ASPE Student Challenge - Error Report

Kushaal Singh

Nanoscale Design and Manufacturing Lab, UT Austin

Introduction

For the American Society of Precision Engineers Student Challenge we were tasked with implementing a low-velocity position feedback control system for a flexure stage for cutting a record. A crucial part of optimizing this system is creating an error budget and characterizing the system itself. Only once the system has been described may optimization occur as it defines the goal to which we are optimizing. This report serves to explain how we went about this process.

Positional Error

One of our initial methodologies was to create a static, positional error budget. In other words, given a target position, how far by a distance metric is our actual position away. When using this early methodology we quantified the distance as a ± offset that each source would contribute then used Root Sum of Squares (RSS) to estimate the total error. We were able to create a physical system very quickly, which allowed for data driven calculation. Assume the following values are true:

Table 1: Positional Error Budget

Error Source	**Magnitude μm**	**Type**
Control Tracking Error	±10.0	Systematic
Sensor Calibration Error	±2.0	Systematic
Mechanical Deflection	±3.0	Systematic
Electronic Noise	±1.0	Random
RSS Total Error	±10.6 μm	

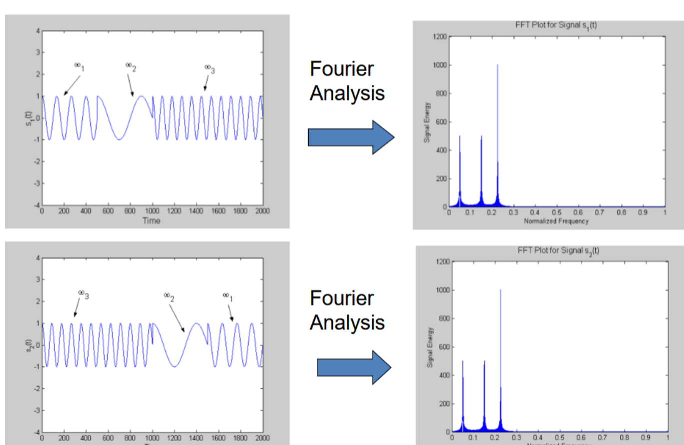
Joint Time-Frequency Analysis

Motivation

Our goal, fundamentally, is to inscribe waves into a physical medium which reveals a few errors with the positional error methodology described above. For example, a constant error offset that might be alarming to the positional system will have no effect on the audio signal. To illustrate this, let us consider a simple example where we seek to inscribe a pure sinusoid. If our target mean is set to be some value, but our system's mean is offset while everything else lines up, the positional description would describe this as problematic while our audio would be exactly correct. As such, we should evaluate dependent on frequency, amplitude, and phase, i.e. the qualities of a wave.

Upon hearing this, an initial reaction might be to use Fourier Analysis. While this is a good instinct, it illuminates another problem that is shared with the positional description, that being time. Using the same example as before, if our signal has a constant delay, which would be described as a constant phase shift or as a sinusoidal positional error, our audio will yet again be fine, while both a positional and a wave description would be calling this problematic. This is an important point, as the judge's feedback to our preliminary presentation inquired about a phase error. The answer isn't as simple as there ought not to be, but rather that a phase error in addition to a frequency and positional description of error still fails to adequately describe our system.

Additionally, Fourier Analysis implicitly assumes the signal to be Wide Sense Stationary. This is an assumption that is very flawed when looking at a song. A Fourier transform simply describes the frequency content present within a signal. As such, a signal with changing frequency content, like a song which changes pitch, is not properly described. As seen in the figure below, two different, non-stationary signals may produce the same Fourier transform. This is because a Fourier transform is only bijective on the domain of stationary signals.



Implementation

As a result of these factors, we made use of joint time-frequency analysis as Cohen [4]. Time-frequency analysis can be done through bilinear distributions, generalized as Cohen's class, or as signal decomposition. We restrict our study to Cohen's class of distributions, as they provide a continuous joint representation of time and frequency without requiring explicit basis functions or iterative transforms. Cohen's general class can be expressed as such:

$$C(t,\omega) = \frac{1}{4\pi^2} \iiint s^*(u - \frac{1}{2}\tau)\ s(u + \frac{1}{2}\tau)\ \phi(\theta,\tau)\ e^{-j\theta t - j\tau\omega + j\tau u} d\theta d\tau du$$

