Modal Estimation on a Warped Frequency Axis with Application to Coupled Piano String Modeling

Orchisama Das and Jonathan S. Abel *Member, IEEE*

***Abstract*—Linear systems such as room acoustics and string oscillations can be modeled as the sum of a set of modes, each characterized by its frequency, damping and amplitude. Currently, the approaches for estimating mode parameters include spectral peak detection followed by energy envelope fitting, and optimization of system pole coefficients. In this paper, the mode frequencies and dampings are estimated directly from the generalized eigenvalues of Hankel matrices of system response samples, similar to ESPRIT. For greater resolution at low frequencies, such as required in audio applications, the estimation is done on a warped frequency axis. The frequencies and decay rates maybe further optimized, and mode amplitudes are fit using least squares. We use this method to model coupled piano strings where frequency warping captures the slightly mistuned triplets, and optimization captures the two-stage decay.**

***Index Terms*—Coupled Strings, ESPRIT, Frequency Warping, Modal Synthesis, Optimization, Piano Modeling.**

I. Introduction

MODAL structures are an efficient way to synthesize acoustic spaces and vibrating systems. Modes originate from standing waves — they are exponentially damped sinusoids that are characterized by their frequencies, amplitudes and decay rates. These damped sinusoids are eigenfunctions of the acoustic transfer function between the input and output pressure waves. Modal analysis aims to estimate these parameters, and modal synthesis resynthesizes the analyzed sound using a bank of parallel biquad filters, each implementing one mode.

In [1], mode frequencies of a room are estimated by finding peaks in the power spectrum of the impulse response, and decay rates are approximated using reverberation times in bands about the estimated mode frequencies. Since rooms have thousands of modes, it is common-practice to estimate them on a band-by-band basis [1], [2]. Mode frequencies of carillon bells are estimated similarly in [3], and the decay rates are found with non-linear optimization [4]. Frequency-zoomed ARMA modeling on filtered groups of resonant frequencies has been used to model noisy string instruments and room responses in [5]. In [6], [7] modal parameters are estimated on a subband basis from the generalized eigenvalues of shifted Hankel matrices formed with impulse response samples. This method is computationally identical to ESPRIT [8], and was first formalized by Hua and Sarkar in [9]. In [7], the number of modes in each band is determined by k-means clustering, whereas in [6], they are determined by considering the singular values of the Hankel matrix. In this paper, we use the technique in [6] to estimate the mode frequencies and dampings.

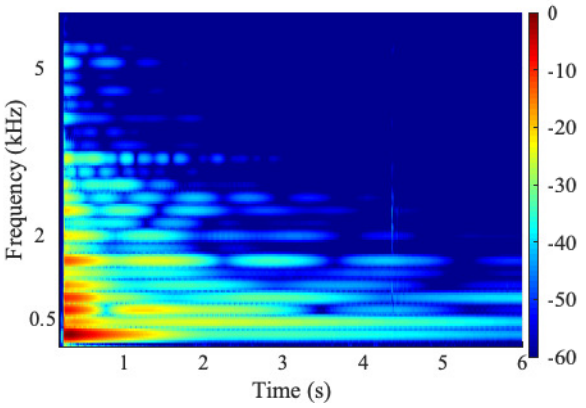
Frequency warping [10] replaces the unit digital delay operator with a first order allpass filter. A cascade of allpass sections introduces non-uniform group delay, where the low-frequency components are delayed in time and the high frequency components get advanced. A warped frequency axis emphasizes psychoacoustic perception with more resolution in low frequencies. Warped digital filters have applications in loudspeaker equalization, linear predictive coding and physical modeling. Warping an impulse response has the effect of spreading out low-frequency modes around the unit circle, so that beating frequencies can be separated and resolved. Resolution of closely spaced sinusoids is important in modeling coupled vibrations. Some methods to address this problem include subspace methods such as [11], or spectral windowing methods such as the one used in [3]. However these methods require parameter tuning or are computationally expensive. In this paper, we perform modal estimation on a warped impulse response to resolve low-frequency beating modes.

Once the mode frequencies and dampings are estimated from the warped signal, we unwarp and optimize them. An iterative time-domain optimization scheme inspired by [12] is used. The mode amplitudes are re-estimated using least squares in each iteration. This is different from [2], [13] where Maestre et. al fine-tune initial estimates of mode parameters with frequency-domain pole optimization.

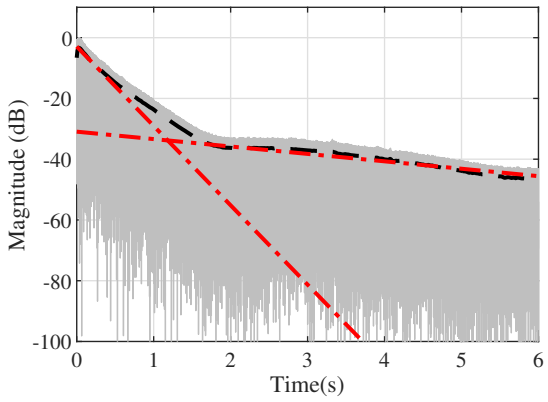
This three-step modal estimation technique is applied to find the modes of coupled piano strings. Each piano key (except the lower keys) is associated with sets of two or three strings coupled at the bridge. These strings are tuned with a small frequency deviation. String coupling leads to two-stage decay and beating [14], [15]. Apart from the piano, coupled strings can be found in many instruments, such as the harpsichord, mandolin and twelve-string guitar. Modal parameters of coupled strings can be used in conjunction with other physical models, such as digital waveguides, for more realistic sound synthesis. A hybrid modal-waveguide model has been used to synthesize coupled strings in [16]. Coupled digital waveguides have been used in [17]. A purely modal real-time piano synthesizer was proposed in [18], where the modes were obtained by solving the string equation. Other approaches include physical modeling of piano strings by solving the fundamental equation of a damped, stiff string struck with a non-linear hammer with numerical finite difference methods, introduced by Chaigne in [19], [20].

The rest of this paper is organized as follows. We first discuss coupled string acoustics in §II. In §III, we introduce the modal model and method for direct estimation of mode parameters from the Hankel matrix of the measurement. In §IV, we discuss the need for frequency warping and choice of warping coefficient. We also discuss band-wise modal estimation in §IV-B as a computationally expensive alternative to frequency

© 2020 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.



(a) Spectrogram of piano note



(b) Two-stage decay of fundamental

Fig. 1: Top - Spectrogram of the measured $A3$ piano note. Bottom - energy envelope of the fundamental showing two-stage decay. Straight lines fit to the data indicate the double slope effect.

warping. In §V, we explain the need for optimization and choice of cost function. We briefly discuss piano hammer action and propose a simple model for the hammer-string interaction in §VI. Finally, we compare the results of different modal estimation methods in §VII and conclude the paper in §VIII.

