Per-Harmonic ADSR Synthesis: Parametric Audio Modeling Through Optimized Envelope Reconstruction

Sekander Ali
sna2151
Digital Signal Processing - ELEN E4810
Columbia University

***Abstract*—This report presents a parametric synthesizer that models musical signals through per-harmonic envelope decomposition with automated greedy selection, enabling pitch-independent timbre reproduction. By employing an FIR filter bank with Kaiser windowing for harmonic isolation and a per-harmonic optimization framework for ADSR (Attack, Decay, Sustain, Release) parameter extraction, the system separates temporal characteristics from fundamental frequency. The optimization process independently tests 26 basis function configurations for each harmonic using mean squared error metrics to select optimal envelope reconstruction parameters. Experimental validation demonstrates faithful reproduction of complex instrumental timbres across arbitrary pitch transpositions, and validate the use of a per-harmonic system.**

I. Introduction

A. Motivation

Traditional time-domain sampling methods suffer from fundamental limitations when pitch transposition is required. Nyquist's sampling theorem dictates a fixed relationship between playback speed and pitch, meaning any frequency shift necessarily alters temporal envelope characteristics. When pitch-shifted upward, the attack becomes sharper and decay accelerates; conversely, lowering the pitch produces sluggish attack and artificially prolonged decay. These artifacts make pitch-shifted samples immediately recognizable as artificial.

This project addresses this limitation through a parametric approach based on additive synthesis. By decomposing the input signal into independent harmonic components, each with its own time-varying envelope, timbre and temporal evolution are decoupled from fundamental frequency, enabling pitch transposition while preserving instrumental character.

B. Related Work

Classical additive synthesis approaches employ global envelope models applied uniformly to all harmonics [2], which fails to capture independent partial evolution due to mode coupling and frequency-dependent damping. McAulay and Quatieri's sinusoidal modeling [3] and Serra and Smith's spectral modeling synthesis [4] provide more general frameworks but with increased complexity. The phase vocoder [5], while widely used for pitch-shifting, can introduce phase discontinuities in transient-rich signals. This work occupies a middle ground through systematic per-harmonic greedy selection, enabling automatic adaptation without manual parameter tuning.

C. Contribution

The primary contributions of this work are:

1) A complete pipeline for per-harmonic envelope extraction using FIR filter banks with Kaiser windowing
2) A per-harmonic greedy selection framework that independently tests 26 configurations for each harmonic using MSE error metrics
3) Systematic phase correction ensuring harmonic coherence across arbitrary pitch transpositions
4) Piecewise envelope reconstruction with explicit continuity constraints to eliminate discontinuities

II. Technical Approach

A. Harmonic Signal Modeling

A periodic musical signal $x[n]$ is represented as a superposition of K harmonically-related sinusoids, each modulated by an independent time-varying envelope:

$$x[n] = \sum_{i=0}^{K} A_i \hat{E}_i[n] \sin\left(2\pi f_i \frac{n}{f_s} + \theta_i\right) \tag{1}$$

where A_i is the magnitude of the i-th harmonic, f_i is the harmonic frequency, θ_i is the phase offset, and $\hat{E}_i[n]$ is the reconstructed envelope. The sampling rate $f_s = 44.1$ kHz is used throughout.

Dynamic Harmonic Count Selection: Rather than using a fixed number of harmonics, K is determined adaptively based on perceptual significance. The fundamental frequency f_0 is detected using FFT peak finding in the range [80, 1000] Hz with a Hanning window for sidelobe suppression. For each potential harmonic i, the FFT bin closest to $i \cdot f_0$ is examined, with a local search within ± 2 bins to account for frequency leakage.

Harmonics are included if they satisfy two criteria:

$$A_i \geq 0.005 \cdot A_1 \quad \text{and} \quad \left|\frac{f_{\text{measured}} - i \cdot f_0}{i \cdot f_0}\right| < 0.10 \tag{2}$$

The 0.5% amplitude threshold ensures that only perceptually significant components are retained, while the 10% frequency

error tolerance accommodates slight inharmonicity and FFT resolution limits. A hard limit of $K \leq 30$ harmonics prevents excessive computational cost while capturing the spectral content of most acoustic instruments.