From this, through selection of a kernel $\phi(\theta,\tau)$ we can construct the Spectrogram, Wigner Distribution, and Garbor Transform, among others. Careful consideration of the kernel is warranted. In Pielemeier et al. it is noted that windowing methods rely heavily on steady-state assumptions to define frequency, which are flawed assumptions in a musical context. The instantaneous energy of the Wigner and Modal Distributions are presented as more faithful representations of muscial signals at computational cost [3]. The Garbor Distribution, while a windowing method, is shown to capture the essential time-frequency information of music signals at little computational cost, and as such, most research for musical signals surrounds Gabor frames [2]. It provides optimal joint time–frequency resolution for linear methods, avoids the cross-term interference inherent to bilinear distributions like Wigner–Ville, and maintains the interpretability necessary for evaluation. Note that compared to Cohen's class, the Gabor frame is linear, as opposed to the generalized bilinear.

To explain the Gabor Transform, first consider the Short Time Fourier Transform (STFT). This works by multiplying our signal $s(\tau)$ by a window function $h(\tau - t)$.

$$s_t(\tau) = s(\tau)h(\tau - t)$$

$$s_t(\tau) = \begin{cases} s(\tau) & for\ \tau\ close\ to\ t \\ 0 & else \end{cases}$$

Thus our STFT becomes:

$$S_t(\omega) = \frac{1}{2\pi} \int e^{-j\omega t} s(\tau)h(\tau - t)d\tau$$

This will create a 2d map of frequency over time. The Gabor Transform is then STFT with a Gaussian, instead of piecewise, window function. We will then perform this transform on both our actual and our target signals. These 2d maps will have all the necessary information we need to calculate 3 different metrics for error, each describing a different aspect of the wave we wish to align.

The Time-Frequency Resolution Trade-off

All joint time-frequency representations face a fundamental limitation expressed by the Gabor Uncertainty Principle, a corollary to Heisenberg's:

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi}$$

where Δt represents time resolution and Δf represents frequency resolution. This means we cannot simultaneously achieve arbitrarily good resolution in both time and frequency domains.

Different time-frequency representations make different compromises from this limit. We specifically make use of the Constant-Q Gabor Transform (CQT). The constant Q Gabor transform is the same as the earlier outlined Gabor transform, except window size scales inversely with frequency. More specifically, this uses logarithmically spaced frequencies with quality factor Q constant across bands using Gabor frames (the Gaussian signal window).

The choice of Constant Q Gabor Transform is also motivated by perceptual considerations. Human hearing exhibits approximately logarithmic frequency resolution. The musical scale itself is logarithmic, as each octave represents a doubling of frequency. The CQT's logarithmic frequency spacing naturally aligns with this perceptual characteristic:

$$f_k = f_{\min} \cdot 2^{\frac{k-1}{B}}$$

where B is the number of bins per octave. This means that frequency resolution relative to center frequency remains constant, matching the behavior of human auditory filters.

For musical signals, which contain harmonic structures spanning multiple octaves, the CQT provides a more efficient representation than linear frequency transforms, such as the normal Gabor transform. The variable window length:

$$N(k) = \frac{f_s}{f_k} \cdot Q$$

where $Q = \frac{1}{2^{1/B}-1}$, ensures that lower frequencies (where finer frequency resolution is needed) are analyzed with longer windows, while higher frequencies (where temporal precision is more important) use shorter windows.

The 3-Metric Framework: A Comprehensive Approach

Error metric 1: Gain / Amplitude Error

The gain error metric addresses the question: *"Does the system reproduce the correct amplitude at each frequency over time?"* This goes beyond simple frequency response measurements by capturing temporal variations in gain.
In more simple terms, this metric quantifies how accurately the system reproduces the amplitude of frequency components over time. A perfect system would have a gain of 1 (0 dB) at all frequencies. If a frequency is supposed to be represented but isn't, that would be found as the amplitude being low when it should be high.
We analyze this per frequency bin over time, quantified on a logarithmic scale. The logarithmic scale is chosen because human loudness perception is approximately logarithmic, and the decibel unit directly relates to perceptual loudness differences.

$$G(t_j, f_i) = 20 \cdot log_{10} \left(\frac{|S_{achieved}(t_j, f_i)|}{|s_{desired}(t_j, f_i)|} \right)$$

This is yet another 2d map, so to extract value from this, we find the mean error and variance across all frequency bins in our range of interest. The final output is a plot of $\mu_G(f)$ and $\sigma_G(f)$ vs. f. An ideal system has $\mu_G(f) = 0$ dB and $\sigma_G(f) = 0$ dB for all f.