II. Coupled piano strings

The seminal work on coupled piano string vibrations was done by Weinreich [14], [15]. The standard full-size piano has 88 keys, with the 68 highest ones having three strings and the 20 lowest with double or single strings. The main purpose of these triplets and pairs of strings is to make the emanated sound louder, but their coupling also causes two-stage decay, i.e., the *prompt* sound (initial fast decay) and *aftersound* (late slow decay), as shown in Fig. 1b. Two-stage decay is also observed in other acoustically coupled systems, such as coupled rooms [21]. Mathematically, two-stage decay can be expressed as a weighted sum of two exponentials.

$$x(t) = c_0 \exp\left(-\frac{t}{\tau_0}\right) + c_1 \exp\left(-\frac{t}{\tau_1}\right) \quad (1)$$

Coupled string systems typically have two or more strings coupled at a non-rigid bridge. Let us consider a bridge with a single vibrating string. If a second string is attached to the bridge at the same location, its presence will affect the impedance of the bridge as seen by the first string. The result is that the frequency of the coupled strings will be different from its vibration frequency before the other string was introduced. The eigenanalysis of two strings coupled at a bridge is explained in [22]. There will be two normal modes of vibration - the "in-phase" vibrations will see a longer effective string length, and will move the bridge vertically a lot, causing a more rapid decay, whereas "anti-phase" vibrations see no length correction because the bridge is rigid in this case. The lengthening of the string due to "in-phase" vibrations makes the piano attack decay faster and flattens it, whereas the *aftersound*, composed of "anti-phase" vibrations, decays slowly and remains in tune.

The piano hammer striking the strings produces motion primarily in the transverse direction, but also in the longitudinal direction. The transverse vibrations are much stronger initially, but attenuate quickly whereas the weaker longitudinal vibrations persist. There is significant coupling between transverse and longitudinal vibrations [23]. Transverse vibrations produce modes at harmonic frequencies. The longitudinal motion is continuously excited by the transverse vibration along the string. The forced response to this excitation gives rise to phantom partials, while the free response produces the components corresponding to the longitudinal modal frequencies. The *aftersound*, or the second stage of a two-stage decay is caused by a combined effect of string coupling and shift from transverse to longitudinal polarization.

Additionally, coupled piano strings are slightly mistuned, and the amount of mistuning either causes audible beating (amplitude modulation), or a beatless *aftersound*. Coupled strings vibrate in phase immediately after impact, producing the *prompt* sound. Because their frequencies are slightly different, they become of phase eventually. Once this phase offset becomes approximately a half period of one of the frequencies, the movement at the bridge almost completely cancels, and sound is sustained. An experienced tuner can adjust the relative mistunings among coupled strings to create a desired *aftersound*.

For this paper, we modeled the $A3$ (220 Hz) note, struck loudly (*forte*) to excite all the modes, on a Yamaha Grand Piano sampled at 44.1 kHz. The spectrogram of this signal is shown in Fig. 1a. The three coupled piano strings associated with this note produce beating in the lower frequencies and two-stage decay. In the following sections, we describe the general modal estimation framework, followed by frequency warped modal estimation.

III. Modal estimation

When a string is struck or plucked, traveling waves move in opposite directions, get reflected at the bridge, and keep traveling back and forth along the string. The resultant motion creates standing waves. The standing waves dissipate energy with time because of scattering at the bridge and absorption.

These standing waves, or modes of vibration, are damped sinusoids vibrating at natural frequencies of the string. Diagonalizing the second order PDE of the one-dimensional traveling wave on a string yields these damped sinusoids as the eigenfunctions of the system. The resulting vibration of an excited string is due to the combination of several modes.

A signal $\hat{h}(t), t = 0, 1, \ldots, T$, represented by a rational system without repeated poles, can be written as a sum of M modes,

$$\begin{aligned} \hat{h}(t) &= \sum_{m=1}^{M} h_m(t) \\ h_m(t) &= \gamma_m \exp\left[(j\omega_m - \alpha_m)t\right] \end{aligned} \tag{2}$$

where ω_m is the angular frequency, α_m is the decay rate and γ_m is the complex amplitude of the mth mode. The goal is to estimate the mode parameters from a noisy measurement of the signal, $h(t)$. The modal reconstruction filter is formed with a parallel bank of M resonant biquad filters, each synthesizing one mode. The biquad filter coefficients are given by

$$\begin{aligned} \hat{H}(z) &= \mathcal{Z}\left(\sum_{m=1}^{M} \mathbb{R}(h_m(t)\right) \\ &= \sum_{m=1}^{M} \frac{\mathbb{R}(\gamma_m) - e^{-\alpha_m}\mathbb{R}(\gamma_m e^{-j\omega_m})z^{-1}}{1 - 2e^{-\alpha_m}\cos\omega_m z^{-1} + e^{-2\alpha_m}z^{-2}} \end{aligned} \tag{3}$$

where $\mathbb{R}$ denotes the real part of a number.

Consider a Vandermonde matrix, $\boldsymbol{V}$, with the mth column, $\boldsymbol{v}_m$, representing the time series of the mth mode, $h_m(t)$

$$\begin{aligned} \boldsymbol{V} &= \begin{bmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \ldots & \boldsymbol{v}_M \end{bmatrix} \\ \boldsymbol{v}_m &= \begin{bmatrix} e^{-(j\omega_m - \alpha_m)0} & \ldots & e^{-(j\omega_m - \alpha_m)T} \end{bmatrix}^\top \end{aligned} \tag{4}$$

where $\top$ denotes Hermitian transpose. The Hankel matrix formed by the signal samples, $\boldsymbol{H}$, can be written as an outer product of the Vandermonde matrix $\boldsymbol{V}$ with a diagonal matrix of mode amplitudes, $\boldsymbol{\Gamma} = \text{diag}\left[\gamma_1, \ldots, \gamma_M\right]$.

$$\boldsymbol{H} = \begin{bmatrix} h(0) & h(1) & \ldots & h(T/2) \\ h(1) & h(2) & \ldots & h(T/2+1) \\ \vdots & \vdots & \ddots & \vdots \\ h(T/2) & h(T/2+1) & \ldots & h(T) \end{bmatrix} \tag{5}$$

$$\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Gamma}\boldsymbol{V}^\top \tag{6}$$

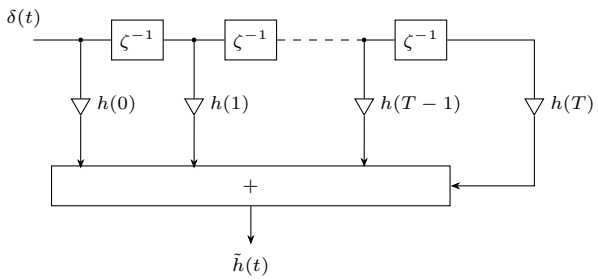
Similar to ESPRIT [8], the Hankel matrix offset by 1 sample, $\boldsymbol{K}$, can be written as

$$\boldsymbol{K} = \begin{bmatrix} h(1) & h(2) & \ldots & h(T/2+1) \\ h(2) & h(3) & \ldots & h(T/2+2) \\ \vdots & \vdots & \ddots & \vdots \\ h(T/2+1) & h(T/2+2) & \ldots & h(T+1) \end{bmatrix} \tag{7}$$

