantaneous amplitude $a_n(t)$ and total phase $\omega_n t + \phi_n(t)$. In general, the modulations $a_n(t)$ and $\phi_n(t)$ are assumed to vary slowly compared to the center frequency ω_n.

$$s(t) = \sum_{n=1}^{N} s_n(t) = \sum_{n=1}^{N} a_n(t)\cos(\omega_n t + \phi_n(t)) \quad (1)$$

For a signal of this form, the instantaneous frequency of the n$^{\text{th}}$ partial, $f_{i,n}(t)$ is logically defined as the time derivative of the total phase, as given in (2). Since the human ear is not sensitive to the absolute phase of a sustained tone, the phase modulation can expressed as an equivalent frequency modulation $\Delta f_n(t) = \frac{1}{2\pi}\frac{d\phi_n(t)}{dt}$.

$$f_{i,n}(t) = \frac{1}{2\pi}\omega_{i,n}(t) = \frac{1}{2\pi}[\omega_n + \frac{d\phi_n(t)}{dt}] \quad (2)$$

Variations of this model have been widely used in both analysis and synthesis applications for many decades due to their relatively compact representation and ease of resynthesis [5, 6]. Although this model does not assume any relationship between the sinusoidal components, many naturally occurring sounds contain sinusoidal components that are harmonically related, in which case $f_n(t) = n \cdot f_1(t)$. Often, such sounds result from the simultaneous excitation of multiple modes of an acoustic generator (such as an air column or tensioned string) and modifying the system parameters affects all of the acoustical modes proportionally [7].

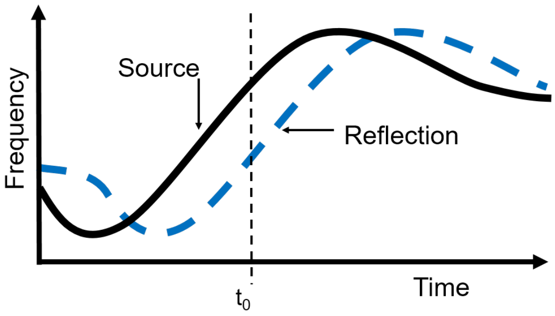


Fig. 1: Instantaneous frequency vs time of a source signal and a single echo

When an instrumentalist performs vibrato, both the amplitude and frequency of the tone are modulated quasiperiodically. While the extent and rate of vibrato as well as the relative amounts of amplitude and frequency modulation vary significantly across instruments and playing styles, the frequency modulation extent generally ranges from about 10 cents to a semitone (between 0.1 and 5 percent change in frequency) and varies at a rate of less than 10 Hz [8, 9].

2 Theory

For the signal described above, each component, $s_n(t)$, consists of a single sinusoid, whose instantaneous frequency can be intuitively defined. This definition, in turn, corresponds well to the perceived pitch of the tone. However, when a signal is filtered, each sinusoidal component is converted into a narrow band multicomponent signal as the frequency content spreads out in time. For example, consider the case of a frequency modulated sinusoid along with a single echo, as shown in Fig. 1. If the frequency of the source changes during the time it takes for the echo to arrive, the resulting sound now contains two sinusoidal components that must be modeled with a single amplitude and frequency trajectory. Since the direct signal and its echo generally cannot be resolved in time or frequency, the corresponding instantaneous frequency estimate is no longer as intuitive as it was in the earlier example.

This inconsistency has been addressed many times in the literature. Often, the instantaneous frequency is defined as the time derivative of the phase of the corresponding analytic signal ([10], ch2). However, defining instantaneous frequency in this manner can quickly lead to non-physical results, including instantaneous frequencies that are unbounded or far outside the bandwidth of the original signal [11]. One method of avoiding these issues is to define instantaneous frequency as a statistical moment of a joint time frequency distribution [12, 13]. Our chosen algorithm includes elements of both of these strategies and estimates the instantaneous frequency of a signal from the phase of its short-time Fourier transform as described below.

