# Model-Based Synthesis of Plucked String Instruments by Using a Class of Scattering Recurrent Networks

Sheng-Fu Liang, Alvin W. Y. Su, *Member, IEEE*, and Chin-Teng Lin, *Senior Member, IEEE*

***Abstract*—A physical modeling method for electronic music synthesis of plucked-string tones by using recurrent networks is proposed. A scattering recurrent network (SRN) which is used to analyze string dynamics is built based on the physics of acoustic strings. The measured vibration of a plucked string is employed as the training data for the supervised learning of the SRN. After the network is well trained, it can be regarded as the virtual model for the measured string and used to generate tones which can be very close to those generated by its acoustic counterpart. The "virtual string" corresponding to the SRN can respond to different "plucks" just like a real string, which is impossible using traditional synthesis techniques such as frequency modulation and wavetable. The simulation of modeling a cello "A"-string demonstrates some encouraging results of the new music synthesis technique. Some aspects of modeling and synthesis procedures are also discussed.**

***Index Terms*—Physical modeling, plucked string instruments, scattering recurrent networks.**

## I. Introduction

ATTEMPTS to analyze and model the dynamics of musical instruments have always been efforts of instrument makers. The most renowned example is research with respect to bowed-string instruments such as violins and cellos done by the master maker of the 18th century, Antonio Stradivari. The research covers topics such as the analysis of the varnish, the physical dimensions, the material used by Stradivari and the responses of his instruments under various excitations [1]–[3]. The purpose of the research is to provide guidance for modern makers such that they can make their instruments sound as good as the master's.

With the introduction of electronic music, many techniques for generating musical tones have been proposed such as *frequency modulation* (FM) synthesis [8] and *wavetable* synthesis [6], [7] which are the two most popular methods used nowadays. The ultimate goal of electronic synthesis is to provide musical tones that sound exactly the same as those generated by their acoustic counterparts. However, the sound quality cannot meet the requirements of the most demanding users, especially the reproduction of the musical dynamics of most instruments. In [9], Karplus and Strong proposed a plucked-string algorithm that used delay lines and simple digital filters. This is a low-cost approach and it successfully generated realistically dynamic performance not seen in other synthesis techniques. However, it is very difficult for the *K–S* algorithm to control the timbre of the sounds simply by modifying the filters such that it can sound like any particular plucked-string instrument. This technique is nevertheless the first step toward the physical modeling approaches of instrumental dynamics for music synthesis.

In order to have the synthesis result closer to the sound generated by acoustic instruments, Smith proposed the so-called *digital waveguide filter* (DWF) technique [12]. This synthesis algorithm starts from the wave equation for musical strings and implements the solution to the equation on a discrete-time system. In his later efforts in the physical modeling synthesis, algorithms for simulating the sounding mechanisms such as reeds for clarinets and bows for violins were studied [10], [11], [13]. These techniques make some very realistic sounds and have become more and more popular in music synthesis related research. These computer-synthesized instruments are sometimes called "virtual instruments." However, it is difficult to find the appropriate model parameters of the DWF's such that the synthetic sounds can be associated with any particular instrument. In our experience with physical modeling techniques, it is found that the problem of obtaining synthesis parameters based on the analysis of the musical instruments themselves have never been addressed.

In order to overcome the difficulty of parameter determination for the physical modeling methods, it is necessary to study how acoustic instruments respond to given excitations. We propose a class of *scattering recurrent networks* (SRN) based on the physical dynamics of musical instruments to model as well as synthesize the vibrations of musical strings. Being a universal approximator [23], the artificial neural network has been widely used in many applications such as pattern recognition, time series analysis, system identification, and so on. If the responses at various positions of an acoustic musical instrument can be measured, it may be possible that a neural network can be trained such that it can reproduce the responses of the same instrument under the identical excitations. The structure of the SRN is first related to the physics as well as the physical shape of the instrument to be modeled. A technique for training the SRN specially designed for the given instrument is developed.

In order to simplify our first attempt, we start from the modeling of a plucked string. A string can be approximately regarded as a one-dimensional instrument. This allows a simpler implementation of the SRN and requires less computation to train the network. The training data are obtained by measuring the vibration of a cello "A"-string excited by a single

Manuscript received July 1, 1997; revised February 2, 1999 and August 23, 1999. This work was supported in part by the National Science Council, Taiwan, R.O.C. under Contract NSC 86-2213-E-216-018.

S.-F. Liang and C.-T. Lin are with the Department of Electrical and Control Engineering, National Chiao-Tung University, Hsin-Chu, Taiwan, R.O.C.

A. W. Y. Su is with the Department of CSIE, Chung-Hwa University, Hsin-Chu, Taiwan, R.O.C.

Publisher Item Identifier S 1045-9227(00)00900-0.

pluck. In order to obtain these measurements, we constructed a steel-string measurement system which consists of seven Dimarzio® Virtual Vintage electromagnetic pickups placed in parallel and equally spaced under the string to measure the vibrations of the chosen cello string. The measured signals from the pickups are sampled and stored in a multitrack digital audio recorder. The *backpropagation-through-time* (BPTT) technique is used for the training of the SRN [19], [21]. In our experiments, the waveforms of the resynthesized outputs of the well-trained SRN are very close to the measurement taken from the plucked string. Though the computation of the training phase is very large, the synthesis processing requires much less computation. Although the proposed synthesis technique needs much more computation than conventional synthesis methods, the sound quality is superior.

In Section II, the dynamics of plucked musical strings is discussed and the scattering junction model for vibrating strings is derived. In Section III, the SRN is proposed. The training of the SRN is also discussed. In Section IV, the construction of a steel-string measurement system to obtain the training data for the SRN is described. In Section V, experiments of the training and the resynthesis of the SRN are presented. Conclusions and future work are given in Section VI.

## II. Dynamics of Musical Strings

Physical-modeling methods for musical instruments have become ubiquitous for their ability to produce realistic and dynamic synthesis sounds without heavy computation compared to the wavetable synthesis method. The basic idea of physical-modeling synthesis methods is to simulate the vibrations of acoustic instruments as closely as possible. In this paper, we propose a class of SRN's that simulate the vibrations of a plucked musical string. The measured vibrations of the string are used as the training data for the SRN which is configured according to the physical properties of an acoustic string. The synthesis processing using the well-trained SRN is discussed in later sections. In order to understand how to relate the physics of a plucked string to the proposed SRN, the physics of an ideal acoustic string is presented first.

### A. The Ideal Vibrating String

The wave equation for an ideal vibrating string was derived by Morse [15]. "Ideal" means lossless, linear, uniform, volumeless, and flexible. Consider the uniform string with linear mass density $\varepsilon$ (kg/m) stretched to a tension $K$ (newtons). A small segment of the string is shown in Fig. 1. Let the restoring force for the segment, $ds$, to its equilibrium position be $dF_y$ which is the difference between the forces with respect to the two ends of the segment in the vertical direction as follows:

$$dF_y = (K \sin\theta)_{x+dx} - (K \sin\theta)_x. \tag{1}$$

By using Taylor's series expansion shown below, we have

$$f(x+dx) = f(x) + \frac{\partial f(x)}{\partial x}\,dx + \frac{1}{2!}\frac{\partial^2 f(x)}{\partial x^2}\,dx + \cdots. \tag{2}$$

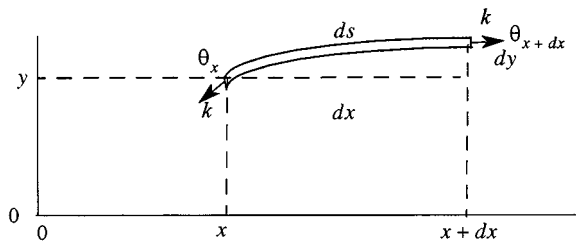



Fig. 1. A segment of an ideal string with tension $K$.

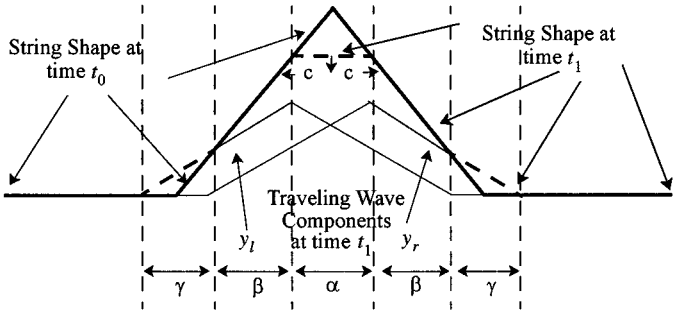



Fig. 2. An infinitely long plucked string simulation.

By applying (2) to (1) and retaining the first two terms, we have

$$dF_y = \left[(K \sin\theta)_x + \frac{\partial(K \sin\theta)}{\partial x}\,dx\right] - (K \sin\theta)_x = \frac{\partial(K \sin\theta)}{\partial x}\,dx. \tag{3}$$

The $\sin\theta$ in (3) can be replaced by $\tan\theta$ if $\theta$ is small. Equation (3) becomes

$$dF_y = \frac{\partial(K\partial y/\partial x)}{\partial x}\,dx = K\frac{\partial^2 y}{\partial x^2}\,dx. \tag{4}$$

Let the mass of the segment $ds$ be $\varepsilon\,ds$. By using Newton's second law of motion, we have

$$K\frac{\partial^2 y}{\partial x^2}\,dx = (\varepsilon\,ds)\frac{\partial^2 y}{\partial t^2}. \tag{5}$$

Since $dy$ is a small quantity for small $\theta$, $ds$ is approximately equal to $dx$. Equation (5) is further reduced to

$$K\frac{\partial^2 y}{\partial x^2} = \varepsilon\frac{\partial^2 y}{\partial t^2}. \tag{6}$$

This is the wave equation for an ideal vibrating string. The general solution of (6) can be written as

$$y(t, x) = y_r(t - x/c) + y_l(t + x/c) \tag{7}$$

where

- $y_r(t - x/c)$ right-going traveling wave with a traveling velocity $c$;
- $y_l(t + x/c)$ left-going traveling wave with the identical velocity.

The transverse wave velocity $c$ is equal to $\sqrt{K/\varepsilon}$.

An ideal plucked string with infinite length is shown in Fig. 2. The initial displacement representing the pluck in this context is

modeled as the sum of two triangular pulses which overlap with each other at time $t_0$. At time $t_1$ shortly after time $t_0$, the two triangular pulses separate and travel in the left direction and the right direction, respectively. The segment $\alpha$ which represents the peak portion becomes flat and gradually reduces to zero. Within the regions denoted by $\beta$, where the two traveling waves overlap, the string displacements remain the same magnitudes as that in the initial condition in the same regions. The two short pieces denoted by $\gamma$ on the left-hand side and the right-hand side are the leading edges of the left-going and the right-going traveling waves. When the traveling waves fully separate, the string will be at rest except for two half-sized triangular pulses heading off to the left and to the right with speed $c$.

