International Journal of Scientific Engineering and Applied Science (IJSEAS) - Volume-1, Issue-9,December 2015
ISSN: 2395-3470
www.ijseas.com

Synthesizing musical notes of *Harmonium* using Spectral Domain Modeling

Samridha Kumar[1], Mihir Narayan Mohanty[2]
[1] Institution of Technical Education and Research, Bhubaneswar, India
[2] Institution of Technical Education and Research, Bhubaneswar, India

Abstract

Digital Sound Synthesis has been an exciting field of research over the last two to three decades. The idea is mainly to generate sound artificially using a system. A lot of work has been done in the past for synthesis of musical tones of stringed instruments. In this work, attempt is made for synthesizing *Harmonium* music using the techniques of Physical Modeling based on Digital Waveguides. In particular, the seven basic musical notes of the *Harmonium* are synthesized using which any *Harmonium* music can be generated. An analysis-synthesis approach is used for model-based good quality music synthesis. Time-frequency representations, particularly STFT are used for analysis of the natural music using which model parameters are estimated. Significant improvement is achieved in the quality of synthesis by extracting the excitation signal from the original music itself.

Keywords: *Physical Modelling, Music Synthesis, Digital Waveguide, STFT, Excitation signal*

1. Introduction

Synthesizing tones of physical musical instruments have been intriguing researchers for more over more than two decades. Basically, generating the music via computers has been claimed to be music synthesis. Many researchers have worked on synthesizing musical sounds based on stringed musical instruments **[1-9]**. In this paper the authors attempt to propose an approach to generate synthetic tones of the *Harmonium*. The approach is based on waveguide modeling which relies on modeling or replicating the source of the sound rather than the sound samples itself. A Graphical User Interface (G.U.I.) for a sample Virtual *Harmonium* is also designed.

Initial work on physical modeling of sound was done by Karplus and Strong for modeling string vibrations based on their physical behaviour **[1]**. The basic Karplus-Strong model is a recursive comb filter initialized with white noise.

The simplest digital waveguide model consists of a Delay line and a filter in feedback known as the loop filter **[1, 2]**. This originates from the discretization of the travelling wave solution of the wave equation **[3]**. The Delay line is for proper tuning of the output synthesized signal and the loop filter is there to account for the losses and dispersion of the string, which are lumped to one point by assuming linearity. Contrary to the original KS model, the loop filter in subsequent research works could match the frequency dependent damping of the physical string.

In **[4]**, the authors have made an attempt to synthesize guitar tones using a model-based approach. The model parameters are estimated from the sound of the acoustic instrument. An excitation signal extracted from the model parameters is fed to the synthesis model as input. In **[5]**, the authors have attempted to discuss the methodology for the digital sound synthesis of the Philippine *Rondalla* through physical modeling. Commuted Waveguide Synthesis has been used to implement the Physical Model of the various stringed instruments in a *Rondalla*. In **[6]**, Karjalainen *et. al* have continued their attempt to investigate Digital waveguide modeling of a nonlinear vibrating string, when the nonlinearity is essentially caused by tension modulation. The various physical modeling techniques of musical instruments were reviewed with main focus on Physical Modeling using Digital Waveguide technique in **[7-8]**. The techniques for synthesizing tones of stringed instruments are incorporated in **[9]** for synthesizing human speech.

In this article we emphasise on the the analysis-synthesis process to synthesise the musical notes of the wind instrument, *Harmonium*. The representations obtained from the analysis provide parameters corresponding to given synthesis models. Physical modeling and commuted waveguide synthesis are considered to develop a spectral domain

model for synthesis of the *Harmonium* music. Synthesis of the basic notes – *Sa*, *Re*, *Ga*, *Ma*, *Pa*, *Dha*, *Nee* and *Saa* and the composite *Sargam* is done. Using these key notes, any music can be produced. Finally, a G.U.I. is developed wherein the user can choose any note he desires to play.
The paper is organized as follows: The Pre-Recording of the musical notes to be synthesised and the corresponding Time-Frequency Analysis is explained in Section II. Section III describes the proposed method for waveguide modeling. Section IV explains the design of G.U.I. Section V discusses the results. Finally, Section VI concludes this piece of work.