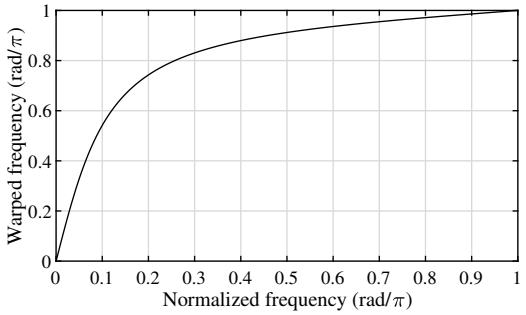
$$\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Psi}\boldsymbol{\Gamma}\boldsymbol{V}^\top \tag{8}$$

where $\boldsymbol{\Psi} = \text{diag}\begin{bmatrix} e^{(j\omega_1 - \alpha_1)} & \ldots & e^{(j\omega_M - \alpha_M)} \end{bmatrix}$. Post multiplying $\boldsymbol{K}$ with the pseudoinverse of $\boldsymbol{H}$, we get

$$\boldsymbol{K}\boldsymbol{H}^+ = \boldsymbol{V}\boldsymbol{\Psi}\boldsymbol{V}^{-1} \tag{9}$$



(a) Warping an impulse response



(b) Frequency warping as conformal mapping

Fig. 2: Frequency Warping with allpass sections

Thus, the diagonal elements of $\boldsymbol{\Psi}$ are the generalized eigenvalues of the matrix pencil $(\boldsymbol{K}, \boldsymbol{H})$. The mode frequency and damping estimates are the imaginary and real parts of the logarithm of the eigenvalues, ψ_m,

$$\ln \psi_m = j\hat{\omega}_m - \hat{\alpha}_m \tag{10}$$

The number of modes, $\hat{M}$, is estimated by the rank of the matrix $\boldsymbol{H}$ (by looking at its largest singular values). The mode amplitudes, $\hat{\boldsymbol{\gamma}}$, are found by least squares fit to the measured signal.

$$\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{V}}^\dagger \boldsymbol{h} \tag{11}$$

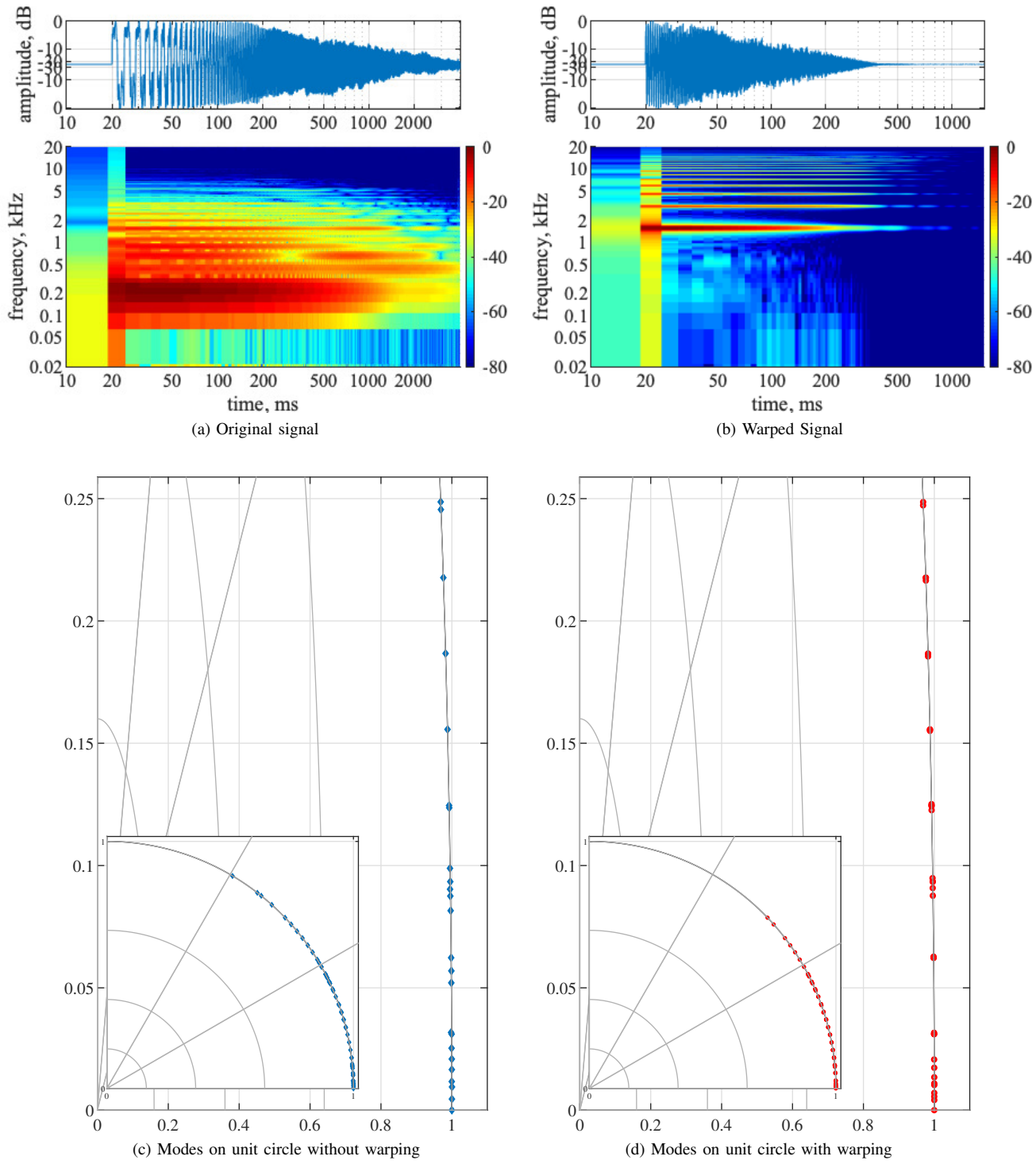
where $\hat{\boldsymbol{V}}^\dagger$ is the pseudoinverse.

In our experiments, the order of the Hankel matrix used is $\mathbb{R}^{T\times T}, T = 2048$. To determine the number of modes, M, the singular values up to -60 dB from the maximum are retained. Human hearing is logarithmic in nature, with more resolution in lower frequencies. This method of modal estimation gives equal emphasis to all frequencies. It cannot detect low frequency beating modes caused due to mistuning of the piano strings. To focus on the low-frequency modes, we use frequency warping, which essentially puts a magnifying glass on the lower frequency spectrum.

