

WAVETABLE MATCHING OF PITCHED INHARMONIC INSTRUMENT TONES

Clifford So, Andrew Horner, and Lydia Ayers

Department of Computer Science
Hong Kong University of Science and Technology
Clear Water Bay, Kowloon, Hong Kong

email:cliffso@cs.ust.hk, horner@cs.ust.hk, layers@cs.ust.hk

ABSTRACT

Wavetable matching is the process of finding the parameters needed to resynthesize a musical instrument tone using wavetable synthesis. The most important parameters are the basis spectra. Previous work using genetic algorithm (GA) determination has assumed the original tone was harmonic or nearly harmonic. This assumption is not satisfied by tones such as those from the plucked strings. This paper introduces a new adaptive and automatic wavetable matching technique that employs a hierarchical grouping method to group the partials with similar normalized frequency deviations. Ordinary wavetable matching is then applied to individual groups to find their basis spectra. Results show that for 11 instrument tones with varying amounts of inharmonicity, the new method improves the perceived match on the pitched inharmonic tones compared to ordinary wavetable matching.

1. INTRODUCTION

Wavetable matching is one of the most successful method to match wavetable parameters for resynthesizing instrument tones [1]-[4]. Previous work on wavetable matching has used genetic algorithms (GAs) [5][6] to determine the best time points to select the *basis spectra* (spectra of the wavetables). It picks spectral snapshots from the original tone as the candidate basis spectra. To judge the fitness of a particular combination of basis spectra, the fitness function uses a least-squares solution to find the amplitude envelopes of each basis spectra, and an average relative amplitude error (RAE) to measure the difference between the original and matched tones:

$$RAE = \frac{1}{N_{frames}} \sum_{j=1}^{N_{frames}} \sqrt{\frac{\sum_{k=1}^{N_{hars}} (b_{k,j} - b'_{k,j})^2}{\sum_{k=1}^{N_{hars}} b_{k,j}^2}} \quad (1)$$

where the N_{frames} is the number of analysis frames, N_{hars} is the number of partials, $b'_{k,j}$ is the k th partial amplitude of the synthesized tone at the j th analysis frame, and $b_{k,j}$ is the original tone's amplitude on the same partial and analysis frame.

However, wavetable synthesis makes a basic assumption: the partials are harmonics, i.e., their frequencies are restricted to integer multiples of the fundamental frequency ($f_k = kf_1$). This causes problems when matching inharmonic instruments, such as the string instruments [7][8]. We define the normalized frequency deviation (NFD) for a specific partial k as the following:

$$NFD_k = \frac{\sum_{t=1}^{N_{frames}} b_{i,t} \times \left(\frac{f_k(t)}{kF} - 1 \right)}{\sum_{t=1}^{N_{frames}} b_{i,t}} \quad (2)$$

where F is the fixed fundamental frequency (specified by the user prior to spectrum analysis), $f_k(t)$ is the actual time-varying frequency of partial k , and t is the analysis frame number.

This paper introduces a new wavetable matching technique that adapts wavetable matching to pitched inharmonic tones. The procedure is fully automatic and requires no intervention from the user about the inharmonicity of the tone, compared to the previous version of the algorithm [9]. The new method employs a hierarchical grouping technique that automatically determines the best parameters for the tones. Section 2 details the new method. Section 3 gives results for the 11 instrument tones. Section 4 concludes our paper.

2. WAVETABLE MATCHING OF PITCHED INHARMONIC TONES

A new version of the wavetable matching is presented in this section. The idea is to group the original tone's partials based on their normalized frequency deviation (NFD), and determine the basis spectra and amplitude envelopes of individual groups by normal GA wavetable matching. The algorithm accepts the required number of

wavetables (N_{tabs}) as input and outputs the optimal parameters that gives the smallest relative amplitude error (RAE) and normalized frequency deviation error ($NFDE$). Section 2.1 discusses the hierarchical grouping method of the partials. Section 2.2 shows how to determine the basis spectra in the tree. Section 2.3 shows how to find the level with the lowest overall error. Section 2.4 describes how to resynthesize the signal.