### *B. The Lossy Vibrating String*

If an ideal string with its ends fixed is plucked, it will vibrate circularly and eternally. As a matter of fact, a string cannot vibrate without any energy loss. Therefore, it is necessary to consider loss factors. In most situations, energy loss is caused by friction of the surrounding air, yielding terminations, and internal friction [26], [30]. Considering the simplest case where the resistive force is linearly proportional to the transverse velocity, we can obtain the modified wave equation as follows:

$$K\frac{\partial^2}{\partial x^2}y(t,\,x) = u\frac{\partial}{\partial t}y(t,\,x) + \varepsilon\frac{\partial^2}{\partial t^2}y(t,\,x). \tag{8}$$

The solution to the above equation can be easily obtained as

$$y(t,\,x) = e^{-(u/2\varepsilon)(x/c)}y_r(t-x/c) + e^{(u/2\varepsilon)(x/c)}y_l(t+x/c). \tag{9}$$

According to Shannon's sampling theorem [16], the traveling wave can be fully expressed by a discrete-time system as long as the sampling interval is small enough. In order to simulate the traveling waves of a plucked string, sampling is performed along the longitudinal direction. The magnitude of a vibrating string at a sampled position is sampled with the sampling period equal to $T$ (s). Let the sampling interval along the string be $\Delta x$ and the transverse wave velocity of the string be $c$. Then, $\Delta x$ is equal to $T \cdot c$. For example, if the sampling frequency is 44.1 KHz and the velocity is 1000 m/s, the sampling interval is 100/44 100 = 0.0227 m. By replacing $x$ with $x_m$ and $t$ with $t_n$ in (9), we have

$$\begin{aligned} y(t_n,\,x_m) &= e^{-(u/2\varepsilon)(x_m/c)}y_r(t_n - x_m/c) \\ &\quad + e^{(u/2\varepsilon)(x_m/c)}y_l(t_n + x_m/c) \\ &= e^{-(u/2\varepsilon)mT}y_r[(n-m)T] \\ &\quad + e^{(u/2\varepsilon)mT}y_l[(n+m)T] \\ &= \varphi_r(tl_n,\,x_m) + \varphi_l(t_n,\,x_m) \end{aligned} \tag{10}$$

where

$t_n \quad = n \cdot T;$
$x_m \quad = m \cdot \Delta x = m \cdot c \cdot T.$

The discrete-time signal representation of (10) is given by

$$\begin{aligned} y(t_n,\,x_m) &= \varphi_r(t_n,\,x_m) + \varphi_l(t_n,\,x_m) \\ &= g^m f_r(n-m) + g^{-m} f_l(n+m) \end{aligned} \tag{11}$$

where

$g \quad \equiv e^{-uT/2\varepsilon};$
$f_r(n-m) \quad \equiv y_r((n-m)T);$
$f_l(n+m) \quad \equiv y_l((n+m)T).$

In practical musical applications, a string is usually fixed at its two ends. If the length of a fixed string is $L$, the boundary condition can be described by using the following two equations:

$$y(0,\,t_n) = 0 = \varphi_r(t_n,\,0) + \varphi_l(t_n,\,0), \qquad \text{for all } t_n \tag{12}$$

$$y(L,\,t_n) = 0 = \varphi_r(t_n,\,L) + \varphi_l(t_n,\,L), \qquad \text{for all } t_n. \tag{13}$$

### *C. The Scattering Junction for the Vibrating String*

In our practical experiments, the traveling waves in an acoustic string cannot be modeled completely by the method shown in the previous section since most strings do not satisfy the uniform-impedance constraint [26], [30]. Therefore, we involve the concept of *scattering junctions* in the situation that a traveling wave may reflect as well as pass through a position if the respective acoustic impedances from the two sides of the position are not identical. The behavior of traveling waves incident with the scattering junction has been proposed [11], [12], [30].

In (7), $y(t,\,x)$ is used to represent the displacement of a vibrating string at position $x$, and at time instant $t$. Let the *transverse velocity wave* be $\dot{y}(t,\,x)$ which is the first time derivative of $y(t,\,x)$, and we have

$$\dot{y}(t,\,x) = \frac{\partial}{\partial t}y(t,\,x) = \dot{y}_r(t-x/c) + \dot{y}_l(t+x/c). \tag{14}$$

By replacing $\dot{y}_r(t-x/c)$ and $\dot{y}_l(t+x/c)$ with $v_r(t,\,x)$ and $v_l(t,\,x)$, respectively, (14) can be rewritten as

$$\dot{y}(t,\,x) = v_r(t,\,x) + v_l(t,\,x) \tag{15}$$

where $v_r$ and $v_l$ are the transverse velocity waves.

Let the *slope wave* be $y'(t,\,x)$ which is the first spatial derivative of $y(t,\,x)$, we have

$$\begin{aligned} y'(t,\,x) &= \frac{\partial}{\partial x}\,y(t,\,x) = y'_r(t-x/c) + y'_l(t+x/c) \\ &= -\frac{1}{c}\,\dot{y}_r(t-x/c) + \frac{1}{c}\,\dot{y}_l(t+x/c) \\ &= -\frac{1}{c}v_r(t,\,x) + \frac{1}{c}\,v_l(t,\,x). \end{aligned} \tag{16}$$

Similarly, $y'_r(t-x/c)$ and $y'_l(t+x/c)$ can be replaced with $u_r(t,\,x)$ and $u_l(t,\,x)$. Therefore, the left-going traveling slope wave can be computed by dividing the left-going traveling velocity wave with the wave velocity $c$, and the right-going slope wave can be computed by dividing the negative right-going traveling velocity wave with the wave velocity $c$ as follows:

$$\begin{cases} u_l(t,\,x) = \dfrac{1}{c}v_l(t,\,x) \\ u_r(t,\,x) = -\dfrac{1}{c}v_r(t,\,x). \end{cases} \tag{17}$$

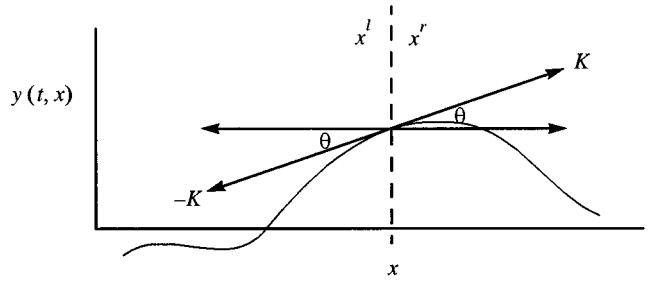



Fig. 3. The transverse force propagation in an ideal string.

In summary, traveling waves in any form can be computed from each other, as long as the left-going and right-going components are available. For example, the transverse velocity wave can be computed by differentiating the displacement wave and the displacement wave can also be computed by integrating the transverse velocity wave if the initial velocity waves are nulls.

To discuss the behavior of traveling waves flowing into a position where two segments of different acoustic characteristics are connected, the force wave propagation property is derived. This is shown in Fig. 3. At any arbitrary position $x$ of a string, the vertical force applied to the left-hand side of the position, denoted by $x^l$, is given by

$$F_l(t, x) = K \sin(\theta) \approx K \tan(\theta) = ky'(t, x) \tag{18}$$

where $K$ is the tension at position $x$. Let $|y'(t, x)| \ll 1$. Similarly, the force applied to the right of the position, denoted by $x^r$, is given by

$$F_r(t, x) = -K \sin(\theta) \approx -Ky'(t, x). \tag{19}$$

These two forces must cancel in order not to produce infinite acceleration to a massless point. Either $F_l$ or $F_r$ can be used to represent the string force wave at the position. Let the vertical force applied to the right of any position $x$, denoted by $F_r$, be the force wave $F$, and we have

$$F(t, x) = F_r(t, x) = -Ky'(t, x). \tag{20}$$

By carrying (16) into the right-hand side of (20) and using (15), we have

$$\begin{aligned} F(t, x) &= \frac{K}{c}\left[\dot{y}_r(t - x/c) - \dot{y}_l(t + x/c)\right] \\ &= \frac{K}{c}\left[v_r(t, x) - v_l(t, x)\right]. \end{aligned} \tag{21}$$

The characteristic impedance is defined as

$$Z \equiv \sqrt{K\varepsilon}. \tag{22}$$

Let the right-going force wave and the left-going force wave be $(K/c)\cdot\dot{y}_r(t-x/c)$ and $(-K/c)\cdot\dot{y}_l(t+x/c)$ which are replaced with $\Phi_r(t, x)$ and $\Phi_l(t, x)$, respectively, as follows:

$$\Phi_r(t, x) = Zv_r(t, x) \tag{23}$$

and

$$\Phi_l(t, x) = -Zv_l(t, x). \tag{24}$$

According to (22)–(24), (21) is rewritten as

$$F(t, x) = \Phi_r(t, x) + \Phi_l(t, x). \tag{25}$$

The physical meaning of (25) is that the vertical force can be computed by summing the right-going and left-going force waves. It is assumed that there is a junction on a vibrating string where the characteristic impedances on the two sides of this junction are different. Let $Z_1$ represent the characteristic impedance in the left-hand-side segment and $\Phi_r^1$ represent the force wave flowing into the junction from the left-hand-side direction. Let $Z_2$ represent the characteristic impedance in the right-hand-side segment and $\Phi_l^2$ represent the force wave flowing into the junction from the right-hand-side direction. Physically, the force wave cannot change instantaneously across the junction and the sum of velocity waves meeting at the junction is zero [12]. According to Kirchoff's node equations [18], there can be only one resultant force wave at the junction which is denoted by $F^J$ and the sum of velocity waves meeting at the junction must be zero if the junction is lossless. Therefore, we have

$$F^1 = F^2 = F^J \tag{26}$$

and

$$v^1 + v^2 = 0. \tag{27}$$

By (15) and (23)–(25), we have

$$\begin{cases} F^i = \Phi_r^i + \Phi_l^i \\ v^i = v_r^i + v_l^i \end{cases} \quad \text{for } i = 1, 2 \tag{28}$$

and

$$\begin{cases} \Phi_r^i = Z_i \cdot v_r^i \\ \Phi_l^i = -Z_i \cdot v_l^i \end{cases} \quad \text{for } i = 1, 2 \tag{29}$$

where

$F^1$ resultant force wave of the left-hand-side segment just beside the junction;

$F^2$ resultant force wave of the right-hand-side segment just beside the same junction.

