Sound Synthesis of the Pipa Based on Computed Timbre Analysis and Physical Modeling

Yi-Huei Chen and Chih-Fang Huang

Abstract—This paper proposes sounds analysis and a synthesis model for the Chinese plucked string instrument called the pipa, one of the oldest Chinese musical instruments with over 2000 years of history. The pipa comprises four strings with 30 frets and individual pear-shaped body so that it can offer a wide chromatic scale around 3.5 octaves and many kinds of tones. The acoustical properties of this instrument are analyzed according to the recorded tones. The most vital playing techniques are synthesized by using both physical and spectral based models with auxiliary rules. Applying the digital waveguide concept, the pipa model was constructed with digital filters and input excitations. Synthetic results are very similar to the recorded one according to the waveforms and spectra comparison and the statistics of listening tests.

Index Terms—Digital waveguide, physical base model, pipa, playing techniques, pluck string, sound synthesis.

I. INTRODUCTION

A. Model-Based Plucked String Sound Synthesis

DUE to the rapid development of digital signal processing (DSP), the physical modeling of a sound object can be realized with the digital waveguide (DWG) [1], [2]. Many different types of instruments have been successfully synthesized with this technique and the sound quality is improving [3]–[7]. Among the instruments, the physical modeling of plucked strings is the earliest one to be developed. After the Karplus–Strong algorithm was proposed [8], besides the western instruments like the guitar and mandolin, the eastern ones like the guqin [9] and the dan trahn [10] had been also modeled in succession. Compared with the sample-based synthesizer, the model-based one is a better approach to generate sufficiently realistic sounds in response to control events without spending a huge amount of computer memory [11], [12]. Moreover, this synthesizer is a conveniently virtual instrument which can be embedded not only in computers but also in many consumer products to provide practically unlimited instrument tones in real-time.

The timbre of the plucked strings is mainly determined by the attack, the sustaining and the decay segments of the sound waveform (corresponding to the playing technique, the string material and the body structure). With the DWG technique, the wave traveling on the string is transferred to a delay line. For generating a continuous and accurate output frequency, a fractional delay filter is required [13]. The characteristic of string stiffness that induces the inharmonicity can be modeled with an allpass filter [7], [14]. A damping filter needed in the waveguide model is designed to fit the tone decay time requirement [4]. All blocks of the string model are important not only for the tone synthesis but for generating the input excitation signals which help to model the attack transience [3], [4]. The components induced by the string of a recorded tone must be canceled as cleanly as possible so that the remainder combining the plucked sound and the impulse response of the body (called an aggregate excitation) can be used generally. This method merges the body response into the original excitation by commuting the order of the string and the body so that it simplifies the whole instrument model. However, one needs to create an independent body model (put after the string) if the length of the string is varied over time to generate a pitch sliding tone for example. Besides increasing the flexibility of the instrument model, the modulation effect to the body response can be avoided [9]. Moreover, when the string length is changed during a pluck period, an integrated transfer function that represents whole string models cannot be used. Otherwise, there must be an obvious dynamic transition in the output sound when a sudden length shift happens. A time-variant glitch-free delay line loop [15] is proposed in the string model to make the pitch slide smoothly.

On the other hand, repeating pluck technique is usually seen in the plucked string instruments. When an already vibrating string is plucked again, it would be damped by the finger temporarily then excited [16]. Several excitations of the typical pluck can be damped in advance then serially combined as a repeated pluck input signal, but if the strings are not perfectly tuned like a real mandolin, using two parallel string models differing in pitch by a few cents and excited at slightly different times may achieve the beating effect [3]. Furthermore, when taking the dynamic variation of each pluck or specific picking pattern into account, the excitation to be generated would become more complex.

B. Prior Works of Pipa Sound Synthesis

The pipa music has been loved by Orientals through the centuries. The special timbre is suitable for both solos and ensembles. In the previous papers, one used nested modulator and feedback FM models to match the harmonics of the pipa tones

[17]. The results sounded pipa-like but still different from the original tones. Moreover, the method lacks for realism and flexibility. A scattering recurrent network method was proposed to implement the physical modeling of plucked string instruments including the pipa. It can generate the parameters of the synthesizer automatically [18]. The signal-to-noise ratio (SNR) of the synthetic tone to the original tone is not quite good because of the high-frequency components error. Besides the typical pluck technique, they tried to model the portamento technique but got a timbre different from the original one. The same authors used another network based on the Kelly–Lochbaum structure with multiple sets of system parameters to simulate the vibrating behavior of a target string [19]. The SNR issue existed again without taking the energy of the harmonics into account. Therefore, using learning algorithms cannot synthesize a plucked string sound efficiently. Basically, it is not a flexible synthesis method either.

An analysis of the pitch perception for inharmonicity [20] provided some important physical structure information and the stiffness characteristics of the pipa strings. For the experimental purpose, a spectral-based additive synthesis was used to control the partial frequency and magnitude accurately. Since the important decay effect was not included, the synthetic tone was not complete, but it proved that the pitch perception is not only affected by the fundamental but also the degree of inharmonicity. Moreover, the quantity of stiffness should be included in the pipa tones synthesis because it directly causes timbre variation.

C. Goals and Organization of the Paper

The playing techniques of Chinese or Asian instruments are versatile. It is very important to investigate their characteristics thoroughly if we want to construct a complete model of the instruments. A few, but important, playing techniques were included in the guqin tone synthesis [9]. However, there is no study of the technique topic in [10] and [19]. The pipa is indeed a good object for research because it owns specific structure and many different playing techniques at the same time. From the computed and physical analysis point of view, the acoustical characteristics of pipa timbres can be perceived. Taking advantage of DWG, a realistic and flexible pipa model can be constructed as well. Meanwhile, this model has no SNR issue because of the physical essence. To save memory usage, the model parameters and input excitations should be minimized. Moreover, a high-efficiency model should be designed for the real-time synthesis purpose. Finally, a pipa synthesizer can be created for easier playing and other applications.

The remainder of this paper is organized as follows. The structure, tuning, and playing techniques of the instrument are discussed in Section II. Section III presents the results of signal analysis of recorded pipa tones to illustrate the characteristics of the timbre of the instrument. In Section IV, the waveguide synthesis algorithm including fixed and varied delay line is introduced. Section V shows the synthetic results compared with recorded ones and the statistics of the listening tests. The discussions are also addressed. Section VI contains the conclusions.

Fig. 1.   Photo of the soloist Luo playing the Chinese instrument pipa.

