A New Recurrent-Network-Based Music Synthesis Method for Chinese Plucked-String Instruments – Pipa and Qin

Sheng-Fu Liang, Alvin W.Y. Su*, and Cheng-Teng Lin
Department of Electrical and Control Engineering, Chiau-Tung University, Hsin-Chu, Taiwan
*Department of Computer Science and Information Eng., Chung-Hwa University, Hsin-Chu, Taiwan
liang@falcon3.cn.nctu.edu.tw, alvin@chu.edu.tw, ctlin@fnn.cn.nctu.edu.tw

Abstract

A new recurrent-network-based synthesis technique for general plucked-string instruments is proposed. It is expected that the proposed technique can serve as a synthesis model as well as a systematic parameter-searching engine such that the synthetic tones can sound as close to the recorded tone as possible. It is also desired that the proposed techniques are low cost.

1. Introduction

A new neural network based approach is presented for synthesizing two different types of Chinese traditional plucked-string instruments, Pipa and Qin (Chin) [1]. The structure of this network is designed based on physical models of stringed instruments and a fast training algorithm is proposed in analysis stage. The basic idea of model-based approaches is to simulate the vibration-transmitting component of an acoustic musical instrument such as membrane of a drum, a string of a stringed instrument, and a bore of a wind instrument. The most famous technique is the so-called Digital Waveguide Filter (DWF) proposed in [2][3][4].

The major problem with model-based synthesis techniques is the determination of the synthesis model parameters. For DWF, this means the process of determining the coefficients of the output filter and feedback filter [2][5]. Usually, spectrum analysis of the original signal is necessary in order to design the filters. This can be time consuming and tedious. A Recurrent Neural Network (RNN) based synthesis model called the Scattering Recurrent Network (SRN) is proposed in [6][7]. The SRN successfully synthesizes plucked-string tones and the corresponding synthesis model parameters can be automatically determined. The supervised training which combines Back-Propagation-Through-Time (BPTT) training algorithm [8] and multi-position synchronous-sampling data is used. The training of SRN requires tremendous computation to converge and computation for the synthesis stage is too large for practical uses.

In this paper, a new recurrent-network-based music synthesis model for plucked-string instruments is proposed. There are some major improvements. First, the computation of the synthesis stage is greatly reduced. Second, the excitation signal is obtained in the training stage. Third, multiple sets of model parameters are used to adapt the changing characteristics of a played musical instrument. Fourth, a hybrid -training algorithm is used to increase the training speed and obtain the better synthesis parameters.

In section 2, two traditional Chinese plucked-string instruments, Pipa and Qin are introduced. In section 3, an analysis/synthesis neural network based model is proposed. In section 4, a multi-stage training procedure and an improved training algorithm for the determination of synthesis parameters and excitation signal is presented. In section 5, the analysis/synthesis results of two Chinese traditional plucked-string instruments, Pipa and Qin, are shown. Conclusion is given in section 6.

2. Pipa and Qin(Chin)

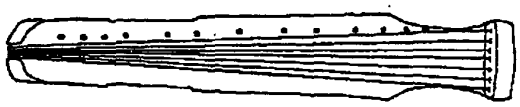
Pipa and Qin (Chin) are two typical Chinese plucked-string musical instruments. Although both belong to plucked-string instruments, the mechanical properties and the playing methods of them are absolutely different. Their pictures are shown in Fig.1.

Pipa contains a neck that serves as a handle and there are four pins on the top of the neck. Each pin holds a string that is stretched beyond its ovoid body. The string is made of solid steel wound with a layer of Nylon wire. The thin top plate has a pair of crescent-shaped sound holes. A thin bridge is installed at the bottom part of the top plate. Qin

0-7803-5529-6/99/$10.00 ©1999 IEEE

is the oldest Chinese plucked-string instrument that consists of a shallow rectangular wooden chamber and seven strings. Qin uses silk strings or silk wound steel strings and the strings are stretched parallel between the two ends of the chamber. The sound holes are on the bottom plate.