2.1 Instantaneous frequency from the short-time Fourier transform

The continuous-time short-time Fourier transform (STFT) of a signal s(t) is defined as $S(t_0,\omega_e)$ and given below in (3), where w(t) is a window function of short duration, centered around $t=0$. Here, ω_e and t_0 correspond to the evaluation frequency and center time of the STFT frame respectively.

$$S(t_0,\omega_e)=\int_{-\infty}^{\infty}\mathrm{w}(t-t_0)s(t)e^{-j\omega_e t}dt \tag{3}$$

Although the choice of window may affect the tracking accuracy and resolution of the spectrogram, the fundamental results presented in this paper are not influenced. For the data presented in later sections, a 20 ms long Hann window was used with a 2 ms hop between frames. The large overlap aids the tracking algorithm by reducing the expected change in frequency between adjacent frames. If the amplitude and frequency modulations vary slowly with respect to the frame length, the STFT of the source signal of (1) can be represented as in (4) where $\mathrm{W}(\omega)$ is the Fourier transform of the analysis window.

$$S(t_0,\omega_e)=\sum_{n=1}^{N}\tfrac{1}{2}a_n(t_0)\mathrm{W}(\omega_{i,n}(t_0)-\omega_e)e^{j\phi_n(t_0)+j(\omega_n-\omega_e)t_0} \tag{4}$$

Provided that the partials are separable given the frequency resolution of the associated window [14], and an initial estimate of the fundamental frequency is known or calculated using a pitch tracker [15], then it is possible to choose an evaluation frequency $\omega_{e,n}$ for each partial to minimize the cross talk between partials. The sum in (4) can then be removed since the magnitude of the window transform becomes negligible for all but the nearest partial. The STFT at each of these chosen evaluation frequencies can then be separated into its magnitude $|S(t_0,\omega_{e,n})|$ and phase $\Phi(t_0,\omega_{e,n})$ as given in (5) and (6).

$$|S(t_0,\omega_{e,n})|=\tfrac{1}{2}|a_n(t_0)\mathrm{W}(\omega_{i,n}(t_0)-\omega_{e,n})| \tag{5}$$

$$\Phi(t_0,\omega_{e,n})=\angle S(t_0,\omega_{e,n})=\phi_n(t_0)+(\omega_n-\omega_{e,n})t_0 \tag{6}$$

For a window with a symmetric spectrum, the instantaneous frequency $\omega_i(t_0)$ can be found as either the peak (mode) or expected value (mean) of the magnitude distribution at time t_0 [10]. This, however, requires calculating the STFT across a large number of frequency bands. Alternatively, the instantaneous frequency can be estimated from the STFT phase as shown in (7). This is the definition used in the remainder of this paper.

$$\omega_{i,n}(t_0)=\omega_{e,n}+\frac{d}{dt_0}\Phi(t_0,\omega_{e,n}) \tag{7}$$

2.2 Convolution and the STFT

We now consider a reverberant signal $x(t)$ which consists of the source signal $s(t)$ described above convolved with an impulse response $h(t)$.

$$x(t)=s(t)*h(t)=\int_{-\infty}^{\infty}h(\tau)s(t-\tau)d\tau \tag{8}$$

When the STFT of this signal is evaluated as above, the resulting expression takes the form of (9).

$$X(t_0,\omega_e)=\int_{\tau}h(\tau)e^{-j\omega_e\tau}\int_{t}s(t-\tau)\mathrm{w}(t-t_0)e^{-j\omega_e(t-\tau)}dtd\tau \tag{9}$$

This can be further simplified by recognizing the inner integration over t as the STFT of the source signal evaluated at $t_0-\tau$, which leads to a representation of the filtered STFT, $X(t_0,\omega_e)$, as a convolution in time of the source STFT with a demodulated version of the impulse response as given in (10).

$$X(t_0,\omega_e)=S(t,\omega_e)\underset{t}{*}\left(h(t)e^{-j\omega_e t}\right) \tag{10}$$

This result, combined with the instantaneous frequency calculation defined above, can be used to estimate the effects of different parametric models for $h(t)$ on the frequency calculations of a known signal. In this way, we can relate the observed instantaneous frequency deviations to a set of relevant acoustical parameters.