II. DESCRIPTION OF THE INSTRUMENT

A. Construction and Tuning

The instrument pipa and the typical playing position are shown in Fig. 1. Unlike the guitar, the pipa is played while holding its body vertically. This difference is for the convenience of the left-hand performing techniques. A standard pipa height is 102 cm and the effective string length is 72.5 cm. Consisting of a wooden body like a dissected pear (round in back, flat in front), the front plate of the pipa is a combination of two or three different thickness and hardness wood plates for the consideration of the string decay time, especially for the A3 string which is usually for solo melody. An older paper [21] shows the wood resonant peaks of the pipa body, which are concentrated at the range of 450 Hz to 650 Hz and are part of the higher pipa tone register. This may be an indication of the sound the ancient Chinese favored. A neck is made with six deep and triangular frets, and a tuning peg head extends from the neck. The tuning pegs are quite large to match the body. Besides the neck frets, there are 24 strips of bamboo on the soundboard of the pipa that also function as frets. Each of the 30 total frets is spaced according to well-tempered tuning. Today, the open tuning of the strings is typically A2-D3-E3-A3 (110.5 Hz–147.33 Hz–165.75 Hz–221 Hz), with the highest A being below middle C. The traditional pipa with silk strings and pentatonic tuning was developed into the modern pipa with steel core strings and chromatic tuning during the first half of the last century [22]. Thus, using the real fingernail becomes almost impossible. Instead, a fake nail made of turtle shell or special plastics is usually attached to each finger of the right hand for plucking harder strings.

B. Playing Techniques

The name of pipa is from the most basic playing technique that means forward (pi) and backwards (pa) plucking the strings with the outward fingernails. In general, the playing technique consists of the right-hand fingers plucking the strings and the

left-hand fingers touching the strings in a variety of ways to create melodies. There are over 60 different techniques that have been developed [22]. Basically, a player needs spectacular finger dexterity to handle such complex performances. The typical pluck is with the left hand touching beside the frets. Besides, the strings can be pushed or pulled like the string bending technique in the electric guitar, and twisted or pressed because of the relatively high frets. The wheel, with the right hand rotating all fingers one by one on the strings, is a unique technique to make an unlimited long note like the tremolo effect. Playing this tone has to focus on the dynamics balance of each finger. This technique of the pipa may be harder than the similar one of the guitar because the tones need to be played with high-speed wheeling of five fingers for different length and strength. Others like rolls (similar to the wheel tone but with typical plucking method), slaps, harmonics, and noises are often used as well.

There used to be a large repertoire of pipa music to describe exciting scenes like battles and lyrical themes inspired by poetry, landscapes, and historical stories. Bai Juyi (772–846 AD), one of the great poets in Tang dynasty wrote the most famous poem for playing the pipa named Pipa Song. It describes the shower of pipa notes by [22]: "... The thicker strings rattled like splatters of sudden rain, the thinner ones hummed like a hushed whisper. Together they shaped strands of melody, like larger and smaller pearls falling on a jade plate."

III. ACOUSTIC SIGNAL ANALYSIS

Next, the essential features in the behavior of recorded pipa tones are illustrated. Their prominent patterns in the time and frequency domains such as decay of tones with different conditions and inharmonicity are discussed. The analyses of push and wheel tones are presented as well. A microphone placed at a distance around 1 m in front of the soundboard is used for recording. This distance can get a good tone quality and avoid recording much ambiance noise or body vibration. The signals were recorded with the 44.1-kHz sampling rate and the 16-bit resolution. Since the chamber is not completely anechoic, to clear the pipa sounds, the background disturbance and the improper echo are captured as a noise profile so that they can be canceled from the recorded sounds.

As for the tones, typical scale sets of four strings used in pipa music were recorded. Two different styles for terminating the string with the left-hand finger were used, touching beside the fret or pressing just on the fret. Two different plucking styles were recorded, plucking with nail inward or outward. In addition, a set of harmonic, push, pull, wheel, and vibrato tones was recorded for different strings and frets.

A. Decay Characteristics of Tones

The T_{60} of four open strings of the pipa with typical plucking are around 5.0 s (A2), 4.3 s (D3), 5.3 s (E3), and 5.7 s (A3). It can be found that the decay times are not proportional to the frequency. Besides the effect of the strings used for recording which are Beijing A3 string (only one steel string) plus three other Shanghai A2, D3, and E3 strings (seven steel strands wound together within a helical coil of covering copper) [20], the specific pipa body structure is an important factor for the
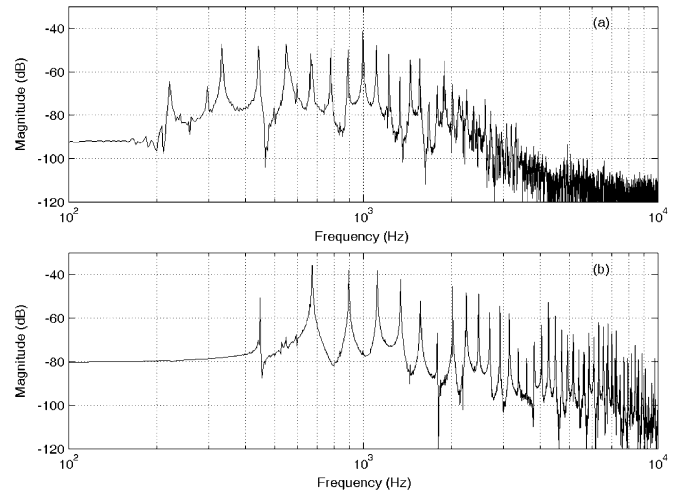

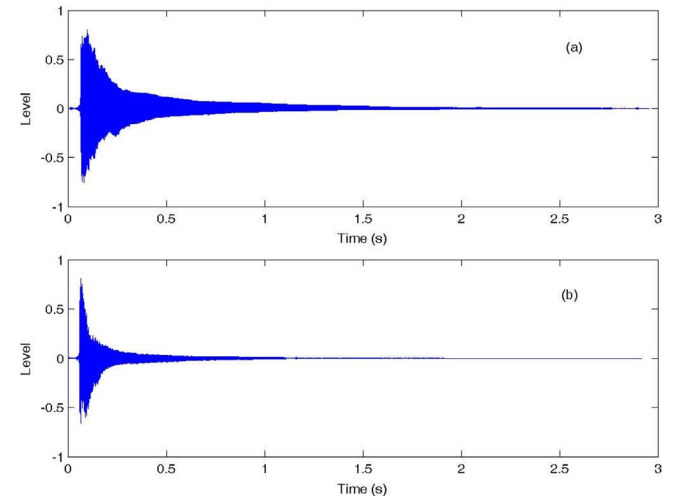Fig. 2.   Spectra of A2 (a) with A3 (b) open string plucked tones.

Fig. 3.   Time responses of pipa notes terminated with the fingertip touching beside the fret (a) and pressing on the fret (b).

tone decay time. While comparing the spectra of A2 with A3 open string plucked tones during the period 0.6 second after the attack as Fig. 2, the higher partials over 2 kHz of A2 tone are much smaller than those of A3 tone. Because of the same string material, two other D3 and E3 strings have a similar characteristic of the A2 string from our observation. This combination makes soft and bright sounds exist at the same time.

On the other hand, the decay times of tones are quite different while the player changes the terminated positions where the fingertip is put. The string vibrations are restrained much more by pressing on the fret than touching beside the fret. Therefore, the tone decays much faster like the example of 370-Hz tones depicted in Fig. 3 and it sounds very crisp. The T_{60} of the tone pressed can be as short as 1.3 s [Fig. 3(b)] compared with the touched one, which is over 4 s [Fig. 3(a)]. Fig. 4 shows the decay period spectra of these two cases, the high-frequency partials beyond the sixth are depressed very much while pressing on the fret.