(a)



(b)

Fig.1. (a) Pipa, (b) Qin.

When Pipa is played, it is enfolded in arms. It can be played like a lute or can be hold upright to accommodate more playing techniques. The left hand presses on the frets and the fingers of right hands strum the strings above the top plate. When Qin is played, it is placed horizontally on a table. The right hand can pluck the string either inward or outward, while the left hand presses heavily to produce notes or lightly to produce overtones. The player is asked to use his/her fingers or fingernails to pluck, which produce very different timbres. In addition, since there are no fret in Qin, the finger of left hand can glide along a string to produce vibrato or portamento effects.

There are some interesting characteristics of these two instruments. For example, the energy corresponding to the fundamental frequency is usually much lower than other lower harmonics and the fundamental component usually gradually disappears. Some harmonics can disappear and then come back after some time. These may be due to the facts that the physical structures and the materials of the instruments are not good enough. Fig.2 and Fig.3 show the Short-Time-Fourier-Transforms (STFT) of one Pipa tone and one Qin tone, respectively. In the past, Pipa and Qin are quite difficult to model. [12]

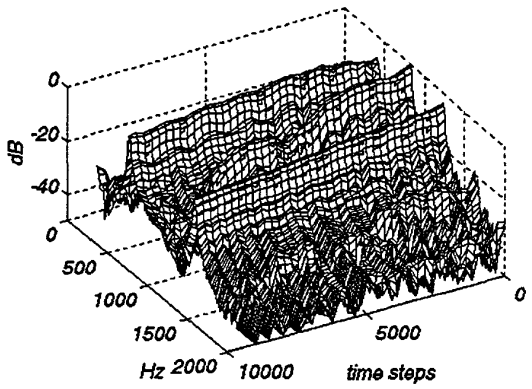


Fig.2. STFT analysis of Pipa tone.

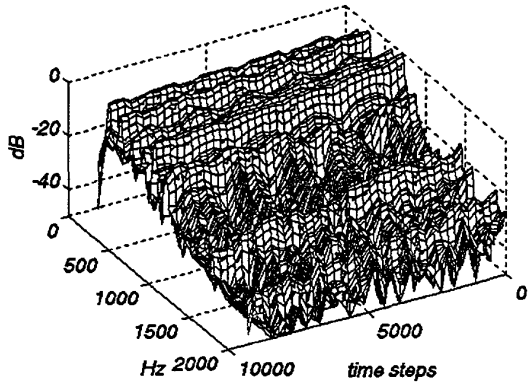


Fig.3. STFT analysis of Qin tone.

3. Recurrent-Network-Based Synthesis Model

When a finger or a pick plucks a string of a string instrument, the coupling vibration of the strings and the body starts. The basic idea of model-based synthesis technique is to simulate the dynamic behavior of a musical instrument. The major problem of model-based synthesis technique is the determination of the synthesis model parameters. Usually, it is necessary to examine the spectrum of the measured signal to design the model [5]. A model-based recurrent neural network synthesis model called Scattering Recurrent Network (SRN) [6][7] was proposed to solving the parameter determination problem.

This structure of the SRN network is constructed based on the physical model of a vibrating string and this network successfully synthesizes the dynamics of a plucked musical string. Electro-magnetic pick-ups are used to pick up

the vibrations of a plucked string at several positions. The measurement data are used as the training vector to obtain the model parameters by supervised training. In synthesis phase, simple triangular-like waveforms that simulate the *"plucks"* are used as the initial excitation.