By computing both mean ($\mu_G[i]$) and standard deviation ($\sigma_G[i]$) across time for each frequency bin, we distinguish between:

- Systematic errors ($\mu_G \neq 0$): Consistent amplification or attenuation at specific frequencies, often due to equalization or filter characteristics
- Variable errors (large σ_G): Time-varying gain changes, potentially indicating compression, automatic gain control, or instability

Pattern	**Likely Cause**	**Perceptual Impact**
$\mu_G > 0$ across all frequencies	System gain too high	Potential clipping, distorted dynamics
$\mu_G < 0$ across all frequencies	System gain too low	Reduced loudness
μ_G frequency-dependent	Equalization issues	Colored sound, unnatural timbre
Large σ_G at all frequencies	Compression or AGC	Pumping effect, lost dynamics
Large σ_G at specific frequencies	Resonances or feedback	Unstable, ringing sound

Error Metric 2: Time Delay / Phase Distortion

Phase relationships affect waveform shape and perceptual quality. This metric answers: *"Does the system preserve correct temporal relationships between frequency components?"* Here we care about the shape of time delay. As noted earlier, a constant time delay is acceptable, so we plot time delay over frequency then curve fit linear regression. if statistically significant non-zero slope (calculated by r^2) or high residuals, we have error.

For a linear, time-invariant system, a pure time delay τ produces a linear phase response:

$$\phi(f) = -2\pi f \tau$$

The derivative of phase with respect to frequency gives the group delay:

$$\tau_g(f) = -\frac{1}{2\pi}\frac{d\phi}{df}$$

For a pure delay, group delay is constant across frequency. When group delay varies with frequency, different frequency components arrive at different times, causing phase distortion.

Direct phase unwrapping and delay calculation can be numerically unstable. Let us define R as the complex CQT matrix of the Reference input signal and A as the complex CQT matrix of the Analyzed output signal.

1. Compute instantaneous phase differences: $\Delta\Phi[i,j] = \arg(\mathbf{A}[i,j] \cdot \mathbf{R}^*[i,j])$
2. Average across time to reduce noise: $\bar{\Delta\Phi}[i] = \mathbb{E}_t[\Delta\Phi[i,j]]$
3. Convert to time delay: $\tau[i] = -\frac{\bar{\Delta\Phi}[i]}{2\pi f_i}$
4. Validate with cross-correlation for ground truth

Delay Pattern	**System Behavior**	**Audio Quality Impact**
Constant τ across frequency	Pure time delay	Usually inaudible (except in multi-channel contexts)
Linear $\tau(f)$ with small slope	Minor phase distortion	Minimal impact on most program material
Nonlinear $\tau(f)$	Significant phase distortion	Smearing, comb filtering, degraded transients
Large σ_τ across time	Jitter or variable latency	Unstable image, fuzzy sound

Error Metric 3: Distortion Products / Harmonic Purity

This metric quantifies the system's linearity by measuring how much it adds new signal content that was not present in the original input. It answers the fundamental question: *"To what extent does the system corrupt the signal by introducing new spectral components?"* We do this by looking for energy that appears at frequencies not present in the input.

To ensure a fair comparison independent of overall loudness changes, the analyzed signal is first normalized to have the same total power as the reference signal. This prevents a simple gain change from being misinterpreted as distortion. The total power of each signal is evaluated as the squared Frobenius norm, corresponding to the L^2 energy of the matrix under the uniform probability measure over its index space.

$$P_{ref} = \|R\|_F^2 = \sum_{k=1}^{K}\sum_{n=1}^{N} |R(k,n)|^2, \quad P_{ach} = \|A\|_F^2 = \sum_{k=1}^{K}\sum_{n=1}^{N} |A(k,n)|^2$$

The scaling factor adjusts the analyzed signal so that $\|A_{norm}\|_F^2 = \|R\|_F^2$. The power-normalized version of the analyzed signal, A_{norm}, is then:

$$A_{norm} = A \cdot \sqrt{\frac{P_{ref}}{P_{ach}}}$$

The core of this metric is the Distortion Matrix, $D_{complex}$, which is defined as the complex difference between the normalized output and the reference. This matrix isolates all new content introduced by the system. A perfect, linear system would result in $D_{complex} = 0$. The instantaneous distortion power at each time-frequency cell is given by $|D_{complex}(k,n)|^2$. The total distortion power is the sum over all cells:

$$D_{complex} = A_{norm} - R$$

$$P_{dist} = \|D_{complex}\|_F^2 = \sum_{k=1}^{K}\sum_{n=1}^{N} |D_{complex}(k,n)|^2$$

Our metric, Total Harmonic Distortion (THD), is then derived from the power of the reference and distortion matrices. THD expresses the total distortion energy as a percentage of the total reference signal energy. A lower THD_{TF} value indicates a purer output. A value of 0% represents a perfect system, while a value of 5% indicates that 5% of the output signal's energy is unwanted distortion.