For some special case such as the double plucking roll tone (plucking two strings at the same time), plucking with outward then inward fingernails fast is required. There is very little difference in the decay time of the tones plucked with either side of
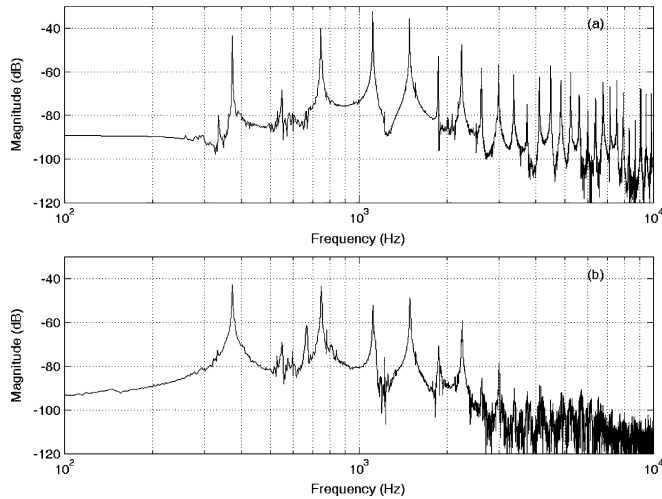
Fig. 4. Spectra of pipa notes terminated with the fingertip touching beside the fret (a) and pressing on the fret (b).

Fig. 5. Estimates of the inharmonicity coefficients of different strings and frets.

TABLE I
DIAMETERS OF SHANGHAI STRINGS FOR [20]

String	A2	D3	E3	A3
Diameter	120.67μm	101.0μm	94.67μm	79.5μm

Fig. 6. Spectrograms of the push tone with (a) the fundamental frequency from 372 Hz to 444 Hz, and (b) the pull tone with the fundamental frequency from 372 Hz to 419 Hz.

the fingernails. Although the dynamic level of the tone with outward fingernail is easily larger than the inward one, the soloists usually adjust the dynamics to equal the levels. Therefore, this effect will not be modeled.

B. Inharmonicity

The stiffness factor of steel strings cannot be ignored in the sound synthesis of the pipa. In order to synthesize the natural timbre and generate general enough input exciting signals of the pipa model, the inharmonicity of pipa tones should be investigated for all strings and frets carefully. The inharmonicity coefficients of the four pipa strings are depicted in Fig. 5 as a function of frequency, on a log-log scale. A partial frequencies deviation method proposed by [23] was used to search the partials of the tone and then evaluate the coefficients. With the help of handed calibration, some wrongly evaluated high partials with lager frequency deviations can be repaired. From the results, two observations can be made. First, the inharmonicity for lower strings is larger than for higher strings. This result can be explained by the inharmonicity coefficient equation as

$$B = \frac{\pi^3 Q d^4}{64 l^2 T} \tag{1}$$

where d is the diameter of the string, l its length, Q is Young's modulus, and T is the tension [13]. A larger diameter will cause the higher inharmonic characteristic. Table I lists the referred diameter values of Shanghai strings. Second, the inharmonicity increases as the length of the string decreases. From (1), the fact can be found that the length decreases while the diameter is fixed.

C. Behavior of Push and Pull Tones

While pushing the pipa string inside or pulling it outside, the tension is increased so that the original plucked pitch will get higher continuously. Basically, the pitch frequency is proportional to the square root of the tension. The interval of the initial and final pitch is flatted thirds or a whole tone typically. The frequency ratios of the intervals are 5:6 and 8:9, respectively.
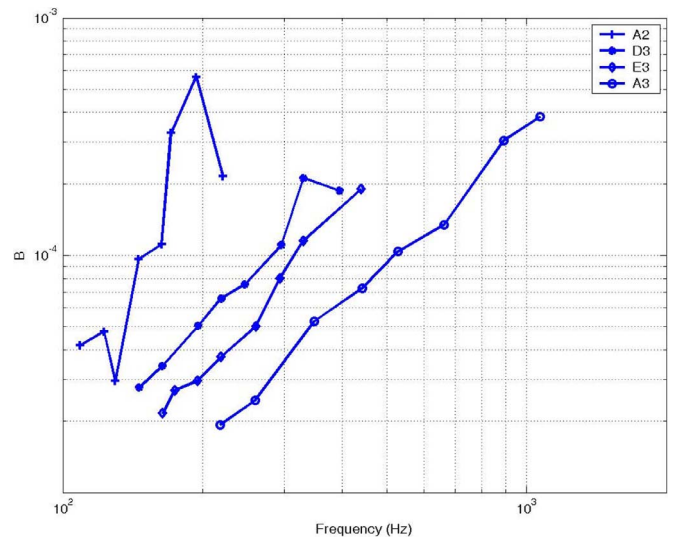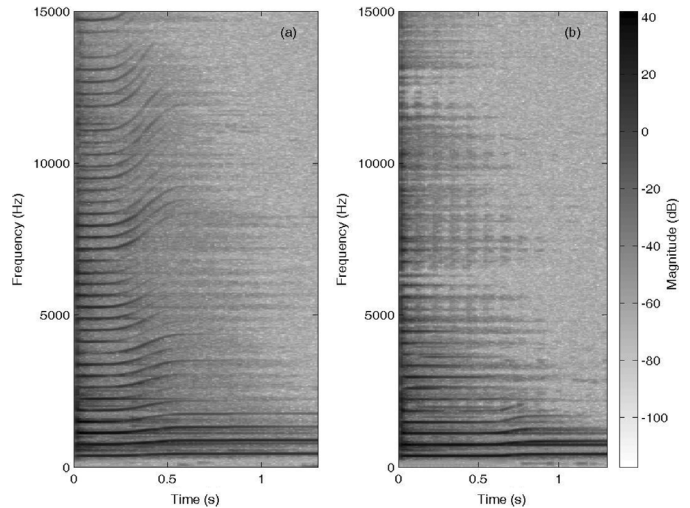
Fig. 6 shows the spectrogram of two push and pull tones. The beginning of pitch slide is around 0.6 s in the pull tone later than the push tone, which is around 0.25 s. The later one seems to make the tone with more poetry. The duration of the pitch sliding is according to the pushing or pulling speed, but since the displacement of the string is small and the technique is usually an attached effect of the main pitch, the duration is usually very short, around 0.23 s. The tone frequency can be also decreased by reversing the string movement.