3 Effects of a set of echoes

3.1 Sum of echoes

The early portion of many room impulse responses is dominated by isolated reflections off of walls in the space. In these cases, the impulse response $h(t)$ is well represented parametrically as a sum of delayed delta functions, each with a corresponding reflection coefficient r_p and time delay d_p as given in (11).

$$h(t) = \sum_p r_p \delta(t - d_p) \tag{11}$$

When this impulse response is substituted into (10), the convolution with each delta function selects out the corresponding time index of the source STFT, allowing $X(t_0, \omega_e)$ to be expressed as in (12).

$$X(t_0, \omega_e) = \sum_p r_p S(t_0 - d_p, \omega_e) e^{-j\omega_e d_p} \tag{12}$$

In order to isolate the change in instantaneous frequency due to the impulse response, it is helpful to consider the ratio of $X(t_0, \omega_e)$ to $S(t_0, \omega_e)$ as in (13). The phase of this quotient then corresponds to the difference in instantaneous phase between $X(t_0, \omega_e)$ and $S(t_0, \omega_e)$, and its derivative produces the change in instantaneous frequency.

$$\frac{X(t_0,\omega_e)}{S(t_0,\omega_e)} = \sum_p r_p \frac{|S(t_0-d_p,\omega_e)|}{|S(t_0,\omega_e)|} e^{j(\Phi(t_0-d_p,\omega_e)-\Phi(t_0,\omega_e)-\omega_e d_p)} \tag{13}$$

At this point, the parameters $\alpha_p(t_0)$ and $\theta_p(t_0)$ are introduced to describe the relative contribution of each echo to the overall amplitude and phase of the STFT.

$$\begin{aligned}\alpha_p(t_0) &= r_p \frac{|S(t_0 - d_p, \omega_{e,n})|}{|S(t_0, \omega_{e,n})|} \\ &= r_p \frac{a_n(t_0 - d_p)}{a_n(t_0)} \frac{\mathrm{W}(\omega_{i,n}(t_0 - d_p) - \omega_{e,n})}{\mathrm{W}(\omega_{i,n}(t_0) - \omega_{e,n})}\end{aligned} \tag{14}$$

$$\begin{aligned}\theta_p(t_0) &= \Phi(t_0 - d_p, \omega_{e,n}) - \Phi(t_0, \omega_{e,n}) - \omega_{e,n} d_p \\ &= \phi_n(t_0 - d_p) - \phi_n(t_0) - \omega_n d_p\end{aligned} \tag{15}$$

Examining the form of (14) and (15) it is noted that $\theta_p(t_0)$ no longer depends on the evaluation frequency ω_e and that $\alpha_p(t_0)$ consists of three distinct components: The reflection coefficient r_p, the change in amplitude of the signal $a_n(t)$, and a term related to the change in amplitude of the window spectrum as a result of the frequency modulation. This allows (13) to be rewritten as (16).

$$\frac{X(t_0, \omega_e)}{S(t_0, \omega_e)} = \sum_p \alpha_p(t_0) e^{j\theta_p(t_0)} \tag{16}$$

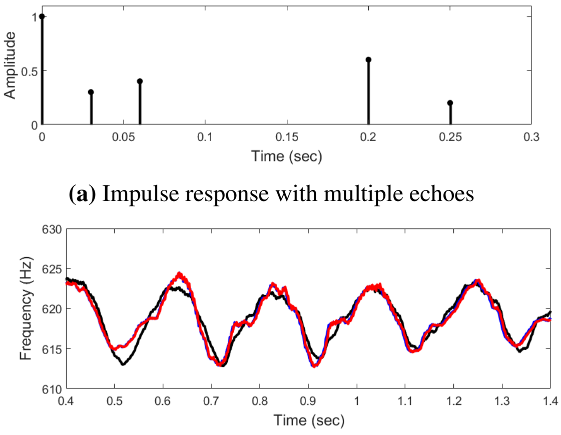
From here, the total phase can be extracted as the inverse tangent of the ratio of the imaginary to real parts of the expression. Taking a derivative in time then yields the following expression for the instantaneous frequency deviation, $\Delta\omega(t_0)$.