B. Harmonic Isolation: FIR Filter Bank Design

Each harmonic's envelope is extracted using narrow-band bandpass FIR filters with Kaiser windowing, providing guaranteed stability and arbitrary stop-band attenuation [1].

1) Filter Design Parameters: **Passband definition:**

$$\Delta f = \min(f_0/4, 50 \text{ Hz}) \tag{3}$$

$$[f_{\text{low}}, f_{\text{high}}] = [f_i - \Delta f, f_i + \Delta f] \tag{4}$$

The choice $\Delta f = f_0/4$ balances harmonic capture with preventing inter-harmonic bleed.

Stopband definition:

$$f_{\text{stop,low}} = \max(0.1, f_i - 2\Delta f) \tag{5}$$

$$f_{\text{stop,high}} = \min(f_s/2 - 0.1, f_i + 2\Delta f) \tag{6}$$

Filter order estimation: The required number of taps N is determined using the Kaiser window formula [1]:

$$N = \left\lceil \frac{A - 7.95}{2.285 \cdot \Delta\omega} \right\rceil + 1 \tag{7}$$

where $A = 60$ dB is the desired stopband attenuation and $\Delta\omega = 2\pi\Delta f/f_s$ is the normalized transition width. The filter order is constrained to the range [51, 1001] taps and forced to be odd to ensure Type I FIR symmetry (linear phase).

Kaiser window parameter: For $A = 60$ dB, $\beta = 0.1102(A - 8.7) \approx 5.65$ [1].

2) Zero-Phase Filtering and Envelope Extraction: Filters are applied using forward-backward filtering (`filtfilt`), doubling the effective filter order and yielding zero-phase response. Envelope extraction uses the Hilbert transform $E_k[n] = |\mathcal{H}\{x[n] * h_k[n]\}|$, producing a $K \times N$ envelope matrix $\mathbf{E}$.

C. Basis Functions for Envelope Analysis

The system employs convolution kernels to extract ADSR characteristics. Multiple options are provided for each stage (full definitions of each are provided in the included code):

1) **Attack (6 options):** Sharp 2-sample to smooth 6-sample differentiators acting as matched filters for rising transients.
2) **Decay (6 options):** Symmetric and asymmetric central differences measuring instantaneous slope.
3) **Sustain (7 options):** Moving average filters from 10-100 ms, plus windowed versions.
4) **Release (7 options):** Applied to log-envelope $L_i[n] = 20\log_{10}(E_i[n] + 10^{-6})$, ranging from 2-sample to 20-sample kernels.

D. Transition Point Detection

ADSR segment boundaries are identified with robustness constraints preventing edge effects:

1) **Attack → Decay:** $n_{i,1} = n_{\text{edge}} + \arg\max_{n \in [n_{\text{edge}}, 0.9N]} E_i[n]$ where $n_{\text{edge}} = 0.01 \cdot f_s$.
2) **Decay → Sustain:** $n_{i,2} = \min\{n > n_{i,1} \mid |\mathbf{DEC}_i[n]| < 0.01 \cdot \max(|\mathbf{DEC}_i|)\}$ with 50 ms minimum.
3) **Sustain → Release:** $n_{i,3} = n_{i,2} + \min\{n \geq 0.05 \cdot f_s \mid \mathcal{E}_{\text{tail}}(n) > 0.95\}$ where $\mathcal{E}_{\text{tail}}(n) = \sum_{k=n}^{N} E_i^2[k] / \sum_{k=n_{i,2}}^{N} E_i^2[k]$.

Transition points are recomputed during optimization as basis functions update.

E. Automated Per-Harmonic Basis Function Optimization

Each harmonic i independently optimizes its basis function indices, capturing that different partials evolve at different rates.

Greedy Sequential Optimization Algorithm: The optimization employs a greedy strategy, sequentially optimizing each ADSR stage using the transition points $n_{i,1}$, $n_{i,2}$, and $n_{i,3}$ defined in the previous section. For each harmonic:

1) **Optimize Attack:** Test all 6 attack kernels for harmonic i on the attack segment $[0, n_{i,1}]$. For each attack kernel j:
   - Extract attack feature: $\mathbf{ATK}_i = E_i[n] * \phi_A^{(j)}$
   - Reconstruct attack segment: $\hat{E}_i[n] = \sum_{k=0}^{n} \alpha_A \cdot \mathbf{ATK}_i[k]$ for $n \in [0, n_{i,1}]$
   - Compute MSE on attack segment only: $\mathcal{E}_i^{\text{attack}}(j) = \frac{1}{n_{i,1}} \sum_{n=0}^{n_{i,1}} |E_i[n] - \hat{E}_i[n]|^2$