Although SRN succeeds in synthesizing the signals of plucked strings, it cannot be employed directly as the synthesis model of a musical instrument. First, the dynamic behavior of a string instrument is the interaction among strings and its body. Therefore, the tone of a string instrument is much more complex than that of a vibrating string. Simple triangular excitation waveforms can not meet the synthesis requirement of string instrument. Second, if the training data is just the tone of the instrument recorded with a microphone, SRN fails since SRN requires a set of multiple measurements. [6]

In this paper, a new recurrent-network-based music synthesis model for general plucked-string instruments is proposed. It is no longer necessary to place multiple sensors over the target instrument to measure the vibrations of various spots. The training vector can be a single recorded tone from a microphone. Furthermore, the computation in the synthesis process is greatly reduced for practical use.

This proposed model shown in Fig.4 consisting of processing blocks (PB), simple delay lines and two reflection ends. This model is used to simulate the situation of general vibration-transmitting-reflecting mechanism of a string instrument. A pair of delay lines connects two adjacent PB's. The function of both left and right reflection ends is:

$$y = f(x) = -x \tag{1}$$

The structure of a PB is shown in Fig.4(b). The PB's simulate the situation that the traveling waves bounce back and forth at the positions of the corresponding PB's. This creates the effect of filtering. There are two types of model parameters to determine, one is the ρ-type parameters and the other is the w-type parameters. In our analysis/synthesis experiments for Pipa and Qin, 7 PB's are used and the structure of each is shown in Fig.4(b). Within each pair of delay lines, the signals pass through them directly without any modification, which is shown in Fig.4(c).

There are two stages in the synthesis phase by using the proposed model. The first stage is the *initialization* stage and the second one is the *propagation* stage.

- *Initialization* stage

In this stage, an initial waveform is provided for each delay unit or the neurons in each PB. Let the initial waveform is defined as follows:

$I^b_{i,j}$: the initial magnitude of the j-th delay unit of the i-th pair of the delay lines

$I^y_{i,j}$: the initial magnitude of the neuron $y_{i,j}$ in the i-th PB.

The operation of *initialization* stage is shown in Fig.5.

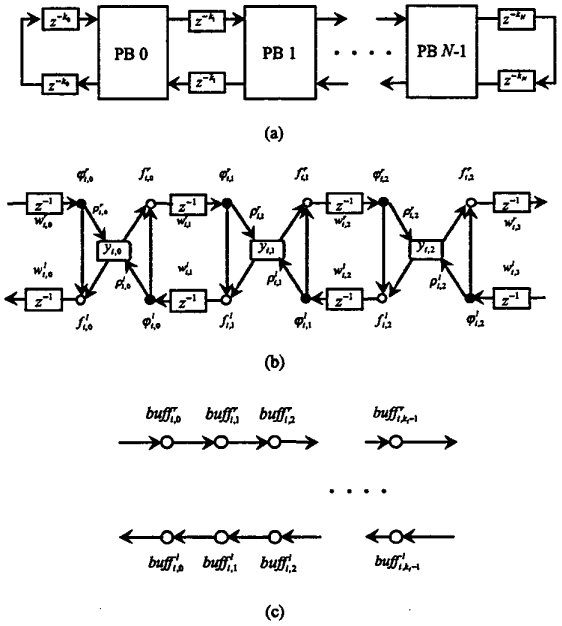


Fig.4. The proposed recurrent-network-based music synthesis model

$$\begin{aligned} &buff^r_{i,j}(0) = buff^l_{i,j}(0) = 0.5 \cdot I^b_{i,j} \\ &y_{i,j}(0) = I^y_{i,j}, \quad j = 0, 1, 2 \\ &f^r_{i,j}(0) = f^l_{i,j}(0) = 0.5 \cdot y_{i,j}(0) \end{aligned}$$

Fig.5. Initialization stage of the proposed synthesis model.

The size of the initial waveform, which equals to that of the synthesis model, depends on the fundamental frequency of the desired output signal and can be computed as

$$L = 0.5 \cdot \frac{f_s}{f} \tag{2}$$

where f is the desired fundamental frequency and f_s is the sampling rate of the synthesis system. The memory cost is much less than that of the wavetable method or the traditional model-based synthesis model, which requires the recorded tone as the excitation signal [5][13].