IV. Frequency warped modal estimation

Frequency warping [10] warps a uniform frequency axis to a non-uniform one. Warping is done by replacing the unit delay z^{-1} with a first-order allpass filter ζ^{-1}, given by:

$$\zeta^{-1} = \frac{z^{-1} - \rho}{1 - \rho z^{-1}} \tag{12}$$



(a) Original signal

(b) Warped Signal

(c) Modes on unit circle without warping

(d) Modes on unit circle with warping

Fig. 3: Top - Original and warped signal waveform and spectrogram on a log time axis. Bottom - Estimated modes with warping (red circles) and without warping (blue diamonds). Left - modes in z-plane without warping. Right - modes in z-plane with warping. The plots are zoomed in up to a polar angle of $\frac{\pi}{12}$ rad (1837.5 Hz) to show the lower frequency modes.

where ρ is the warping factor. Figure 2b shows the mapping between natural frequencies, ω, and warped frequencies, $\tilde{\omega}$, determined by the phase of the allpass filter in (12). The negative group delay of this allpass filter gives the slope of the mapping function as a function of frequency,

$$\frac{d\tilde{\omega}}{d\omega} = -\frac{1-\rho^2}{1+\rho^2-2\rho\cos\omega} \tag{13}$$

By choosing an appropriate warping factor, ρ^*, we can map uniformly spaced points on the frequency axis to a non-uniformly spaced warped frequency axis approximating a Bark scale [24], as shown in Fig. 2b. For a sampling frequency of f_s Hz, the optimal warping factor is given by

$$\rho^* = 1.0674\left[\frac{2}{\pi}\arctan\left(0.6583 f_s\right)\right]^{0.5} - 0.1916 \tag{14}$$

which approximately equals 0.75 for $f_s = 44.1$ kHz.

For modes near DC, warping increases the frequency by a factor of $(1+\rho)/(1-\rho)$. High-frequency modes near the bandedge get compressed by a factor of $(1-\rho)/(1+\rho)$. Similarly, the decay times of the low-frequency modes are decreased by a factor of $(1+\rho)/(1-\rho)$, and that of the high-frequency modes are increased by a factor of $(1-\rho)/(1+\rho)$.

To warp a signal or an impulse response, $h(t)$, the filter structure in Fig. 2a needs to be implemented. As a result of the warping, the signal frequencies have non-uniform group delay. The original and warped piano signal with $\rho = \rho^*$ is shown in Fig. 3a, 3b. As visible in both the time and frequency domains, the modes in the warped response have been shifted to higher and further spaced frequencies, and their dampings have been increased (approximately by a factor of 7).

A. Warped Processing

To estimate the modes, we pre-warp the response, estimate the warped modal parameters, and unwarp the frequencies and dampings before estimating the amplitudes using the unwarped impulse response. The measured piano response is warped with a warping factor of ρ^* (14) and its modes are estimated using the method described in §III. Modal estimation with frequency warping misses some high frequency modes because of compression. These modes are required for the accurate reconstruction of the piano transient. As a workaround, two sets of modes are calculated from the signal, one with and one without frequency warping. A combined set of modes is used, with lower frequency modes from the warped set and higher frequency modes from the non-warped set. A frequency cutoff threshold, ω_c, for combining the modes can be selected by setting the slope of the mapping to 1, where the warped frequency is equal to the non-warped frequency. This gives $\omega_c = \cos^{-1}(\rho)$ (5 kHz in this case). To unwarp the mode frequencies and dampings, we do the following, and then use (10).

$$\begin{aligned}\tilde{\psi}_m &= e^{(j\tilde{\omega}_m - \tilde{\alpha}_m)} \\ \psi_m &= \frac{-\rho + \tilde{\psi}_m^*}{1-\rho\tilde{\psi}_m^*}\end{aligned} \tag{15}$$

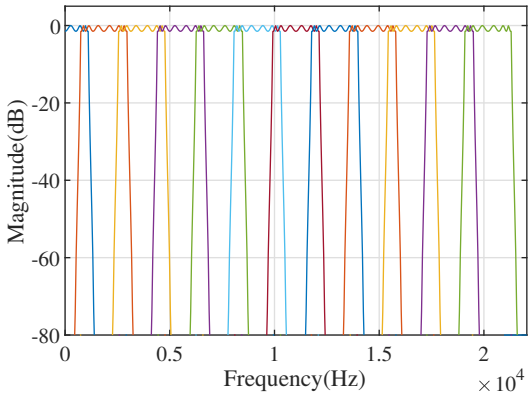


Fig. 4: Frequency subbands for modal estimation.

The modes estimated without and with frequency warping are plotted as poles on the z-plane in Figs. 3c, 3d respectively. Beating modes can be clearly seen in Fig. 3d, where overlapping circles in the zoomed plot indicate triplets caused by the three coupled strings vibrating together. Low-frequency beating modes are captured by frequency warping, which are otherwise missed in the non-warped case. However, due to reasons explained, frequency warping misses some of the high frequency modes.

As an alternative to frequency-warping, we can use subband processing which also gives high resolution in the lower frequencies. Subband processing involves filtering into overlapping frequency bands and downsampling, which spreads out the frequencies of beating modes. The details of this process is described in the following subsection.

B. Subband Processing

The modal estimation method described in §III can be done on a frequency band-by-band basis. The signal, $h(t)$ is filtered into overlapping frequency bands, and downsampled. This process ensures that each band has relatively few, spread-out modes to be estimated. Furthermore, downsampling effectively increases the damping, allowing more accurate estimation of long lasting modes. Let us consider N_b bands at a sampling rate of f_s Hz, each with band center at f_n Hz, $f_n = (2n-1)f_s/4N_b$, $n = 1, 2, \cdots, N_b$. The impulse response is heterodyned by multiplying with a complex exponential tuned to the band center frequencies.

$$h_n(t) = h(t)\exp\left(-2\pi j f_n t\right) \tag{16}$$

The heterodyned signal, $h_n(t)$, which is now centered at DC, is then filtered with a high-order LPF with a narrow passband. The filtered signal is decimated by a factor of r. The modes are then estimated for each band as $\hat{f}_{m,n}$ and $\hat{\alpha}_{m,n}$. Mode frequencies are moved up the spectrum to $\hat{f}_{m,n} + f_n$ to undo the effect of heterodyne, and mode amplitudes are adjusted as $\sqrt[r]{\hat{\alpha}_{m,n}}$ to undo the effect of downsampling.

Modes from subsequent bands are concatenated after discarding modes in overlapping passbands. The advantage of subband processing is that it gives equal emphasis to all frequency bands (all bands are of the same width), so there

is no compression of high frequencies. However, compared to frequency warping it is more computationally expensive, since it scales proportionally to the number of frequency bands.

To compare with frequency warped modal estimation, we implement subband processing with a downsampling factor of $r = 8$ for each of a series of $N_b = 12$ bands. This downsampling factor was selected to keep the results comparable with that of frequency-warped processing. The downsampling LPF is an elliptic filter of order 8, with a passband ripple of 1.5 dB and a stopband of -80 dB. The resulting subbands are shown in in Fig. 4.

V. Mode Optimization

Frequency-warped modal estimation successfully captures the beating modes in coupled piano strings. However, another factor that produces the characteristic *aftersound* is two-stage decay. Although frequency warping increases mode damping, the decay rates of the unwarped modes are underestimated. Neither is two-stage decay observed. To mend this, we propose a new time-domain iterative mode optimization method.