By defining the characteristic admittance of segment $i$ to be $\Gamma_i = Z_i^{-1}$ for $i = 1, 2$, and solving (26)–(29), the resultant junction force wave is obtained as

$$F^J = 2 \cdot \frac{\Gamma_1\Phi_r^1 + \Gamma_2\Phi_l^2}{\Gamma_1 + \Gamma_2}. \tag{30}$$

Define the reflection coefficient as

$$\rho = \frac{Z_2 - Z_1}{Z_1 + Z_2}. \tag{31}$$

The outgoing force waves can be obtained by the following equations:

$$\Phi_l^1 = F^J - \Phi_r^1 = \rho\Phi_r^1 + (1 - \rho)\Phi_l^2 \tag{32}$$

$$\Phi_r^2 = F^J - \Phi_l^2 = (1 + \rho)\Phi_r^1 - \rho\Phi_l^2. \tag{33}$$

Since the wave equations shown in (7) are derived for the displacement waves instead of the force waves, it is necessary to modify (30), (32), and (33) such that the displacement representation of a vibrating nonuniform string used can be applied. By carrying (29) into (32), we have

$$-Z_1 v_l^1 = \rho Z_1 v_r^1 + (1-\rho)\left(-Z_2 v_l^2\right). \tag{34}$$

By dividing (34) by $-Z_1$, and using (31), we have

$$\begin{aligned} v_l^1 &= (-\rho) v_r^1 + (1-\rho)\frac{Z_2}{Z_1} v_l^2 \\ &= (-\rho) v_r^1 + \left(\frac{2Z_1}{Z_1+Z_2}\cdot\frac{Z_2}{Z_1}\right) v_l^2 \\ &= (-\rho) v_r^1 + \frac{2Z_2}{Z_1+Z_2} v_l^2 \\ &= (-\rho) v_r^1 + (1+\rho) v_l^2. \end{aligned} \tag{35}$$

Similarly, (33) can be rewritten as

$$v_r^2 = (1-\rho) v_r^1 + \rho v_l^2. \tag{36}$$

The displacement wave representation of the string can be obtained by integrating (35) and (36) with respect to time, because the velocity wave is the first-order time derivative of the string displacement according to (14). By integrating (35), we have

$$\int_0^t \dot{y}_l^1(t, x) = \int_0^t (1+\rho)\dot{y}_l^2(t, x) - \int_0^t \rho \dot{y}_r^1(t, x). \tag{37}$$

By solving (37), we have

$$\begin{aligned} y_l^1(t, x) - y_l^1(0, x) = (1+\rho)\left[y_l^2(t, x) - y_l^2(0, x)\right] \\ - \rho\left[y_r^1(t, x) - y_r^1(0, x)\right]. \end{aligned} \tag{38}$$

The initial magnitude of the right-going wave of a plucked string is the same as that of the left-going wave and their sum is equal to the magnitude of the initial displacement at any position of the string [11]. To be specific, we have

$$y_r^i(0, x) = y_l^i(0, x) = \tfrac{1}{2} y^i(0, x), \qquad \text{for } i = 1, 2. \tag{39}$$

Moreover, the string displacement must be a continuous function of position. The displacement at the left-hand side of a position is equal to the displacement at the right-hand side of this position, i.e.,

$$y^J(t, x) = y^1(t, x) = y^2(t, x), \qquad \text{for all } t \tag{40}$$

where $y^J(t, x)$ is the initial displacement. Substituting (39) and (40) into (38), we have

$$y_l^1(t, x) = -\rho y_r^1(t, x) + (1+\rho) y_l^2(t, x). \tag{41}$$

Similarly, (33) can be converted into

$$y_r^2(t, x) = (1-\rho) y_r^1(t, x) + \rho y_l^2(t, x). \tag{42}$$

By combining (40)—(42), it is easy to show that

$$y^J = y_r^1 + y_l^1 = y_r^2 + y_l^2 = (1-\rho) y_r^1 + (l+\rho) y_l^2 \tag{43}$$

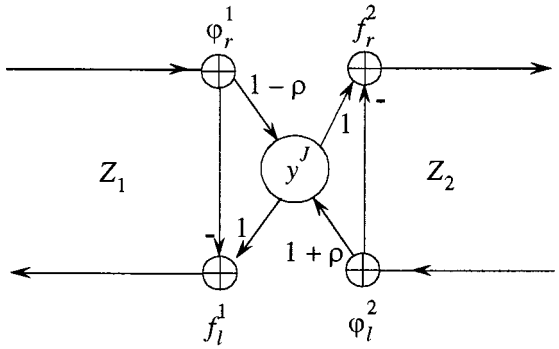



Fig. 4. An alternative model for the scattering junction within a string.

where $(t, x)$ is omitted for notational simplicity. According to (41) and (42), the reflecting wave and the passing wave at a scattering junction can be obtained. The discrete-time representations of (41) and (42) can be derived in the way similar to the derivation of (10) and (11). Let $\varphi_r^1$ and $\varphi_l^2$ be the right-going traveling wave flowing to the junction from the left-hand side and the left-going traveling wave flowing to the junction from the right-hand side, respectively. Let $f_l^1$ and $f_r^2$ be the traveling waves departing from the junction and flowing to the two sides, respectively. We have

$$f_l^1(n, m) = -\rho\varphi_r^1(n, m) + (1+\rho)\varphi_l^2(n, m) \tag{44}$$

and

$$f_r^2(n, m) = (1-\rho)\varphi_r^1(n, m) + \rho\varphi_l^2(n, m). \tag{45}$$

For simplicity, $(n, m)$ can be omitted. The discrete-time representation of (43) can be obtained as

$$y^J = (1-\rho)\varphi_r^1 + (1+\rho)\varphi_l^2. \tag{46}$$

By substituting (46) into (44) and (45), the right-going and the left-going traveling waves departing from the junction can be obtained as follows:

$$f_l^1 = -\rho\varphi_r^1 + (1+\rho)\varphi_l^2 = y^J - \varphi_r^1 \tag{47}$$

and

$$f_r^2 = (1-\rho)\varphi_r^1 + \rho\varphi_l^2 = y^J - \varphi_l^2. \tag{48}$$

Note that only the displacements of a vibration string at various positions are measurable. It is impossible to measure externally the departing traveling waves $f_l$ and $f_r$, and the arrival traveling waves $\varphi_l$ and $\varphi_r$. Therefore, (46)–(48) have to be combined with (39), and the initial string displacement in order to obtain the model parameters of an acoustic string, the reflection coefficients and the loss factors. We have more details on this subject in the later sections.

Based on (46)–(48), the model of a scattering junction of a string is shown in Fig. 4. The output of $y^J$ represents the displacement of a plucked string. This model means that the traveling wave departing from the junction can be computed by subtracting the traveling wave belonging to the same segment flowing into the junction from the displacement magnitude of the junction, and the displacement magnitude can be computed

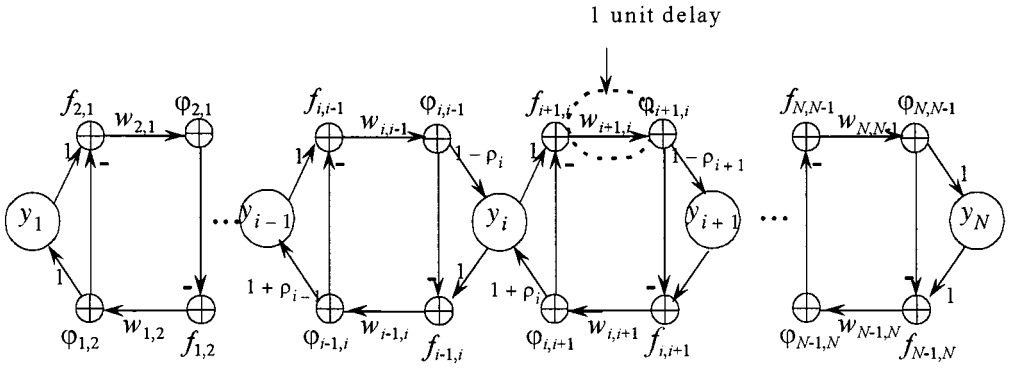



Fig. 5. SRN model of the plucked string with fixed ends.

by summing the arrival traveling waves multiplied by their individual factors $(1-\rho)$ and $(1+\rho)$.

## III. Scattering Recurrent Networks for the Modeling of Musical Strings

### A. Scattering Recurrent Network Model

In this section, an analysis/synthesis model for acoustic strings called the SRN is proposed and is shown in Fig. 5. The SRN is configured based on the theory explored in the previous section. There are three kinds of nodes in the SRN: the *displacement nodes* denoted by $y$, the *arrival nodes* denoted by $\varphi$, and the *departure nodes* denoted by $f$, which are borrowed from (46)–(48). Let the recurrent network have $N$ displacement nodes denoted by $y_i$ and the output of $y_i$ represents the displacement of the $i$th sampling position for $i=1, 2, \cdots, N\cdot y_1$ and $y_N$ represent the displacement nodes of the two fixed ends and the outputs are zeros all the time. Based on the discussion in Section II, each of displacement nodes is associated with the physical position of a small string segment whose length is $\Delta x$. For most complex situations, the characteristic impedance of each segment which is represented by a unit delay is different from its adjacent segments such that all physical positions represented by the displacement nodes except for the two ends can be regarded as scattering junctions. The arrival node $\varphi_{i,j}$ represents the traveling wave flowing into the junction $y_i$ from the junction $y_j$. Similarly, the departure node $f_{i,j}$ represents the traveling wave departing from the junction $y_j$ to the junction $y_i$. A traveling wave departs from the junction and passes through a link which contains $w_{i,j}$, representing the corresponding loss factor, and a unit-sample delay and the result becomes a traveling wave flowing into the adjacent junction.

The initial displacement of the string is normalized such that the largest magnitude is bounded by unity. In practice, the magnitude at any position throughout the entire period of vibrations cannot exceed the largest magnitude in the initial condition since the string is assumed to be lossy. Although other types of activation functions can be used, in order to reduce the amount of computation required in the resynthesis stage, we choose the linear function as our activation function

$$a(x)=x, \qquad \text{for all } x. \tag{49}$$

This is a reasonable assumption because a musical string is usually quite linear unless it is overly stretched. The linear function works fine in our experiments and the results can be seen in Section V.

According to Fig. 5 and (49), the outputs of the arrival nodes in the upper track and the lower track of the model can be computed as follows, respectively,

$$\varphi_{i,i-1}(t+1)=a\left[\text{net}^{\varphi}_{i,i-1}(t)\right]=a[w_{i,i-1}\cdot f_{i,i-1}(t)] \tag{50}$$

and

$$\varphi_{i,i+1}(t+1)=a\left[\text{net}^{\varphi}_{i,i+1}(t)\right]=a[w_{i,i+1}\cdot f_{i,i+1}(t)]. \tag{51}$$

The magnitudes of the displacement nodes of a plucked string at time $(t+1)$ can be obtained by

$$y_i(t+1)=\begin{cases} a[\text{net}^{y}_{i}(t+1)], & i=2,\cdots,N-1 \\ 0, & i=1 \text{ or } i=N. \end{cases} \tag{52}$$

where

$$\text{net}^{y}_{i}(t+1)=(1-\rho_i)\varphi_{i,i-1}(t+1)+(1+\rho_i)\varphi_{i,i+1}(t+1). \tag{53}$$

The outputs of the departure nodes can be computed by

$$f_{i+1,i}(t+1)=a\left(\text{net}^{f}_{i+1,i}(t+1)\right), \qquad i=1,\cdots,N-1 \tag{54}$$

where

$$\text{net}^{f}_{i+1,i}(t+1)=y_i(t+1)-\varphi_{i,i+1}(t+1) \tag{55}$$

and

$$f_{i-1,i}(t+1)=a\left(\text{net}^{f}_{i-1,i}(t+1)\right), \qquad i=2,\cdots,N \tag{56}$$

where

$$\text{net}^{f}_{i-1,i}(t+1)=y_i(t+1)-\varphi_{i,i-1}(t+1). \tag{57}$$

Equations (50)–(57) represent the approximated model of wave propagation of a musical string by using the proposed SRN model.