- *Propagation* stage

After the neurons are initialized in the *initialization* stage, this network starts the operation to generate the desired synthetic data automatically without any external signals. According to Fig.4, the dynamics of this synthesis model is described in Fig.6. The synthetic output is the magnitude of a chosen y-type neuron within one of the PB's. In this paper, the activation functions of the neurons are all identity function in order to reduce the computational cost and it is successful in synthesizing the tones of Pipa and Qin. Nonlinear functions can also be used though this may not be necessary. In this stage, multiple sets of synthesis parameters are used to imitate the changing characteristics of a plucked-string instrument.

$t = 0$

$i = 0, 1, 2, \ldots, N-1$

$a(\cdot)$: activation function

Repeat

for $i = 0, 1, 2, \ldots, N-1$

$$\varphi^r_{i,j}(t+1) = \begin{cases} a(w^r_{i,j} \cdot f^r_{i,j-1}(t)), & j = 1, 2 \\ a(w^r_{i,j} \cdot buff^{\ r}_{i,k_i-1}(t)), & j = 0 \end{cases}$$

$$\varphi^l_{i,j}(t+1) = \begin{cases} a(w^l_{i,j+1} \cdot f^l_{i,j+1}(t)), & j = 0, 1 \\ a(w^l_{i,j+1} \cdot buff^{\ l}_{i+1,0}(t)), & j = 2 \end{cases}$$

$$y_{i,j}(t+1) = a(\rho^r_{i,j} \cdot \varphi^r_{i,j}(t+1) + \rho^l_{i,j} \cdot \varphi^l_{i,j}(t+1)), \quad j = 0, 1, 2$$

$$f^r_{i,j}(t+1) = a(y_{i,j}(t+1) \cdot \varphi^l_{i,j}(t+1)), \quad j = 0, 1, 2$$

$$f^l_{i,j}(t+1) = a(y_{i,j}(t+1) \cdot \varphi^r_{i,j}(t+1)), \quad j = 0, 1, 2$$

$$buff^{\ r}_{i,j}(t+1) = \begin{cases} buff^{\ r}_{i,j-1}(t), & 0 \le i \le N, 1 \le j \le k_i - 1 \\ w^r_{i-1,3} \cdot f^r_{i-1,2}(t), & 1 \le i \le N, j = 0 \\ -buff^{\ l}_{0,0}(t), & i = 0, j = 0 \end{cases}$$

$$buff^{\ l}_{i,j}(t+1) = \begin{cases} buff^{\ l}_{i,j+1}(t), & 0 \le i \le N, 0 \le j \le k_i - 1 \\ w^l_{i,0} \cdot f^l_{i,0}(t), & 0 \le i \le N-1, j = k_i - 1 \\ -buff^{\ r}_{N,k_N-1}(t), & i = N, j = k_N - 1 \end{cases}$$

end for

$t = t + 1$

Untile (stops)

Fig.6. Propagation stage of the proposed music synthesis model.

4. Multi-Stage Training Procedure

A multi-stage training procedure for the proposed music synthesis model is shown in Fig.7. In the training phase, the initial pluck waveform and the synthesis parameters should be determined. There are two types of synthesis parameters, the ρ-type ones and the w-type ones. Since the string length, the string tension and other factors keep changing during the vibration of a plucked string instrument, this plucked instrument acts like a time-varying system. A synthesis model with time-varying model parameters is too costly to be practical. Therefore, multi-sets of synthesis parameters are necessary to simulate these dynamic characteristics.