The measured signal, $\hat{h}(t)$. (2) can be written in a different form as

$$\hat{h}(t) = \sum_{m=1}^{M} e^{-\alpha_m t} \left[\gamma_{s_m} \sin(\omega_m t) + \gamma_{c_m} \cos(\omega_m t)\right] \quad (17)$$

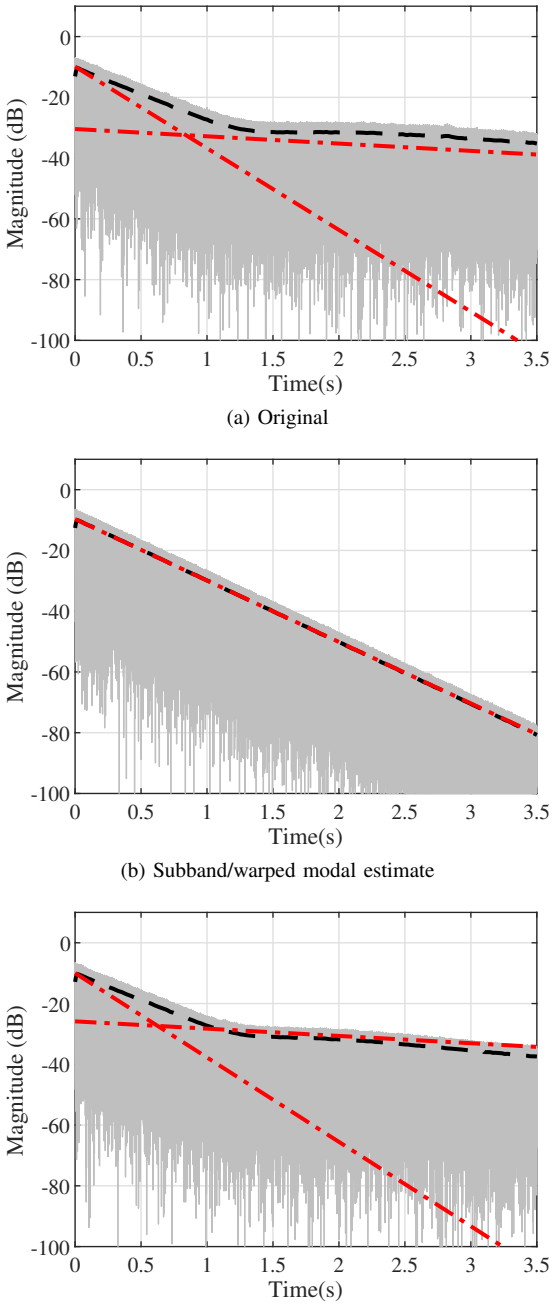
where $\gamma_{s_m}, \gamma_{c_m}$ are the real mode amplitudes associated with the sine and cosine components of the mth mode. Using (17) has the advantage of making the cost function real. In vector form, (17) can be written as

$$\begin{aligned}\hat{\boldsymbol{h}} &= \begin{bmatrix}\Im(\boldsymbol{V}) & \Re(\boldsymbol{V})\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma_s}\\ \boldsymbol{\gamma_c}\end{bmatrix} \\ &= \begin{bmatrix} e^{-\alpha_1 0}\sin(\omega_1 0) & \cdots & e^{-\alpha_M 0}\cos(\omega_M 0)\\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ e^{-\alpha_1 T}\sin(\omega_1 T) & \cdots & e^{-\alpha_M T}\cos(\omega_M T)\end{bmatrix}\begin{bmatrix}\gamma_{s_1}\\ \vdots \\ \gamma_{s_M}\\ \gamma_{c_1}\\ \vdots \\ \gamma_{c_M}\end{bmatrix}\end{aligned} \quad (18)$$

We can now form a vector of mode dampings and frequencies, $\theta = [\boldsymbol{\alpha}^T\ \boldsymbol{\omega}^T]^\top$ to optimize. The time-domain non-linear cost function and its gradient are

$$\begin{aligned} J(\theta) &= \frac{1}{2}||\boldsymbol{h} - \hat{\boldsymbol{h}}(\theta)||_2^2 \\ &= \frac{1}{2}\left(\boldsymbol{h} - \hat{\boldsymbol{h}}(\theta)\right)^\top \left(\boldsymbol{h} - \hat{\boldsymbol{h}}(\theta)\right) \\ \nabla_\theta J(\theta) &= D_\theta(\hat{\boldsymbol{h}}(\theta))^\top \left[\hat{\boldsymbol{h}}(\theta) - \boldsymbol{h}\right] \end{aligned} \quad (19)$$

where $\boldsymbol{h}$ is the measured signal vector, $\hat{\boldsymbol{h}}(\theta)$ is the modeled signal vector and $D_\theta(\hat{\boldsymbol{h}}(\theta)) \in \mathbb{R}^{T\times 2M}$ is the Jacobian matrix of $\hat{\boldsymbol{h}}(\theta)$ with respect to θ. Each entry of the Jacobian matrix is given by



(a) Original

(b) Subband/warped modal estimate

(c) Warped modal estimate with optimization

Fig. 5: Optimization helps capture 2-stage decay.

$$\begin{aligned} D_\theta(t,m) &= \frac{\partial \hat{h}(t)}{\partial \alpha_m} = -te^{-\alpha_m t}\left[\gamma_{s_m}\sin(\omega_m t) + \gamma_{c_m}\cos(\omega_m t)\right] \\ D_\theta(t,2m) &= \frac{\partial \hat{h}(t)}{\partial \omega_m} = te^{-\alpha_m t}\left[\gamma_{s_m}\cos(\omega_m t) - \gamma_{c_m}\sin(\omega_m t)\right] \end{aligned} \quad (20)$$

We use MATLAB's `fmincon` optimizer with initial mode parameter estimates calculated from §III after unwarping. The amplitudes $\gamma_{s_m}, \gamma_{c_m}$ are re-calculated in each iteration of the optimization using least squares. Providing `fmincon` with the gradient of the cost function speeds up computation significantly. Additionally, we set constraints such that

$-\pi < \omega_m < \pi$ and specify T_{60} limits for the mode dampings. The decay rate is related to the T_{60} as

$$\alpha = \frac{\ln 1000}{t_{60}/f_s} \tag{21}$$

Algorithm 1 Mode optimization pseudocode

Require: $T_{60_l} \leq \theta_{1:\mathrm{M}} \leq T_{60_u}, -\pi \leq \theta_{\mathrm{M+1:2M}} \leq \pi$

repeat

$\hat{\boldsymbol{h}}(\theta_i) \leftarrow \begin{bmatrix} \Im(\boldsymbol{V}(\theta_i)) & \Re(\boldsymbol{V}(\theta_i)) \end{bmatrix} \begin{bmatrix} \boldsymbol{\gamma_s} \\ \boldsymbol{\gamma_c} \end{bmatrix}$

$J(\theta_i) \leftarrow 0.5 \; ||\boldsymbol{h} - \hat{\boldsymbol{h}}(\theta_i)||^2$

$\theta_{i+1} \leftarrow \arg\min_\theta J(\theta_i)$

$\begin{bmatrix} \boldsymbol{\gamma_s} \\ \boldsymbol{\gamma_c} \end{bmatrix} \leftarrow \begin{bmatrix} \Im(\boldsymbol{V}(\theta_{i+1})) & \Re(\boldsymbol{V}(\theta_{i+1})) \end{bmatrix}^\dagger \boldsymbol{h}$

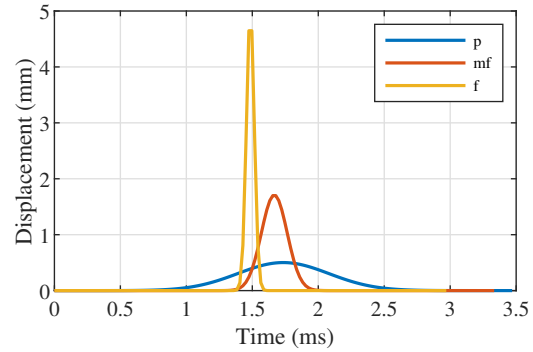
until convergence

We optimize the mode frequencies and dampings obtained from frequency-warped modal estimation using the first 2 s of the measured piano signal. The mode T_{60}s are constrained to be between 10 ms and 25 s. Final mode amplitude estimates are calculated after optimization. To show the benefits of mode optimization, the energy envelopes of the original, subband/warped modal, and optimized modal model are plotted in Fig. 5. Two-stage decay is not captured by subband or warped modal estimation, but mode optimization tunes the mode dampings to capture the double-slope effect.