### B. Training of SRN's

The temporal operation of the proposed SRN can be unfolded into the multilayer feedforward architecture with synchronous update. Fig. 6 illustrates the unfolded version of the SRN shown

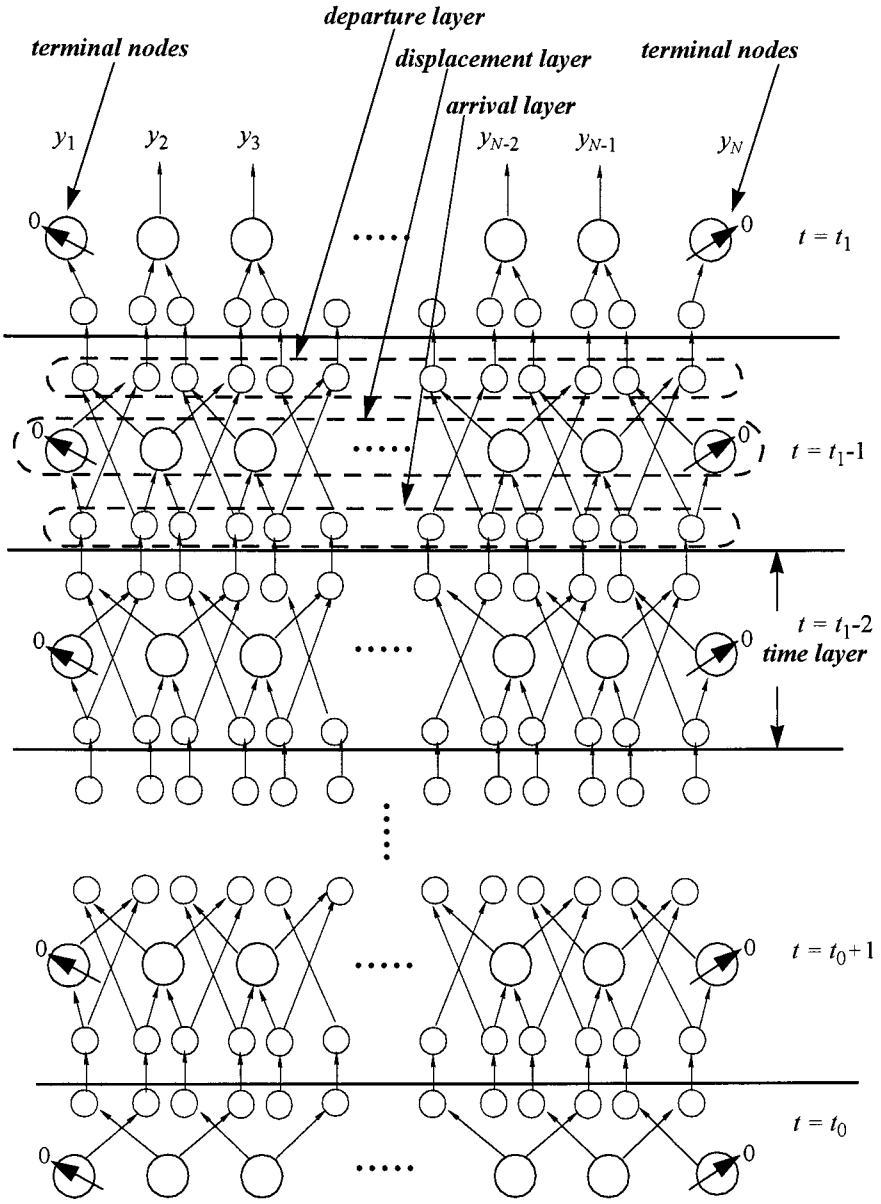



Fig. 6. The feedforward architecture behaves as the SRN of a plucked string shown in Fig. 5.

in Fig. 5. The unfolding procedure is described as follows. In Fig. 6, there is a *time layer* for each time instant. Each time layer contains a *displacement layer*, an *arrival layer*, and a *departure layer*. At time layer $t_1 - 1$ corresponding to time instant $t_1 - 1$, the displacement node $y_i$ receives the weighted signals from two arrival nodes, $\varphi_{i,\,i-1}$ and $\varphi_{i,\,i+1}$. The output signal of $y_i$ and the output signal of $\varphi_{i,\,i-1}$ are sent to the departure node $f_{i-1,\,i}$. The output signal of $y_i$ and the output signal of $\varphi_{i,\,i+1}$ are sent to the departure node $f_{i+1,\,i}$. The signals of $f_{i-1,\,i}$ and $f_{i+1,\,i}$ have to be sent to the next time layer through the corresponding loss factors and one unit delay to the arrival nodes $\varphi_{i-1,\,i}$ and $\varphi_{i+1,\,i}$ of time layer $t_1$. The weighted output signals of $\varphi_{i-1,\,i}$ and $\varphi_{i+1,\,i}$ of time layer $t_1$ are sent to $y_{i-1}$ and $y_{i+1}$ of time layer $t_1$, respectively. This means that after subtracting the signals of the associated arrival nodes, the output signal of $y_i$ at time instant $t_1 - 1$ arrives at $y_{i-1}$ and $y_{i+1}$ at time instant $t_1$, respectively. This is exactly the situation that the traveling waves travel in the upper track and the lower track and meet with each other at the scattering junctions. The output signals of the displacement nodes are the response of the SRN that is used to imitate the string vibration. Continue this procedure and an unfolded version is obtained. Note that the first time layer, time layer $t_0$, has no arrival layer and the outputs of the displacement nodes are magnitudes of the initial *pluck* at the corresponding positions. The BPTT method [19], [21] can be used to train the unfolded version of the SRN once the training vectors are available.

In Fig. 6, a neural-network layer called a *time layer* is assign for each time instant. Each time layer consists of three sublayers, *displacement layer*, *departure layer*, and *arrival layer*, which contain the displacement nodes, the departure nodes, and the arrival nodes, respectively. Because only the string displacements are measurable, the training vector is actually the set of sampled

displacement magnitudes of the string under excitations such as plucks at several preset positions. It is unlikely to measure the displacement for each physical position on the string in order to supply the training data for the corresponding displacement node. It is unavoidable that most of the positions corresponding to the displacement nodes are absent in the measurement. As a matter of fact, we measure only seven positions in our simulations. Those nodes with measured data are *visible nodes*. Those positions without measured data are *invisible nodes*. Note that the visible nodes are also displacement nodes expressed by (52). If $d_i(t)$ denotes the desired displacement of the $i$th sampling position at time $t$ and $A(t)$ denotes the set of visible nodes, the sampled data is employed to train the SRN by using the BPTT method such that the generated outputs of the visible nodes can be as close to the sampled data as possible. The remaining nodes of this network, the departure nodes, arrival nodes, and the invisible nodes, are called *hidden nodes*. The error signals at any time instant $t$ are defined as

$$e_i(t) = \begin{cases} d_i(t) - y_i(t), & \text{if } i \in A(t) \\ 0, & \text{otherwise.} \end{cases} \tag{58}$$

where

$e_i(t)$ error signal of the $i$th displacement node at time $t$;
$d_i(t)$ desired response of the $i$th displacement node at time $t$;
$y_i(t)$ actual output of the $i$th displacement node at time $t$.

The error function at time $t$ is defined as

$$E(t) = 1/2 \sum_{i \in A(t)} e_i^2(t) \tag{59}$$

and we have the cost function

$$E^{\text{total}}(t_0, t_1) = \sum_{t=t_0+1}^{t_1} E(t) \tag{60}$$

to be minimized over one epoch $[t_0, t_1]$, where $t_1$ is the last time step and $t_0$ is the initial time step. It is necessary to adjust the weights of the SRN to approach the desired loss factors and the desired reflection coefficients to minimize the cost function. These weights corresponding to the loss factors should change along the negative gradient of the cost function as follows:

$$\Delta w_{i+1,\,i} = -\eta \frac{\partial E^{\text{total}}(t_0, t_1)}{\partial w_{i+1,\,i}}, \qquad \text{for } i = 1, \cdots, N-1 \tag{61}$$

$$\Delta w_{i-1,\,i} = -\eta \frac{\partial E^{\text{total}}(t_0, t_1)}{\partial w_{i-1,\,i}}, \qquad \text{for } i = 2, \cdots, N. \tag{62}$$

The weights corresponding to the reflection coefficients should change along the negative gradient of the cost function as follows:

$$\Delta \rho_i = -\eta \frac{\partial E^{\text{total}}(t_0, t_1)}{\partial \rho_i}, \qquad \text{for } i = 2, \cdots, N-1 \tag{63}$$

where $\eta$ is the learning constant. After the SRN runs through each epoch, the backpropagated error signals are used to compute the weight changes at the corresponding backward propagated time layer. The total weight change for each connecting weight is computed by accumulating the individual weight change for it at each time layer. According to (50), –(57), at any time layer $t$, the weight changes corresponding to the reflection coefficients are computed by

$$\begin{aligned} \Delta \rho_i(t) &= -\eta \frac{\partial E^{\text{total}}(t_0, t_1)}{\partial \operatorname{net}_i^y(t)} \\ &\quad \cdot \frac{\partial \operatorname{net}_i^y(t)}{\partial \rho_i}, \qquad \text{for } i = 2, \cdots, N-1 \\ &= \eta \cdot \delta_i^y(t) \cdot (\varphi_{i,\,i+1}(t) - \varphi_{i,\,i-1}(t)) \end{aligned} \tag{64}$$

where $\delta_i^y(t)$ is the gradient values of the displacement nodes.