$$\Delta\omega(t_0) = \frac{\sum_p \sum_q \alpha_p \alpha'_q sin(\theta_q - \theta_p) - \alpha_p \alpha_q \theta'_q cos(\theta_q - \theta_p)}{\sum_p \sum_q \alpha_p \alpha_q cos(\theta_q - \theta_p)} \tag{17}$$

3.2 Verifying the multi echo model

In order to verify the model described above, a series of saxophone vibrato tones recorded in an anechoic environment were used [16]. The saxophone examples were chosen for their pronounced vibrato and a bright timbre which produces many loud partials. Additionally, the frequencies of saxophone partials are not affected by any large body resonances, as is typical of string instrument tones [2]. For each tone, the amplitudes and frequencies were tracked for the first 5 partials using the phase of the STFT as described above. At each frame, the values for $\omega_{e,n}$ were chosen for each partial to track the yin pitch estimate and magnitude peak in the STFT. These tones were then combined with a set of synthetic impulse responses in order to verify the accuracy of the model and compare the effects of different echo parameters.

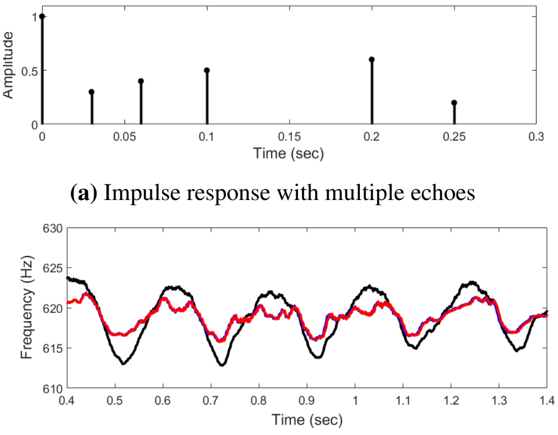
For the first example, shown in fig 2, an impulse response was generated to include the direct sound and



(a) Impulse response with multiple echoes

(b) Resulting instantaneous frequency track

Fig. 2: Instantaneous frequency track of the anechoic sound (black) compared with the filtered version (blue) and model predictions(red) for the impulse response shown. No significant deviations occur



(a) Impulse response with multiple echoes

(b) Resulting instantaneous frequency track

Fig. 3: Instantaneous frequency track of the anechoic sound (black) compared with the filtered version (blue) and model predictions(red) for the impulse response shown. When a reflection is added at half the modulation period, significant deviations are introduced