2. Pre- Recording and Time-Frequency Analysis

The seven musical notes of the *Harmonium* and the combined *Sargam* were individually recorded.
In the time-frequency analysis, we have heavily relied on short-time processing techniques, mainly Short-Time Fourier Transform (STFT). This is essential in the case of non-stationary signals like EEG, speech signals and audio signals whose properties change markedly with respect to time. Hence a joint time-frequency representation is needed to describe the nature of non-stationary signals completely.
The STFT maps a signal into a two dimensional functions of time and frequency and provides some information about when and at what frequencies a signal event occurs **[10-12]**.

STFT of a signal x(n) as given in **[7]** is:

$$X_m(w) = \sum_{n=-\infty}^{\infty} x(n)w(n-mR)e^{-jwn} \quad (1)$$

where;
x(n) = input signal at time n
w(n) = length M window function
$X_m(w)$ = DTFT of windowed data centered about time mR.
R = hop size, in samples, between successive DTFT samples.

Using the above relation the STFT for each recording was calculated as follows.

1. The recorded signal was divided into overlapping frames using windowing technique. A Hamming window with 50% overlap was used for our purpose.
2. Each overlapping frame was fourier transformed (using FFT) and the complex result was added to a matrix, which records the magnitude and phase for each point in time and frequency.

The STFT frames were plotted as a function of frequency and peak-detection was done to measure the detectable harmonics and the fundamental frequency for each recording. For the recorded *Harmonium* notes, the fundamental frequency was 215 Hz for the lowest note *Sa* and gradually increased as the notes became higher reaching up to 425 Hz for *Saa*.
Further, the Energy Decay Relief using STFT has been computed for use in the Waveguide Modeling methodology. The Energy Decay Relief is a time-frequency distribution and can be represented as:

$$EDR(t_n, f_k) \cong \sum_{m=n}^{M} |H(m,k)|^2 \quad (2)$$

where *H(m,k)* denotes bin *k* of STFT at time frame *m*, and *M* denotes the total no. of time frames. $EDR(t_n,f_k)$ is the total amount of signal energy remaining at time $t_n=nT$ in a frequency band centered about $f_k=k*Fs/N$ where, *N* denotes the FFT length **[7].**

3. Proposed Method for Waveguide Modelling

In the proposed model, the time-frequency features have been used for synthesis. This is implemented through the loop filter design taking into consideration the delay line. The design of delay line and loop filter is explained as follows.

3.1 Delay line design

The length of the delay line controls the frequency of oscillation and consequently the pitch of the output signal. The length of the delay line is easily calculated as:

$$L = \frac{SamplingFrequency}{FundamentalFrequency} \qquad (3)$$

where;
L = no. of integer delays

i.e. the no. of delays equals the sampling rate divided by the fundamental frequency of the recorded music signal (or the pitch of desired output).
The problem in modelling the delay line arises when the no. of delays turns out to be a fraction. To fix this problem, the value at the fractional point was interpolated using a 3rd order lagrange's interpolator. This fractional delay was used in filter, cascaded with the delay line. Efficient techniques for fractional delay filtering were proposed in **[13].**
Lagrange's interpolation was chosen because the resulting fractional delay filter was a linear phase filter and hence, the output wasn't distorted.
This was followed by the design of the loop filter.

3.2 Loop Filter Design

It has been proposed to design a loop filter based on Energy Decay Relief (EDR). Similar methodology for design of loop filters for speech synthesis has been adopted in **[9]**.
The EDR values were calculated from the computed STFT values and then converted to dB scale. For each harmonic frequency, the EDR values were plotted as a function of time. The damping of each harmonic was measured by fitting a straight line to the decay and calculating the slope of the straight-line fits. Finally, slopes of decay of the EDR values (in dB) were used to compute the loop gain for different harmonic frequencies using the relation:

$$|H_L(z)| = 10^{(slopes*overlap)/(20*F_s)} \qquad (4)$$

where;
slopes = slopes of time decay of EDR values (in dB).
overlap = overlap time (in sec.) between adjoining STFT frames.
F_s = sampling frequency (in samples/sec.)

The above relation is derived from the recursion equation for successive EDR time-slices as given in **[7]** which is stated here as:

$$E_{m+1}(w_k) = |H_L(z)|^2 \times E_m(w_k) \qquad (5)$$

where;
$E_m(w_k)$ = EDR value at m^{th} time frame and k^{th} frequency.