2.1. Hierarchical grouping of partials

At the beginning, a short-time Fourier analysis such as the phase vocoder [10][11] transforms the tone from the time domain to the frequency domain. To track the severe pitch changes, we apply the McAulay-Quatieri [12] analysis. Beauchamp [13] gives more detail about the procedures. The tone's spectrum and its corresponding NFD (from Equation 2) of each partial are obtained.

We measure the “goodness” of a grouping of partials by its normalized frequency deviation error ($NFDE$), which is:

$$NFDE = \sqrt{\sum_{i=1}^{N_{groups}} NFDE_i^2} = \sqrt{\sum_{i=1}^{N_{groups}} \sum_{k \in group_i} (NFD_k - GNFD_i)^2} \quad (3)$$

where N_{groups} is the total number of groups, $NFDE_i$ is the internal $NFDE$ of group i . $GNFD_i$ is the average NFD_k in group i , which is defined as

$$GNFD_i = \overline{NFD_{k \in group_i}} = \frac{1}{H_i} \sum_{k \in group_i} NFD_k \quad (4)$$

where H_i is the number of partials in group i . $NFDE$ basically measures the squared error of every partial's NFD to the group center. The lower the $NFDE$ of the overall grouping, the better the grouping is.

We propose a hierarchical grouping algorithm that gives as low an $NFDE$ as possible for a given total number of groups N_{groups} . We first sort all the NFD_k values along an axis. In the beginning, every consecutive pair of values are connected as edge and regarded as single group as a whole. Test every edge for the largest internal $NFDE$ reduction after its removal. That is to give the smallest sum of $NFDE$ of its left sub-group and right sub-group. Break this edge and continue the process until all the partials are separated in its own group. The decomposition history is stored in a tree. Figure 1 shows an example of the decomposition tree with an example tone of 10 partials.

2.2. Determination of the group basis spectra

After the group decomposition tree is formed, ordinary GA wavetable matching can be applied to each group to determine the basis spectra and amplitude envelopes. The next problem is that given the total number of wavetables (N_{tabs}), how do we distribute N_{tabs} to each group in each level?

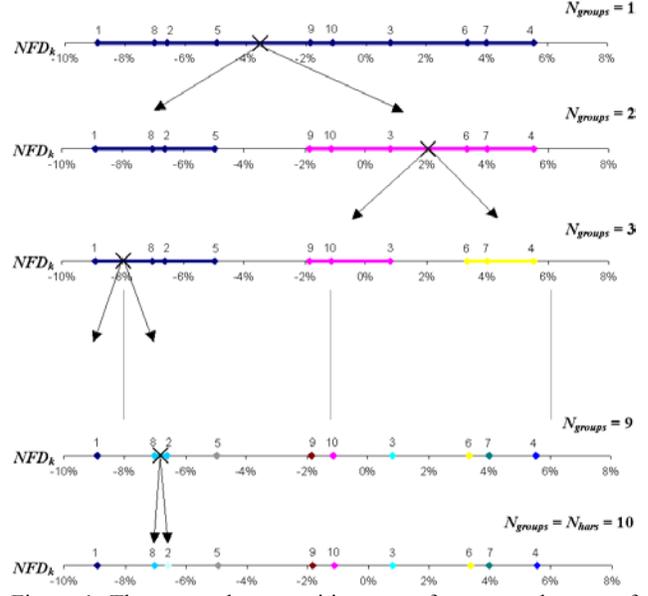


Figure 1. The group decomposition tree of an example tone of 10 partials.

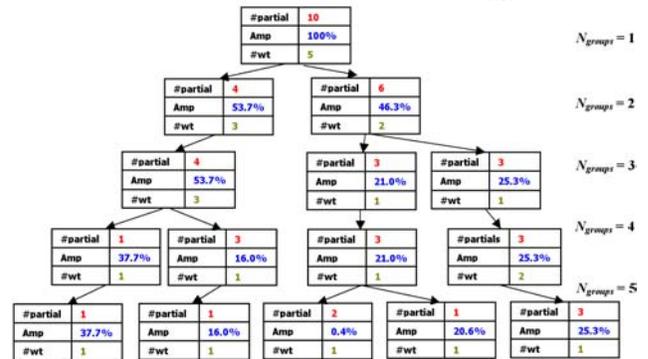


Figure 2. The example wavetable allocation tree based on the tree in Figure 1 with $N_{tabs} = 5$.

