

國立交通大學
音樂研究所 音樂科技組
碩士論文

以物理模型為方法之琵琶聲音合成

**PHYSICAL MODEL BASED SOUND SYNTHESIS
OF THE PIPA**

研究生：陳宜惠

指導教授：黃志方

曾毓忠

中華民國一百年一月

以物理模型為方法之琵琶聲音合成

PHYSICAL MODEL BASED SOUND SYNTHESIS
OF THE PIPA

研究生：陳宜惠 Student：Yi-Huei Chen

指導教授：黃志方/曾毓忠 Advisor：Chih-Fang Huang/Yu-Chung Tseng

國立交通大學

音樂研究所 音樂科技組

碩士論文

A THESIS SUBMITTED TO
THE INSTITUTE OF MUSIC
COLLEGE OF HUMANITIES AND SOCIAL SCIENCES
NATIONAL CHIAO TUNG UNIVERSITY
IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
(MUSIC TECHNOLOGY)

HSINCHU, TAIWAN

JANUARY 2011

中華民國一百年一月

以物理模型為方法之琵琶聲音合成

學生：陳宜惠

指導教授：黃志方博士

曾毓忠博士

國立交通大學 音樂研究所 音樂科技組

摘要

本論文欲對琵琶樂器的物理模型合成法做一探討。由於數位信號處理的發展，我們可以直接使用數位波導 (Digital waveguide) 的方式來實現物理模型，多種樂器的物理模型合成已經有相當長時間的發展，許多西方管弦樂器的音色已經成功地實現，包括了提琴、吉他、鋼琴等，在近幾年，中國的古琴，也有物理模型的文獻報告。運用此合成方法不但可以大大減少記憶體的使用，且可以趨近錄音樣本一樣，呈現出樂器獨特的音色。

琵琶在中國屬於歷史悠久的樂器，其本身音色即非常具有特色，加上種類繁多的運音指法，更讓其琵琶成為中國代表樂器之一，目前琵琶按照十二平均律調音，標準的琵琶有四弦六相二十四品，除了品相的增加，在琵琶制作方面，原來用的絲質弦也改成了尼龍或銅包覆的鋼絲弦，加大了琵琶音量和共鳴。

根據琵琶的錄音分析所得到的琴弦聲學特性會以物理模型來實現，另外根據不同的琴弦特性及運指法製做所需的輸入刺激信號；琴身的共鳴腔模型則以其在頻譜上的特徵來設計，經由模擬所得到的琵琶音色合成結果與錄音結果比較，發現本論文所提出的模型可以幾乎即時的合成出非常接近琵琶若干指法所演奏出的音色，其中包括彈挑、悶音、輪指、及推拉等，而透過參數的控制，一些真實琵琶所達不到的音色，也可以容易的被合成出來，最終希望能建立一個仿真且有效率的虛擬樂器。

Physical Model Based Sound Synthesis of the Pipa

Student: Yi-Huei Chen

Advisor : Dr. Chih-Fang Huang

Dr. Yu-Chung Tseng

Institute of Music
National Chiao Tung University

Abstract

This thesis proposes sounds analysis and a synthesis model for the Chinese plucked string instrument called pipa, one of the oldest Chinese musical instruments with over 2000 years of history. 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. 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.

致 謝

謝謝我最愛的家人們；謝謝這幾年在我身邊的朋友們；謝謝所有音樂所及聲音與音樂創意科技碩士學位學程的師長及同學們；因為你們的幫忙，讓我能順利完成人生中第二個碩士學業和實現一個夢想，真的非常感激。

Acknowledgements

I am thankful to Professors Chih-Fang Huang, Phil Winsor, and Yu-Chung Tseng for their advice and for challenging my thinking over the past three years. I would like to appreciate the demonstrative recording from two pipa soloists Chao Yun Luo and Ming Fang Chen. Dr. S. Van Duyne, Professors Pao-Chi Chang and Yi-Wen Liu are thanked for their helpful comments. Mr. Wei-Gang Hong is also thanked for his work of SPSS data collection and analysis.

Contents

1. Introduction	1
1.1 Background	1
1.2 Prior Works of Pipa Sound Synthesis	3
1.3 Goals and Organization of the Thesis	3
2. Model-Based Plucked String Sound Synthesis Overview	5
2.1 Digital Waveguide Modeling	5
2.2 Physical Modeling Implementation Using Digital Filtering Techniques	6
2.3 Fractional Delay Filter Implementation	10
2.4 Stiffness Dispersion Filter Design	13
2.5 Damping Filter Design	14
2.6 Excitation Signals and Body Modeling	15
3. Description of the Pipa Instrument	17
3.1 Construction and Tuning	17
3.2 Playing Techniques	19
4. Acoustic Signal Analyses	21
4.1 Decay Characteristics of Tones	21
4.2 Inharmonicity	24
4.3 Behavior of Push and Pull Tones	25
4.4 Analysis of Wheel Tone	27
5. Sound Synthesis of the Pipa	28
5.1 Description of the Synthesis Model	28
5.2 String Model and Excitation Signals Generation	28
5.3 Body Model	31

5.4 Time-Variant String Model	33
5.5 Wheel Tone and Other Pipa Techniques Models.....	34
6. Synthetic Results	36
6.1 Plucked Tones	36
6.2 Push Tones	38
6.3 Wheel Tones	40
6.4 Listening Tests	41
6.5 Discussions	42
7. Conclusions and Future Works	44
References	45

Chapter 1

Introduction

1.1 Background

Due to the rapidly 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]. The research results from CCRMA in Stanford University [8] and the Acoustic Lab in Helsinki University [9] have become very important for this field. Among the instruments, the physical modeling of plucked strings is the earliest one to be developed. From the Karplus-Strong algorithm was proposed [10], besides the western instruments like the guitar and mandolin, the eastern ones like the guqin [11], dan trahn [12] 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 [13], [14]. 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 string material, the body structure, and the playing techniques. 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 [15]. The characteristic of string stiffness that induces the inharmonicity can be modeled with an allpass filter [7], [16]. 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]. Therefore, the components induced by the string of a recorded tone must be cancelled as cleanly as possible so that the remainder combining the plucked sound and the impulse response of the body can be used generally.