In Stage #1 as shown in Fig.7, both initial pluck signal and the fist set of synthesis parameters have to be determined by using the recorded tone from $t_{1,0}$ to $t_{1,1}$ as the training vector. This set of parameters is used for synthesis processing from $t_{1,0}$ to $t_{2,0}$. It is no longer necessary to have the initial pluck signal updated after this stage. The Stage #2 training begins at $t_{2,0}$ by using the recorded tone from $t_{2,0}$ to $t_{2,1}$ as the training vector and the second set of parameters is obtained. This procedure continues until the recorded tone is finished. A preset threshold is used to determine whether the difference between the synthetic tone and the recorded tone is too large and a new stage is necessary.

For a recurrent neural network, Backpropagation-Through-Time (BPTT) [8] is a widely used training algorithm that is an extension of the standard backpropagation algorithm for training a feedforward network. It may be derived by unfolding the temporal operation of the network into a multi-layer feedforward network and the topology grows by one layer at every time step.[7][8] The parameters are adjusted based on the accumulated magnitude of corresponding gradient calculated from each time step. However, this algorithm requires tens of thousand epochs to converge for the training procedure of our synthesis model and the results are sometimes unsatisfactory.

In [9], the Simulated Annealing Resilient Back-Propagation (SAPROP) method is proposed for feedforward neural networks. This algorithm combines a quick gradient descent algorithm, resilient backpropagation (RPROP) [10], with the global search technique of simulated annealing (SA) [11]. The RPROP takes into account the sign of the gradient as seen by a particular parameter instead of the magnitude of the gradient. The SA involves the addition of random noise to the parameter updates as well as applies to decrease the weights of the updates in the training process gradually.

In this paper, a hybrid-training algorithm consisting of BPTT and SARPROP is used for the training procedure of the proposed music synthesis network. Since this synthesis network is a recurrent neural network, BPTT is used to

calculate the magnitude of gradient for each parameter to obtain the sign of the gradient, and the corresponding parameter update value is obtained by SARPROP. Particularly, this music synthesis network is constructed based on the physical model of a musical instrument; therefore, the values of synthesis parameters reflect the physical behavior of the instrument. The ρ-type parameters simulate the non-uniform characteristics at various physical positions of the mechanism and the w-type ones simulate the energy decay factors of a played instrument. The initial values of the synthesis parameters have to be assigned to reasonable values according to the physical characteristic of the target instrument instead of random values such that the training performance will be better and the training speed will be faster.

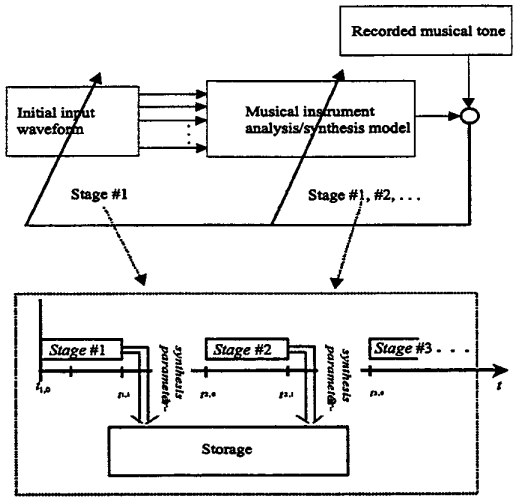


Fig.7. Multi-stage training procedure.

5. Experimental Results

In our experiments, the tones of Pipa and Qin are recorded by using an AKG-C414ULS microphone and a TASCAM DA-P1 DAT. The signals are transferred to a PC by using an EVENT Layla through its SPDIF interface. The followings are the synthesis experiments with respect to Pipa and Qin. Fig.8(a) shows the original tone of Pipa. The synthetic tone of the proposed model is shown in Fig. 8(b). The STFT of the original Pipa tone is shown in Fig. 2, and the STFT of the synthetic Pipa tone is shown in Fig. 9. The initial pluck waveform obtained form Stage #1 is shown in Fig. 10. Fig. 11(a) shows the original tone of Qin. The synthetic Qin tone of the proposed model is shown in Fig. 11(b). The STFT of the original Qin tone is shown in Fig. 3, and the STFT of the synthetic Qin tone is shown in Fig. 12. The initial pluck waveform obtained form Stage #1 is shown in Fig. 13.