D. Analysis of Wheel Tone

A wheel tone made by continuously plucking with all fingers is a unique timbre. Fig. 7 shows the five-pluck wheel tone wave-
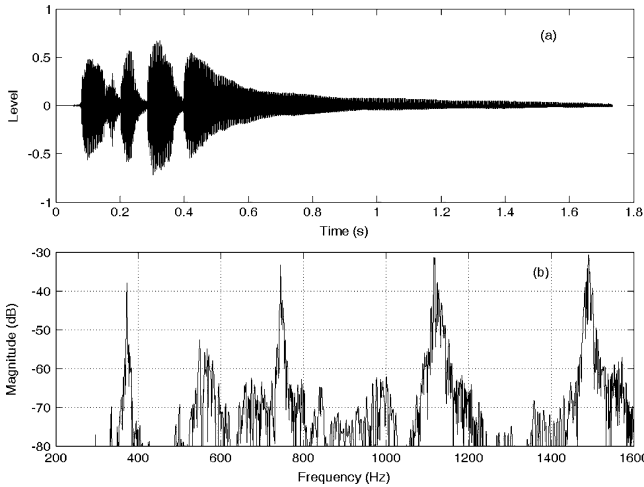
Fig. 7. (a) Waveform and (b) the early four partials spectrum of a wheel tone.
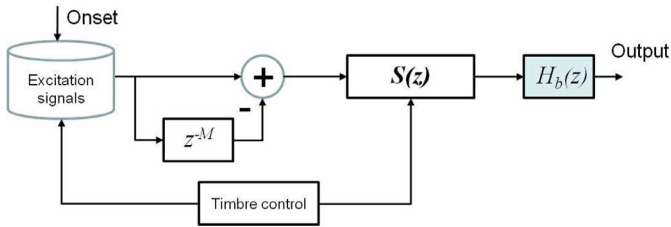

Fig. 8. Block diagram of the pipa synthesizer.

form of the frequency 372 Hz and the earlier four partials spectrum extracted from the whole tone duration fast Fourier transform (FFT) result. The dynamics of each pluck cannot the same in fact. The time interval between two plucks can be smaller than 0.06 s. Moreover, it can be found in the spectrum that each partial is not a pure peak but with a little extension. In fact, not only the mandolin, the strings of the pipa are also never perfectly tuned. With the pluck variation of the different fingers in terms of pluck position and force, the disturbance can even spread a partial peak over 10 Hz.

IV. SOUND SYNTHESIS OF THE PIPA

A. Description of the Synthesis Model

The block diagram of the synthesis model is as Fig. 8. The pipa model essentially is comprised of a delay loop (DL) digital waveguide (DWG) string model, $S(z)$, a body model filter, $H_b(z)$, and a comb filter placed at the input for simulating the effect of plucking position [3]. The input signal is read from the excitation database. The different timbres are selected by the controller which also sends the parameters (according to the timbre) to the string model. Finally, the output signals are sent through a dynamic control to make the tones change smoothly and naturally.

B. String Model and Excitation Signals Generation

The DL string model, $S(z)$, synthesizes the transversal vibrations with the inharmonicity effect shown in Fig. 9. In this model, the z^{-L} block implements the integral delay of the DL. Regarding the sampling theory, L is the integral result of f_s/f,
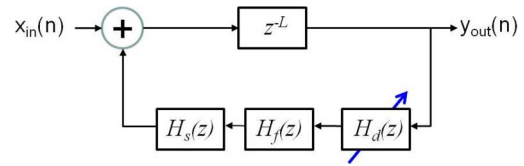

Fig. 9. Block diagram of the string model.

where f_s is sampling frequency and f is the desired note frequency. The remainder fractional part of the loop delay is realized by $H_f(z)$, a third-order Lagrange FIR filter [13]. The coefficients are shown as

$$h(n) = \prod_{k=0,k\neq n}^{N} \frac{D-k}{n-k} \quad \text{for } n = 0,1,2,\ldots\ldots,N$$

(2)

where D is the desired total fractional delay and N is the order of the filter. This filter works as an interpolator. It suffers from high-frequency response degradation. Higher order filters with some help to this issue are increasingly more difficult to compute. After making a tradeoff, the third-order filter is used for the string model. The $H_d(z)$ block is the damping filter implementing the frequency dependent decay due to losses in the string. The filter coefficients are determined with the method illustrated in [4]. The algorithm consists of fitting a straight line to the temporal envelopes of a number of early harmonics then using the slopes of the lines as the attenuation factors for those harmonics. However, the attenuation factors for remaining partials cannot be set to zero but held with gradual degraded values because of A4 string timbre consideration. The damping filter is then generated as a first-order infinite impulse response (IIR) structure. It should also pass a calibration process to make sure not only that the resulting magnitude response does not exceed unity but also that the match is best for the partials whose attenuation rate can be heard easily. Equation (3) is the final version of the damping filter where g_1 is an adjustable parameter for the slight damping variation of different tones or the gain compensation described later. Of course, changing this parameter also can stretch or shorten the decay time of the same tone for a special effect:

$$H_{d1}(z) = \frac{g_1(0.9015 + 0.1275z^{-1})}{1 + 0.0331z^{-1}}.$$

(3)

On the other hand, $H_s(z)$ is the dispersion response for the stiffness characteristic made of a cascade of four first-order allpass filters [7] with the coefficient a_1 computing from the Thiran allpass filter method as

$$a_1 = \frac{1 - D_1}{D_1 + 1}$$

(4)

where the D_1 is the desired delay value at DC. This filter will generate the extra phase delay (i.e., frequency shifting) added on each partial of the tone according to the inharmonicity coefficient B. Using cascaded filters can achieve the accuracy and

the stability at the same time. The values of L used in the integral delay and the remainder fractional part must be modified as well because of the extra delay. Then, the string model blocks can be integrated to one transfer function $S_i(z)$, the computation instruction will be reduced to one time:

$$S_i(z) = \frac{z^{-L}}{1 - z^{-L}H_d(z)H_f(z)H_s(z)}. \tag{5}$$

Once the $S_i(z)$ has been determined, an excitation signal of a string pluck can be generated by putting the recorded signal through the inverse waveguide filter $A(z) = 1/S_i(z)$ [4]. The waveform of the result is then truncated to be an excitation signal as a short burst that dies away rather quickly. A single excitation signal is not enough to synthesize each note of the pipa because many factors, especially different strings and playing techniques, will affect the timbre. As in the previous discussion, except for the A4 string, the other three strings have a weak response at higher frequencies. In the meantime, due to the termination by the fingertip pressing directly on the fret, the high-frequency partials are also greatly filtered out. Therefore, canceling the zero which boosts the high-frequency partials but using just a one-pole filter $H_{d2}(z)$ can achieve the requirement of these two situations

$$H_{d2}(z) = \frac{g_2 0.6502}{1 - 0.3412z^{-1}} \tag{6}$$

where g_2 works as g_1. The exciting signals need to be generated again with the same process but with a different damping filter.

C. Body Model

The body response included in the excitation signal may be modulated while running the string length varied in time. An independent body model is needed to avoid the effect. Using a 2-D mesh digital waveguide can model the body resonator, but the computing complexity becomes another issue [24]. A spectral-based body model [9] instead of the physical model can not only simplify the computing but also maintain the flexibility requirement.