If the whole instrument model is a time-invariant system, an integrated transfer function representing the string model can be used to save the computation time. Merging the body response in the excitation to be an aggregate one also simplifies the model [2]. However, an independent body model needs to be created if the length of the string is varied over time. Besides increasing the flexibility of the instrument model and avoiding the modulation effect to the body response, the excitation signals can therefore be whitened further [11]. Moreover, when the string length is changed during a pluck period for generating a pitch sliding tone for example, the integrated string model 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 [17] 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 and then excited [18]. Several damped excitations of the typical pluck can be 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, if taking the dynamic variation of each pluck or specific picking pattern into account, the excitation would become more complex to be generated.

1.2 Prior Works of Pipa Sound Synthesis

The pipa music has been loved by Orientals through 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 [19]. The results were sounded pipa-like but still different from the original tones. Moreover, the method lacked 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 [20]. The SNR of the synthetic tone to the original tone is not quite good because of the high frequency part 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 [21]. The SNR issue existed again without taking the energy of the harmonics into account. These learning algorithms need recurrent computation for training the system so that the complexity cannot be reduced.

An analysis of the pitch perception for inharmonicity [22] provided some important physical structure information and the stiffness characteristics of the pipa strings. For the experimental purpose, an 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. Therefore the inharmonicity effect for pipa tones definitely should be included in the instrument sound synthesis model.

1.3 Goals and Organization of the Thesis

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 [11]. But there is no study of the technique topic in [12] and [21]. 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. 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 more other applications.

The remainder of this thesis is organized as follows. In Chapter 2, an overview of digital waveguide and digital filter design related to the plucked string sound synthesis is presented. The spectrum-based but not model-based method adopted only to model the body response is explained as well. The structure, tuning, and playing techniques of the pipa are discussed in Chapter 3. Chapter 4 presents the results of signal analysis of recorded pipa tones to illustrate the characteristics of the timbre of the instrument. In Chapter 5, the waveguide synthesis algorithm including fixed and varied delay line is introduced. Besides the string model, generating the body model and the input excitation signal is also included. Chapter 6 shows the synthetic results compared with recorded ones and the statistics of the listening tests. The discussions are also addressed. The last Chapter contains the conclusions and future works.

Chapter 2

Model-Based Plucked String Sound Synthesis Overview

2.1 Digital Waveguide Modeling

The solution of the one dimension lossless wave equation can be expressed as a combination of a right-going and a left-going traveling wave [2]. After sampled the waves, the vibration of a string can be modeled simply using a digital waveguide with two delay lines. A digital simulation diagram for a rigidly terminated ideal string is shown in Fig. 2.1. The value of N is defined by $2L/X$ where the L is the string length. The spatial sampling interval $X = cT$ with traveling speed c and time sampling interval T . The waves reflecting at either end of the string due to rigidly terminations can be modeled by negating each sample after it reaches the end of a delay line, before feeding it into the next delay line. By summing the values of two delay lines at a certain location along the delay lines, a total output displacement can be obtained.

Fig. 2.2 shows an initial excitation of an ideal pluck string with doubly rigid terminations. The delay elements are initialized with a shape corresponding to the initial displacement of the string. Using a smooth triangular shape but not a sharp corner is for avoiding the aliasing when sampled. The length of the delay lines controls the frequency of oscillation, and consequently the pitch of the output signal. This changes the wavelength of the travelling waves, which in turn changes the pitch of the sound. Therefore, if the desired frequency of the output is f and the sampling frequency is f_s , then the N mentioned before equals the value of f_s/f .

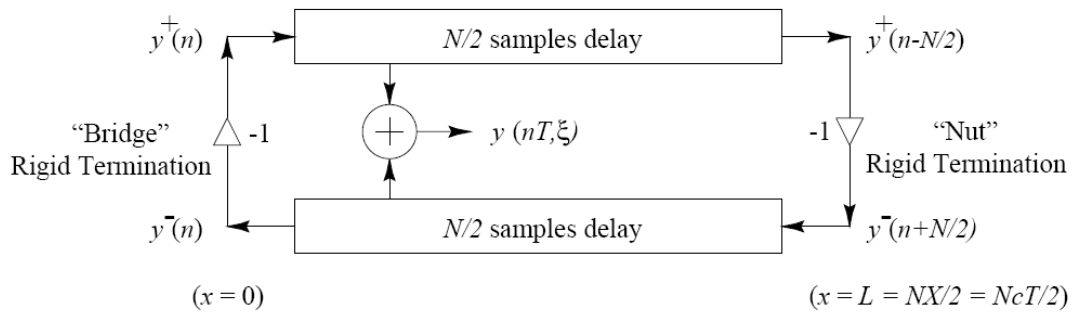


Figure 2.1: The rigidly terminated ideal string with a displacement output indicated at position $x=\xi$ [1].