Thus, the desired frequency response of the loop filter, sampled at the harmonic frequencies was prepared. The nearest partial peak-frequency was used to calculate the desired phase delay of the filter.
The desired frequency response over the entire frequency band till the Nyquist frequency, $F_s/2$ was calculated by designing a one-pole filter for the specification as suggested in **[14].** The magnitude response of the one-pole filter was used as a specification for the high frequencies.
To minimise the error in the desired frequency response and the approximated frequency response, following steps were followed.

1. The impulse response of the filter, $h(n)$ was evaluated and modeled as the output of the loop filter for unit sample input.
2. The sum of squares of error $e(n) = h(n) - \hat{h}(n)$ were minimized using signal modeling approach. Shank's method was utilized as explained in **[15]**.

A recursive IIR filter of first order was chosen for loop filter design. The higher order coefficients were not considered as they may not be useful for the generation of the signal; however the above methodology can be extended to higher order coefficients as well.

3.3 Extraction of Excitation signal

The excitation signals for the actual musical notes were generated from the recordings via. Inverse Filtering which involves putting the recorded signal through inverse waveguide filter, as shown in Fig.1

$$A(z) = 1 - H_L(z) * z^{-L} \qquad (6)$$

where $H_L(z)$ is the transfer function of loop filter and z^{-L} is for representation of delay element.

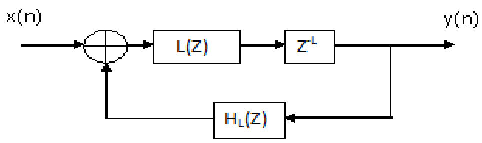

Fig. 1 The model for extraction of excitation signal via. Inverse filtering.

where, *L(Z)* is the lagrange's interpolating filter.
The excitation signal generated using this technique had a considerable improvement in sound quality over white noise or some other artificial excitation signal used as input for the synthesis model.

3. Design of Graphical User Interface (G.U.I)

A Graphical User Interface was designed for playing the *Harmonium* music. It consists of a pop-up menu consisting the list of various notes to play namely *Sa, Re, Ga, Ma, Pa, Dha, Nee and Saa*. The user can choose any note he desires to play. To play the selected note, a push button named 'PLAY' is given. A 2-D axis has also been included in the G.U.I. to show the plot of the synthesized note which is being played.

4. Results and Discussions

The results of our work are presented in this section. In particular, we present the results obtained during synthesis of the composite *Sargam.*

The delay line length was calculated to be:
L = 352 with a fractional delay of 0.8

The coefficients of loop filter were obtained as:
B = [0.8446 -0.3834]
A = [1.0000 -0.4540]
The transfer function of the filter hence can be written as:

$$H_L(z) = \frac{0.8446 - 0.3834z^{-1}}{1 - 0.4540z^{-1}} \qquad (7)$$

The plot of the original *Harmonium* music and results of its time-frequency analysis are shown in Fig. 2-4. The loop gain estimates at the measured harmonics and the approximated frequency response of the loop filter are shown in Fig 5-6. The excitation signal extracted is shown in Fig. 7. A comparison among the original music and synthesised music is shown in Fig. 8. Finally, Fig. 9 shows the G.U.I. designed.

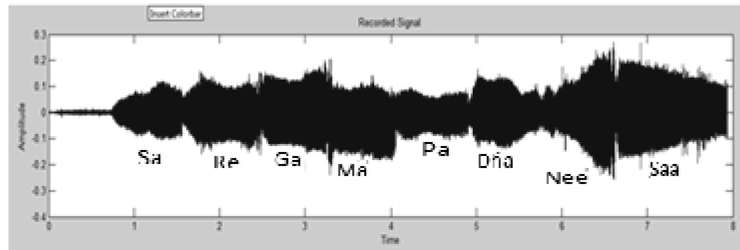

Fig. 2 The original *Harmonium* music – *Sargam*

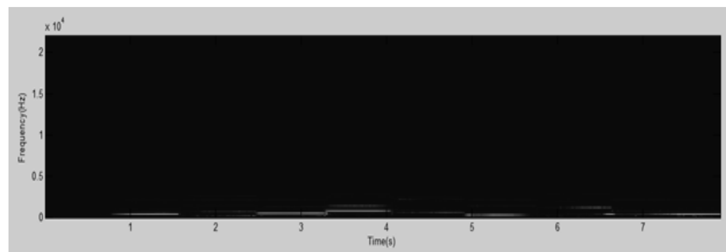

Fig. 4 Time- frequency representation of the recorded music.