The weight changes corresponding to the loss factors are

$$\begin{aligned} \Delta w_{i+1,\,i}(t) &= -\eta \frac{\partial E^{\text{total}}(t_0, t_1)}{\partial \operatorname{net}_{i+1,\,i}^{\varphi}(t-1)} \cdot \frac{\partial \operatorname{net}_{i+1,\,i}^{\varphi}(t-1)}{\partial w_{i+1,\,i}} \\ &= \eta \cdot \delta_{i+1,\,i}^{\varphi}(t) \cdot f_{i+1,\,i}(t-1) \end{aligned} \tag{65}$$

and

$$\begin{aligned} \Delta w_{i-1,\,i}(t) &= -\eta \frac{\partial E^{\text{total}}(t_0, t_1)}{\partial \operatorname{net}_{i-1,\,i}^{\varphi}(t-1)} \cdot \frac{\partial \operatorname{net}_{i-1,\,i}^{\varphi}(t-1)}{\partial w_{i-1,\,i}} \\ &= \eta \cdot \delta_{i-1,\,i}^{\varphi}(t) \cdot f_{i-1,\,i}(t-1) \end{aligned} \tag{66}$$

where

$\delta_{i+1,\,i}^{\varphi}(t)$ gradient values of the right-going arrival nodes;

$\delta_{i-1,\,i}^{\varphi}(t)$ gradient values of the left-going arrival nodes.

The gradient values of the displacement nodes can be obtained as follows:

$$\begin{aligned} \delta_i^y(t) &= -\frac{\partial E^{\text{total}}(t_0, t_1)}{\partial \operatorname{net}_i^y(t)}, \qquad i = 2, \cdots, N-1,\ t_0 < t \le t_1 \\ &= -\frac{\partial E(t)}{\partial y_i(t)} \cdot \frac{\partial y_i(t)}{\partial \operatorname{net}_i^y(t)} \\ &\quad \underbrace{-\frac{\partial E^{\text{total}}(t_0, t_1)}{\partial \operatorname{net}_{i-1,\,i}^{f}(t)} \cdot \frac{\partial \operatorname{net}_{i-1,\,i}^{f}(t)}{\partial y_i(t)}} \cdot \frac{\partial y_i(t)}{\partial \operatorname{net}_i^y(t)} \\ &\quad \underbrace{-\frac{\partial E^{\text{total}}(t_0, t_1)}{\partial \operatorname{net}_{i+1,\,i}^{f}(t)} \cdot \frac{\partial \operatorname{net}_{i+1,\,i}^{f}(t)}{\partial y_i(t)}} \cdot \frac{\partial y_i(t)}{\partial \operatorname{net}_i^y(t)}. \end{aligned} \tag{67}$$

According to (52), (55), and (57), (67) can be rewritten as

$$\delta_i^y(t) = \begin{cases} e_i(t) \cdot a'\left(\operatorname{net}_i^y(t)\right), & t = t_1 \\ \left(e_i(t) + \delta_{i-1,\,i}^{f}(t) + \delta_{i+1,i}^{f}(t)\right) & \\ \quad \cdot a'\left(\operatorname{net}_i^y(t)\right), & t_0 < t < t_1. \end{cases} \tag{68}$$

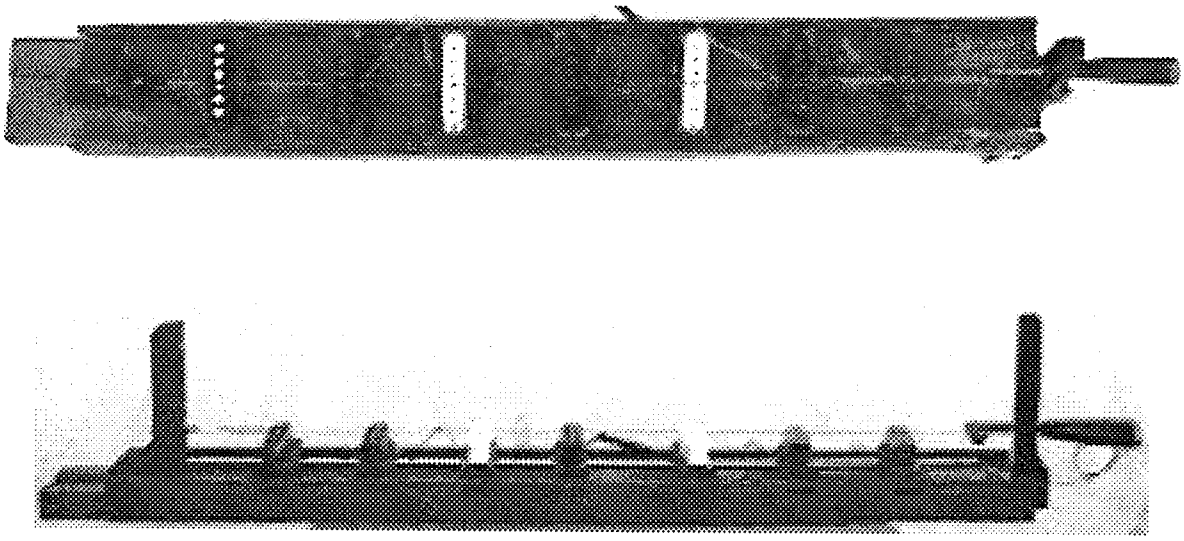


Fig. 7. The steel-string measurement device constructed to obtain the training data for the SRN model.

Similarly, the gradient values of the left-going departure nodes and the right-going departure nodes are computed as follows, respectively,

$$\begin{aligned}\delta^{f}_{i-1,\,i}(t) &= -\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial\,\text{net}^{f}_{i-1,\,i}(t)}\\ &= \left(\underbrace{-\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial\varphi_{i-1,\,i}(t+1)}\cdot\frac{\partial\varphi_{i-1,\,i}(t+1)}{\partial\,\text{net}^{\varphi}_{i-1,\,i}(t)}}\right.\\ &\qquad\left.\cdot\frac{\partial\,\text{net}^{\varphi}_{i-1,\,i}(t)}{\partial f_{i-1,\,i}(t)}\cdot\frac{\partial f_{i-1,\,i}(t)}{\partial\,\text{net}^{f}_{i-1,\,i}(t)}\right)\\ &= \delta^{\varphi}_{i-1,\,i}(t+1)\cdot w_{i-1,\,i}\cdot a'\left(\text{net}^{f}_{i-1,\,i}(t)\right)\end{aligned}\tag{69}$$

for $i=2,\cdots,N$, $t_0<t\le t_1-1$.

$$\begin{aligned}\delta^{f}_{i+1,\,i}(t) &= -\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial\,\text{net}^{f}_{i+1,\,i}(t)}\\ &= \left(\underbrace{-\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial\varphi_{i+1,\,i}(t+1)}\cdot\frac{\partial\varphi_{i+1,\,i}(t+1)}{\partial\,\text{net}^{\varphi}_{i+1,\,i}(t)}}\right.\\ &\qquad\left.\cdot\frac{\partial\,\text{net}^{\varphi}_{i+1,\,i}(t)}{\partial f_{i+1,\,i}(t)}\cdot\frac{\partial f_{i+1,\,i}(t)}{\partial\,\text{net}^{f}_{i+1,\,i}(t)}\right)\\ &= \delta^{\varphi}_{i+1,\,i}(t+1)\cdot w_{i+1,\,i}\cdot a'\left(\text{net}^{f}_{i+1,\,i}(t)\right)\end{aligned}\tag{70}$$

for $i=1,\cdots,N-1$, $t_0<t\le t_1-1$.

The gradient values of the right-going arrival nodes and the left-going arrival nodes can be computed as follows:

$$\begin{aligned}\delta^{\varphi}_{i+1,\,i}(t) &= -\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial\,\text{net}^{\varphi}_{i+1,\,i}(t-1)}\\ &= \underbrace{-\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial y_{i+1}(t)}\cdot\frac{\partial y_{i+1}(t)}{\partial\,\text{net}^{y}_{i+1}(t)}}\\ &\qquad\cdot\frac{\partial\,\text{net}^{y}_{i+1}(t)}{\partial\varphi_{i+1,\,i}(t)}\cdot\frac{\partial\varphi_{i+1,\,i}(t)}{\partial\,\text{net}^{\varphi}_{i+1,\,i}(t-1)}\\ &\qquad\underbrace{-\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial f_{i,\,i+1}(t)}\cdot\frac{\partial f_{i,\,i+1}(t)}{\partial\,\text{net}^{f}_{i,\,i+1}(t)}}\\ &\qquad\cdot\frac{\partial\,\text{net}^{f}_{i,\,i+1}(t)}{\partial\varphi_{i+1,\,i}(t)}\cdot\frac{\partial\varphi_{i+1,\,i}(t)}{\partial\,\text{net}^{\varphi}_{i+1,\,i}(t-1)}\\ &= \left(\delta^{y}_{i+1}(t)\cdot(1-\rho_{i+1})-\delta^{f}_{i,\,i+1}(t)\right)\\ &\qquad\cdot a'\left(\text{net}^{\varphi}_{i+1,\,i}(t-1)\right)\end{aligned}\tag{71}$$

for $i=1,\cdots,N-1$, $t_0<t\le t_1$.

$$\begin{aligned}\delta^{\varphi}_{i-1,\,i}(t) &= -\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial\,\text{net}^{\varphi}_{i-1,\,i}(t-1)}\\ &= \underbrace{-\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial y_{i-1}(t)}\cdot\frac{\partial y_{i-1}(t)}{\partial\,\text{net}^{y}_{i-1}(t)}}\\ &\qquad\cdot\frac{\partial\,\text{net}^{y}_{i-1}(t)}{\partial\varphi_{i-1,\,i}(t)}\cdot\frac{\partial\varphi_{i-1,\,i}(t)}{\partial\,\text{net}^{\varphi}_{i-1,\,i}(t-1)}\\ &\qquad\underbrace{-\frac{\partial E^{\text{total}}(t_0,\,t_1)}{\partial f_{i,\,i-1}(t)}\cdot\frac{\partial f_{i,\,i-1}(t)}{\partial\,\text{net}^{f}_{i,\,i-1}(t)}}\\ &\qquad\cdot\frac{\partial\,\text{net}^{f}_{i,\,i-1}(t)}{\partial\varphi_{i-1,\,i}(t)}\cdot\frac{\partial\varphi_{i-1,\,i}(t)}{\partial\,\text{net}^{\varphi}_{i-1,\,i}(t-1)}\\ &= \left(\delta^{y}_{i-1}(t)\cdot(1+\rho_{i-1})-\delta^{f}_{i,\,i-1}(t)\right)\\ &\qquad\cdot a'\left(\text{net}^{\varphi}_{i-1,\,i}(t-1)\right)\end{aligned}\tag{72}$$

for $i=2,\cdots,N$, $t_0<t\le t_1$.