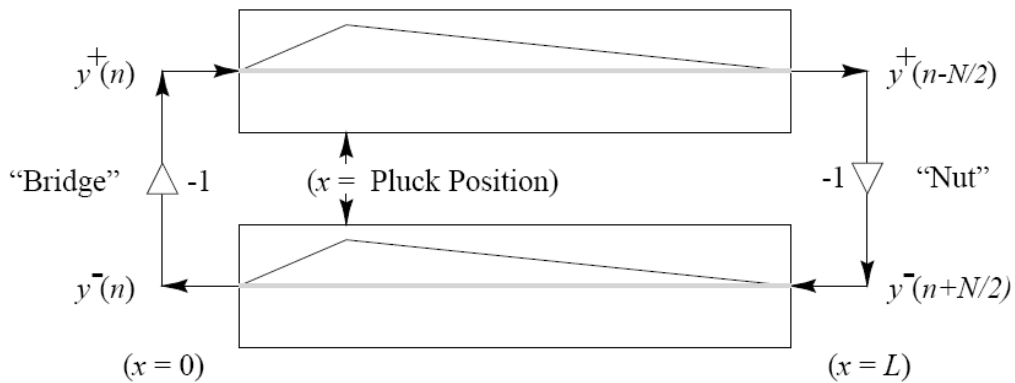


Figure 2.2: Digital waveguide with initial conditions of delay lines set to triangular waves. The sum of the upper and lower delay lines gives the actual initial string displacement [1].

2.2 Physical Modeling Implementation Using Digital Filtering Techniques

To simplify the implementation of the waveguide, the two delay lines can be combined into one as shown in Fig. 2.3. The negative multipliers cancel each other out, and the two delay lines can be combined leaving only a length N delay line. In the real world, due to friction and air resistance, the amplitude of the string vibrations decay over time, so it is important to model this effect in the digital waveguide. To attenuate the output, a simple damping factor g ($|g| \leq 1$) is added with each delay in Fig. 2.1. So that the values are damped before being fed

into the other delay line. The total N damping factors can also be lumped together into one damping factor g^N to construct a simple loop with N samples delay.

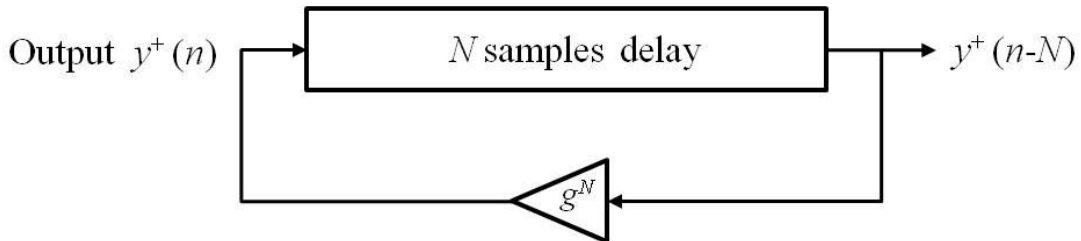


Figure 2.3: Simplified digital waveguide after combining delay lines and damping factors [1].

Besides, the damping of real vibrating strings is typically with a characteristic that increases with frequency for a variety of physical reasons. Therefore, for further realism the lumped damping factor is replaced by a filter that damps each frequency differently. This loop filter always has a low pass characteristic. Fig. 2.4 is the simplest frequency dependent loss filter proposed in the Karplus-Strong algorithm [10]. This loop filter is a single zero FIR filter that averages the N th and $N-1$ th sample. The difference equation between the input and the output is:

$$y(n) = 0.5 \cdot (y(n - N) + y(n - N - 1)) \quad (2.1)$$

The original Karplus-Strong algorithm is based on the wavetable synthesis. The table is made by a specific length of buffer filled with samples and then read by a pointer from the beginning of the table to the end and back to the beginning. It finally creates a periodic sound. The way is very simple but unfortunately the sound doesn't present any variation over time. Most synthesis techniques remedy to this situation by modifying the sound after the sound synthesis [23]. The Karplus-Strong algorithm differs from those techniques because it directly modifies the wavetable at each of its iteration, so that it can therefore be seen as a delay-line.

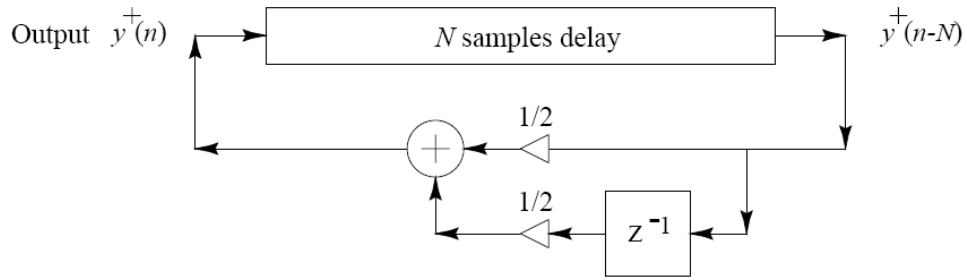


Figure 2.4: Simplified digital waveguide after combining delay lines and damping factors [1].

In the Karplus-Strong algorithm, the pluck, which in real string can be considered to contain energy at any frequency, is simulated by filling the delay-line with random noise at each note beginning. At the end, the output will almost be a periodic waveform corresponding to the fundamental frequency of the string. The non-harmonic elements are depressed a lot at the end with introducing the average of two last outputs in the loop again and again which also decays higher frequencies more than lower frequencies. Like in the case of a real string, the delay-line contents finally decay to a small value to silence the sound.

One problem with implementing this system is that the size of the delay lines is an integer. If we wish to always use a set sampling frequency, then the delay line lengths will not always be integers. Besides, the characteristic of string stiffness induces the higher partials to disperse which can affect both music timbre and pitch perception, for example, a moderate amount of inharmonicity provide a sense of warmth. The prior research [22] has proposed a systematic method for measuring the threshold at which inharmonicity affects perceived pitch. To solve these issues, Fig. 2.5 shows a block diagram of the extended Karplus-Strong algorithm and Table 2.1 is the description of the model respectively [3]. The fractional delay filter and the stiffness dispersion filter in the model are discussed in the next sections. In this extended model, pluck direction and position are included.