Fig. 10 shows the two spectra representing the pipa body with the average of the excitation signal spectra of open strings in the commuted model and the impulse response by knocking on the front plate. These two diagrams are with similar shapes. This can prove the hypothesis that the plucking impulse is like a white noise so that the excitation signal spectrum can be also treated as a body resonant response. From the two spectra, the resonation of the body can be observed that focuses on the frequency range covering the main register of the pipa. The average excitations example is the better one to be used because it contains some phenomena not included in the knocking plate instance, such as the open strings sympathetic effect. Four cascaded second-order shelf and peak-notch filters are appropriate to implement the response. The frequency response and the coefficients of the filters are shown in Fig. 11 and Table II. The body model is then put in back of the string model and the excitation signal can be whitened by passing through the inversed filter response.
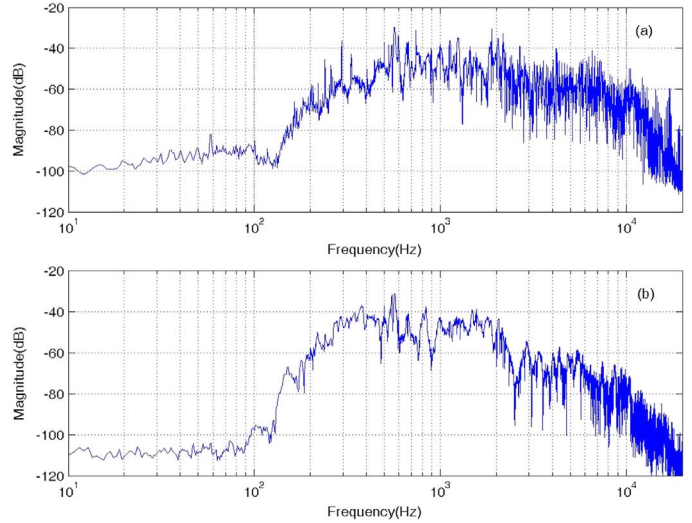

Fig. 10. Spectra representing the pipa body with the (a) average of the excitation spectra of open strings and (b) the impulse response of a knocking sound.
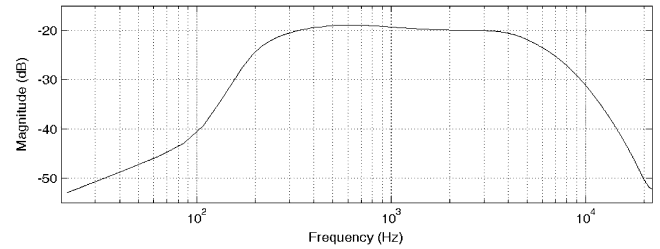

Fig. 11. Frequency response of the body filter.

TABLE II
COEFFICIENTS OF THE FOUR CASCADED BODY FILTERS

Coefficient	$H_{b1(\text{low-shelf})}$	$H_{b2(\text{peak-notch})}$	$H_{b3(\text{peak-notch})}$	$H_{b4(\text{high-shelf})}$
b_0	0.2485	0.5888	0.7701	0.1753
b_1	-0.4937	-0.4687	-0.1817	-0.1708
b_2	0.2453	-0.1190	-0.4351	0.0743
a_0	1	1	1	1
a_1	-1.9747	-1.8746	-0.7269	-1.1763
a_2	0.9755	0.8794	0.3400	0.4913
$H_b = (b_0 + b_1 z^{-1} + b_2 z^{-2})/(a_0 + a_1 z^{-1} + a_2 z^{-2})$				

D. Time Varying String Model

Pushing or pulling the string along the frets during a pluck period can produce a pitch sliding tone. Although this is a motion varying the string tension, for easier interpretation and computation, the timbre is modeled from a varying-length point of view. A string model with the circular buffer delay line is shown in Fig. 12. The length of the buffer is set with a little longer than the maximum string length so that the possible elongation can be accounted for. The string model is then separated into several blocks and run with a one-by-one sample. As a result, the glitch due to the integrated string model can be reduced.

To simplify the computing, a first-order Lagrange filter is used and the four first-order allpass dispersion filters are reduced to one for this case. According to (2), the Lagrange filter with $N = 1$, $h(0) = 1 - D$ and $h(1) = D$ is a linear interpolator. Except for the exact delay of initial and final frequency, the fractional parts (D) can be set as 0.5 and then 1 between two integral delays. After gradually shifting the delay line position, the tone
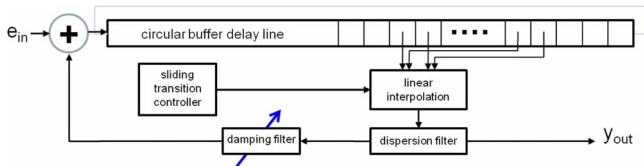
Fig. 12.   String model with circular buffer delay line and linear interpolation.

pitch can be increased or decreased. It saves computing time and avoids a noise induced if only integral jump is adopted as well.

The degradation of high-frequency response is a serious issue while using linear interpolation. To overcome the disadvantage, the damping filter gain [g_1 and g_2 in (3) and (6)] has to be increased to compensate for the additional attenuation. According to [25], the sound energy is also decreased by changing the delay line length. Therefore, an energy enhancement function should be implemented while running the varied delay length as well. According to the experimental result, the gain only needs to be adjusted with five steps distributed averagely before and during the frequency sliding. It is unnecessary to change the gain per sample interval. This method is easy and the tone can be sustained to a long enough length.

E. Wheel Tone and Other Pipa Techniques

For the other techniques of pipa, except the harmonic, some auxiliary algorithms are needed. One example is the frequency modulation technique, which can be used to do the vibrato effect.

Moreover, using excitations combination or parallel string models mentioned in the introduction cannot synthesize a natural pipa wheel tone easily. In order to take the dynamics into account, we use a wheel tone (rotating five fingers one time) as one sample put into the inversed pipa model to get an excitation signal. However, because of the disturbance of each pluck, the inharmonicity coefficient B of the wheel tone is different from that of the same pitch typical pluck tone. After doing some calibration, the higher frequency partials of the recorded tone still cannot be canceled thoroughly. These wrong terms in the excitation signal would be enlarged inadequately while synthesizing the other wheel tone pitch, so that the output energy is accumulated wrongly while the pluck triggers repeatedly. All attack moments could be mixed up and the wheel timbre was consequently weakened in the output result. (Doing offline reduction of the wrong terms can modify the excitation to a certain extent.) This signal is then extended to a required length with a loop computation or reorganized for different pluck speed, and finally sent into the instrument model for a long note synthesis.

V. SYNTHETIC RESULTS

The waveform and the spectrum of the recorded and synthesized signals are compared below. The synthesized results are all from the Matlab simulator.