At the two fixed end points, the gradient values are set to be zero as follows:

$$\delta^{y}_{1}(t)=\delta^{y}_{N}(t)=0,\ t_0<t\le t_1.\tag{73}$$

In (68)–(72), $a'(\cdot)$ is the derivative of the activation function. The training process begins at the last time step $t_1$ and finishes at the beginning time step $t_0$. When the backpropagation computation is performed back to time $t_0+1$, the total weight changes for the weight factors in the SRN are obtained by summing the

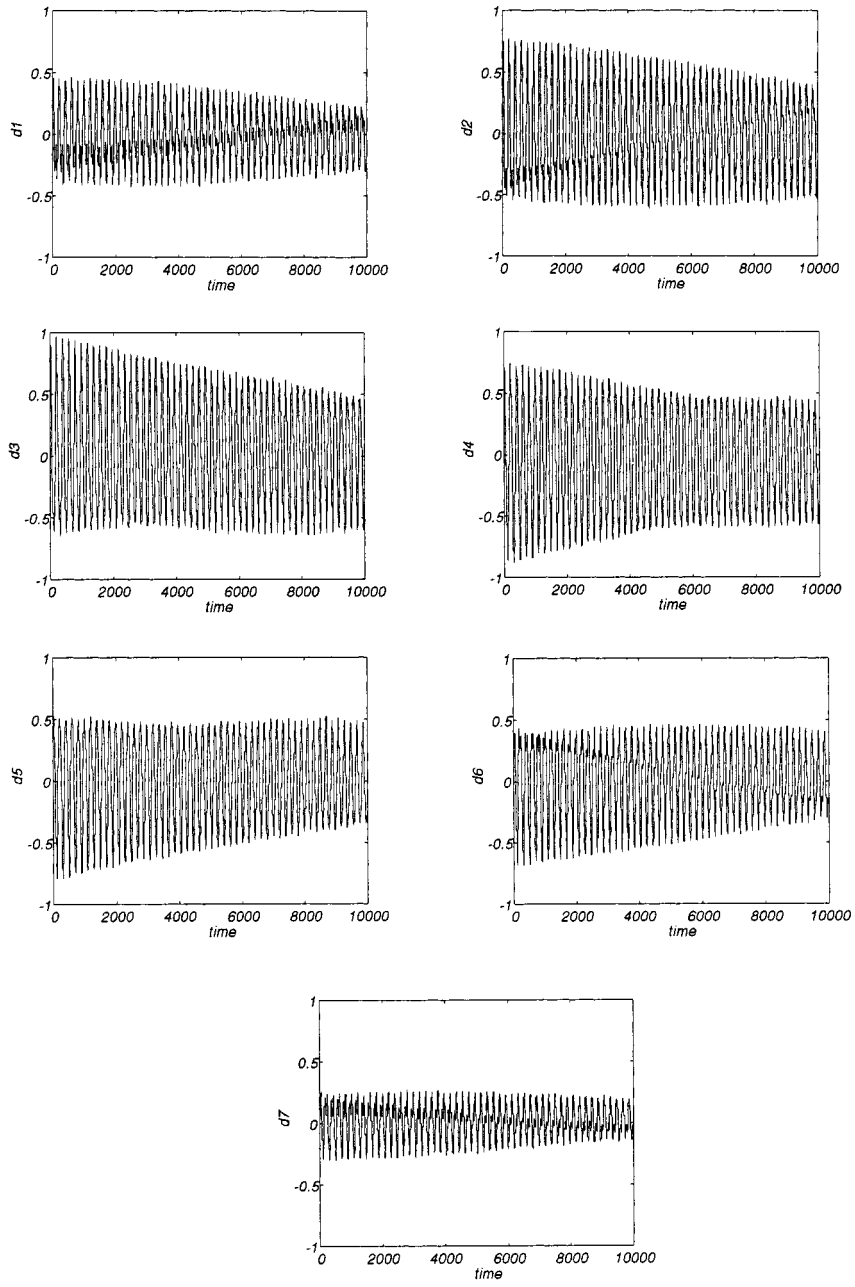



Fig. 8. The vibrations of a cello "A" string at various sampling positions.

weight changes in (64)–(66) at each backward propagated *time layer* together individually, as follows:

$$\Delta\rho_i = -\eta\frac{\partial E^{\text{total}}(t_0, t_1)}{\partial \rho_i} = \sum_{t=t_0+1}^{t_i}\Delta\rho_i(t)$$
$$= \eta\sum_{t=t_0+1}^{t_1}\delta_i^y(t)\cdot(\varphi_{i,i+1}(t) - \varphi_{i,,i-1}(t)),$$
$$\text{for } i = 2, \cdots, N-1 \tag{74}$$

$$\Delta w_{i+1,i} = -\eta\frac{\partial E^{\text{total}}(t_0, t_1)}{\partial w_{i+1,i}} = \sum_{t=t_0+1}^{t_i}\Delta w_{i+1,i}(t)$$
$$= \eta\sum_{t=t_0+1}^{t_1}\delta_{i+1,i}^{\varphi}(t)\cdot f_{i+1,i}(t-1),$$
$$\text{for } i = 1, \cdots, N-1 \tag{75}$$

$$\Delta w_{i-1,i} = -\eta\frac{\partial E^{\text{total}}(t_0, t_1)}{\partial w_{i-1,i}} = \sum_{t=t_0+1}^{t_i}\Delta w_{i-1,i}(t)$$
$$= \eta\sum_{t=t_0+1}^{t_1}\delta_{i-1,i}^{\varphi}(t)\cdot f_{i-1,i}(t-1),$$
$$\text{for } i = 2, \cdots, N \tag{76}$$

where $\delta_{i,i+1}^{f}(t_1) = \delta_{i,i-1}^{f}(t_1) = 0$. Repeating the *epoch-wise backpropagation* procedure until the cost function is less than

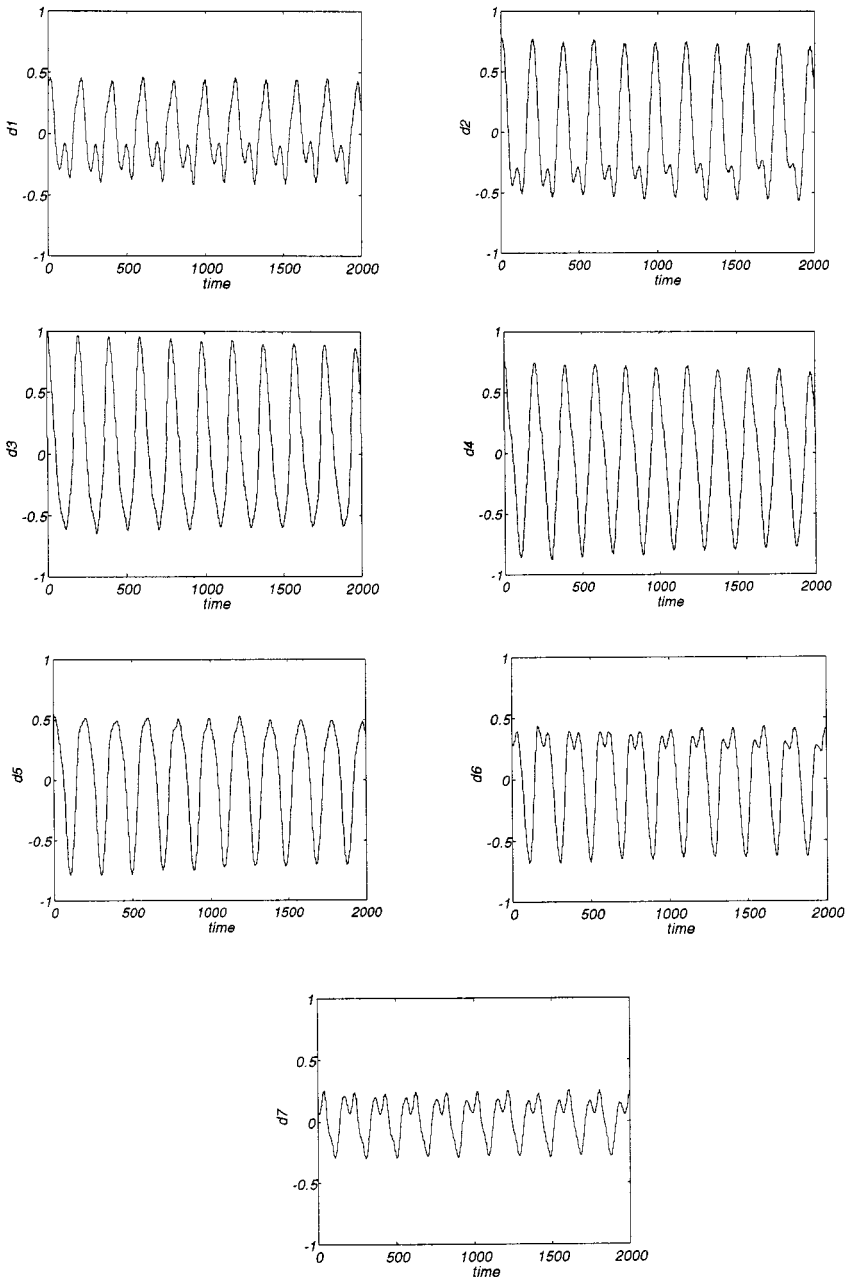



Fig. 9. The vibrations of a cello "A" string at various sampling positions in the interval [0, 2000].

a preset threshold, the corresponding model parameters of the SRN for a musical string is obtained.

## IV. Measurements of Musical Strings

In order to obtain the training data for the SRN model, a steel-string measurement system was constructed. It is in Fig. 7. Seven Dimarzio® Virtual Vintage Solo electromagnetic pickups usually found in electric guitars are placed in parallel and equally spaced to measure the vibrations of the chosen steel string in different sampling positions. An electromagnetic pickup consists of a coil with a permanent magnet. The plucked steel string causes magnetic flux changes for the pickups and electrical signals are induced through the coil. The more pickup units used in the system, the more measured data can be obtained for training the SRN model. Furthermore, the pickup units are combined with their sliding seats so that they can be moved along the track under the string to different positions whenever required.