Another disadvantage in Karplus-Strong model which is no dynamics variation of the output wave due to fixed initial noise amplitude is considered as well. In a real string, the pluck force not only varies the amplitude but also the energy of the higher frequency content. A hard pluck usually creates a sound with more energy in the higher frequency range than a soft pluck which is a non-linearity phenomenon [23]. If varying the input noise amplitude only, it creates an effect closer to a variation of the source location than that of a variation in the plucking force. Therefore, a variable bandwidth low pass filter is used to achieve the dynamic level adjustment. However, after adjusting the source, the output levels become different at all pitch frequencies. A level control filter is put at the end of the string model to balance the dynamics of different pitch tones.

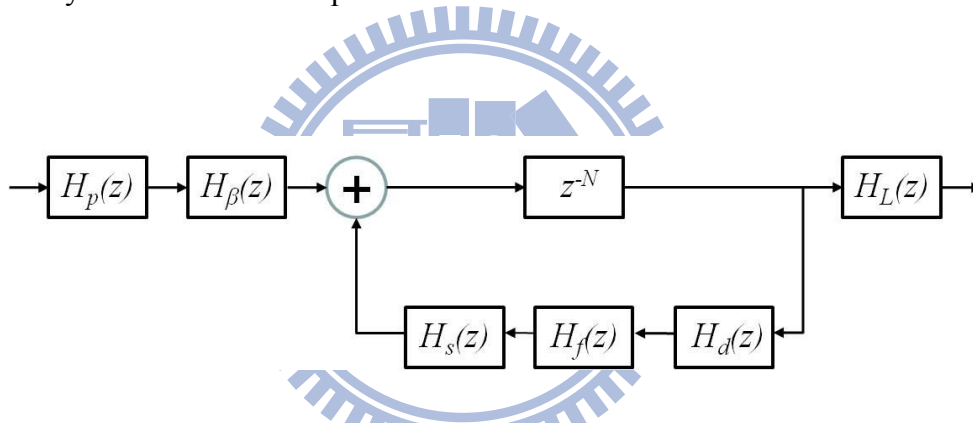


Figure 2.5: Simplified digital waveguide after combining delay lines and damping factors.

Table 2.1
Filter description of the extended Karplus-Strong model

<i>Filter</i>	<i>Description</i>
$H_p(z)$	Pick-direction low pass filter
$H_s(z)$	Pick-position comb filter
$H_d(z)$	String-damping filter (one/two poles/zeros typical)
$H_f(z)$	String tuning (fractional delay) filter
$H_i(z)$	String-stiffness allpass filter (several poles and zeros)
$H_L(z)$	Dynamic-level lowpass filter

2.3 Fractional Delay Filter Implementation

A non-integral number of samples long loop delays is very important for modeling a string system. Fortunately, some filtering techniques can be adopted to implement the fractional delay function in time-sampled environment [24]. The linear interpolation is the easiest and inexpensive way of the finite impulse response (FIR) forms which effectively draws a straight line between two neighboring samples and returns the appropriate point along that line. It straightforward deals with output signals only and the results are very good when the signal bandwidth is small compared with half the sampling rate. The difference equation is depicted as

$$y(n+D) = (1-D) \cdot y(n) + D \cdot y(n+1) \quad (2.2)$$

where D is the desired fraction delay. The magnitude and phase delay response of a linear interpolator for eleven different fractional delay values ($D = 0, 0.1, 0.2, \dots, 1.0$) are shown in Fig 2.6. The phase delay gives the time delay in sample intervals experienced by each sinusoidal component of the input signal [25]. For all fractional delays, the accuracy is higher at low frequencies and zero at DC. Note that there are only six different curves in the upper figure (not eleven), because the magnitude responses for fractional delays d and $1-d$ are the same [15]. When $D = 0.5$, it can be found that the gain degrade seriously at the high frequency which is the same effect of Karplus-Strong model. Therefore, it would cause extra output loss if used in the string model with a loop damping filter.

A first-order infinite impulse response (IIR) allpass interpolation is sometimes a better choice since it costs almost the same as linear interpolation in the first-order case and has no gain distortion. But it operates with both input and output signals depicted as the (2.3) so that needs more response time than the FIR form.

$$y(n) = a \cdot x(n) + x(n-1) - a \cdot y(n-1) \quad (2.3)$$

The value a is around the value $(1-D)/(1+D)$. The phase delay of the allpass filter for a variety of desired delays at DC is shown in Fig. 2.4. Since the amplitude response of any allpass is 1 at all frequencies, there is no need to plot it.

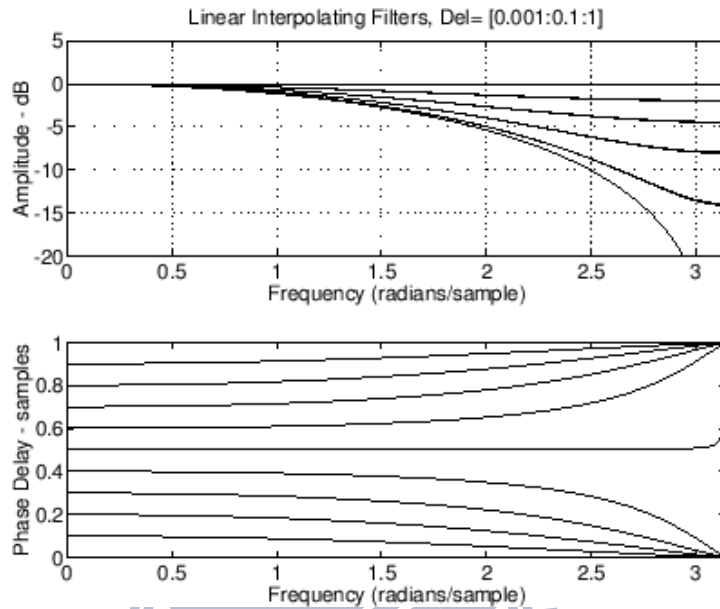


Figure 2.6: Frequency and phase responses of the linear interpolator [2].

