



Title	Time domain analysis/resynthesis of musical tones based on polynomial interpolation techniques
Author(s)	Hung, Roy; Yung, NHC; Cheung, PYS
Citation	International Conference On Signal Processing Proceedings, Icsp, 1996, v. 2, p. 1511-1513
Issued Date	1996
URL	http://hdl.handle.net/10722/45999
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Time Domain Analysis/Resynthesis of Musical Tones Based on Polynomial Interpolation Techniques

Roy Hung¹, N.H.C. Yung², P.Y.S. Cheung³
Dept. of Electrical and Electronics Engineering,
The University of Hong Kong,
Hong Kong.

E-mail:- rhung@hkueee.hku.hk¹, nyung@hkueee.hku.hk², cheung@hkueee.hku.hk³

ABSTRACT

In this paper, a novel algorithm employing polynomial interpolation techniques is proposed for the analysis and resynthesis of musical tones, based on time domain information. This algorithm models a single period of oscillation as a series of features, with curves joining such features together. The trajectories of these features and the shape of the curves across the whole input signal can be parameterised, such that the signal can be analysed and resynthesised as a close approximation to the original. The current research introduced an alternative approach to analyse/resynthesise sampled musical signals in the frequency domain, one which characterises a signal by its physical structure rather than its frequency components. The results from this analysis can be used to further refine existing physical models of musical instruments.

1. INTRODUCTION

Traditionally, the analysis and resynthesis of musical signals have been carried out based on the function performed by the basilar membrane of the cochlea, that of separating the incoming signal into frequency bands. Fourier analysis techniques[1,2,3] fit perfectly into this paradigm since they model periodic waveforms as a series of sine and cosine functions of varying amplitudes, frequencies and phases, with residue components modeled as noise.

In this paper, a novel algorithm is proposed to analyse and resynthesise acoustical signals solely employing time domain information, from musical instruments which accepts as input a continuous flow of energy. The signals from such instruments are said to be "sustainable", since their amplitudes and tonal characteristics can be controlled by the player after the initial onset of energy input. This class of instruments can be functionally divided into two distinctive parts:- one which is directly manipulated by the player, and which involves nonlinear interactions with the player, and

another which amplifies these oscillations into the surrounding space, which can be treated as linear. Here, we concentrate our efforts on the former part (the excitation), because different mechanisms of excitation produce characteristic acoustical outputs. For the class of signals under consideration, each cycle of oscillation is composed of distinct states, each a result of different physical processes. In the time domain such processes can be easily identified and analysed, which is not possible in the frequency domain, due to the process of functional approximation.

2. OVERVIEW OF ALGORITHM

The principle from which this algorithm is developed is the Weierstrass Approximation Theorem[4], which proves that for a function of one variable (time in this case), there exist a polynomial which can interpolate an arbitrary number of points of this function. The present algorithm is composed of two blocks:- Analysis and Resynthesis (Fig. 1). Analysis is carried out on two levels. Firstly, significant features are selected from each period of the input signal so as to identify its general structure. Secondly, significant features from all periods are matched to find those that are present in a large number of periods throughout the input signal, and they are said to be common features of this signal. The common features are used as starting points for the interpolation algorithm which approximates the shape of each period to within an error percentage of each sample. At this stage, we have a reconstructed version of the input where the abstraction is made on the waveshape of the periods, with no modifications to the evolution of the common features. The changes to the common features are then analysed with straight line interpolation, so that the trend of such changes can be readily identified. Finally, the simplified trajectories of the common features are combined with the abstracted waveshape from the interpolation algorithm to form a resynthesised version of the input, where the overall evolution of the entire input signal is abstracted.

As stated in section 1, this research is interested in the excitation signal of an instrument. Therefore, the input

signals to the algorithm were recorded from piezo-electric sensors located at the bridge of a cello at 22.05kHz. The sensors captures the nonlinear interactions between the bow and the string, before this signal is transmitted to the body of the cello.