A. Plucked Tones

Fig. 13 depicts the time responses of synthesized typical pluck (touching beside the fret) and damped (pressing on
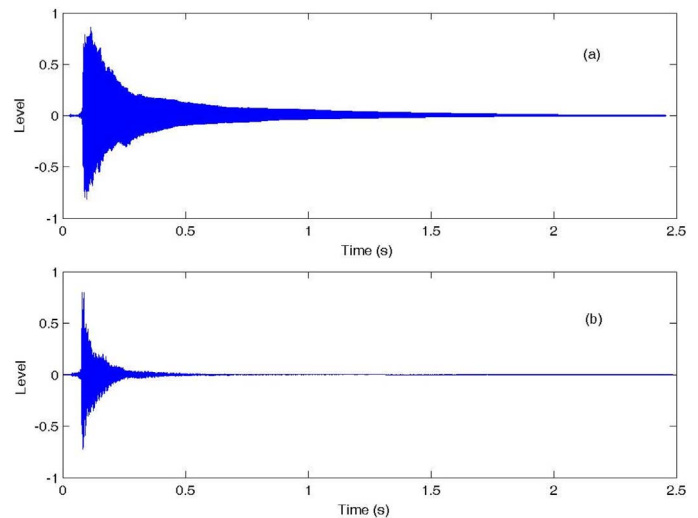

Fig. 13.   Pluck synthetic results with (a) touching beside the fret and (b) pressing on the fret terminations.
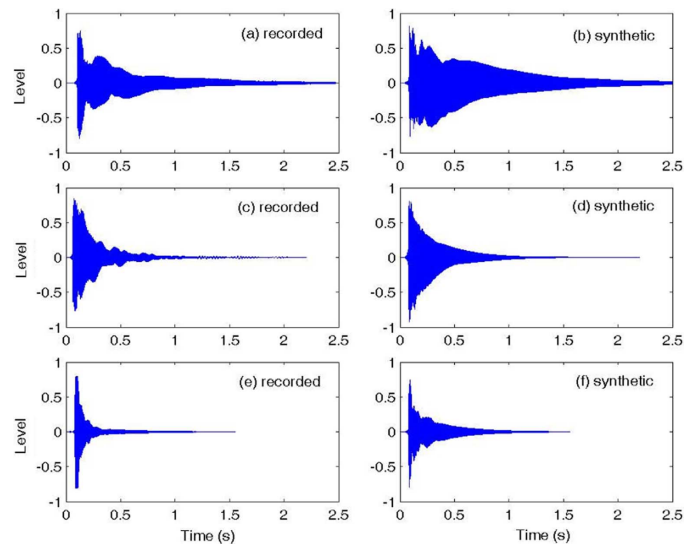

Fig. 14.   Waveforms of synthetic and recorded typical pluck tones of (a), (b) 166 Hz, (c), (d) 660 Hz, and (e), (f) 1109 Hz.

the fret) tones of 372 Hz with A3 string. The decay time responses correspond fairly well (compared with Fig. 3). Since the damped tones are shorter and sharper, it can be found that there is stronger variation in the spectrum of different pitch tones. For the generality, the excitation signal of the pressing tone is averaged from two adjacent pitch tones. On the other hand, due to different types of strings, two general excitation signals of typically plucked events with two types of damping filters are generated for the A3 string and three other strings. Therefore, besides the 372 Hz, three other low-, middle-, and high-frequency tones are also synthesized as test patterns. They are 166 Hz, 660 Hz, and 1109 Hz, respectively. Figs. 14 and 15 illustrate the characteristics of the three synthetic tones along with recorded ones. The attack-decay (AD) curves of the three tones are similar. From the observation of these spectra, the inharmonicity of the synthesis has followed the coefficients extracted from the recorded tones. The amounts of damping at high-frequency match each other as well.
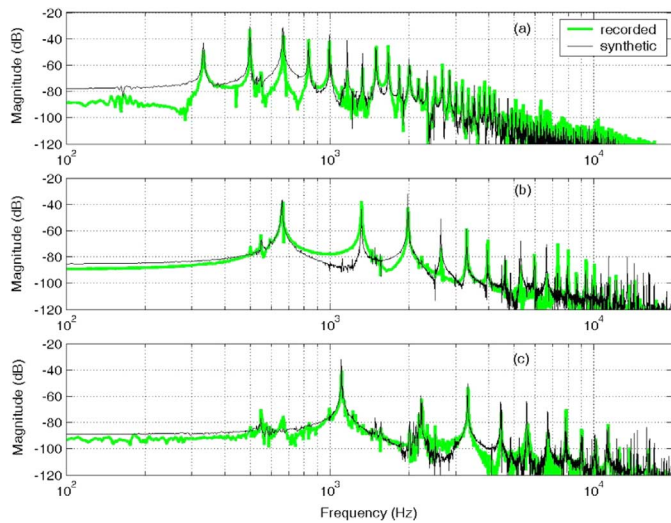
Fig. 15. Spectra of three synthetic and recorded tones of (a) 166 Hz, (b) 660 Hz, and (c) 1109 Hz.
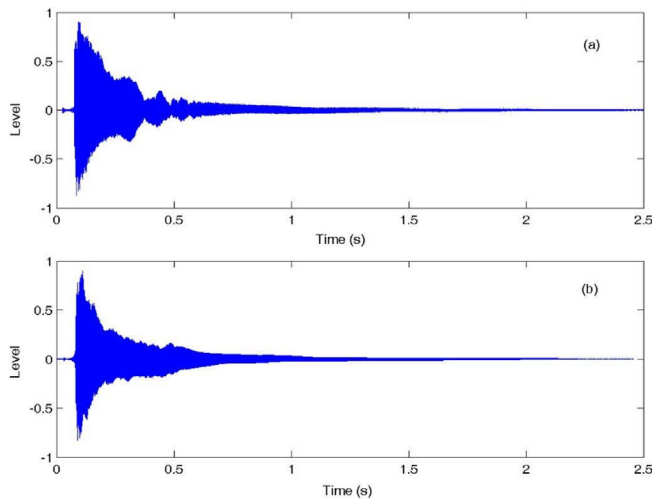

Fig. 16. Waveform of a synthetic push tone (from 370 Hz to 444 Hz) (b) compared with the recorded one (a).

Fig. 17. Spectrogram of the synthetic push tone.

Fig. 18. Spectra of synthetic and recorded push tones (from 370 Hz to 444 Hz).