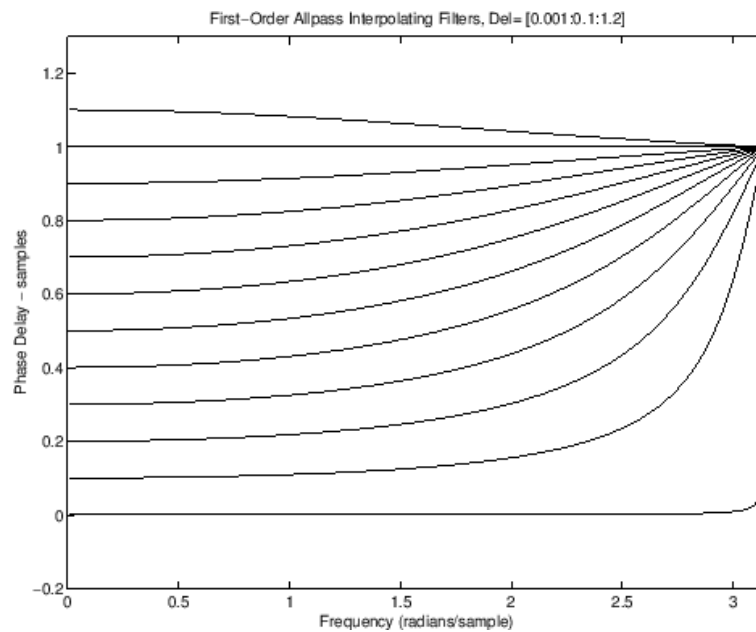


Figure 2.7: Phase response of a first-order allpass filter [2].

Comparing the two phase responses of the linear interpolation and first-order allpass filter, the characteristics at low frequency are similar which can achieve the fractional delay accurately. However, when the frequency increases, the phase delays diverge from different directions.

Since the linear interpolator suffers from high frequency response degradation, higher order FIR filter can also be used to improve some. A useful FIR filter approximation for the fractional delay is obtained by setting the error function to zero at zero frequency. This is the maximally flat gain design at DC. It is found that the coefficients of this filter as (2.4) correspond to the weighting coefficients in the classical Lagrange interpolation.

$$h(n) = \prod_{k=0, k \neq n}^{N_f} \frac{D-k}{n-k} \text{ for } n = 0, 1, 2, \dots, N_f \quad (2.4)$$

where N_f is the order of the filter. There is a condition to make sure the magnitude response of the Lagrange interpolator is less than or equal to one for the mentioned values of D when the delay has been chosen so that $(N_f - 1)/2 \leq D \leq (N_f + 1)/2$ when N_f is odd and $(N_f/2) - 1 \leq D \leq (N_f/2) + 1$ when N_f is even [15]. This property is advantageous because in digital waveguide models the interpolator is normally used inside a feedback loop and then it is extremely important to preserve the loop gain less than unity. Otherwise the system may become unstable. Since a Lagrange interpolator is a passive filter, the interpolation error only decreases the loop gain but never increases it.

Regarding the higher order allpass filters, can increase the accuracy of phase delay at higher frequency. The transfer function follows the Thiran method as (2.5) - (2.7). The filter has a characteristic with maximally flat group delay (phase slop) at DC [2].

$$H(z) = \frac{a_{N_f} + a_{N_f-1}z^{-1} + \dots + a_1z^{-(N_f-1)} + z^{-N_f}}{1 + a_1z^{-1} + \dots + a_{N_f-1}z^{-(N_f-1)} + a_{N_f}z^{-N_f}} \quad (2.5)$$

$$a_k = (-1)^k \binom{N_f}{k} \prod_{n=0}^{N_f-k} \frac{D - N_f + n}{D - N_f + k + n}, \quad k = 0, 1, 2, \dots, N_f \quad (2.6)$$

$$\binom{N_f}{k} = \frac{N_f!}{k!(N_f - k)!} \quad (2.7)$$

2.4 Stiffness Dispersion Filter Design

On the other hand, $H_s(z)$ in Fig. 2.5 is the frequency dispersion response for the stiffness characteristic. The relationship of the k th partial frequency f_k and the fundamental frequency f_0 becomes to:

$$f_k = kf_0 \sqrt{1 + Bk^2} \quad (2.8)$$

Since the function is for delay adjustment only, the Thiran allpass filter method described in last section is a good way to achieve the requirement and get no more gain loss at the same time. This filter generates the extra phase delay (i.e. frequency shifting) added on each partial of the tone. Using cascaded filters can achieve the accuracy and the stability at the same time if more than one order is needed. Moreover, this filter is different from the fractional delay one because the required delay D_1 formulated in (2.9) comes from the calculation of the pitch frequency with the inharmonicity coefficient B which is depicted in section 4.2 [7].

$$D_1(I_{key}, B) = e^{(C_d(B) - I_{key})k_d(B)} \quad (2.9)$$

This D_1 is approximated as the delay at DC for the filter design. The value I_{key} is the logarithmic representation of the desired fundamental frequency as (2.10), so that the $\ln(D_1)$ as a function of I_{key} is approximately a straight line in the lower frequency range. The other

values C_d and k_d are two predefined constants [26]. Fig. 2.8 shows a phase delay characteristic of a cascaded four first order dispersion filter which fits the A2 (110 Hz) tone. To fit the dispersion curve, the extra delay at DC generated by this filter can be very large according to the D_1 formula.