In our experiments, the vibrations of the steel cello strings at various sampling positions are measured synchronously. All measurements are carried out through preamplifiers, analog-to-digital converters and stored on a multitrack real-time digital audio storage device, Audio Engine® (Sprectral Co. USA), at

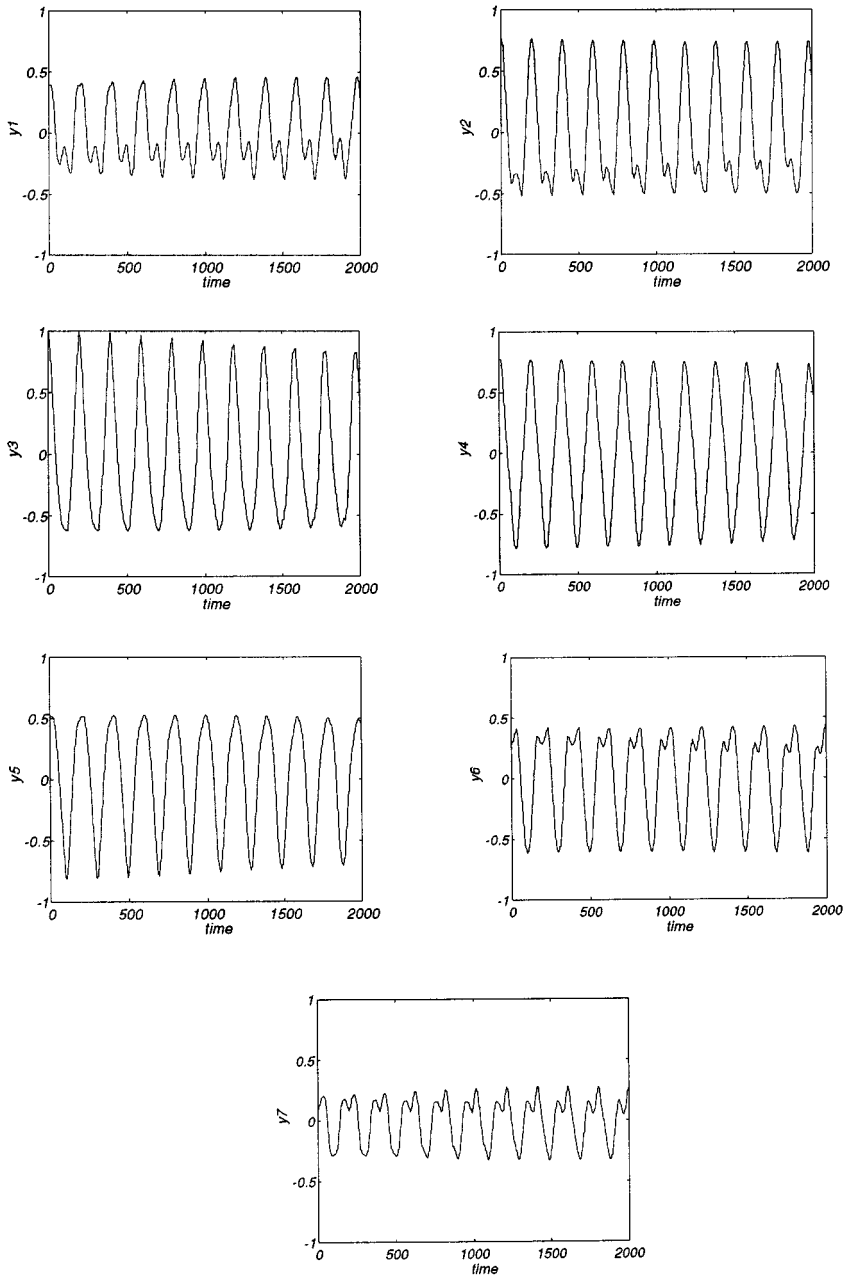



Fig. 10. The analysis outputs of the SRN at various sampling positions after 10 000-epoch learning.

32-kHz sampling rate and 16-bit quantization level. These data are used as the training and testing patterns for the SRN's.

## V. Experiments

Following are the experiments using the proposed SRN's to simulate the vibrations of a cello "A" string. The steel-string measurement system shown in Fig. 7 is used to measure the response of the plucked string such that the training vectors for the SRN's can be obtained. Fig. 8 shows the measured responses of the plucked "Jargar"® cello "A"-string at the seven sampling positions. The magnitudes of the training vectors are normalized into the interval [1, −1]. The 10 000 samples from the measurement for each sampling position are used as the desired data. In each subplot, the first 2000 samples are used as training data and the next 8000 samples are used for the comparison with the resynthesis data generated by the trained SRN at the displacement nodes corresponding to the physical sampling positions. The modeling work consists of two steps: training and resynthesis. The SRN is trained by using the method proposed in Section III. The resynthesis part utilizes the trained SRN to artificially generate the vibrations of the seven visible nodes by a single "pluck" shown in Fig. 11 on the "*virtual string*." The accuracy of the synthesis result depends on whether the synthetic outputs generated by the SRN are close enough to the measured data. Note that it is not necessary to have any limitation on the

training data as long as the string is not overly stretched because the proposed approach is designed to achieve accurate synthesis for any plucked-string sounds measured by using our setup.

### A. Analysis and Modeling

There are 100 displacement nodes in the SRN and the two ends are fixed. The string is assumed to be ideal in the beginning, i.e., the weights for the loss factors of the SRN are assigned to be unity and the weights for the reflection coefficients are assigned to be zero. All of the loss factors and the reflection coefficients can be adjusted in the training phase. The learning constant is $1 \times 10^{-7}$ which is chosen empirically. Because it is desired that the output of the SRN be as close to the desired signals as possible, the training is performed until there is no more change of the error signal. This is due to the fact that the analysis part can be done off-line and the purpose is to obtain better SRN parameters to achieve accurate synthesis. Usually, 10 000 epochs are enough because another 10 000 epochs of training can produce only less than 0.3 db SNR improvement. Fig. 10 shows the output of the SRN after 10 000-epoch learning at each sampling position. By comparing Figs. 9 and 10, the outputs of the visible nodes in the SRN are indeed close to the desired training data of the corresponding sampling positions in the interval [0, 2000]. The best signal-to-noise-ratio (SNR) that can be achieved by using the proposed approach is 22.16 db. The equation used to compute the SNR is (77)

$$SNR = 10 \cdot \log \left( \frac{\sum_{t=t_0+1}^{t_1} d^2(t)}{\sum_{t=t_0+1}^{t_1} e^2(t)} \right). \tag{77}$$

### B. Resynthesis

In the resynthesis part, the synthetic output generated by the proposed SRN is compared with the measured data of the vibrating string. Unlike the technique introduced in [29] which requires a short period of the time-domain response taken from the acoustic plucked-string instrument to be synthesized as its initial input, the proposed SRN requires only the initial magnitudes of the visible nodes at the corresponding plucking positions. The magnitudes of the rest of the displacement nodes are calculated by simple interpolation. This significantly reduces the required memory space if hardware implementation is desired. The "*virtual pluck*" to the SRN obtained in the above analysis part is shown in Fig. 11. The time-domain responses of the measured data are shown in Fig. 8, and the synthetic data are shown in Fig. 12. It can be seen that these two figures are very close in the interval [0, 6000]. From [6000, 10 000], the waveforms of the figures are still quite close except for $y_1$ and $y_7$. This is within our expectation since the acoustic characteristic of a vibrating string keeps changing with time, especially at the two end positions. This is to say that it may be necessary to perform the analysis again if the characteristic changes too much and the model parameters of the SRN have to be changed during the synthesis phase, too. Critical subjective listening tests are conducted with six subjects through a pair of Spendor S20 near field monitors driven by a Crown K-2 power amplifier. All subjects

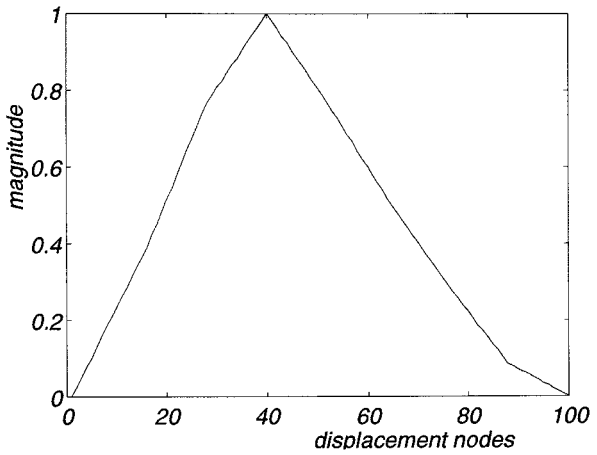



Fig. 11. The *virtual pluck* to the SRN.

can still tell the difference though they judge that the difference is very minor. However, if low-price multimedia speakers usually come with personal computers are used, they cannot tell the difference. For most synthesizers available on the market, listeners can easily tell that the initial transient signal does not sound like a real pluck. Furthermore, these synthesizers need more memory space to store the information for synthesizing different parts of the sampled tone in order to imitate the beginning part, the sustaining part and the decaying part of the sounds to be synthesized.

## VI. Conclusion and Future Work

A new physical modeling electronic music synthesis technique based on physics of musical strings and recurrent neural networks is presented. By using sampled magnitudes of plucked-string vibrations measured by the electromagnetic pick-up system as the training data, the proposed SRN is able to imitate the sounds of its acoustic counterpart after sufficient training. The "*virtual string*" constructed by the SRN can accept arbitrary "*plucks*" just like a real string. The proposed approach can synthesize very realistic starting transient responses of a plucked string. This is not possible for traditional approaches such as the FM method and wavetable method.

There are still some unsolved problems, however. First of all, most musical instruments have to be analyzed by using 2-D or even 3-D structures, and thus it is necessary to extend our current technique for multidimensional modeling. Second, if the shape of a musical instrument is complicated, it may be very difficult to design the structure of the corresponding SRN so that the analysis/synthesis can be performed. For example, the bridges of violins are very difficult to model for their complex shapes. Third, the assumption of the proposed technique is that we are able to measure the responses of excited musical instruments so that the SRN based on the sampled magnitudes of the vibrations can be trained. However, this may be very difficult for many acoustic instruments. For example, it is almost impossible to measure the vibrations of reeds for woodwind instruments and bows for violins without affecting their dynamic behaviors because of the attached sensors. Finally, the required computation for the synthesis part is still too large to be used in most