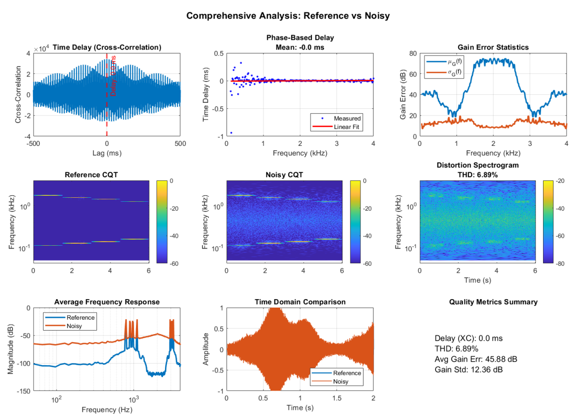
$$\text{THD}_{TF} = \frac{\|D_{complex}\|_F}{\|R\|_F} \times 100\% = \sqrt{\frac{P_{dist}}{P_{ref}}} \times 100\%$$

Reading Error Metrics

General Distortion Types and Their Signatures

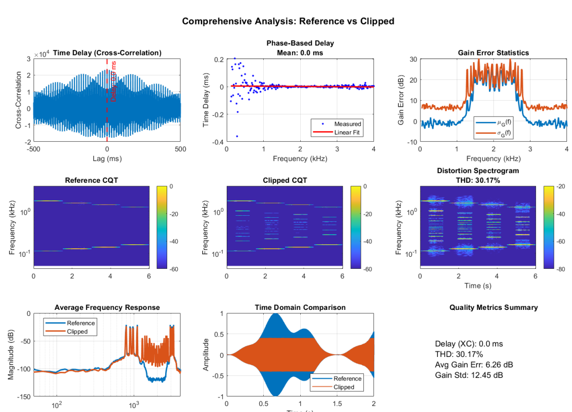
Additive Noise:

- Gain error: Near-zero μ_G but increased σ_G due to random fluctuations
- Delay: Minimal effect
- Distortion: Elevated THD, broadband distortion spectrum



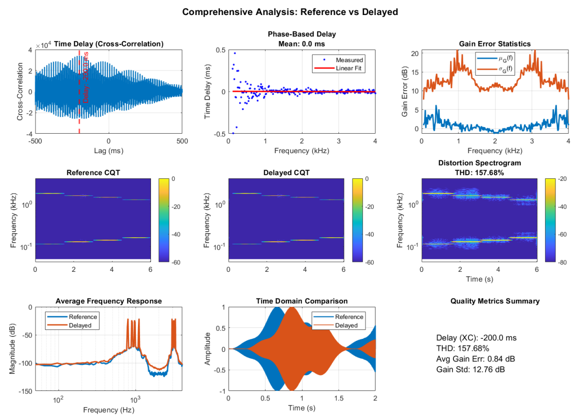
Clipping:

- Gain error: Compression-like pattern with $\mu_G < 0$ during peaks
- Delay: Minimal effect
- Distortion: High THD, distortion concentrated at signal peaks and high frequencies



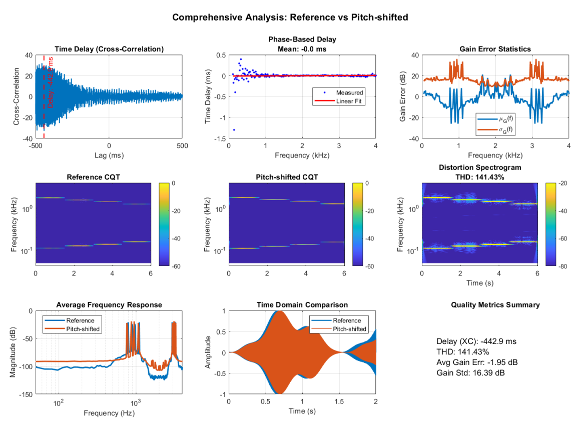
Pure Time Delay:

- Gain error: No effect
- Delay: Constant τ across frequency
- Distortion: No effect on THD (theoretical)



Pitch Shifting:

- Gain error: Frequency-dependent patterns due to spectral misalignment
- Delay: Complex frequency-dependent phase relationships
- Distortion: Elevated THD from interpolation artifacts and spectral smearing



Diagnostic Summary Table

Mechanical Error	Delay Signature	Gain Error Signature	Distortion Signature
Hysteresis	Frequency-dependent $\tau(f)$	History-dependent μ_G	Odd-order harmonics
Thermal Effects	Slow $\tau(t)$ drift	$\mu_G(t)$ correlates with P_{avg}	THD increases with temperature
Deflection	Resonant $\tau(f)$ peaks	Resonant $\mu_G(f)$ peaks	THD peaks at mechanical resonances
Calibration Error	Constant τ offset	Constant μ_G offset	Minimal unless saturated
Sensor Noise	No coherent effect	Elevated white σ_G	Broadband noise floor

Quick Diagnosis Flow

1. **Check Delay Metrics First**: High XC delay? → Start with time alignment
2. **Examine THD**:
   - High THD + peak distortion → Clipping
   - High THD + broadband noise → Additive noise
   - High THD + spectral smearing → Pitch-shifting artifacts
3. **Analyze Gain Patterns**:
   - Systematic μ_G errors → EQ/filtering issues
   - Random σ_G spikes → Connection/interference problems
   - Frequency-dependent patterns → Phase/algorithm issues
4. **Verify with CQT Visualization**: Confirm diagnosis with time-frequency patterns

Mechanical Sources of Error

It would be nice to quantify our minimum errors and their sources, but this is a problem outside the scope of this project. Below are some error types to consider:

Hysteresis:
The output depends on the history of the input, creating path-dependent non-linearities.

Thermal Effects
Temperature-dependent parameter changes causing slow, time-varying system response. Thermal expansion also changes system dynamics.

Deflection at Scribe/Contact
Nonlinear spring behavior from mechanical compliance and play in linkages.

Calibration Error
Systematic offsets from improper calibration or alignment.

Sensor Noise
Stochastic processes from mechanical-thermal interactions.

IMPORTANT Clipping due to downsampling
We had to use, day of, a lower sampling rate for our voicecoil due to memory issues, which would cause clipping and missing some high frequencies

Due to all of this, we had a different methodology for creating our error budget: Targets were set by taking a 4 pure-tone signal (C4–E4–G4–C5) and analyzing error metrics on augmented versions. We tested the effects of constant delay, clipped audio, noisy data, and tone shift. Because we can't have a true 0 of all metrics, we need to set targets of "good enough". We intentionally added set amounts of each type of error until we found that the musical signal was so bad.

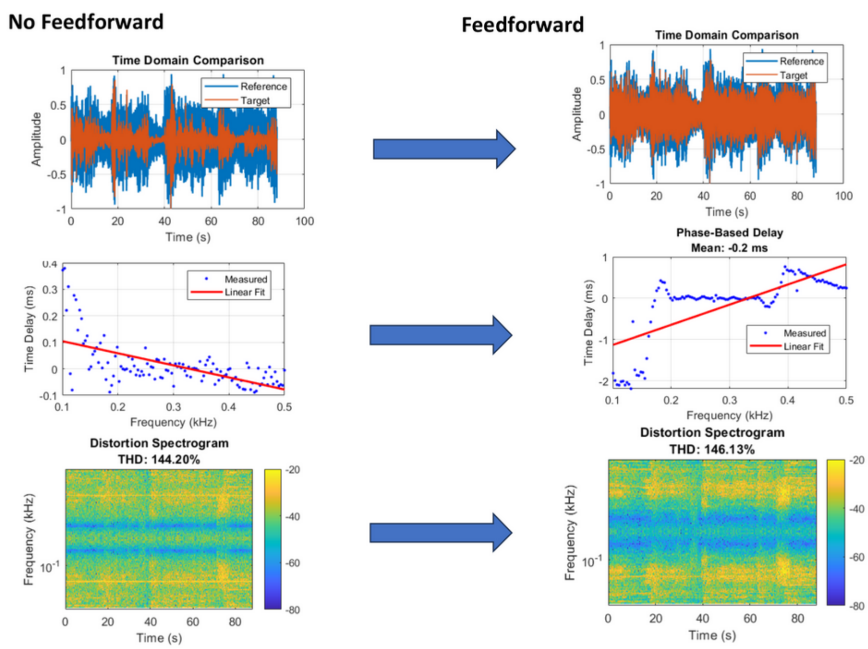
I tried finding a mathematical estimation for the minimums, but I am not confident in these calculations and have left them out.