3. DETAILED DESCRIPTION OF ALGORITHM

The input signal is first divided into individual periods by visual inspection. Subsequently, each cycle of oscillation is searched for local maxima, local minima and zero crossings because such features are indicative of the physical process within a cycle (Fig. 1). A samples is labeled as maximum if its amplitude is greater than its nearest neighbour, a minimum if less than its neighbour. Zero crossing is defined as the point before a sign change in the value of the amplitude occurs[5]. The initial sets are processed by Feature Tuning (Fig. 1) which identifies and deletes insignificant features according to the following rules:-

1) a distance measure d is specified in terms of the number of samples, such that given n samples, and $0 < j < n-1, j < k < n-1$:-

$$\text{if } (x_k^s) - (x_j^s) < d \text{ (superscript } s \text{ stands for feature)} \quad (1)$$

then both x_k^s and x_j^s are considered too close together;

2) A group of features that are deemed to be close are then examined for the difference between their amplitudes, given a percentage deviation dev :-

$$\text{if } \frac{f(x_j^s) - \text{mean}\{f(x_j^s), f(x_k^s)\}}{\text{mean}\{f(x_j^s), f(x_k^s)\}} \times 100 < dev \quad (2)$$

then x_j^s and x_k^s are considered as insignificant.

The optimised feature sets are used as input to the Common Feature Detection module (Fig. 1), so that features can be found which is significant throughout the input signal. Features are matched by their type and distance deviation (in number of samples) from each other, starting from the present period to the last period. If more than one possible match is available, the feature with minimum distance deviation is chosen. If the number of matched features is larger than or equal to a percentage of total number of cycles, then this list of features are labeled as common[5]. After all possible common features are found, the remaining features are deleted from the feature sets.

The common feature sets are used as starting points for the interpolation of each period's waveform; using Hermite polynomials, to a specified percentage error for each sample. Hermite is a member of osculating polynomials, which interpolates not only the values of a

function at given points, but also a specified number of derivatives. Hermite polynomials interpolates to the first derivative, and are guaranteed to generate curves with continuous first derivatives, given the sample values and their slopes. Without such continuity, artificial high frequencies will be generated in the interpolated signals. Also, in a prior study, Hermite polynomial is found to perform well in the number of samples needed to adequately approximate a waveform, and its mean square error performance as the interpolation error is relaxed[5].

The interpolation of the input signal is carried out on a cycle by cycle basis. The curves between common features are first approximated using amplitude and slope information at the end points. Every point within this segment has an error percentage calculated:-

$$E(x_k) = \left| \frac{P(x_k) - f(x_k)}{f(x_k)} \right| \times 100\% \text{ for all } k = 0, 1, \dots, n, \quad (3)$$

Any subsequent interpolated points are taken as the point with worst percentage error that is larger than that specified. Interpolation ends when all samples are within the error percentage specified. The interpolated periods are appended to form the reconstructed signal, which differs from the original in the number of samples used and the shape of each period. We can assess the quality of the reconstructed signal by listening tests and the mean square error between this and the original signal.

The lists of common features are further processed by Envelope Parameterisation (Fig. 1), whose function is to perform straight line fitting of the trajectories of the common features, so as to give an intuitive indication to the evolution of the features across the entire input signal, and to access the subjective effect of parameterising the envelopes of the waveform. At present, such trajectories contain the amplitudes of the common features. Within each trajectory, regions without null entries are found, and the first and last points of this region are joined with straight lines. Every entry within this region has an error percentage calculated using Eq. 3 and further interpolation points are found using the same procedures and criteria in Hermite Interpolation.

From the parameterised trajectories of the common features and the curve shapes from the Hermite Interpolation module, the signal is resynthesised by scaling the parameters of the interpolants such that the resynthesised curve will have the same shape as the original, according to the following definition:-

$f(x_k)$ for $k = 0, 1, \dots, n-1$ $f(x_0) = 0$, is the original curve;

we seek $\tilde{f}(x_k)$ for $k = 0, 1, \dots, n-1$ $\tilde{f}(x_0) = 0$ such that

$$f(x_k) = a\tilde{f}(x_k) \text{ for } k = 0, 1, \dots, n-1 \quad (4)$$

where a is the ratio of the slope as found by joining a straight line between two adjacent common features, to

the slope found before the same features were processed by Envelope Parameterisation. It can be seen that if we multiply the coefficients of the Hermite polynomials with this ratio, a new curve conforming to the above definition will be formed.

4. RESULTS AND CONCLUSIONS

In order to test the performance of the analysis/resynthesis algorithm, twelve tones were used as input, ranging in frequency from 70Hz to 441Hz (see Fig.3, Fig. 4 for all frequency values). Both subjective listening tests and objective MSE tests were used.