$$I_{key}(f_0) = \log_{\sqrt[12]{2}} \frac{f_0 \sqrt[12]{2}}{27.5} \quad (2.10)$$

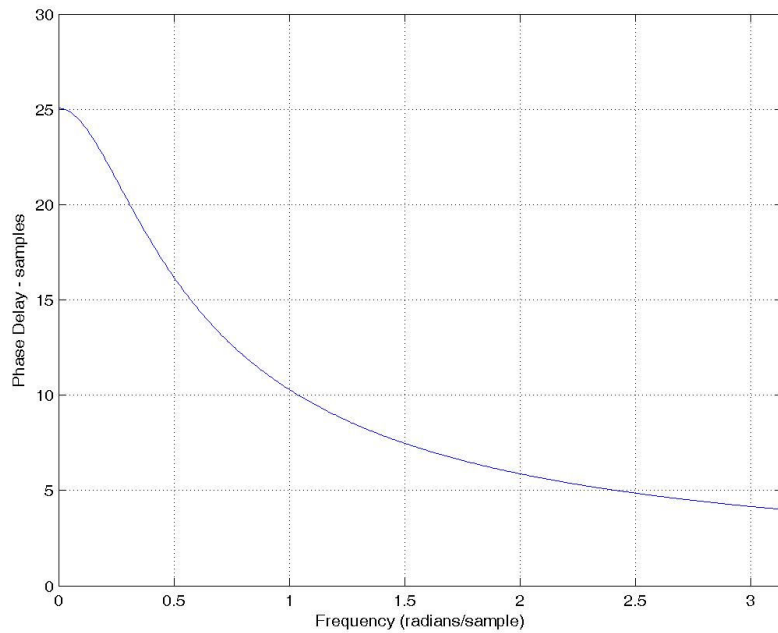


Figure 2.8: Phase delay of an allpass dispersion filter for a 110Hz pipa tone.

2.5 Damping Filter Design

The damping filter ($H_d(z)$ in Fig. 2.5) is designed for 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 several mainly lower 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 for sound natural characteristic [27]. The damping filter is then generated to fit the magnitude spectrum as the example in Fig. 2.9. The number of order should not be high for perfect fitting because of the stability and computation speed considerations. After the filter generated, it should 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 lowest harmonics whose attenuation rate can be heard easily. As long as the filter design is proper, the synthesized sound would be satisfied because the string model has included most essential blocks.

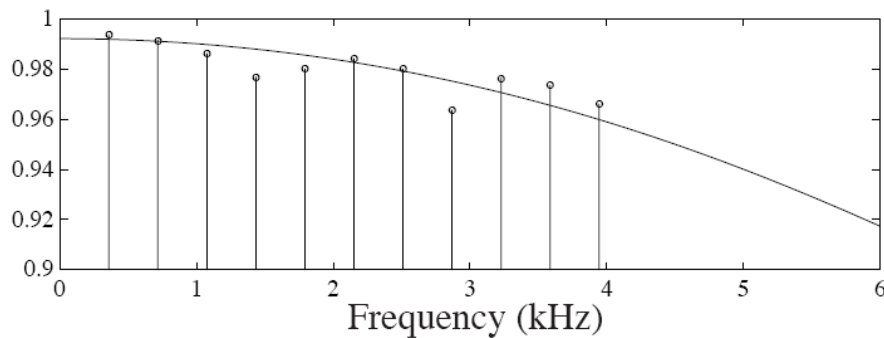


Figure 2.9: An example of estimating magnitude spectrum (circles) and magnitude response of a first-order IIR filter [4].

2.6 Excitation Signals and Body Model

Although the pluck moment of each plucked instrument is like an impulse with very wide frequency band, there are some differences among them. Using a simple random noise cannot represent these characteristics. Besides the string materials, the different playing techniques also affect the sound of the initial pluck. Even plucking the same string with the same technique, the timbre could vary a little dependent on the string open or not. Therefore, the excitation signals extracted from the recorded tones of different strings and techniques replace

the noise after constructing the string model [4]. The extraction process would be introduced in detail in section 5.2.

Besides the string and the input excitation, another important part of the plucked string instrument is the sound resonator. Using a physical-based 2-D or even 3-D mesh digital waveguide can model an instrument body [28]. But the computing complexity will become another issue to be considered. A spectrum-based body model [11] instead of the physical model can not only simplify the computing but also maintain the flexibility requirement. It is to design a set of filters to fit the spectral shape of the body impulse response. Using cascaded biquad filters [25] can get more accurate fitting and better stability at the same time.



Chapter 3

Description of the Pipa Instrument

3.1 Construction and Tuning

The instrument pipa and the typical playing position are shown in Fig. 3.1. Unlike the guitar, the pipa is played by holding its body vertically. This difference is for the convenience of the left hand performing techniques. Fig. 3.2 shows the front and back views of a pipa. A standard pipa height is 102cm and the effective string length is 72.5cm. With a wood body like a halved 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 [29] shows the wood resonant peaks of pipa body are concentrated at the range of 450 Hz to 650 Hz and are part of higher pipa tone register. This may be an indication of the sound the ancient Chinese favored. A neck is made as six deep and triangular frets called ledges, 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 [30]. 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. 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 [31].



Figure 3.1: Photo of the soloist Luo playing the Chinese instrument pipa.