We define the number of wavetables in group i ($\#wt_i$) proportional to the amplitude percentage (Amp_i) of group i , that is

$$\#wt_i \approx Amp_i \times N_{tabs} \quad (5)$$

where,

$$\#wt_i \leq \#partial_i \quad (6)$$

and the amplitude percentage (Amp_i) is from the group decomposition tree. $\#wt_i$ will be rounded to the nearest integer. Equation 5 allows stronger group (higher Amp_i) has a higher matching accuracy. Stronger group is more important because the total relative amplitude error (RAE) (see Equation 1) will be adversely affected if the strong group does not match well. It gives a fair distribution of wavetables according to the amplitude percentage of each group. Equation 6 restricts that $\#wt_i$ should be no more than $\#partial_i$, because when $\#wt_i = \#partial_i$, we can perform additive synthesis directly in the group, which gives 100% match to the group (We call this kind of group as perfect group in the later sections). The surplus

wavetable will be passed to the next strongest group and so on. Figure 2 shows an example of wavetable allocation tree with $N_{tabs} = 5$ based on the example decomposition tree from Figure 1.

Each group (box in Figure 2) is then passed to the ordinary GA-based wavetable determination process. The corresponding basis spectra, amplitude envelopes and internal amplitude error (RAE) of each group are obtained. The total amplitude error (RAE) and frequency error ($NFDE$) of each level will be calculated by summing up the internal group errors across the level.

2.3. Determination of the optimal grouping

After the amplitude error (RAE) and the frequency error ($NFDE$) are obtained in each level, we need to determine which level gives the best grouping. We plot a graph of RAE versus $NFDE$ of all the levels of the tree. Figure 3 shows the plot of the tree in Figure 2. However, there are only 5 points in our curve since there are only 5 levels in our example tree ($N_{tabs} = 5$).

We need to find a point that gives both amplitude error and frequency error as low as possible. In other words, a point is a best match if it is the closest to the origin. It can be measured by using the normalized Euclidian distance of the point to the origin. The matching quality (MAQ) of a point is then defined as:

$$MAQ = \sqrt{\left(\frac{NFDE}{maxNFDE}\right)^2 + \left(\frac{RAE}{maxRAE}\right)^2} \quad (7)$$

where the $maxNFDE$ and $maxRAE$ are the maximum $NFDE$ and RAE of the curve. A normalized distance is more preferable than an absolute distance because the $NFDE$ and RAE always differ in a large scale. In our example, the 3rd level point ($N_{groups} = 3$) is selected as the optimal grouping as it gives the lowest normalized MAQ in this case. All of its grouping parameters will be output for signal resynthesis.

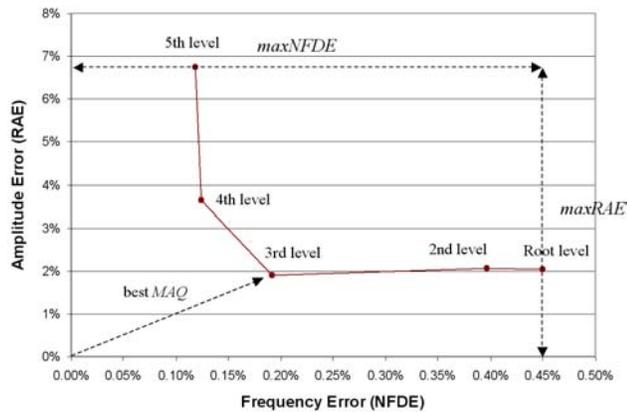


Figure 3. The RAE versus $NFDE$ plot of the tree in Figure 2. The 3rd level ($N_{groups} = 3$) gives the best matching quality.

2.4. Resynthesis of the signal

To resynthesize the signal, we first synthesize the signal of each group $x_i(t)$ using the ordinary multiple wavetable synthesis and sum up all the signals to produce the final synthetic signal $x(t)$.