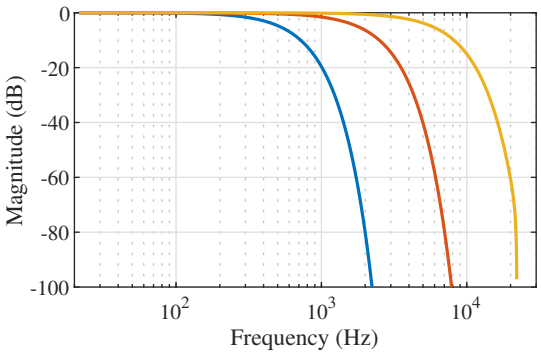
VI. Piano Hammer Action

If a digital piano uses a parallel bank of modal filters for synthesis, the mode frequencies and decay rates only need to be calculated once per key, irrespective of the force with which the key is struck. The hammer action can be represented as an external force that drives the filters. Assuming a linear system, an appropriate hammer action can simply be convolved with the modal filterbank to generate different dynamic levels. In reality, the hammer-string interaction is modeled by a mass striking a non-linear spring [22], but for our purposes we develop a simple linear model. We know that a *fortissimo (ff)* will sound brighter (high frequency mode amplitudes will be louder) than a *pianissimo (pp)* [25]. According to [26], the displacement curve of the strike point as a function of hammer strike velocity shows slower rise times associated with weaker impacts. The curve, as a function of time, gets narrower with a faster peak time as velocity of impact increases. We model this with a scaled Gaussian window, with the window parameters (a and window length L s) parameterized by the hammer strike velocity, v mm/s.

$$\begin{aligned} x_h(t) &= 0.1 \; a(v) \exp\left(-\frac{n^2}{2\sigma(v)^2}\right) \\ \sigma(v) &= \frac{\lceil L(v) f_s \rceil - 1}{2a(v)} \\ a(v) &= a_{min} + \frac{a_{max} - a_{min}}{1 + e^{-v}} \\ L(v) &= L_{min} + \frac{L_{max} - L_{min}}{1 + e^{v}} \end{aligned} \tag{22}$$



(a) Time domain



(b) Frequency domain

Fig. 6: Hammer excitation function used to drive the modal model for 3 different dynamic levels of increasing loudness.

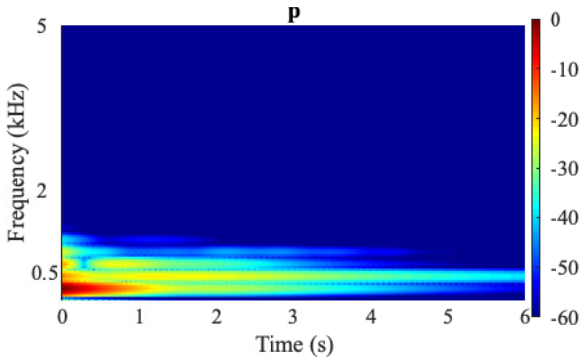
where $\lceil \; \rceil$ is the ceiling function. a controls the cutoff-frequency of the low-pass filter in the frequency domain. With increasing hammer velocity, the cut-off frequency increases (sound becomes brighter), while the impact time decreases. The sigmoid non-linearity ensures there is saturation at extreme velocities. The parametric hammer excitation function in the time and frequency domains, for three different dynamic levels of increasing loudness are shown in Fig. 6. The spectrograms generated after filtering the synthesized modal piano note with these hammer action functions are shown in Fig. 7. As expected, with greater hammer velocity, higher frequency partials start appearing.

This model designed to be simple. While it does not accurately capture hammer-string interaction, the results sound plausible. More physically accurate hammer-string interaction models exist in the literature [27], [28]. A different approach can also be used. We know that the hammer force affects the mode amplitudes only. A parametric function that relates mode amplitudes to hammer strike velocity can be fit to data measured at a range of dynamic levels. This function should depend on which piano key is being struck by the hammer. We leave this open for future work.

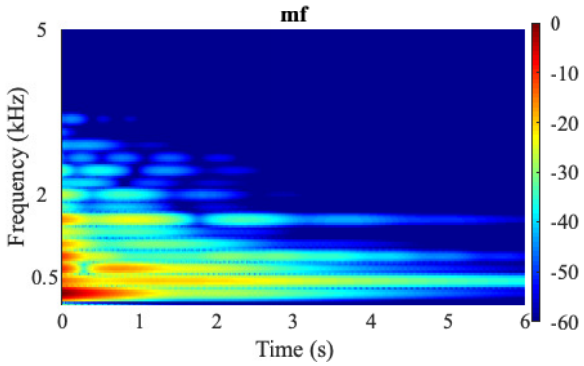
VII. Results and Discussion

Sound examples of the measured and modeled piano signal are available at https://ccrma.stanford.edu/~orchi/Modal/warped_modal.html.

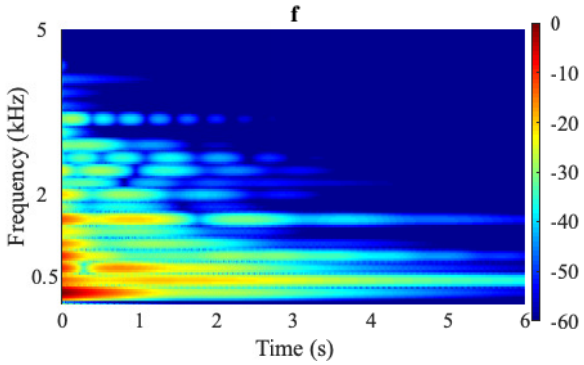
To show how the estimated modes differ with and without frequency warping, with optimization and with subband



(a) *Piano*



(b) *Mezzo forte*



(c) *Forte*

Fig. 7: Piano note spectrograms for different hammer strike velocities (softer to louder from top to bottom).

processing, mode frequencies versus mode T_{60}s are plotted in Fig. 8a. The grey vertical lines represent the harmonics of the fundamental at 220 Hz. Direct modal estimation without any warping (blue diamonds) estimates mode frequencies that are off by a large margin. Frequency warped modal estimation (red circles) does a better job at detecting mode frequencies and finding the triplets, but underestimates the T_{60}s. Subband modal estimation (yellow squares) performs best out of the three in finding mode frequencies, but it often misses the triplets. Frequency warped modal estimation followed by optimization (purple crosses) shifts the frequencies to be at the harmonics of the fundamental, finds triplets and fixes the issue of underestimated decay rates. The mode frequencies deviate from the harmonics beyond 2 kHz. This is because of the inharmonicity of the piano strings caused by string stiffness [29]. Although optimization is computationally expensive, this plot shows that it should be used whenever possible to get the best results.