   Select attack index: $a_i^* = \arg\min_j \mathcal{E}_i^{\text{attack}}(j)$
2) **Optimize Decay:** With attack basis fixed to a_i^*, test all 6 decay kernels on the decay segment $[n_{i,1}, n_{i,2}]$. Recompute transition point $n_{i,2}$ with the optimized attack basis and select based on decay segment error. Select: d_i^*
3) **Optimize Sustain:** With attack a_i^* and decay d_i^* fixed, test all 7 sustain kernels on the sustain segment $[n_{i,2}, n_{i,3}]$. Select: s_i^*
4) **Optimize Release:** With attack, decay, and sustain fixed to their optimal values, test all 7 release kernels on the release segment $[n_{i,3}, N]$. Select: r_i^*

The result is K sets of indices minimizing local segment error: $\{a_1^*, d_1^*, s_1^*, r_1^*\}, \ldots, \{a_K^*, d_K^*, s_K^*, r_K^*\}$.

Error Metric: Mean squared error quantifies reconstruction quality: $\text{MSE}_i^{\text{segment}} = \frac{1}{N_{\text{segment}}} \sum_{n \in \text{segment}} |E_i[n] - \hat{E}_i[n]|^2$.

F. Envelope Reconstruction

Piecewise segments reconstruct envelopes::

1) **Attack:** Cumulative sum of attack features scaled by $\alpha_A = 2.0$
2) **Decay:** Exponential decay from attack peak to detected sustain level
3) **Sustain:** Smoothed envelope using selected sustain kernel
4) **Release:** Exponential decay from sustain level to zero

G. Phase Reconstruction and Synthesis

Accurate phase information is essential for waveform reconstruction. The phase θ_k must be referenced to the note onset ($n = 0$) rather than the FFT window center. Given an analysis window starting at sample $n_{\text{start}} = 0.05 \cdot f_s$ (50 ms onset detection), the corrected phase is:

$$\theta_k^{\text{corrected}} = \theta_k^{\text{FFT}} - 2\pi f_k \frac{n_{\text{start}}}{f_s} \tag{8}$$

For arbitrary target frequency f_{new}, synthesis proceeds by maintaining the harmonic relationship:

$$x_{\text{new}}[n] = \sum_{i=1}^{K} A_i \hat{E}_i[n] \sin\left(2\pi i f_{\text{new}} \frac{n}{f_s} + \theta_i^{\text{corrected}}\right) \tag{9}$$

Note that each harmonic frequency is $i \cdot f_{\text{new}}$, preserving the harmonic series relationship while the envelopes $\hat{E}_i[n]$ remain unchanged, thus decoupling pitch from temporal evolution.

III. Experimental Validation

Three experiments validate the system across a test set spanning 14 total bass guitar, piano, violin, clap, and voice samples with varying timbral and envelope characteristics.

A. Experiment 1: Harmonic Count Sensitivity Analysis

Objective: Determine how reconstruction quality varies as a function of the number of harmonics K retained in the synthesis, with detailed analysis of the two samples with the highest number of detected harmonics.

Method: For all samples, test $K = 1, 2, \ldots, K_{\text{max}}$, where for each K:

- Reconstruct signal using only the first K harmonics
- Optimize basis functions for these K harmonics
- Compute MSE between reconstructed and original waveforms

Metric:

- MSE $= \frac{1}{N}\sum_{n=1}^{N}(x[n] - \hat{x}[n])^2$

Results: The `Clap` sample ($K_{\text{max}} = 30$) exhibits monotonic error reduction from $K = 1$ to $K = 30$, demonstrating the benefit of extensive harmonic content for percussive timbres. `Violin A4` ($K_{\text{max}} = 16$) shows greater initial sensitivity, with MSE decreasing from $K = 1$ to $K = 5$, then plateauing around 0.16 for $K \geq 6$. This reflects the violin's complex harmonic structure with significant energy in higher partials.