To synthesize the group signal $x_i(t)$, every wavetable basis spectra in group i will be synthesized with a common time-varying fundamental frequency $g_i(t)$ to generate the time-varying wavetable $y_i(t)$. $g_i(t)$ is defined as:

$$g_i(t) = F \times (1 + v_i(t)) = F \times \left(1 + \frac{\sum_{k \in Group_i} b_{k,t} \times NFD_k(t)}{\sum_{k \in Group_i} b_{k,t}} \right) \quad (8)$$

where F is the fixed fundamental frequency (specified by the user prior to spectrum analysis), and $v_i(t)$ is the time-varying NFD of group i , which is equal to the weighted-average of $NFD_k(t)$ of every partial k in group i by the partial amplitude $b_{k,t}$.

Afterwards, we synthesize the group signal $x_i(t)$ using multiple wavetable synthesis.

$$x_i(t) = \sum_{j=1}^{\#wt_i} w_{i,j}(t) \times y_{i,j}(t) \quad (9)$$

where $y_{i,j}(t)$ is the time-varying j -th wavetable for group i generated from $g_i(t)$, $w_{i,j}(t)$ is the time-varying weight of the j -th wavetable for group i , and $\#wt_i$ is the number of wavetables in group i . Finally, we sum up the signals of all the group $x_i(t)$ to produce the resynthesized signal $x(t)$.

$$x(t) = \sum_{i=1}^{N_{groups}} x_i(t) \quad (10)$$

3. RESULTS

This section gives the results of the new method compared to the original wavetable matching method. Figure 4 shows the results of the 11 pitched inharmonic instrument tones with varies inharmonicity. Each tone is tested with 1-12 number of wavetables (N_{tabs}). We decide the level of indistinguishability to be less than 5% in the amplitude error (RAE) and less than 1% in the frequency error ($NFDE$). The levels are shown as dotted line in the figures.

The amplitude error of the new method decreases faster than the original method along the N_{tabs} . Since when large number of wavetables are distributed in the tree, more perfect groups (see Section 2.2) result. That lowers the overall amplitude error. On the other hand, the frequency error of the new method decreases with N_{tabs} but the original remains unchanged, because the original method does not involve any group decomposition mechanism and the frequency error stays the same throughout the process.

The new method gives overall lower errors so that better match can be done using fewer wavetables using the new method. The new method is also capable of

matching extreme inharmonic tones like the Qin, Zheng and Yangqin, in which the ordinary method fails.