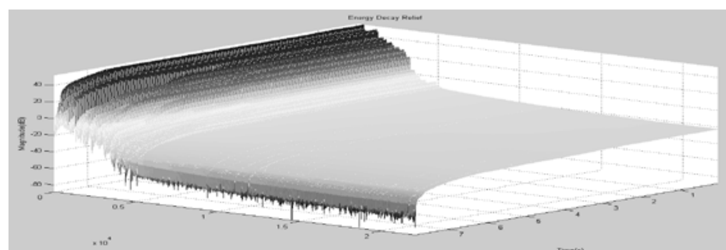

Fig 5. The EDR of the recorded music obtained using Time-Frequency features.

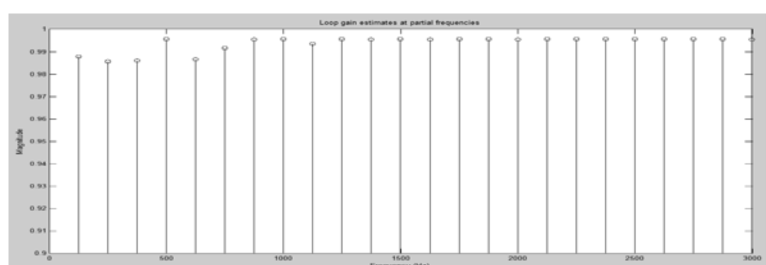

Fig. 6. Loop gain estimates at the detected harmonics.

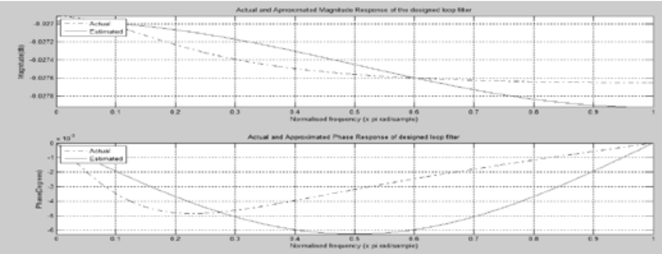

Fig. 7. The approximated magnitude and phase response of designed loop filter.

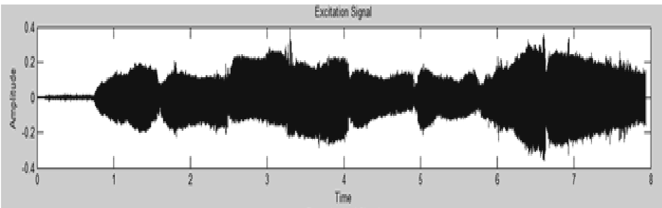

Fig. 8 Excitation signal extracted using Inverse Filtering.

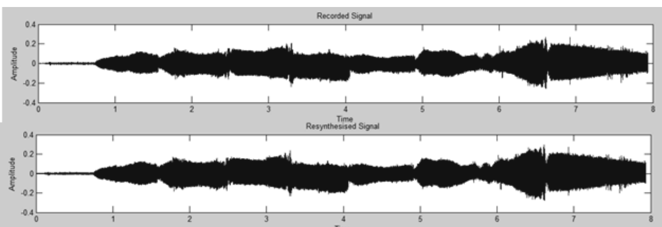

Fig 9. Comparison among original recorded music and synthesised music

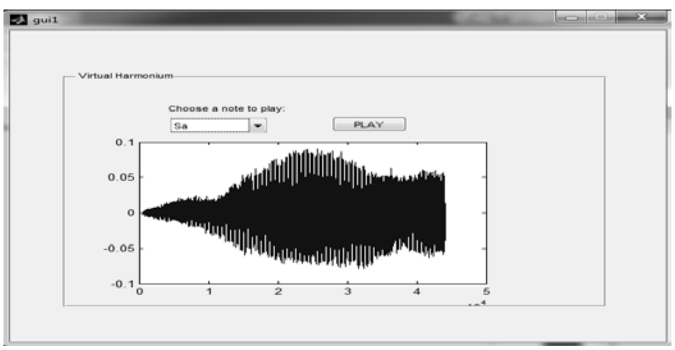

Fig. 10 Graphical User Interface for playing the synthesised music.