B. Push Tones

The synthetic push tone is generated by the typical pluck excitation signal compared with the recorded one shown in Fig. 16. If the energy compensation is omitted, the tone would damp very fast. The time response of a synthetic tone is made similar to the recorded one by increasing the damping filter gain initially and also in the decay stage. The total rate of the gain increased is 1.10855 times according to the extra damping. Fig. 17 illustrates the spectrogram of this synthetic tone. Compared with Fig. 6(a), the frequency sliding can be found to be not so smooth, but the duration of the pitch sliding is controlled accurately. Besides, it can be found in the recorded one that the partials of the initial pitch are like remaining partials along with the partials of the sliding pitch. That is why that the larger magnitude of most initial pitch partials in the recorded tone spectrum after the attack also can be observed as in Fig. 18. Fortunately, first-order interpolation and dispersion filters seem to work well on the partial frequencies match.
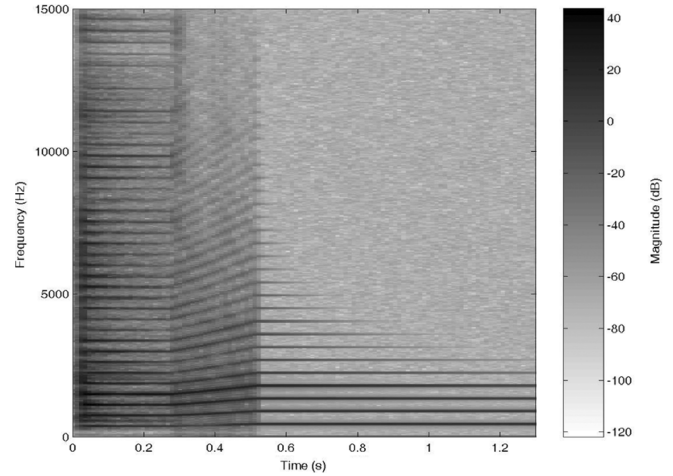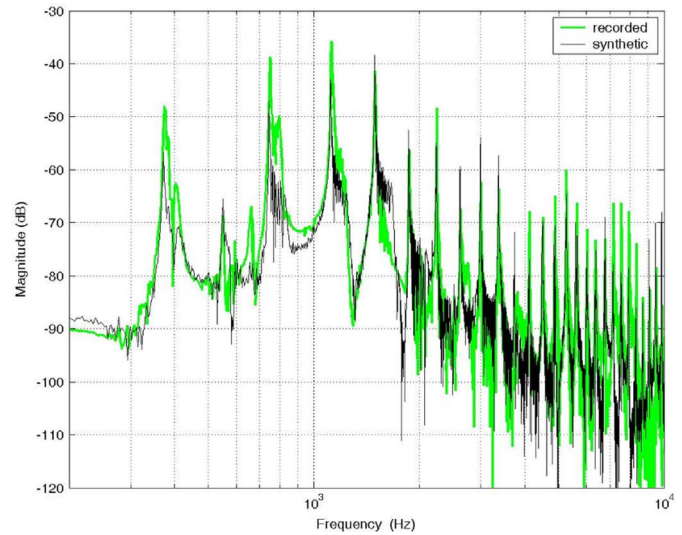
C. Wheel Tones

The synthetic wheel tone 660 Hz compared with the recorded one is shown in Fig. 19. Although the synthetic one still has a phenomenon of the energy accumulation, each pluck can be identified from the waveform and heard clearly from the tone and the tone color is also bright. However, the pluck speed of the two is still not exactly the same. Fortunately, it is not the key point of the tone.

D. Listening Tests

The listening tests are separated into two parts which are the similarity between synthetic and recorded tones and the synthesis acceptability. Any judgment about the synthesis results is also pleased to provide. The SPSS software is use to get the statistic results. Among the total 13 subjects, there are 69.2% who have more than five years music experience and 53.8% who have practiced the piano. Although there are 15.4% listening to Chinese music frequently, the reliability coefficient (Cronbach's α) is 0.856 which shows the good quality of the subjects' timbre discrimination. Tables III and IV list the statistic results with
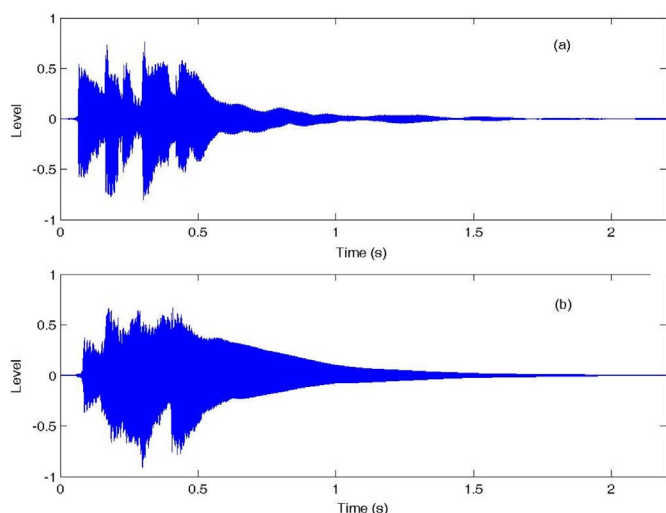
Fig. 19. Waveform of a synthetic wheel tone (660 Hz) (b) compared with the recorded one (a).

TABLE III
STATISTIC RESULTS OF THE SIMILARITY

Similarity	Typical pluck tone (five pitches)	Damped tone (one pitch)	Push tone (one pitch)	Wheel tone (two pitches)
Mean	4.46	3.15	3.38	3.85
Std. Deviation	0.776	1.068	0.870	0.801

TABLE IV
STATISTIC RESULTS OF THE ACCEPTABILITY

Acceptability	Typical pluck tone	Damped tone	Push tone	Wheel tone
Mean	4.54	3.31	3.54	3.77
Std. Deviation	0.660	1.032	0.967	1.013

Likert 5-point scale. It can be found that the pipa tones synthesis has got an acceptable estimation, especially for the typical pluck tone. The reasons for some testing results not quite good are discussed in next section. All tones for listening tests are available on the Internet at http://web2.cc.nctu.edu.tw/~ea-music/music_tech_lab/pipa/pipa.html.

E. Discussion

The recorded tones are played by human fingers while the synthetic ones are generated with few general excitation signals, there are more variations in the former. Therefore, there are some variances between the synthetic and recorded tones in the magnitude and the level of the spectra and the waveforms as shown in Figs. 14 and 15, but because the model has included most acoustic characteristics of the pipa, there are not big differences in auditory timbre between the two. However, a minor common comment of the listening test subjects is about the timbre of higher pitch synthetic tones (660 Hz and 1109 Hz). Therefore, increasing the number of excitations could be a compromised solution if needed. The method may help for other techniques as well as for the typical plucked tones.

Another issue happening in push tones is also discussed here. Because of the remaining partials of the initial pitch, the timbre of the recorded one is more saturated or fuller than that of the

synthetic one after the pitch sliding. It's unlike the nonlinear mixing partials occurring because of the longitudinal string vibration [26], but to solve the issue, the string model may still need to be divided into two paths for generating different partials (which is similar to the guqin model [9]). Or the resonant characteristic of the pipa body should be further investigated so that it can preserve (reverberate) the initial pitch waves longer. Besides, although the just-noticeable-difference (JND) for pitch change is typically around 1/10th of a semitone [27], we almost cannot distinguish the continuous frequency varying due to the short process time and the small interval of the pitch sliding. Therefore, each shift interval with a 0.5 delay length is enough for synthesis.

Regarding the wheel tone, since the attack transient is a very important characteristic for the timbre of this tone, continuing to search out other ways to get better excitation signals may be required.

Finally, more samples should be provided to the subjects while doing the listening test. Since there is only one damped tone sample, a larger standard deviation compared with others is shown in the statistical result.