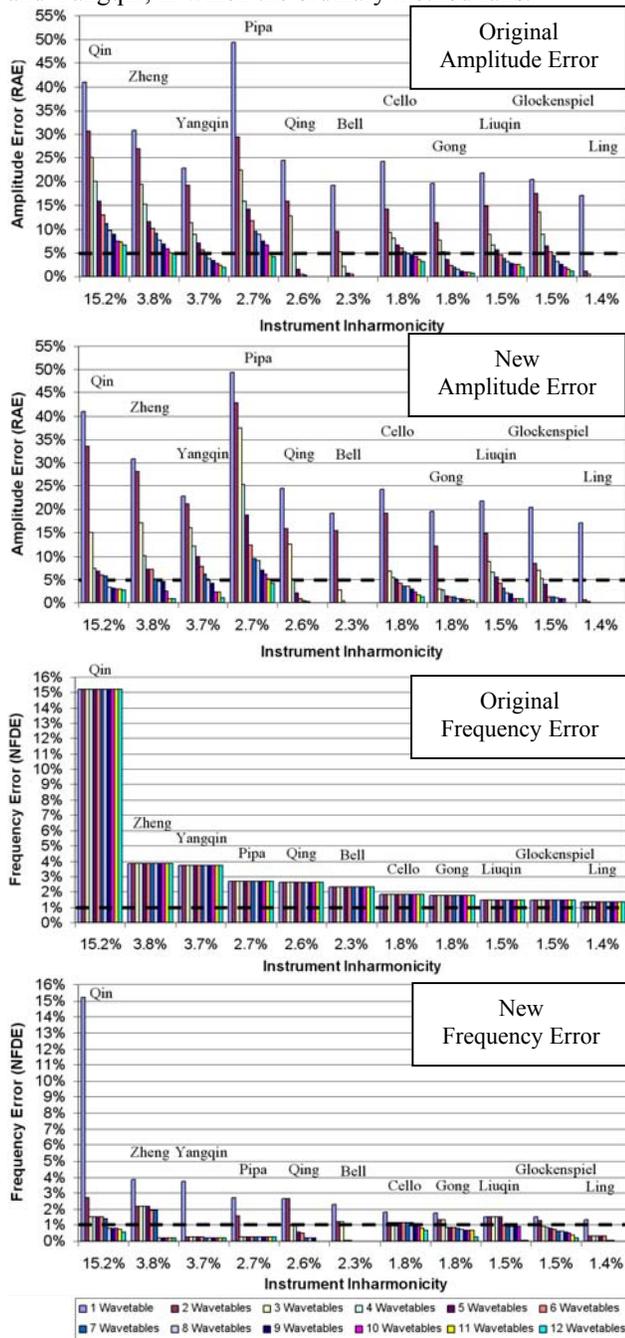


Figure 4. The amplitude error (RAE) and frequency error (NFDE) of the original and new method of the 11 pitched inharmonic tones.

4. CONCLUSION

Ordinary wavetable matching method cannot effectively resynthesize pitched inharmonic tones such as the plucked strings. To remedy this problem, an adapted method has been presented to match pitched inharmonic tones. The

new method separates the partials into groups to give better frequency resolution. The new method uses fewer wavetables due to perfect groups.

The new method does not handle non-pitched instruments such as the cymbal, or general sounds because their partials are much varied and unstable over time, so no quasi-periodic waveform can be extracted. Future work might take into account masking and the high sensitivity of the human ear in the frequency range of 250 to 3000Hz [18]. New wavetable allocation strategies will be needed to take into account these issues.

5. REFERENCES

- [1] M.-H. Serra, D. Rubine and R. Dannenberg, "Analysis and Synthesis of Tones by Spectral Interpolation," *Journal of Audio Eng. Soc.*, vol. 38, no. 3, pp. 111-128, 1990.
- [2] A. B. Horner, J. Beauchamp, and L. Haken, "Methods for Multiple Wavetable Synthesis of Musical Instrument Tones," *Journal of Audio Eng. Soc.*, vol. 41, no. 5, pp. 336-356, May 1993.
- [3] A. B. Horner and J. Beauchamp, "Synthesis of Trumpet Tones Using a Wavetable and Dynamic Filter," *Journal of Audio Eng. Soc.*, vol. 43, no. 10, pp. 799-812, 1995.
- [4] A. B. Horner, "Computation and Memory Tradeoffs with Multiple Wavetable Interpolation," *Journal of Audio Eng. Soc.*, vol. 44, no. 6, pp. 481-496, 1996.
- [5] J. H. Holland, *Adaptation in Natural and Artificial Systems*. Ann Arbor: The University of Michigan Press.
- [6] D. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading, MA: Addison-Wesley.
- [7] S. Dubnov and N. Tishby, "Influence of Frequency Modulating Jitter on Higher Order Moments of Sound Residual with Applications to Synthesis and Classification," *Proceedings of the 1996 ICMC*, pp. 378-385, 1996.
- [8] J. C. Brown, "Frequency Ratios of Spectral Components of Musical Sounds," *Journal of Acoust. Soc. of America*, vol. 99, pp. 1210-1218, 1996.
- [9] C. So and A. B. Horner, "Wavetable Matching of Inharmonic String Tones," *Journal of the Audio Eng. Soc.*, April 2002.
- [10] J. B. Allen, "Short Term Spectral Analysis, Synthesis, and Modification by Discrete Fourier Transform," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-25, pp. 235-238, 1977.
- [11] M. Dolson, "The Phase Vocoder: A Tutorial," *Computer Music Journal*, vol. 10, no. 4, pp. 14-27, 1986.
- [12] R. McAulay and T. Quatieri, "Speech Analysis/Synthesis based on a Sinusoidal Representation," *IEEE Trans. Acoust., Speech, Signal Proc.*, vol. ASSP-34, pp. 744-754, 1986.
- [13] J. Beauchamp, "Unix Workstation Software for Analysis, Graphics, Modification, and Synthesis of Musical Sounds", presented at the 94th Convention of the Audio Engineering Society, *Journal of the Audio Eng. Soc. (Abstracts)*, vol. 41, preprint no. 3479, pp. 387, 1993 May.
- [14] C. Dodge, T. A. Jerse, *Computer Music Synthesis, Composition, and Performance*, Schirmer Books, 1985.