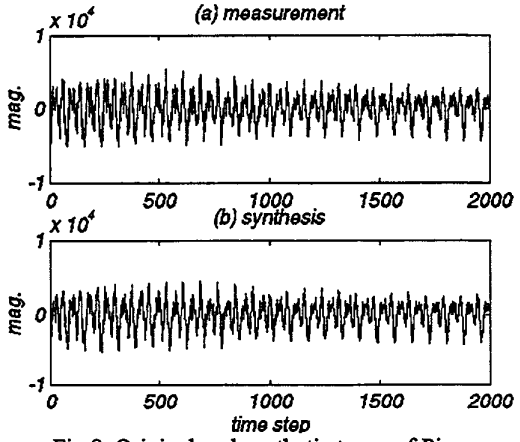


Fig.8. Original and synthetic tones of Pipa.

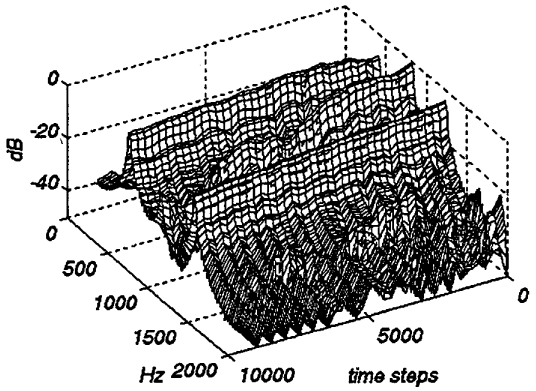


Fig.9. STFT of synthetic Pipa tone.

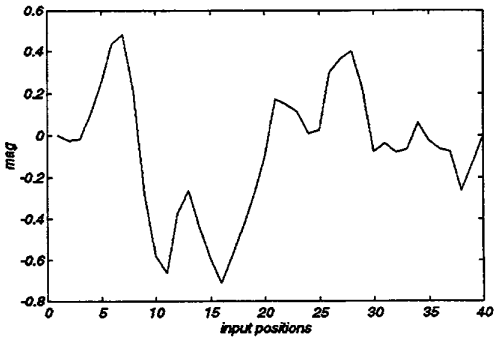


Fig.10. Initial pluck waveform for pipa.

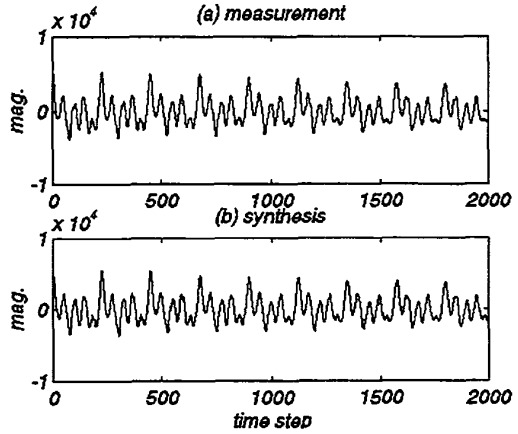


Fig.11. Original and synthetic tones of Qin.

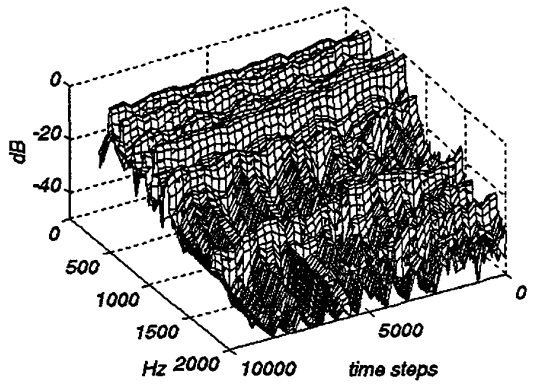


Fig.12. STFT of synthetic Qin tone.