	Delay Signature	**Gain Error Signature**	**Distortion Signature**
Target Error	$\leq$2.0 ms	$\pm$2.0 dB	50%
Error with No control	-16.0 ms	-3.95 $\pm$ 9.27 dB	144%
Error with FF	-0.2 ms mean	-2.97 $\pm$ 6.7 dB	145%

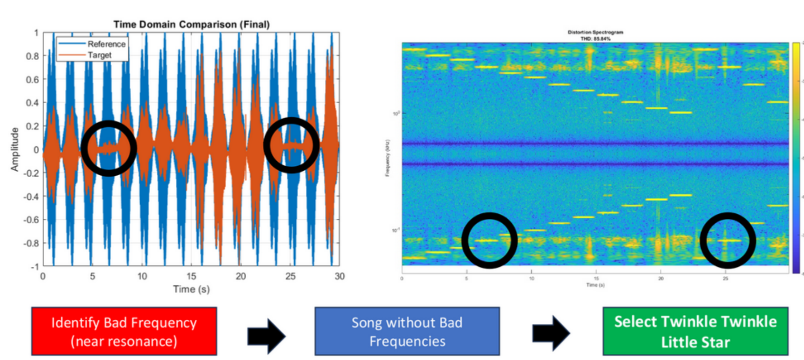
Real data

Controls selection

We implemented a Feed Forward algorithm and improved our amplitude matching, changed our phase based delay, and identified resonance using the CQGT.

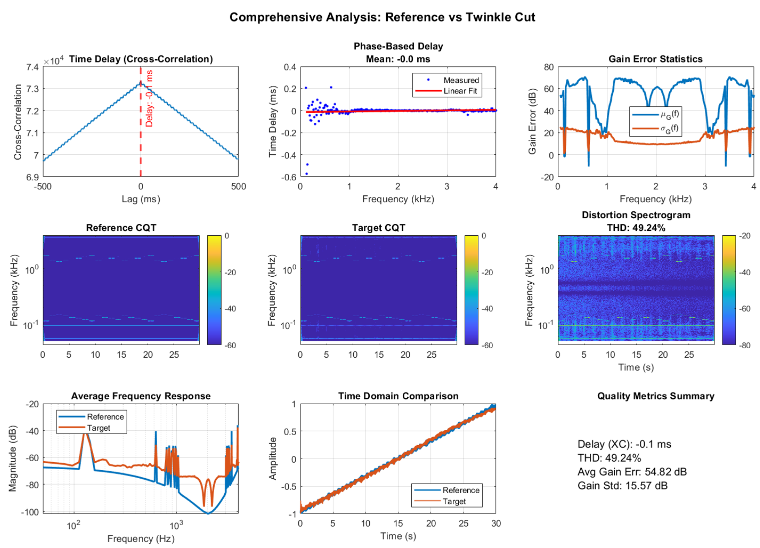


In our frequency sweep, using pure sin waves at different tones going through the arpeggio of a C Major scale, we identified a problematic frequency of C3, which is right at resonance. Our controller struggled with commanding this frequency, so we selected a song without this frequency. This can be thought as a notch filter eliminating that frequency, but in reality we just restrict to a song that doesn't have that frequency at all.



Final Cut Metrics

Attached is the final error window for our cut record. We were also able to reduce distortion significantly by commanding a sine wave at resonance as a form of noise canceling.



References

[1] Alm et al. "Time-Frequency Analysis of Musical Instruments". In: *SIAM REVIEW* 44.3 (2002), pp. 457–476.

[2] Ottosen et al. "A Characterization of Sparse Nonstationary Gabor Expansions". In: *Journal of Fourier Analysis and Applications* 24 (2017), pp. 1048–1071.

[3] Pielemeier et al. "Time Frequency Analysis of Music Signals". In: *IEEE* 84.9 (1996), pp. 1216–1230.

[4] Leon Cohen. *Time Frequency Analysis*. Prentice Hall PTR, 1995. ISBN: 013594532-I.

[5] Monika Dörfler. "Time Frequency Analysis for Music Signals: A Mathematical Approach". In: *Journal of New Music Research* 30.1 (2001), pp. 3–12.