The time domain waveforms and spectrograms of the original and modeled signals are shown in Fig. 9. In Fig. 9a, the signal h is normalized to have a maximum amplitude of 1, according to $\tanh(\beta h)/\tanh(\beta), \beta = 3$ to reveal small amplitude modeling errors. The top-most plot shows the measured signal, followed by a plot showing the modeled signal with modes estimated from the original signal without any warping. Here, the mode decay rates are severely underestimated. The next two plots, with subband modal estimation and frequency warping, are a much better fit. The results of subband v/s frequency warped modal estimation are comparable in lower frequencies, but frequency warping omits some of the high-frequency modes. This is because frequency warping compromises on resolution in high frequencies, whereas subband processing gives equal resolution in all frequency bands. However, in both these methods, long decay rates of low-frequency modes are underestimated. Using a combined set of optimized modes gives the best fit and also captures fast decaying high frequency modes in the transient. However, the fundamental's decay rate is underestimated even in this best case scenario.

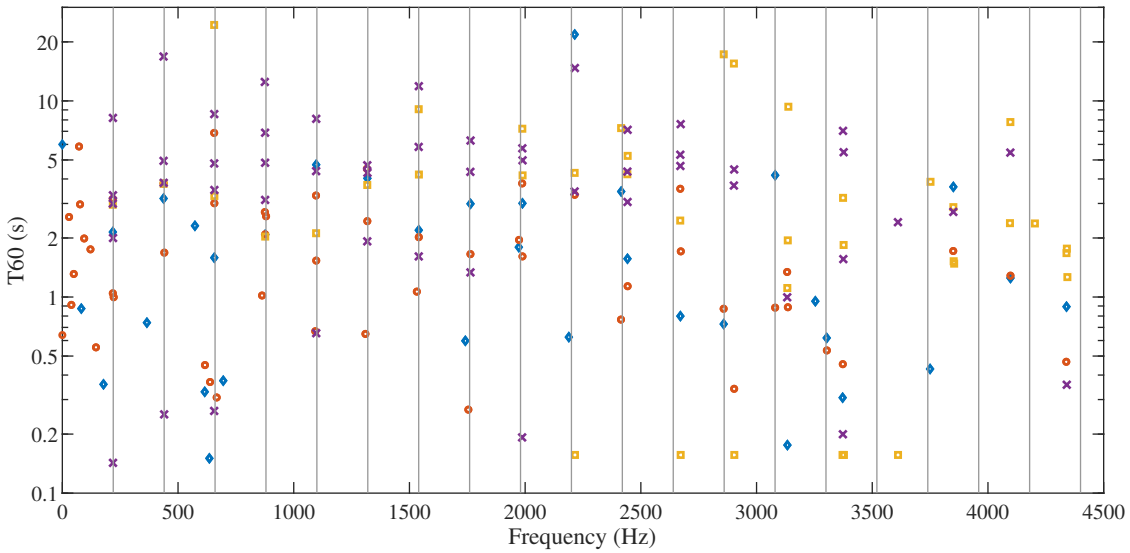
Using frequency warping over the subband method comes with a computational advantage, because subband modal estimation repeats the operations described in §III $N_b = 12$ times. Even with a combined set of modes (obtained with and without frequency warping), our proposed method (without optimization) is 6 times more efficient than subband modal estimation.

VIII. Conclusion

In this paper, we have proposed a modal estimation technique which operates on a warped frequency axis, and is well-suited to modeling coupled strings that have low-frequency beating modes. This method achieves results comparable to subband modal estimation with a significant computational advantage. To fine-tune the estimated mode parameters and capture details of the mode decays, we have optimized the error between our model and the measured response.

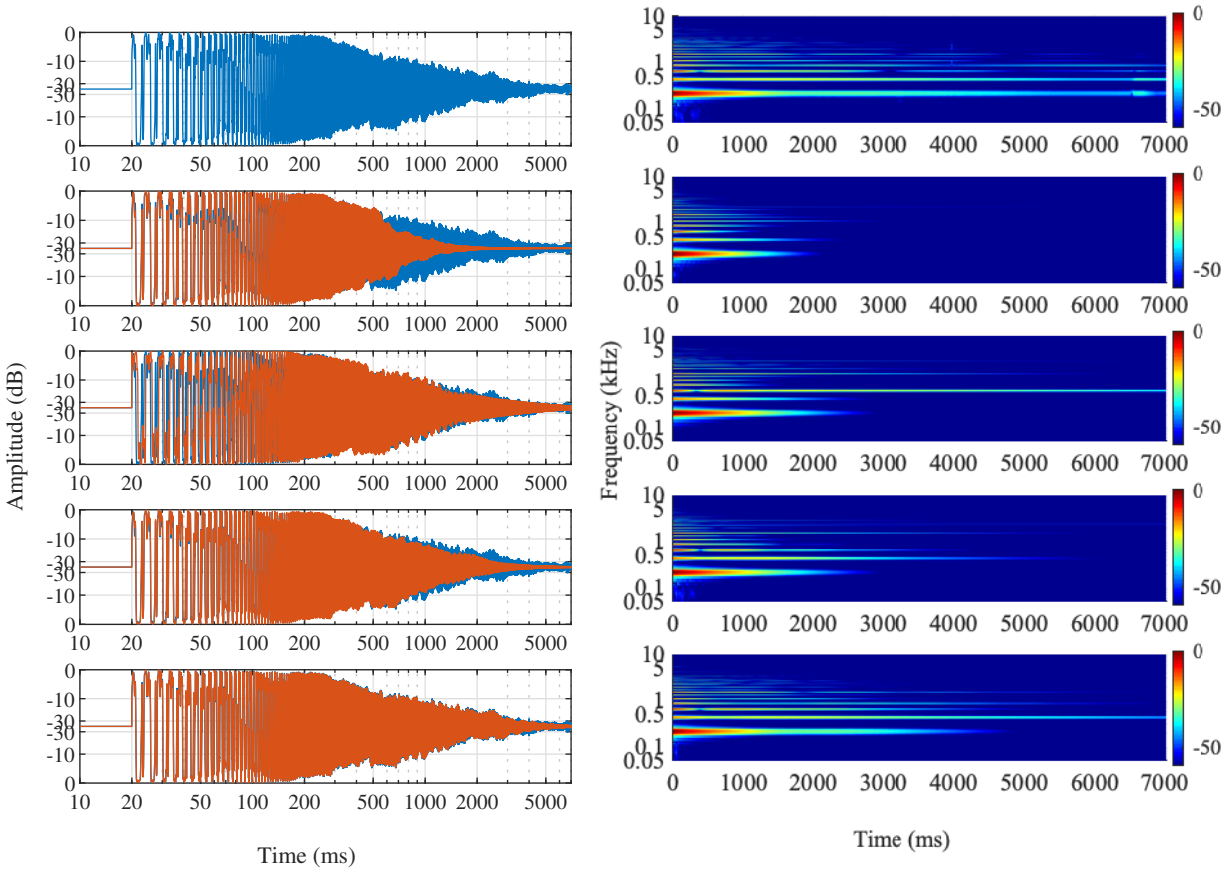
To test the proposed method, we have successfully modeled a piano note which has a few hundred modes appearing in triplets that exhibit two-stage decay as a result of string coupling. The results of the modal estimator align with Weinreich's theory of coupled piano strings. An empirical piano hammer excitation function filters the coupled string modes to generate different dynamic levels.