Across all samples, the average MSE delta analysis reveals adding the 2nd harmonic provides the largest benefit, while subsequent harmonics contribute progressively smaller improvements (typically 1% to 4%). Beyond $K \approx 0.5 \times K_{\text{max}}$, additional harmonics provide minimal improvement, suggesting an optimal operating point for compression applications.

B. Experiment 2: Basis Function Selection Analysis

Objective: Analyze which basis functions are selected for different instrument types and harmonic indices to identify patterns and validate the need for per-harmonic optimization.

Method: Run per-harmonic optimization on all test samples and analyze selection patterns across 102 total harmonics.

Results: All 102 harmonics unanimously selected ATK0 and DEC1 (100%), indicating universal transient capture. SUS3 (10ms uniform) is most common at 50%, followed by SUS0 (1ms) at 32.4%. REL1 (3-sample difference) dominates other release options with 82.4% prevalence. Higher harmonics generally prefer shorter sustain windows. Bass guitar shows most variation while piano and violin are highly consistent, though this may reflect the sample composition. Table I and Figure 2 show the distribution of basis functions across all harmonics and per-harmonic selections for representative samples.

TABLE I
PER-HARMONIC BASIS FUNCTION SELECTION FOR THREE REPRESENTATIVE SAMPLES

Sample	Harm.	Freq (Hz)	SUS	REL
`Bass Guitar D2` (f_0=221.2 Hz, K=5)	1	221.2	SUS2	REL1
	2	442.3	SUS2	REL1
	3	663.9	SUS0	**REL6**
	4	885.1	SUS0	REL1
	5	1105.8	SUS3	REL1
`Violin A4` (f_0=439.8 Hz, K=16)	1	439.8	SUS0	REL1
	2	879.3	SUS3	REL1
	3	1319.2	SUS3	REL1
	6	2638.3	SUS2	**REL3**
	7	3078.2	SUS3	**REL2**
	11	4838.3	SUS3	**REL2**
	14	6157.5	SUS4	REL1
	15	6596.9	SUS3	**REL2**
`Piano Middle C` (f_0=296.9 Hz, K=9)	1	296.9	SUS4	**REL2**
	2	591.3	SUS3	**REL2**
	3	888.2	SUS3	REL1
	5	1484.4	SUS3	REL1
	7	2083.2	SUS3	**REL0**
	9	2674.5	SUS3	REL1

C. Experiment 3: Most Common vs. Optimized Basis Function Comparison

Objective: Quantify the reconstruction quality improvement achieved by per-harmonic greedy selection compared to using the most commonly selected basis functions (from Experiment 2) for all harmonics.

Method: Process all test samples with most common basis functions (ATK0, DEC1, SUS3, REL1) used on every harmonic vs. per-harmonic optimized selection. Compute total MSE across all ADSR segments.

Results: Figure 3 reveals a critical limitation of greedy sequential optimization. Only piano samples benefit from optimization, while bass and violin samples show degradation; others remain unchanged due to selecting identical basis functions in both approaches. This counterintuitive result occurs because the optimizer *computes transitions using different decay features at each optimization stage*: attack optimized