4. Conclusions

The proposed method of synthesis uses Digital Waveguide model Approach. This method results in much more realistic synthesis than the earlier methods. The synthesis may still be enhanced by a more carefully designed or a higher order loop filter. In future work, further improvements can be done using filter-bank. However, there is a huge correlation between the synthesised music and the natural music. Furthermore, the GUI designed can be integrated to a mobile application. This leads to developments in the design of Virtual Instruments.

References

[1] Kevin Karplus and Alex Strong. “Digital synthesis of plucked-string and drum timbres” *Computer Music Journal*, **7**(2): pp. 43-55, 1983.

[2] David Jaffe and Julius O. Smith. “Extensions of the Karplus-Strong plucked string algorithm”, *Computer Music J.*, **7**(2): pp. 56-69, 1983.

[3] J. O. Smith, “Physical modeling using digital waveguides,” *Comput. Music J.*, vol. 16, no. 4, pp. 74–91, Winter 1992.

[4] Matti Karjalainen, Vesa Valimaki, Zoltan Janosy, “Towards High-Quality Sound Synthesis of the Guitar and String Instruments” in *International Computer Music Conference, September 10-15, 1993, Tokyo, Japan*, 1993.

[5] Crisron Rudolf G. Lucas, Clariza Joy L. Soriano,” Digital Sound Synthesis of *Rondalla* using Physical Modeling Implemented on PC and iPad” in *IEEE 2013 TENCON-SPRING*,2013.

[6] Matti Karjalainen, Vesa Valimaki,Tero Tolonen, “Plucked-string synthesis algorithms with tension modulation nonlinearity” *in IEEE International Conference on Acoustics, Speech, and Signal Processing, Phoenix, Arizona, March* ,1999.

[7] Julius O. Smith III, “Physical Audio Signal Processing for Virtual Musical Instruments and Audio Effects”, *Centre for Computer Research in Music and Acoustics.*

[8] Matti Karjalainen, Vesa Valimaki, Tero Tolonen, “Plucked-String Models: From the Karplus-Strong Algorithm to Digital Waveguides and beyond.” in *Computer Music Journal, 22:3, pp. 17-32*, 1998.

[9] Samridha Kumar, Mihir N. Mohanty, “Waveguide Synthesis for Speech Signal using Spectral Domain Model”, *Proceedings of the Joint International Conference ICCCPT 2015 & ICAIECES 2015*,Chennai, pp. 59.

[10] L. Cohen, Time Frequency Analysis, Prentice-Hall, Englewood Cliffs, NJ,1995.

[11] Mihir N. Mohanty, Aurobinda Routray, P. Kabisatpathy, “Time frequency Kernel based classification of EEG signals using support vector

machines", IJEEE, vol.4, No.6, pp-130-138,2009, May-2009.

[12] Mihir Narayan Mohanty, AurobindaRoutray and Ashok Kumar Pradhan, PrithvirajKabisatpathy, ''Power quality disturbances classification using support vector machines with optimized time-frequency kernels'', Int. J. Power Electronics, vol. 4, no. 2, pp. 181-196, 2012.

[13] VesaV.lim.ki, Matti Karjalainen, and Timo I. Laakso, "Fractional delay digital filters", *Proc. IEEE Int. Symp. on Circuits and Systems*, pp. 355-358, Chicago, 1993.

[14] Balazs Bank, Vesa Valimaki, "Robust Loss Filter Design for Digital Waveguide Synthesis of string tones" in *IEEE SIGNAL PROCESSING LETTERS, VOL.10, NO.1, JANUARY 2003.*

[15] Monson H. Hayes, "Statistical Digital Signal Processing and Modelling"

Samridha Kumar completed his B.Tech. from Institute of Technical Education and Research, Bhubaneswar, India in 2015. He has worked in the area of Speech and Sound Analysis and Synthesis as part of his B.Tech Curriculum. He presented his research work "Waveguide Synthesis for Speech Signal using Spectral Domain Model" at the Joint International Conference, ICCCPT 2015 & ICAIECES 2015. His other research areas include Digital Signal Processing, Embedded Systems.

Mihir Narayan Mohanty received his Ph.D. degree from the Biju Patnaik University of Technology, Odisha, India. His Ph.D. research area was Signal Processing. Presently he is an Associate Professor at the Institute of Technical Education and Research, Bhubaneswar, India. He has more than 150 publications in reputed journals worldwide. His research interests include Antenna Design, Computational Intelligence, Signal Processing and Wireless Communications.