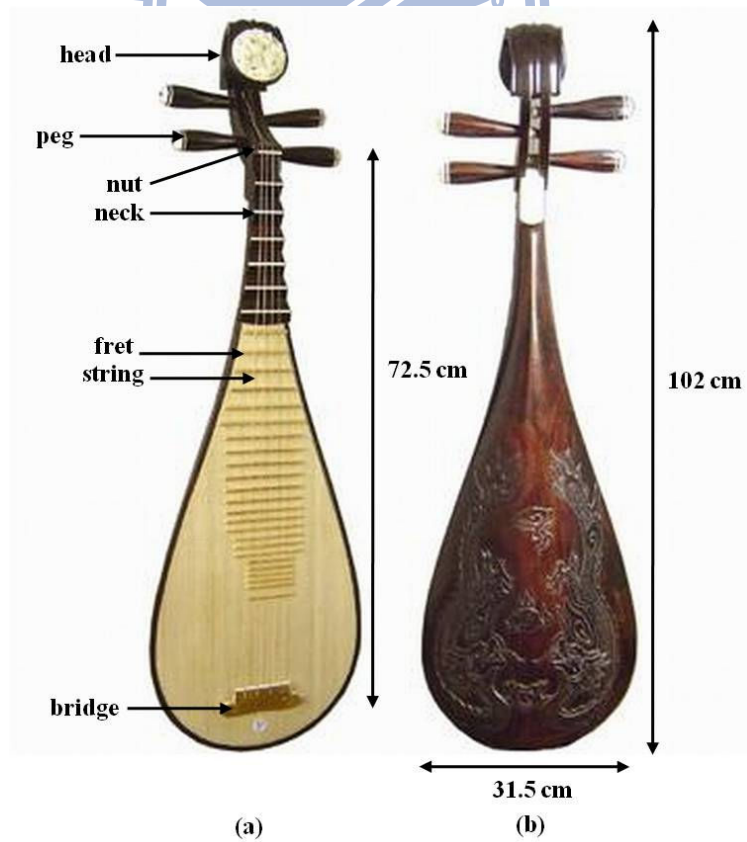


Figure 3.2: Photo of the pipa front (a) and back (b) views.

3.2 Playing Techniques

The name of pipa is from the most basic playing technique which 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 through the centuries [31]. Table 3.1 is the summary of the techniques. Basically, they need spectacular finger dexterity to handle such complex performances. The typical pluck is with 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 pretty high frets. The wheel (also called finger ring), with the right hand rotating all fingers one by one on the strings, is a unique technique which is able 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 [31]: "... 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."

Table 3.1
Summary of common playing techniques of the pipa

<i>Technique¹</i>	<i>Description</i>
彈挑 (typical pluck)	Left hand finger press beside the frets
滾 (roll) 輪指 (wheel)	High speed pluck with fingers rotation
撥分 (double pluck)	Two fingers pluck two strings inward or outward
掃/拂 (sweep/brush)	Pluck four strings at the same time with different direction
泛音 (harmonic)	Left hand finger touch the string
推 (push)	Left hand finger bend the string outward to shift the pitch with a flat third interval
拉 (pull)	Left hand finger bend the string inward to shift the pitch with a whole tone interval
絞弦 (wring)	Left hand fingers twist the string, like noise sound
揉弦 (knead)	Vibrato effect
悶音(damped)	Left hand finger press on the frets
木魚音(woodblock)	Left hand press string beyond the frets
拍(slap)	



¹

<i>Technique</i>	<i>Translation of Chinese Conservatory</i>
彈	Prod
挑	Stir
滾	Slide
輪	Finger Ring / Ring

Chapter 4

Acoustic Signal Analyses

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 analysis of push and wheel tones are presented as well. A microphone placed at a distance around 1 meter in front of the sound board is used for recording. This distance can get a good tone quality and avoid recording much ambient 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 cancelled 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, pressing beside the fret or on the fret. Two different plucking types were recorded, plucking with nail inward or outward. In addition, a set of harmonics, push, pull, wheel, or vibrato tones was recorded for different strings and frets.

4.1 Decay Characteristics of Tones

The T_{60} of four open strings of the pipa with typical plucking are around 5.0 second (A2), 4.3 second (D3), 5.3 second (E3), and 5.7 second (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 strings (seven steel strands wound together within a helical coil of covering copper) [22], the specific pipa body structure is an important factor for the tone release time. While comparing the spectra of

A2 with A3 open string plucked tones after the attack as Fig. 4.1, 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 strings D3 and E3 have a similar characteristic of the A2 string from the observation. This combination makes soft and loud sounds exist at the same time.

On the other hand, the decay times of tones terminated with the positions of frets are quite different. The string vibrations are restrained much more by the fingertip pressing on the fret than beside the fret. Therefore, the sound decays much faster like the example of 370 Hz tones depicted in Fig. 3. The T_{60} of the tone pressed on the fret can be as short as 1.3 second (Fig. 4.2 (b)) compared with another one, which is over 4 seconds (Fig. 4.2 (a)). Fig. 3 shows the decay spectra of these two cases, the high frequency partials beyond the sixth are depressed very much while pressing on the fret.