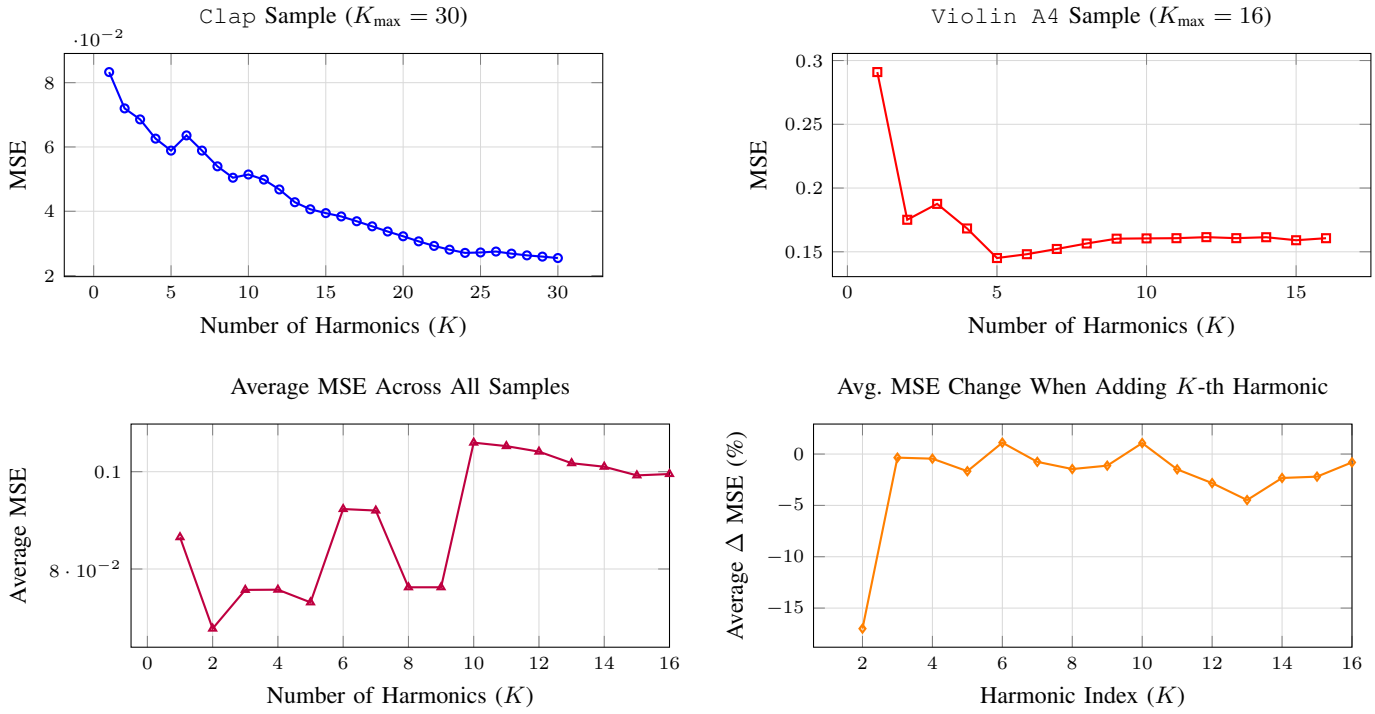

Fig. 1. Experiment 1 Results: (Top) MSE vs. K for `Clap` and `Violin A4`, the two samples with highest harmonic counts. (Bottom Left) Average MSE across all samples shows sharp improvement at $K = 2$ with diminishing returns thereafter. (Bottom Right) Percentage change in MSE when adding the K-th harmonic reveals that the 2nd harmonic provides the largest benefit (17%), with subsequent harmonics contributing progressively smaller improvements.

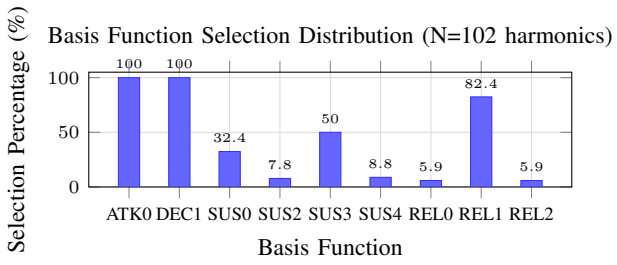

Fig. 2. Basis function selection distribution across all 102 harmonics from the test set. Attack and decay show unanimous selection (ATK0, DEC1 at 100%), while sustain and release exhibit more variation. SUS3 (50%) and REL1 (82.4%) are most common in their respective stages.

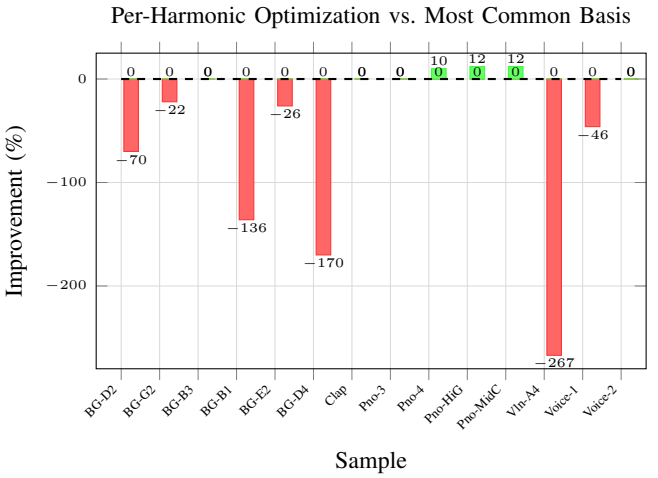

Fig. 3. Reconstruction quality improvement of per-harmonic optimized basis functions compared to most common basis (ATK0, DEC1, SUS3, REL1). Green bars indicate improvement, red bars indicate degradation. Only piano samples benefit (10-12%), while bass guitar and violin degrade significantly (22% to 267%) due to transition mismatch in greedy optimization.