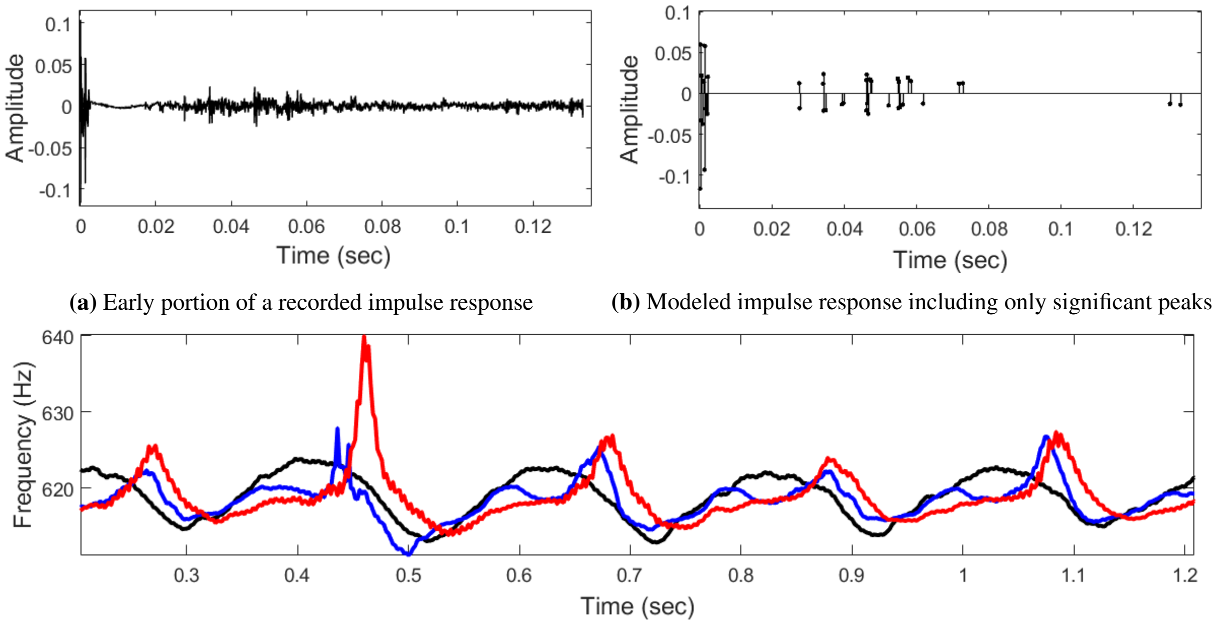
four echoes at varying intensities. The modeled frequency track, calculated using (17) is shown in red and coincides well with the results obtained from analyzing the filtered signal, which are shown in blue. Although some deviations in instantaneous frequency can be observed for this impulse response, the effects are minimal when compared with later examples. This results from the fact that the reflections of the impulse response correspond to delays where the frequency of the original signal has not changed significantly. The first pair of reflections occur very shortly after the direct sound and the later reflections coincide roughly to one period of the vibrato, where the source frequency has returned to its previous value.

In contrast, when an additional echo is added to the impulse response with a delay of 0.1 sec, corresponding to half the modulation period, larger instantaneous frequency deviations appear, as shown in fig 3. This result can also be understood from the form or (17). When a reflection is placed so as to maximize the difference in frequency, this corresponds to a maximum in $\theta_q'(t_0)$ in the numerator and dominates the contribution to the overall deviation.

3.3 Early reflections: model vs actual

Given the importance of isolated reflections to the sound of the early portion of many impulse responses, it is worth investigating their contribution to the instantaneous frequency deviations associated with real acoustic environments. For this purpose, a set of impulse responses with prominent reflections were analyzed [17]. Here, a prominent reflection was defined as any peak in the impulse response whose amplitude was greater than 10 percent of the amplitude of the direct sound. The early portion of the impulse response was then defined to include all samples prior to the last significant reflection. The 10 percent threshold was selected so as to capture many salient features of the early response without including samples from the later portion of the response. In the example shown in fig 4, this choice results in the first 130 ms of the impulse response being used in the full analysis. Although this is somewhat longer that the more commonly used 80 ms threshold, it remains plausible, given the particularly strong reflections in this example.

In order to asses the relative importance of early reflections in generating instantaneous frequency deviations, the source signal was convolved with the early portion of the response shown in fig 4a and the multi echo model of the previous section was calculated using the amplitudes and delays of only the significant peaks as seen in fig 4b. These results are shown fig 4c where the black line represents the instantaneous frequency of the



(a) Early portion of a recorded impulse response

(b) Modeled impulse response including only significant peaks

(c) Resulting instantaneous frequency trajectories

Fig. 4: Instantaneous frequency tracks of an anechoic saxophone tone (black) compared with those of the same tone after it was filtered with a recorded impulse response (blue) or a reduced model(red)

anechoic source signal and the red and blue frequency tracks correspond to the model results and observed deviations due to the full early response respectively. Although the results do not agree completely, many of the largest frequency deviations are well predicted using the early reflection model.

