Analysis and Critical-Band-Based Group Wavetable Synthesis of Piano Tones

Zheng, Hua and James W. Beauchamp
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
{huazheng, j-beau}

Abstract
From the time-frequency analysis of a rich library of piano tones recorded specifically for this study, we have found many piano tone characteristics that may lead to efficient representation and synthesis. Based on the analysis results and elementary psychoacoustic principals, we have developed a wavetable synthesis algorithm using a group synthesis model which yields perceptually accurate sounds and computational efficiency.

1 Introduction
While the physics of piano sound production has been studied extensively, little has been done to explore the temporal and spectral characteristics of piano tones for efficient and perceptually accurate synthesis. Synthesis methods such as additive, physical modeling, FM, and sampling synthesis have been constrained by tradeoffs between efficiency, accuracy, variety, and flexibility. A recent study [Lee and Horner, 1999] reported good results of synthetic piano tones using a group wavetable synthesis model with a genetic algorithm applied for the grouping procedure. This study employs a similar model with a different grouping algorithm.

2 Piano Tone Acquisition
This study adopts an analysis-synthesis paradigm that requires a variety of high-quality recorded piano tones. Commercial piano samples are usually made available for specific synthesizers and samplers and lack the variety and precision in performance parameters, such as dynamic and articulation. For the purposes of this study, a complete set of piano tones covering wide spans of pitch, dynamic, and duration, and pedal usage was recorded under repeatable conditions. The tones were produced by a Yamaha Disklavier studio grand piano controlled by MIDI messages. Therefore, all of the tone parameters are explicit and quantified. The set of tones includes 22 pitches from A0 to A7 in major third intervals, at six dynamic levels, and with three durations. Also, tones with selected pitches, dynamics, and duration were recorded with the sustain pedal and soft pedal individually applied.

3 Piano Tone Characteristics

3.1 Analysis of piano tones
On this rich library of piano tones, the SNDAN software package [Beauchamp, 1993] was used for pitch-synchronous phase vocoder analysis, parameter estimation, and graphical display. The bin frequencies of the phase vocoder are inherently harmonically spaced. Because some high partials of a piano sound are inharmonic, they will fall between adjacent analysis bins and be represented by two pseudo-harmonics. This misrepresentation is not a problem for the synthesis, the reason to be discussed in Sec. 4.1. The sinusoid-plus-noise analysis method of SMS [Serra, 1997] was used to supplement the phase vocoder analysis to obtain the partial frequencies and the action noise of piano tones. Some results from these analysis are discussed here.

3.2 Piano tone characteristics

3.2.1 Duration
A typical piano tone can be divided into three parts: attack, prompt-sound, and after-sound. The theory of string coupling and vibration modes of the strings of a unison indicates dual decay rates in piano envelopes [Weinreich, 1977]. In this study, piano decay was found to have two different stages with regard to its duration and amplitude, but few examples of conspicuous dual decay rates exist in the partial time envelopes and their RMS amplitudes of all pitches at all dynamics. The attack times of all pitches at all dynamics are within the range of 10–20 ms. The prompt-sound duration is in reverse proportion of dynamic, i.e. the louder the tone, the shorter the prompt-sound. Intuitively, the after-sound duration increases with dynamic and decreases with pitch.
3.2.2 Dynamic

A loud piano tone sounds brighter than a soft one, which is the case for almost all instruments. Fig. 1 shows the normalized average spectral envelopes of four pitches at six dynamics. For higher dynamics, some high partials have relatively greater strength. The same result holds for the partial envelopes and their RMS amplitudes of the same pitch at different dynamics. The RMS amplitudes of A0 and A2 at six dynamics are shown in Fig. 2. Four representative normalized partial envelopes of A2 are shown in Fig. 3. They share a significant similarity even at fine scales. For some partials, like partials 3 and 10, their normalized envelopes at different dynamics coincide perfectly. Other partials, like partials 1 and 13, as well as the normalized RMS amplitudes, show a dynamic balance between the onset (attack and prompt-sound) and the after-sound: a louder tone has a stronger onset but a relatively weaker after-sound. The relationship between dynamic and onset strength is illustrated in Fig. 4, where all the envelopes are normalized with respect to their values at about 1.1 seconds so that the after-sounds have a best match.

3.2.3 Frequency characteristics

1. Inharmonicity

The partial frequencies of most pitches are found to satisfy the classical inharmonic formula [Fletcher, 1964] fairly well:

\[ f_k = k f_1 \sqrt{\frac{1 + B k^2}{1 + B}} \]

The inharmonicity coefficient \( B \) varies with pitch but not with dynamic for the same pitch. The perceptual effect of changing the value of \( B \) differs from pitch to pitch. The investigation of the proper ranges of \( B \) across pitches for realistic versus agreeable sound is left for future listening tests.

2. Frequency deviation

The frequencies of the most significant partials of all pitches at all dynamics are found to have only minor and random fluctuations. Weak partials have much larger time-varying frequency deviation mostly due to noise and analysis artifacts. Therefore, constant inharmonicity can be used in the synthesis.

3.2.4 Pedal

1. Sustain pedal

A significant amount of fluctuations is found in the spectra of sustained piano tones because of the sympathetic vibration of all the strings. Strong sympathetic vibration only exists in the neighboring strings of the strings that are struck, resulting in some form of beating, and the vibration diminishes as the energy of the struck strings dissipates.