VI. CONCLUSION

The acoustic characteristics and a synthetic method of performing the traditional Chinese instrument pipa are presented in this paper. The proposed synthesis model has demonstrated the important techniques of the instrument (such as plucking with two termination types, push, pull, and wheel) successfully. The results can capture the prime features of the pipa, and the insufficient parts are discussed as well. More important, the model has been proved efficient enough for computation and provides high flexibility to let people play simulated and even surreal pipa tones with this synthesis technique in real-time. Of course, this prototype model can be improved further for the timbre quality, and more playing techniques should be covered. In addition, as long as the excitation database is completed for all kinds of timbre, a complete pipa synthesizer based on the model can be achieved with an attached control interface.

REFERENCES

[1] J. O. Smith III, "Physical modeling using digital waveguides," *Comput. Music J.*, vol. 16, no. 4, pp. 74–91, 1992.

[2] J. O. Smith III, *Physical Audio Signal Processing for Virtual Musical Instruments and Audio Effects* W3K, Aug. 2007 [Online]. Available: http://www.w3k.org/

[3] D. Jaffe and J. Smith, "Extensions of the Karplus-strong plucked-string algorithm," *Comput. Music J.*, vol. 7, no. 2, pp. 56–69.

[4] M. Karjalainen, V. Välimäki, and Z. Janosy, "Towards high-quality sound synthesis of guitar and string instruments," in *Proc. Int. Comput. Music Conf.*, Tokyo, Japan, Sep. 1993, pp. 56–63 [Online]. Available: http://www.acoustics.hut.fi/~vpv/publications/icmc93-guitar.htm

[5] G. D. Scavone, "Digital waveguide modeling of the non-linear excitation of single-reed woodwind instruments," in *Proc. Int. Comput. Music Conf.*, 1995, pp. 512–524.

[6] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Jánosy, "Physical modeling of plucked string instruments with application to real-time sound synthesis," *J. Audio Eng. Soc.*, vol. 44, no. 5, pp. 331–353, May 1996.

[7] J. Rauhala and V. Välimäki, "Tunable dispersion filter design for piano synthesis," *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 253–256, May 2006.

[8] K. Karplus and A. Strong, "Digital synthesis of pluck-string and drum timbres," *Comput. Music J.*, vol. 7, no. 2, pp. 43–55, 1983.

[9] H. Penttinen *et al.*, "Model-based sound synthesis of the guqin," *J. Acoust. Soc. Amer.*, vol. 120, no. 6, pp. 4052–4063, Dec. 2006.

[10] S.-J. Cho, U.-P. Chong, and S.-B. Cho, "Synthesis of the Dan Trahn based on a parameter extraction system," *J. Audio Eng. Soc.*, vol. 58, no. 6, pp. 498–507, Jun. 2010.

[11] V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen, "Discrete-time modelling of musical instruments," *Rep. Progr. Phys.*, vol. 69, no. 1, pp. 1–78, Jan. 2006.

[12] M. Laurson, V. Norilo, and M. Kuuskankare, "PWGLSynth: A visual synthesis language for virtual instrument design and control," *Comput. Music J.*, vol. 29, pp. 29–41, 2005.

[13] V. Välimäki, "Discrete-time modeling of acoustic tubes using fractional delay filters," Dr. Tech. thesis, Lab. of Acoust. and Audio Signal Process., Helsinki Univ. of Technol., Helsinki, Finland, 1995, Report 37.

[14] J. S. Abel, V. Välimäki, and J. O. Smith, "Robust, efficient design of all-pass filters for dispersive string sound synthesis," *IEEE Signal Process. Lett.*, vol. 17, no. 4, pp. 406–409, Apr. 2010.

[15] S. A. Van Duyne, D. A. Jaffe, G. P. Scandalis, and T. S. Stilson, "A lossless, click-free, pitchbend-able delay line loop interpolation scheme," in *Proc. Int. Comput. Music Conf.*, Thessaloniki, Greece, Sep. 1997, pp. 252–255.

[16] C. Erkut, V. Valimaki, M. Karjalainen, and M. Laurson, "Extraction of physical and expressive parameters for model-based sound synthesis of the classical guitar," in *AES108th Conv.*, Paris, France, Feb. 2000, preprint no. 5114.

[17] A. Horner, "Nested modulator and feedback FM matching of instrument tones," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 4, pp. 398–409, Jul. 1998.

[18] A. W. Y. Su and S.-F. Liang, "A class of physical modeling recurrent networks for analysis/synthesis of plucked string instruments," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1137–1148, Sep. 2002.

[19] S.-F. Liang and A. W. Y. Su, "Modeling and analysis of acoustic musical strings using Kelly-Lochbaum lattice networks," *J. Inf. Sci. Eng.*, vol. 20, pp. 1161–1182, 2004.

[20] S. Hui, L. Chin, and J. Berger, "Analysis of pitch perception of inharmonicity in pipa string using response surface methodology," *J. New Music Res.*, vol. 39, no. 1, pp. 63–73, 2010.

[21] F. Liu [Online]. Available: http://www.liufangmusic.net/

[22] S. Feng, "Some acoustical measurements on the Chinese musical instrument p'i-p'a," *J. Acoust. Soc. Amer.*, vol. 75, no. 2, pp. 599–602, 1984.

[23] J. Rauhala, H. M. Lehtonen, and V. Välimäki, "Fast automatic inharmonicity estimation algorithm," *J. Acoust. Soc. Amer.*, vol. 121, no. 5, pp. EL184–EL189, 2007.

[24] S. A. Van Duyne and J. O. Smith III, "Physical modeling with the 2-D digital waveguide mesh," in *Proc. Int. Comput. Music Conf.*, Tokyo, Japan, Sep. 1993, pp. 40–47.

[25] J. Pakarinen, M. Karjalainen, V. Valimaki, and S. Bilbao, "Energy behavior in time-varying fractional delay filters for physical modeling synthesis of musical instruments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 18–23, 2005, vol. 3, pp. III/1–III/4.

[26] H. A. Conklin, "Generation of partials due to nonlinear mixing in a stringed instrument," *J. Acoust. Soc. Amer.*, vol. 105, no. 1, pp. 536–545, 1999.

[27] T. D. Rossing, *The Science of Sound*, 2nd ed. Reading, MA: Addison-Wesley, 1990.

Yi-Huei Chen was born in Taiwan in 1973. She received the B.S. and M.S. degrees in electrical engineering from National Central University, Jhongli City, Taiwan, in 1996 and 1998, respectively, and the M.A. degree from National Chiao Tung University, Hsinchu, Taiwan, in 2011.

From 1998 to 2007, she had been with various research institute and IC design companies, working on the RF and mixed-signal IC design for wireless communication applications. After 2007, she started to study at the Institute of Music, National Chiao Tung University. Her major research is the music technology including sound synthesis of instruments and electroacoustic music composition.

Chih-Fang Huang was born in Taiwan in 1965. He received the M.A. degree in music composition and the Ph.D. degree in mechanical engineering from National Chiao Tung University, Hsinchu, Taiwan, in 2001 and 2003, respectively.

He is an Assistant Professor in the Department of Information Communication, Yuan Ze University, Jhongli City, Taiwan, and also currently the Chairman of Taiwan Computer Music Association (TCMA). His research papers include the fields such as automated music composition, virtual reality, and the intermedia integration.