This modal analysis technique can be a visual tool for piano tuners to help them inspect and adjust the relative frequencies and decay rates of coupled strings for a desired *aftersound*. It can also be used in digital piano synthesizers. Additionally, accurate modal synthesis gives a lot of of flexibility to composers and sound designers to manipulate the sound of the piano by adding effects such as pitch shift and vibrato [30].



(a) Mode frequencies v/s T_{60}

Fig. 8: Mode T60 v/s mode frequencies with vertical grid lines at harmonic frequencies. Legend - modes with warping (red circles), without warping (blue diamonds), with warping and optimization (purple crosses), with subband processing (yellow squares).



(a) Time domain waveforms

(b) Spectrograms

Fig. 9: From top - measured (blue, underlaid), modeled without warping, subband model, warped model, modeled with warping and optimization over combined modes.

REFERENCES

[1] J. S. Abel, S. Coffin, and K. Spratt, "A modal architecture for artificial reverberation with application to room acoustics modeling," in *Audio Engineering Society Convention 137*. Audio Engineering Society, 2014.

[2] E. Maestre, J. S. Abel, J. O. Smith, and G. P. Scavone, "Constrained pole optimization for modal reverberation," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2017.

[3] M. Rau, O. Das, and E. K. Canfield-Dafilou, "Improved carillon synthesis," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2019.

[4] P. Antsalo, A. Makivirta, V. Valimaki, T. Peltonen, and M. Karjalainen, "Estimation of modal decay parameters from noisy response measurements," in *Audio Engineering Society Convention 110*. Audio Engineering Society, 2001.

[5] M. Karjalainen, P. A. Esquef, P. Antsalo, A. Mäkivirta, and V. Välimäki, "Frequency-zooming arma modeling of resonant and reverberant systems," *Journal of the Audio Engineering Society*, vol. 50, no. 12, pp. 1012–1029, 2002.

[6] J. S. Abel, "Method and system for designing a modal filter," September 3 2018, uS Patent App. No. 62-726325 (originally private communication, November 27 2017).

[7] C. Kereliuk, W. Herman, R. Wedelich, and D. J. Gillespie, "Modal analysis of room impulse responses using subband esprit," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2018.

[8] R. Roy and T. Kailath, "Esprit-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, no. 7, pp. 984–995, 1989.

[9] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814–824, 1990.

[10] A. Härmä, M. Karjalainen, L. Savioja, V. Välimäki, U. K. Laine, and J. Huopaniemi, "Frequency-warped signal processing for audio applications," *Journal of the audio engineering society*, vol. 48, no. 11, pp. 1011–1031, 2000.

[11] O. Das, J. S. Abel, and J. O. Smith III, "Fast music–an efficient implementation of the music algorithm for frequency estimation of approximately periodic signals," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2018.

[12] P. Ainsleigh and J. George, "Modeling exponential signals in a dispersive multipath environment," in *Proceedings of ICASSP-92: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 5. IEEE, 1992, pp. 457–460.

[13] E. Maestre, G. P. Scavone, and J. O. Smith, "Design of recursive digital filters in parallel form by linearly constrained pole optimization," *IEEE Signal Processing Letters*, vol. 23, no. 11, pp. 1547–1550, 2016.

[14] G. Weinreich, "Coupled piano strings," *The Journal of the Acoustical Society of America*, vol. 62, no. 6, pp. 1474–1484, 1977.

[15] ——, "The coupled motions of piano strings," *Scientific American*, vol. 240, no. 1, pp. 118–127, 1979.

[16] N. Lee, J. O. Smith, and V. Valimaki, "Analysis and synthesis of coupled vibrating strings using a hybrid modal-waveguide synthesis model," *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 4, pp. 833–842, 2009.

[17] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kronland-Martinet, "Resynthesis of coupled piano string vibrations based on physical modeling," *Journal of New Music Research*, vol. 30, no. 3, pp. 213–226, 2001.

[18] B. Bank, S. Zambon, and F. Fontana, "A modal-based real-time piano synthesizer," *IEEE transactions on audio, speech, and language processing*, vol. 18, no. 4, pp. 809–821, 2010.

[19] A. Chaigne and A. Askenfelt, "Numerical simulations of piano strings. i. a physical model for a struck string using finite difference methods," *The Journal of the Acoustical Society of America*, vol. 95, no. 2, pp. 1112–1118, 1994.

[20] ——, "Numerical simulations of piano strings. ii. comparisons with measurements and systematic exploration of some hammer-string parameters," *The Journal of the Acoustical Society of America*, vol. 95, no. 3, pp. 1631–1640, 1994.

[21] O. Das, E. K. Canfield-Dafilou, and J. S. Abel, "Delay network architectures for room and coupled space modeling," in *Proceedings of the International Conference on Digital Audio Effects (DAFx)*, 2020, under review.

[22] J. O. Smith, *Physical audio signal processing: For virtual musical instruments and audio effects*. W3K publishing, 2010.

[23] B. Bank and L. Sujbert, "Generation of longitudinal vibrations in piano strings: From physics to sound synthesis," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2268–2278, 2005.

[24] J. O. Smith and J. S. Abel, "Bark and erb bilinear transforms," *IEEE Transactions on speech and Audio Processing*, vol. 7, no. 6, pp. 697–708, 1999.

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

[26] S. Birkett, "Experimental investigation of the piano hammer-string interaction," *The Journal of the Acoustical Society of America*, vol. 133, no. 4, pp. 2467–2478, 2013.

[27] X. Boutillon, "Model for piano hammers: Experimental determination and digital simulation," *The Journal of the Acoustical Society of America*, vol. 83, no. 2, pp. 746–754, 1968.

[28] J. Bensa, K. Jensen, and R. Kronland-Martinet, "A hybrid resynthesis model for hammer-string interaction of piano tones," *EURASIP Journal on Advances in Signal Processing*, vol. 2004, no. 7, p. 805171, 2004.

[29] R. W. Young, "Inharmonicity of plain wire piano strings," *The Journal of the Acoustical Society of America*, vol. 24, no. 3, pp. 267–273, 1952.

[30] J. S. Abel and K. J. Werner, "Distortion and pitch processing using a modal reverberator architecture," in *Proceedings of the International Conference on Digital Audio Effects*, 2015.