4 Effects of a sum of resonant modes

The low frequency regions of many rooms are dominated by the resonant modes of the space. Each of these modes can be represented as a decaying sinusoid defined by a decay time τ_m, resonant frequency ω_m as well as an initial amplitude b_m and phase ϕ_m. The impulse response corresponding to this set of modes is then given by (18) where $U(t)$ is the unit step function to ensure that the impulse response is causal.

$$h(t) = \sum_m b_m e^{-t/\tau_m} \cos(\omega_m t + \phi_m) U(t) \qquad (18)$$

When this expression is substituted into (10) and a factor of $S(t_0, \omega_e)$ is removed as before, the result is given in (19). In contrast to the case of reflections discussed previously, the convolution integral can no longer be solved analytically, but can be evaluated numerically when the model is implemented.

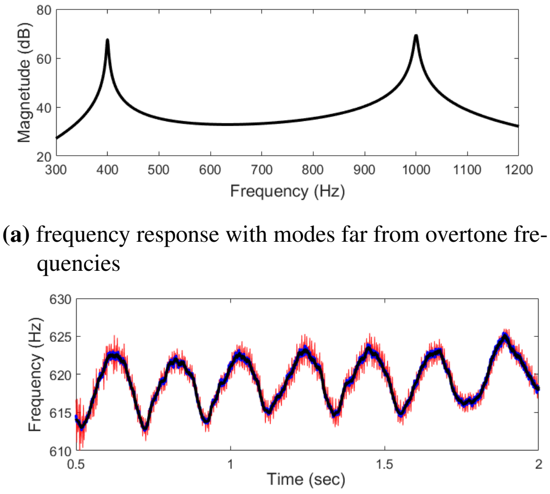
$$\frac{X(t_0,\omega_e)}{S(t_0,\omega_e)} = \int_{-\infty}^{t_0} \frac{|S(t,\omega_e)|}{|S(t_0,\omega_e)|} \sum_m \alpha_m(t,t_0) e^{j\theta_m(t,t_0)} dt \qquad (19)$$

The relative contribution of each mode to the overall STFT can be divided into amplitude and phase terms $\alpha_m(t,t_0)$ and $\theta_m(t,t_0)$ as defined in (20) and (21).

$$\alpha_m(t,t_0) = b_m e^{j\phi_m} e^{\frac{(t-t_0)}{\tau_m}} \qquad (20)$$

$$\theta_m(t,t_0) = (\omega_n - \omega_m)(t - t_0) + \phi_n(t) - \phi_n(t_0) \qquad (21)$$

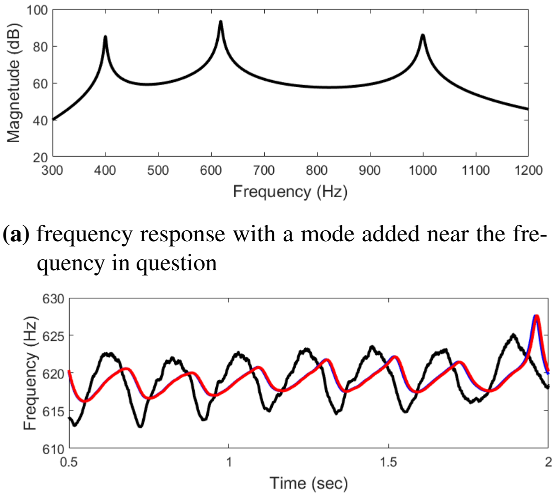
The relative amplitude contribution $\alpha_m(t,t_0)$, now inside the convolution integral, incorporates both the magnitude and decay rate of the relevant mode. Once again, the phase term, does not depend on ω_e. Instead, a factor



(a) frequency response with modes far from overtone frequencies

(b) Resulting instantaneous frequency track

Fig. 5: resonant modes do not affect the overtones far away from their center frequencies



(a) frequency response with a mode added near the frequency in question

(b) Resulting instantaneous frequency track

Fig. 6: When a mode is added near the center frequency of a given overtone, significant deviations are introduced

of $\omega_n - \omega_m$ appears in the exponent. Since the integral of e^{ax} is always proportional to $\frac{1}{a}$, this term implies that the modes with resonant frequencies closest to ω_n will have the most pronounced effect on the frequency deviations, which corresponds with the observed trends.