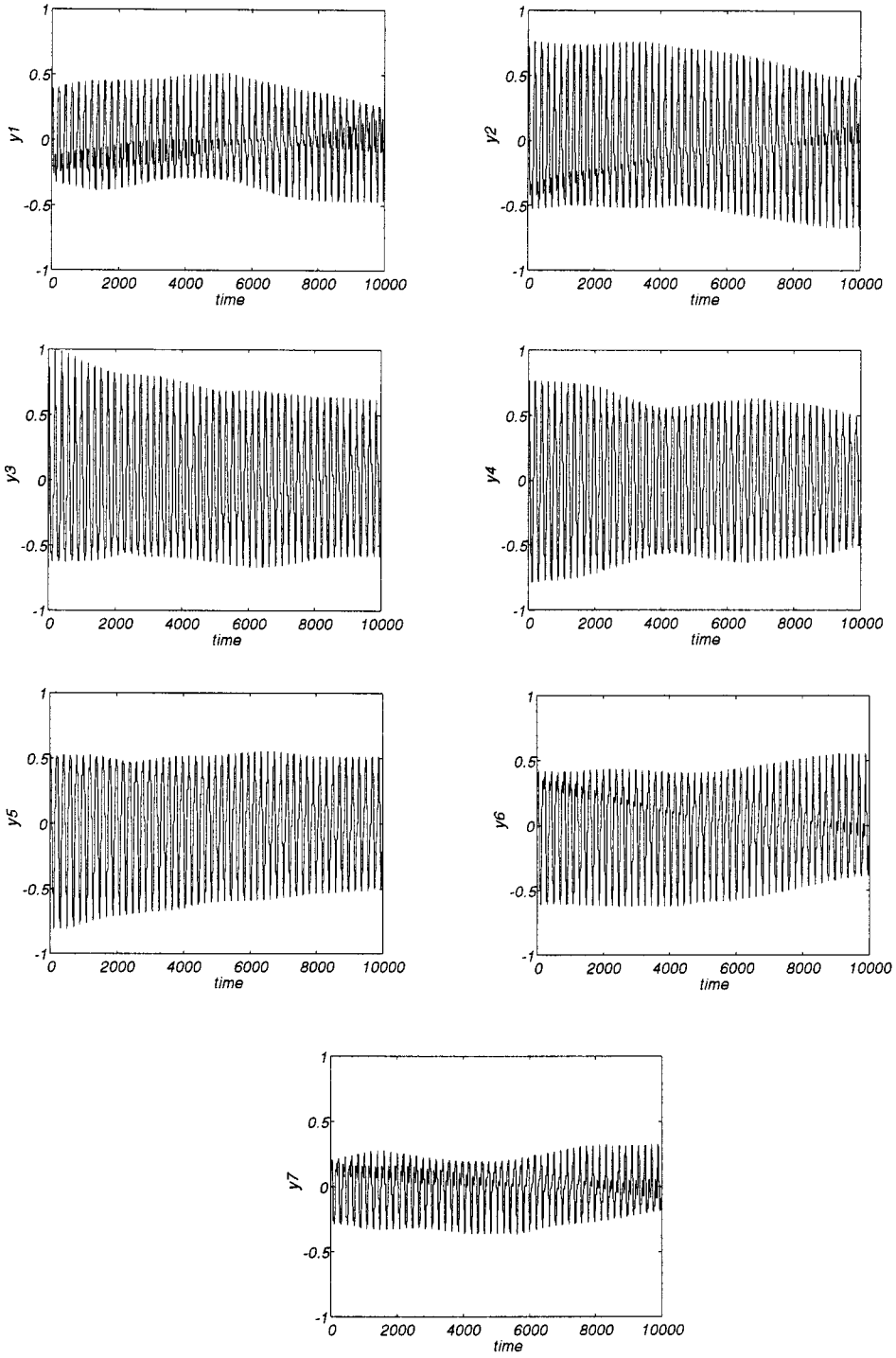



Fig. 12. The resynthesis data generated by the trained SRN at various sampling positions.

real-time applications, and thus it is necessary to improve the computational efficiency of the proposed SRN in the synthesis stage. It is also desired to develop a more efficient training algorithm to improve the SRN and the training speed.

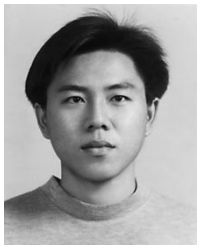
## ACKNOWLEDGMENT

The authors would like to thank Dr. J. O. Smith of CCRMA, Stanford University, Stanford, CA, for numerous inspiring discussions on physical modeling synthesis, L. W. Wang of ITRI, Taiwan, who performed the string measurements, and the editor and the reviewers for their helpful opinions.

## REFERENCES

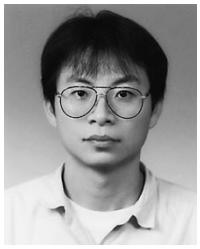
[1] C. M. Hutchins, "Stradivarius plate tests," *Catgut Acoust. Soc. J.*, no. 37, p. 30, 1982.
[2] L. Cremer, *The Physics of the Violin*. Cambridge, MA: MIT Press, 1984.

[3] A. D. Lindsay and R. A. Willgoss, "Finite-element analyzes of "Stradivari" and "Modern" top plates," *Catgut Acoust. Soc. J.*, vol. 3, no. 4, 1997.
[4] S. F. Liang and A.W. Y. Su, "Dynamics modeling of musical string by linear scattering recurrent network," in *ICS'96 Proc. Artificial Intell.*, Taiwan, 1996, pp. 263–270.
[5] S. F. Liang, "Dynamics modeling of musical string by ANN," Master thesis, Dept. Contr. Eng., Chiao-Tung Univ., Taiwan, 1996.
[6] E. Holsinger, *How Music and Computers Work*. Chicago, IL: Ziff-Davis Press, 1994.
[7] S. Pellman, *An Introduction to the Creation of Electroacoustic Music*. Belmont, CA: Wadsworth, 1994.
[8] J. M. Chowning, "The synthesis of complex audio spectra by means of frequency modulation," *J. Audio Eng. Soc.*, vol. 21, no. 7, pp. 526–534, 1973.
[9] K. Karplus and A. Strong, "Digital synthesis of plucked-string and drum timbres," *Comput. Music J.*, vol. 7, no. 2, pp. 43–55, 1983.
[10] J. O. Smith, "Efficient synthesis of stringed musical instruments," in *Proc. Int. Comput. Music Conf.*, 1993, pp. 64–71.
[11] ——, "Physical modeling using digital waveguides," *Comput. Music J.*, vol. 16, no. 4, pp. 74–87, 1992.
[12] ——, "Music application of digital waveguide," Stanford Univ., CCRMA Tech. Rep. STAN-M-67.
[13] ——, "Techniques for digital filter design and system identification with application to violin," Ph.D. dissertation, Stanford UNIV., 1983.
[14] P. R. Cook, "Identification of control parameters in an articulatory vocal tract model with applications to the synthesis of singing," Ph.D. dissertation, Stanford Univ., 1990.
[15] P. M. Morse, *Vibration and Sound*. Woodbury, NY: Amer. Inst. Phys. Acoust. Soc. Amer., 1936.
[16] A. J. Terri, "The Shannon sampling theorem—Its various extensions and applications: A tutorial review," *Proc. IEEE*, vol. 65, pp. 1565–1596, Nov. 1977.
[17] L. Hill and P. Ruiz, "Synthesizing musical sounds by solving the wave equation for vibrating objects," *J Audio Eng. Soc.* , pt. 1/2, vol. 19, no. 6/7, pp. 462–470, 1971.
[18] L. O. Chua, C. A. Desoer, and E. S. Kuh, *Linear and Nonlinear Circuits*. New York: McGraw-Hill, 1987.
[19] R. J. Williams and J. Peng, "An efficient gradient-based algorithm for on-line training of recurrent network trajectories," *Neural Comput.*, vol. 2, pp. 490–501, 1990.
[20] R. J. Williams and D. Zipser, "A learning algorithm for continually running fully recurrent neural networks," *Neural Comput.*, pp. 270–280, 1989.
[21] F. J. Pineda, "Recurrent backpropagation and the dynamical approach to adaptive neural computation," *Neural Comput.*, pp. 161–172, 1989.
[22] D. E. Rumelhart, D. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing*, vol. 1, D. E. Rumelhart and J. L. McClelland, Eds., 1986, pp. 318–362.
[23] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation*. New York: Addison-Wesley, 1991.
[24] M. V. Mathews and J. R. Pierce, *Current Directions in Computer Music Research*. Cambridge, MA: MIT Press, 1991.
[25] F. R. Moore, *Elements of Computer Music*. Englewood Cliffs, NJ: Prentice-Hall, 1990.
[26] N. H. Fletcher and T. D. Rossing, *The Physics of Musical Instruments*. New York: Springer-Verlag, 1991.
[27] A. V. Oppenhim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
[28] B. C. J. Moore, *An Introduction to the Psychology of Hearing*. New York: Academic, 1989.
[29] M. Karjalainen *et al.*, "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.
[30] L. E. Kinsler *et al.*, *Fundamentals of Acoustics*, 3rd ed. New York: Wiley, 1982.



**Sheng-Fu Liang** was born in Tainan, Taiwan, in 1971. He received the B.S. and M.S. degrees in control engineering from the National Chiao-Tung University, Taiwan, in 1994 and 1996, respectively. He is currently a Ph.D. student there.

His research activities include model-based music synthesis, neural networks, and signal processing. His current projects include analysis/synthesis of traditional Chinese plucked-string instruments and transducer-dependent digital audio signal processing. He is interested in oriental musical instruments and is also a Moon-Chin player in the Taiwanese Opera.



**Alvin W. Y. Su** (S'90–M'96) was born in Taiwan, in 1964. He received the B.S. degree in control engineering from the National Chiao-Tung University in 1986, and the M.S. and Ph.D. degrees in electrical engineering from Polytechnic University, Brooklyn, New York, in 1990 and 1993, respectively.

From 1993 to 1994, he was with the Center for Computer Research in Music and Acoustics (CCRMA), Stanford University, CA. From 1994 to 1995, he was with Computer Communication Lab. of Industrial Technology Research Institute (CCL. ITRI.), Taiwan. In 1995, he joined the Department of Information Engineering and Computer Engineering at Chung-Hwa University where he serves as an Associate Professor. His research interests cover the areas of digital audio signal processing, physical modeling of acoustic musical instruments, human computer interface design, video and color image signal processing, and VLSI signal processing. His research projects include computer composition and synthesis system for traditional Chinese plucked-string instruments, automatic computer music accompaniment system, resizing of still images and video images, and color correction and enhancement of imaging devices. He is currently conducting several industrial projects on VLSI design for digital audio and video signal processing.

Dr. Su is a member of the IEEE Computer Society and Signal Processing Society. He is also a member of the Acoustical Society of America.

**Chin-Teng Lin** (S'88–M'91–SM'99) received the B.S. degree in control engineering from the National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., in 1986 and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, in 1989 and 1992, respectively.

Since August 1992, he has been with the College of Electrical Engineering and Computer Science, National Chiao-Tung University, Hsinchu, Taiwan, R.O.C., where he is a Professor of electrical and control engineering. He has been the Deputy Dean of the Research and Development Office of the National Chiao-Tung University since 1998. His current research interests are fuzzy systems, neural networks, intelligent control, human-machine interface, and video and audio processing. He is coauthor of *Neural Fuzzy Systems—A Neuro-Fuzzy Synergism to Intelligent Systems* (Englewood Cliffs, NJ: Prentice-Hall, 1996), and the author of *Neural Fuzzy Control Systems with Structure and Parameter Learning* (Singapore: World Scientific, 1994). He has published more than 40 journal papers in the areas of neural networks and fuzzy systems.

Dr. Lin is a member of Tau Beta Pi and Eta Kappa Nu. He is also a member of the IEEE Computer Society, the IEEE Robotics and Automation Society, and the IEEE Systems, Man, and Cybernetics Society. He has been the Executive Council Member of the Chinese Fuzzy System Association (CFSA) since 1998. He was the Vice Chairman of the IEEE Robotics and Automation Taipei Chapter in 1996 and 1997. He won the Outstanding Research Award granted by the National Science Council (NSC), Taiwan, in 1997 and 1999, and the Outstanding Electrical Engineering Professor Award granted by the Chinese Institute of Electrical Engineering (CIEE) in 1997.