using decay features from (0,0,0,0), decay from (A*,0,0,0), sustain from (A*,D*,0,0), and release from (A*,D*,S*,0). However, final reconstruction uses transitions computed from the complete basis (A*,D*,S*,R*). Since transition point $n_{i,2}$ depends on the DEC feature (which changes with each decay kernel selection), segment boundaries shift during optimization, creating a mismatch between optimization and evaluation conditions that leads to local minima. The strong performance on piano samples suggests these basis functions align well with sharp transients, while bass and violin samples suffer from the transition mismatch problem.

IV. Discussion

A. Analysis: Validation of Per-Harmonic Optimization

The critical finding validating per-harmonic optimization is that different harmonics of the same require different basis functions. Table I demonstrates this for three representative samples. This per-harmonic diversity demonstrates that

treating all harmonics identically would fail to capture the independent partial evolution characteristic of real instruments, where mode coupling and material properties cause frequency-dependent temporal behavior.

Experiment 3 provides further validation that per-harmonic envelope optimization does improve reconstruction quality when harmonics exhibit independent temporal evolution, as demonstrated by piano samples which achieve improvements exceeding 10%. This validates the core hypothesis that different harmonics require different basis functions.

Proof of Concept (Piano Samples): Piano samples provide the clearest evidence for per-harmonic optimization. Crucially, `Piano Middle C` exhibits per-harmonic variation where harmonic 1 uses (SUS4, REL2), while harmonics 2-9 vary between (SUS3, REL2), (SUS3, REL1), and (SUS3, REL0), with harmonic 7 uniquely selecting REL0. Due to this heterogeneity, optimization achieves 12% improvement over uniform basis functions. `Piano High G` and `Piano 4` show similar gains, with `Piano High G` varying between REL1 and REL0 across harmonics. These results conclusively demonstrate that when harmonics evolve differently, per-harmonic optimization captures this behavior and improves reconstruction. The fact that piano harmonics require different release functions reflects the physical reality of frequency-dependent damping in struck strings, and the optimization successfully exploits this structure.

Limitations of Greedy Sequential Optimization: Bass guitar and violin samples reveal an important caveat: the greedy sequential optimization strategy introduces a transition point dependency problem for instruments with extreme per-harmonic heterogeneity. `Bass Guitar D2` (three sustain variants, two release variants across 5 harmonics) and `Violin A4` (four sustain variants, three release variants across 16 harmonics) suffer degradation because transition point $n_{i,2}$ shifts unpredictably as decay kernels update during sequential optimization stages. However, this degradation reflects a limitation of the greedy algorithm, not a failure of the per-harmonic modeling concept. The fact that these samples select such diverse basis functions (Table I) confirms that different harmonics genuinely evolve at different rates, validating the underlying physical model.

Thee greedy sequential algorithm can be improved upon by implementing an exhaustive search or using machine learning techniques. Machine learning algorithms could also be used to find optimal basis functions for each type of sound and harmonic, potentially allowing for real-time synthesis with a pre-determined set of basis functions and number of harmonics.

Implication: The experimental evidence supports per-harmonic envelope modeling: piano samples with moderate per-harmonic variation demonstrate measurable reconstruction improvements when optimization adapts to each harmonic's temporal characteristics. For instruments with extreme heterogeneity, the concept remains valid but requires more sophisticated optimization strategies, either joint optimization of all basis functions simultaneously or fixed transition points independent of basis selection. The unanimous selection of different sustain and release functions across harmonics (Experiment 2) provides independent confirmation that per-harmonic adaptation captures real physical phenomena in musical instrument acoustics. The fact that only sustain and release basis functions varied further support this claim. Different harmonics die out at different rates meaning the difference is most pronounced further after the onset of the note.