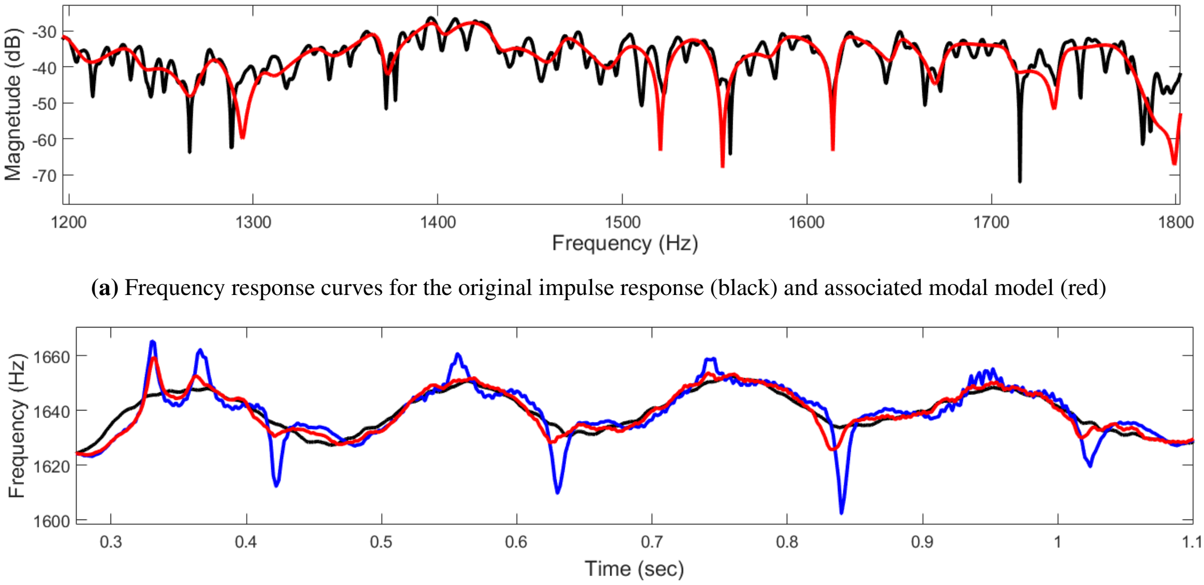
4.1 Verification of the multi mode model

To evaluate the multi mode model, the same saxophone tone as above was filtered with 2 different modal impulse responses. The first impulse response consisted of only 2 modes whose resonant frequencies did not coincide with any of the tone's partials. The frequency deviations predicted by (19) were then compared with the results obtained by convolving the source with the resulting impulse response and calculating the resulting instantaneous frequencies. As predicted, this filter has minimal impact on the instantaneous frequency tracks for this tone, as seen in fig 5. Neither the model predictions (shown in red) or the observed frequency track (shown in blue) deviate from the original frequency modulation (in black).

In the second example, a third mode was added with a resonant frequency close to the center frequency of the displayed partial. The resulting large instantaneous frequency deviations are shown in fig 6. Once again, the model predictions (in red) coincide well with the observed frequencies (blue). Although not shown in the figure, the other partials were not affected by the addition of the third mode.

4.2 Low frequency response: modal model vs low-pass IR

Next, the modal model described above was evaluated using a recorded impulse response with significant modal content [18]. In order to isolate the modal region of the spectrum, the impulse response was low pass filtered with a cutoff of 2500 Hz. Prony's method was then used to fit a set of resonant modes to the resulting frequency response [19]. As can be seen in fig 7a, a large number of modes are required to accurately model this transfer function. For this example, the number of modes was set to 150. This number of modes models the dominant features of the frequency response while still significantly reducing the order of the system. The frequency deviations calculated from the modeled transfer function were then compared to the observed results when the signal was convolved with the low frequency portion of the original impulse response, as shown in fig 7. Although the modeled deviations (red) do not trace the observed deviations (blue) exactly, they retain many of the pertinent features including the locations of maximum deviation.



(a) Frequency response curves for the original impulse response (black) and associated modal model (red)

(b) Instantaneous frequency tracks of the original sound (black), filtered sound (blue) and modal model predictions (red).