2. Soft pedal

Depressing the pedal shifts the action horizontally, at a distance that depends on the piano brand and the degree by which the pedal is depressed. For a long shift, the leftmost string among of the strings of a unison will not be struck. The string coupling will be different, leading to a timbre change. For a shorter shift, the part of the hammer that is softer due to lack of impact will hit the strings. The resulting tone has less significant high partials, which is similar to a tone of a low dynamic.

4 Group Wavetable Synthesis

4.1 Synthesis algorithm

The critical-band-based group wavetable synthesis offers intuitive interpretation and flexible control like additive synthesis, with great computation efficiency which does not compromise perceptual accuracy. Fig. 5 shows the analysis-synthesis diagram. The synthesis part is formulated as:

\[ s(n) = \sum_{i=1}^{N_g} c_i(n) \sum_{k \in G_i} p_{ik} \cos \left( 2\pi (k f_a \beta_i) n + \theta_{0k} \right), \]

where

\[
\begin{align*}
  n & = \text{sample number}, \\
  i & = \text{group number}, \\
  k & = \text{partial number}, \\
  s(n) & = \text{final signal}, \\
  N_g & = \text{number of groups}, \\
  c_i(n) & = \text{the envelope of group } i, \\
  G_i & = \text{the set of partials within group } i, \\
  p_{ik} & = \text{the weight of partial } k \text{ within group } i, \\
  f_a & = \text{the analysis fundamental frequency}, \\
  \beta_i & = \text{the stretching factor of group } i, \\
  \theta_{0k} & = \text{the initial phase of partial } k.
\end{align*}
\]

The second sum representing the group signal is stored in a wavetable \( t_i(m) \). The crucial part of the algorithm is the grouping procedure: how to get a \( G_i \) set that yields a small \( N_g \) and minimum perceptual difference between the original and synthetic sounds. Critical bandwidth is used to find
a disjoint set of critical-band groups (CBGs). Those consecutive partials within a critical band centered at the middle partial are included in one CBG. Each pitch up to C3 yields 26 CBGs, and the number of CBGs decreases to six for A7.

The assumption for this grouping procedure is that the partials within one CBG are perceived not individually but as one acoustical entity. The overall presence of those partials is much more important than their individual micro structures. This also validates the use of phase vocoder on the inharmonic piano tones, the problem mentioned in Sec. 3.1. An inharmonic partial and its associated pseudo-harmonics belong to the same CBG, so the two pseudo-harmonics grouped together are perceptually equivalent to the original inharmonic partial.

For greater data reduction and more efficient implementation, a second grouping step combining consecutive CBGs can be applied. The criterion is that two strong CBGs cannot be combined, because grouping partials across critical-bands would lead to a perceptible change if their envelopes differ greatly. The result is a set of hypergroups (HPGs), each containing a number of consecutive partials from one or more CBGs. Tradeoff can be made between the number of HPGs and the perceptual quality.

4.2 Implementations

The synthesis algorithm is divided into a pre-synthesis stage and a wavetable synthesis stage. The first stage generates all the data required for the wavetable synthesis. The complete algorithm is programmed as a subroutine for the SNDAN package. The wavetable synthesis part has been implemented as a Music 4C instrument [Beauchamp, 1993] and a plug-in actor for a real-time software sound server [Bargar et al., 1994]. Both implementations utilize the data files generated by the pre-synthesis stage. In the data files, group envelopes $e_i(n)$ are normalized to one, and the normalizing factors $w_i$ for available pitches and dynamics are tabulated. Tone duration is also tabulated. All of the piano characteristics described in Sec. 3.2 can be implemented:

1. Duration

Group envelopes $e_i(n)$ are divided into onset and after-sound. The after-sound part is time-stretched to achieve the appropriate duration calculated from the duration table for arbitrary pitch and dynamic.

2. Dynamic and soft pedal

The appropriate normalizing factors of group envelopes are calculated from the $w_i$ table for arbitrary pitch and dynamic. To implement the dynamic balance, the onset and after-sound parts of the group envelopes are scaled differently.

3. Sustain pedal

A decaying sinusoid with pseudo-random amplitude and frequency variation can be superimposed on the group envelopes as a crude approximation of the sustain effect.

4. Inharmonicity

Constant inharmonicity is applied to groups of partials because they are encapsulated in the group wavetables. This is not a problem since those partials in a CBG or HPG are consecutive so that their frequency deviation from perfect harmonics are close. A recent study [Scalcon et al., 1998] also showed that above certain cutoff frequency, equal- and stretched-distance positioning of the partials does not lead to perceptual difference.

Informal listening tests showed that the piano tones synthesized using full CBGs cannot be distinguished from the original sounds. Using HPGs less than full CBGs for real-time performance yields good sound quality, but the synthetic sounds are more likely to be discriminated from the original.

5 Conclusion

The analysis of recorded piano tones with explicit parameters gives a complete representation of the sounds, from which a number of piano tone characteristics are discovered. The critical-band-based group wavetable synthesis uses the analysis results and produces convincing piano tones. This framework is also promising for the analysis-synthesis of other musical instrument sounds.

References


---

Figure 3: Normalized partial envelopes of A2

Figure 1: Normalized average spectral envelopes

Figure 2: RMS amplitudes of A0 and A2

Figure 4: Renormalized partial envelopes

Figure 5: Analysis-synthesis diagram