The listening tests were carried out with a few participants of no formal musical background. Both the reconstructed signals and the resynthesis signals (Fig. 1) were evaluated for subjective closeness to the original. The general impression was that both types of signals bore a close resemblance to the original tone, even at high percentage of error. However some high frequency contents were found to be missing due to the smoothing effect of the interpolating polynomials.

As for the closeness between the reconstructed and resynthesised signals, it is agreed that even for large interpolation errors in the Analysis module, both signals are difficult to tell apart. However, the durations of the test signals varies from 0.3 to 0.8 seconds, and such subjective results therefore does not mean that the temporal evolution of the signals are unimportant in their identification. Fig. 2 shows a original tone against its resynthesised version.

Fig.3 shows the percentage of the total number of samples used in interpolation against frequency, for 5%, 15% and 25% of allowed interpolation error. The 15% line and 25% line have very similar shapes, due to the small reduction in interpolated points between these two values of interpolation error, roughly 5%. They have slopes roughly in the region of two parts in twelve, while the 5% error line has a slope of about eight parts in twelve. More significantly, the percentage of points used is very close for the 70Hz tone for all three error percentages, but diverges as frequency increases. For example, for the 70Hz tone the number of interpolated points used to achieve 25% (1094) error is reduced by 50% to the number needed to achieve 5% (2228) error, while for the 275Hz tone, a 75% reduction is achieved as the interpolation error is relaxed from 5% (2349) to 25% (657). This suggests that the data reduction that can be achieved is greater for high frequency tones. However, for 5% error, the percentage of total samples used for the 441Hz tone is roughly three and a half times that of the 70Hz tone.

Fig. 4 shows the MSE per interpolated point of the

reconstructed tones slope from Fig. 4. If we examine the MSE of a group of tones where the frequencies are doubled, we can see for the 70Hz group, 79Hz group and 92Hz group, the MSE increases with frequency in a closely linear trend for the different errors shown. The general trend is that at 5% error, the slope is essentially negligible for all three groups. The slopes for 15% error is about one part in three, while at 25% it is roughly four times greater. This suggests that the MSE performance is poorer at higher frequencies and as interpolation error is relaxed. For example, the MSE increased by about thirty times for the 70Hz signal, while at worst (368Hz signal), it increased by forty times as the interpolation error is changed from 5% to 25%. Also, the MSE values at 25% error for the 368 tone is four times that of the 70Hz tone.

Taking both Fig. 3 and Fig. 4 into account, we can conclude that for lower frequency tones, a greater value of interpolation error can be tolerated and employed since the absolute values and rate of increase of MSE is small as the interpolation error is relaxed. However at higher frequencies, there is a considerable tradeoff between the amount of data reduction and the MSE. Considering the fact that significant data reduction is achieved between 5% and 15% interpolation error for high frequency tones, while relatively small amount of reduction can be gleaned from 15% to 25% error, and of the great increase in MSE between these two values of interpolation error, it is concluded that for higher frequency signals, an interpolation error of about 15% should achieve optimal performance on both counts.

REFERENCES:-

- [1] J-C. Risset & M. Matthews. "Analysis of Musical Instrument Tones", *Physics Today*. 1969 vol.22 no.2 page 23-30.
- [2] J. Strawn. "Approximation and Syntactic Analysis of Amplitude and Frequency functions for Digital Sound Synthesis", *Computer Music Journal* vol.4 no.5 Fall 1980 page 671-692.
- [3] M-H Serra, D. Rubine & R. Dannenburg. "Analysis and Synthesis of Tones by Spectral Interpolation", *Journal of the Audio Engineering Society* vol.38 no.3 1990 March, page 111-127.
- [4] D. Kincaid & W. Cheney. "Numerical Analysis : Mathematics of Scientific Computing", Brook/Cole Publishing Company, 1991, Pacific Grove, California, USA.
- [5] R. Hung, N.H.C. Yung, P.Y.S. Cheung. "The Analysis and Resynthesis of Sustained Musical Signals in the Time Domain", *Proceedings of the International Computer Music Conference*, to be published.