B. Design Evolution and Challenges

This project evolved significantly during development. Initial Wiener deconvolution attempts proved inadequate due to treating all harmonics identically, ignoring the physical reality that different partials evolve at different rates. This led to the per-harmonic architecture.

Filter design presented the primary technical challenge. Elliptic IIR filters proved unstable at low frequencies; Butterworth filters required prohibitively high orders. FIR filters with Kaiser windowing ultimately proved ideal, offering guaranteed stability and arbitrary stopband attenuation despite computational cost.

The automated greedy selection evolved through three versions: fixed single functions, manual selection from 3 options per stage, and per-harmonic automated greedy selection from 26 options. The greedy sequential strategy was chosen for efficiency ($26K$ trials vs. 1764^K exhaustive search). However, Experiment 3 reveals a fundamental limitation: each stage optimizes using transitions from *partial* basis functions, but final reconstruction uses transitions from the *complete* basis, leading to local minima where optimization underperforms fixed global choices.

C. Limitations and Assumptions

Several assumptions constrain the model's applicability:

- **Harmonic periodicity:** The model assumes perfect integer-multiple relationships $f_i = i \cdot f_0$. Inharmonic instruments (piano, bells) with stretched or compressed partials are not accurately captured, though the 10% frequency error tolerance provides some robustness.
- **Constant fundamental:** The current implementation assumes static f_0 throughout the sample. Pitch glides or vibrato that changes the fundamental frequency (as opposed to amplitude modulation) require time-varying fundamental tracking.
- **Exponential release:** The release model assumes approximately exponential decay. Some instruments (bowed strings with after-ringing, sustained wind instruments) exhibit more complex release profiles.
- **Monophonic signals:** Polyphonic input causes harmonic overlap, violating the independence assumption of the filter bank approach. Multiple simultaneous notes produce intermodulation that cannot be cleanly separated.

D. Computational Considerations and Future Work

Per-harmonic optimization is expensive, but produces a compact parametric representation enabling real-time synthe-

sis after the initial computation. Applications include pitch-shifting for DAW plugins, timbre morphing, and envelope manipulation. Future work includes inharmonicity modeling ($f_i = i \cdot f_0 + \delta_i$), time-varying fundamental tracking, adaptive release models, polyphonic source separation, and more robust basis selection and creation algorithms.

E. Perceptual Evaluation and Limitations of MSE Metrics

While mean squared error provides a quantitative measure of waveform reconstruction accuracy, it treats all frequency components equally, whereas human auditory perception exhibits frequency-dependent sensitivity.

Future work should incorporate formal perceptual evaluation using standardized perceptual audio quality metrics or listening test methodologies. Additionally, spectral distance metrics could be used in conjunction with MSE metrics to capture more perceptually-relevant differences.

If you wish to determine timbral quality for yourself, included in the code submission is a `synth.py` python script. Running gives a demonstration of the algorithm, see the provided `readme` file for further information.

V. Conclusion

This work presents a parametric synthesizer for monophonic musical signals at arbitrary pitches through per-harmonic ADSR envelope modeling. The key innovation is per-harmonic optimization where each of K harmonics independently selects optimal basis functions from 26 configurations using MSE metrics. Experiment 3 reveals that greedy sequential optimization produces suboptimal results due to transition mismatch: piano samples show improvements exceeding 10%, while bass and violin samples degrade. The variance in sustain and release basis function selection as well as the results from experiment 1 validate the per-harmonic approach, capturing that different partials evolve at different rates. Future work should investigate different selection algorithms for both basis functions and transition points, potentially exploring machine learning or other adaptive methods. This parametric representation enables pitch-shifting applications in music production while preserving instrumental character and demonstrates a per-harmonic approach yields significant improvements over a universal envelope approach.

References

[1] Oppenheim, A. V., and Schafer, R. W. *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ: Prentice Hall, 2010.

[2] Smith, J. O. *Spectral Audio Signal Processing*. W3K Publishing, 2011. [Online]. Available: https://ccrma.stanford.edu/~jos/sasp/

[3] McAulay, R. J., and Quatieri, T. F. "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 4, pp. 744–754, Aug. 1986.

[4] Serra, X., and Smith, J. O. "Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, Winter 1990.

[5] Flanagan, J. L., and Golden, R. M. "Phase vocoder," *Bell System Technical Journal*, vol. 45, no. 9, pp. 1493–1509, Nov. 1966.