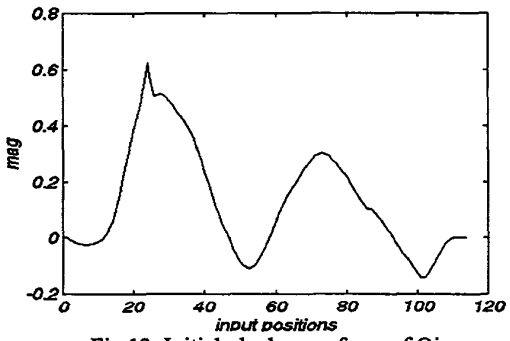


Fig.13. Initial pluck waveform of Qin.

6. Conclusion

A model-based music synthesis technique by using recurrent networks is proposed. The new approach provides not only an accurate synthesis of most plucked-string instruments but also an automatic way of searching the necessary synthesis parameters. The synthetic sounds possess the timbre of the target instrument to be synthesized, which is not possible in previous model-based approaches. What is needed is the recorded tone of the instrument to be used as the training vector. The research provides a new way of playing with electronic music. In addition, its synthesis processing is very computationally efficient and it also has a regular structure for low cost hardware implementation.

References

[1] H. D. Bodman, Chinese Musical Iconography: A History of Musical Instrument Depicted in Chinese Art, Asian-Pacific Cultural Center, Taipei, 1987.

[2] Smith, J. O., "Physical Modeling Using Digital Waveguides", Computer Music Journal, Vol. 16, No.4, pp. 74-87, 1992.

[3] Smith, J. O., "Efficient Synthesis of Stringed Musical Instruments", ICMC 1993.

[4] Smith, J. O., "Music Application of Digital Wavrguide" CCRMA Technical Report STAN-M-67. Stanford University.

[5] V. Valimaki, J. Huopaniemi, M. Karjalainen, Z. Janosy, "Physical Modeling of Plucked String Instruments with Application to Real-Time Sound Synthesis", J. Audio Eng. Soc., Vol. 44, No. 5, pp. 331-353, 1996.

[6] Alvin. W. Su and S. F. Liang, "Synthesis of Plucked-String Tones by Physical Modeling with Recurrent Neural Networks," in Proceedings of the IEEE. 1997 Workshop on Multimedia Signal Processing, (Princeton, NJ, 1997), pp. 71-76.

[7] Alvin W. Su, S. F. Liang, and C. T. Lin, "Model-Based Synthesis of Plucked String Instruments by Using a Class of Scattering Recurrent Networks," submitted to IEEE Trans. on Neural Networks, under revision, 1997.

[8] Ronald J. Williams, and Jing Peng, "An Efficient Gradient-Based Algorithm for On-Line Training of Recurrent Network Trajectories", Neural Computation, Vol. 2, p490-501, 1990.

[9] N. K. Treadgold and T. D. Gedeon, "Simulated Annealing and Weight Decay in Adaptive Learning: The SARPROP Algorithm", IEEE Trans. on Neural Networks, Vol. 9, No. 4, 1998.

[10] M. Riedmiller and H. Braun, " A direct adaptive method for faster backpropagation learning: The RPROP algorithm," in Proc. ICNN 93, San Francisco, CA, 1993, pp. 586-591.

[11] H. Szu, "Nonconvex optimization by fast simulated annealing," in Proc. IEEE, 1987,vol. 75, pp. 1538-1540.

[12] N.M. Cheung and A. Horner, "Group Synthesis with Genetic Algorithm," J. Audio Eng. Soc., Vol. 44, No. 3, pp. 130-147, 1996.

[13] A. Horner, J. Beauchamp, and L. Haken, "Methods for Multiple Wavetable Synthesis of Musical Instrument Tones," J. Audio Eng. Soc., Vol. 41, pp. 336-356, 1993.