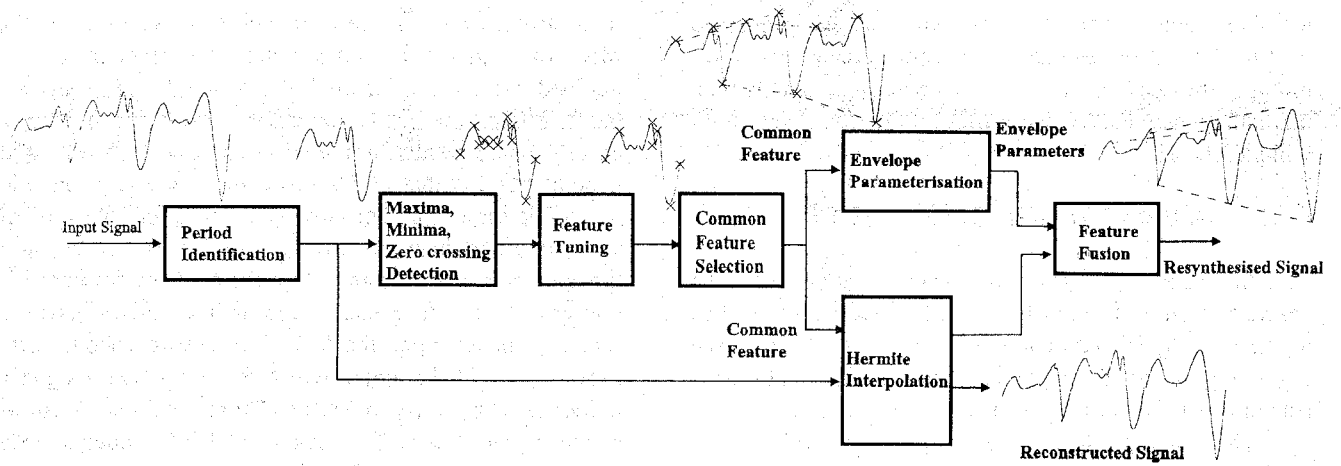


Fig. 1 Graphical Representation of Analysis/Resynthesis Algorithm

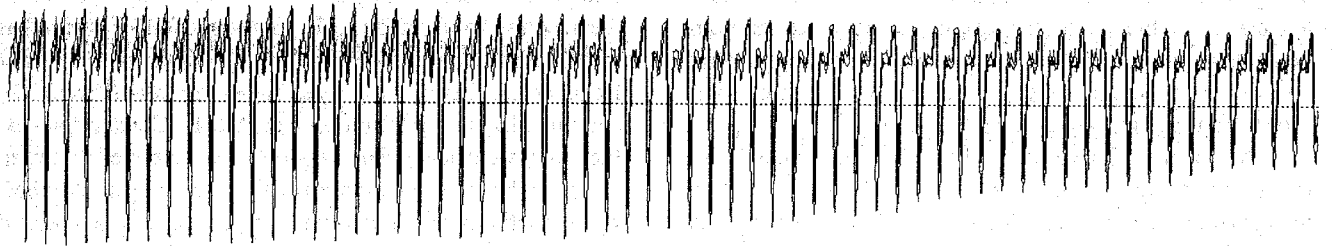


Fig. 2a The original 140Hz tone

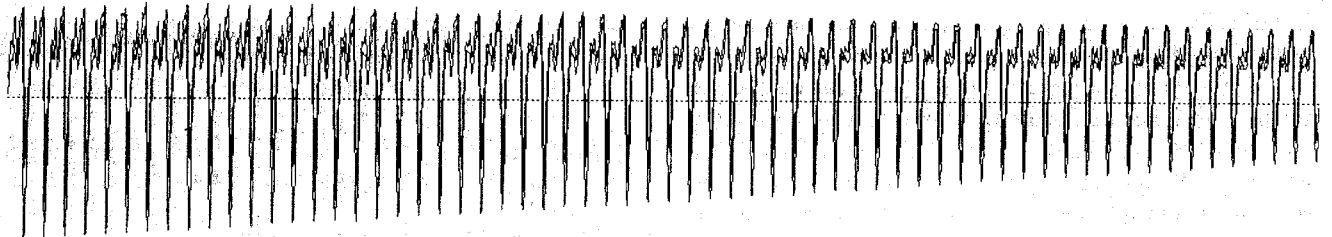


Fig. 2b Same tone resynthesised with 30% error in Analysis. Notice the change in overall envelope.