Fig. 7: The modal model predicts the form and location of many of the deviations, while sometimes underestimating their extent.

5 Conclusions and future work

This paper presents a framework for analyzing the effects of different impulse responses on the instantaneous frequencies of modulated tones. In particular, parametric models for a set of early reflections or sum of resonant modes are presented and the conditions which maximize the instantaneous frequency deviations are discussed. These results are evaluated using musical vibrato tones and effects of the early and low frequency portions of recorded impulse responses are compared with the modeled predictions. These models are found to accurately predict many of the observed frequency deviations.

Future work remains to model the impacts of late reverberation on instantaneous frequency tracking with a statistical model. Additionally, it is hoped that these results will be able to inform methods to estimate the acoustic parameters of a space using natural recordings.

References

[1] Arroabarren, I., Rodet, X., and Carlosena, A., “On the measurement of the instantaneous frequency and amplitude of partials in vocal vibrato,” *IEEE Transactions on Audio, Speech, and Language Processing*, 14(4), pp. 1413–1421, 2006.

[2] Smith, S. R. and Bocko, M. F., “Effect of Reverberation on Overtone Correlations in Speech and Music,” in *Audio Engineering Society Convention 139*, Audio Engineering Society, 2015.

[3] Dubnov, S. and Rodet, X., “Investigation of phase coupling phenomena in sustained portion of musical instruments sound,” *The Journal of the Acoustical Society of America*, 113(1), pp. 348–359, 2003.

[4] Brown, J. C., “Frequency ratios of spectral components of musical sounds,” *The Journal of the Acoustical Society of America*, 99(2), pp. 1210–1218, 1996.

[5] McAulay, R. J. and Quatieri, T., “Sinusoidal coding,” 1995.

[6] Gerkmann, T., Krawczyk-Becker, M., and Le Roux, J., “Phase processing for single-channel speech enhancement: History and recent advances,” *IEEE signal processing Magazine*, 32(2), pp. 55–66, 2015.

[7] Fletcher, N. H. and Rossing, T., *The physics of musical instruments*, Springer Science & Business Media, 2012.

[8] Papich, G. and Rainbow, E., "A pilot study of performance practices of twentieth-century musicians," *Journal of Research in Music Education*, 22(1), pp. 24–34, 1974.

[9] Fletcher, H. and Sanders, L. C., "Quality of violin vibrato tones," *The Journal of the Acoustical Society of America*, 41(6), pp. 1534–1544, 1967.

[10] Cohen, L., *Time-frequency analysis*, Prentice hall, 1995.

[11] Loughlin, P. J. and Tacer, B., "On the amplitude- and frequency-modulation decomposition of signals," *The Journal of the Acoustical Society of America*, 100(3), pp. 1594–1601, 1996.

[12] Cohen, L. and Lee, C., "Instantaneous frequency, its standard deviation and multicomponent signals," in *32nd Annual Technical Symposium*, pp. 186–208, International Society for Optics and Photonics, 1988.

[13] Boashash, B., "Estimating and interpreting the instantaneous frequency of a signal. II. Algorithms and applications," *Proceedings of the IEEE*, 80(4), pp. 540–568, 1992.

[14] Smith III, J. O., *Spectral audio signal processing*, W3K publishing, 2011.

[15] De Cheveigné, A. and Kawahara, H., "YIN, a fundamental frequency estimator for speech and music," *The Journal of the Acoustical Society of America*, 111(4), pp. 1917–1930, 2002.

[16] Fritts, L., "University of Iowa musical instrument samples," *on-line at http://theremin. music. uiowa. edu/MIS. html*, 1997.

[17] Stewart, R. and Sandler, M., "Database of omnidirectional and b-format room impulse responses," in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pp. 165–168, IEEE, 2010.

[18] Murphy, D. T. and Shelley, S., "Openair: An interactive auralization web resource and database," in *Audio Engineering Society Convention 129*, Audio Engineering Society, 2010.

[19] Markel, J. D. and Gray, A. J., *Linear prediction of speech*, volume 12, Springer Science & Business Media, 2013.