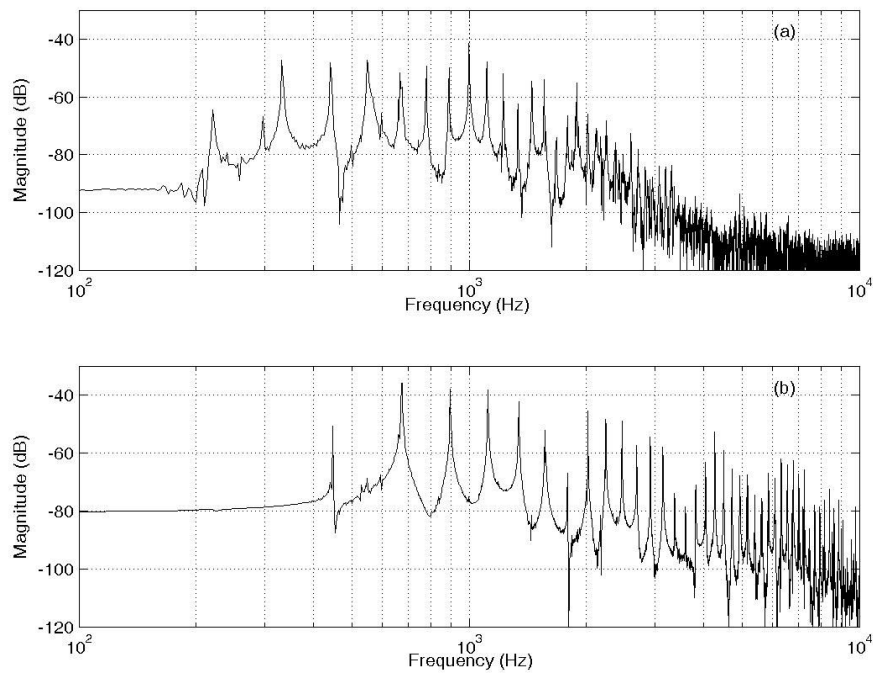


Figure 4.1: The spectra of A2 (a) with A3 (b) open string plucked tones.

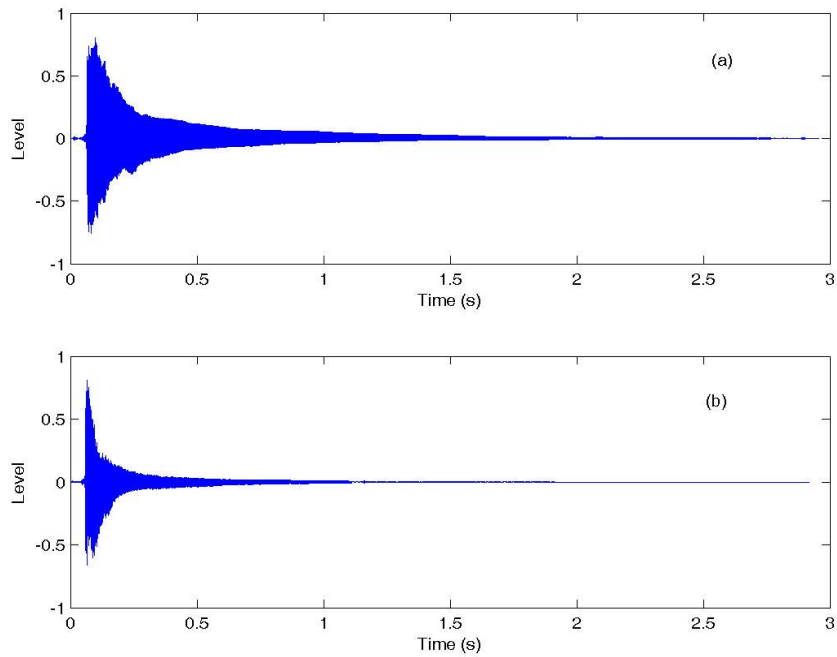


Figure 4.2: Time responses of pipa notes terminated with the fingertip pressing beside the fret (a) and on the fret (b).

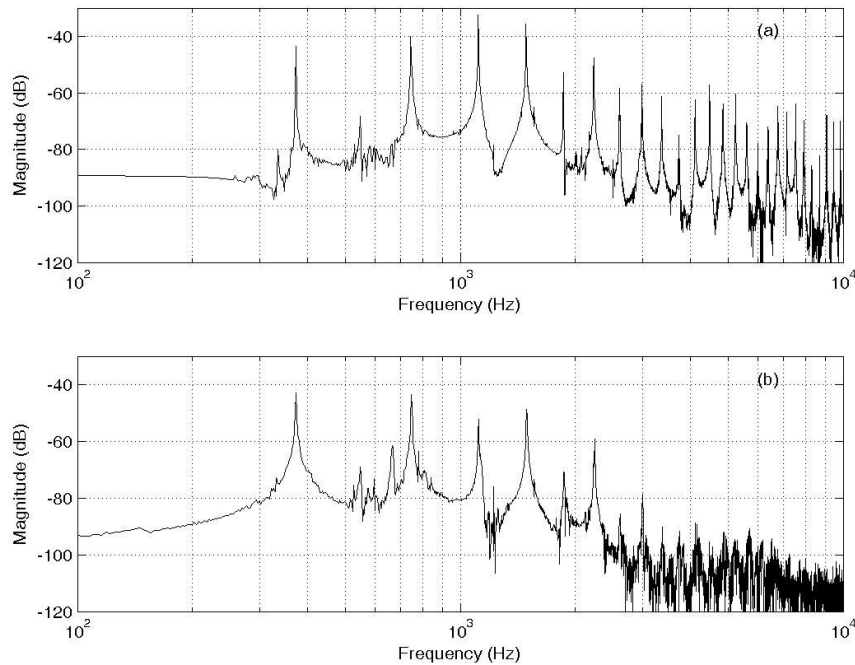


Figure 4.3: Spectra of pipa notes terminated with the fingertip pressing beside the fret (a) and on the fret (b).

For some special case such as 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 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.

4.2 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 is investigated for all strings and frets carefully. The inharmonicity coefficients of the four pipa strings are depicted in Fig. 4.4 as a function of frequency, on a log-log scale. A partial frequencies deviation method proposed by [32] was used to search the partials of the tone and then evaluate the coefficients. With the help of handed calibration, some very high frequency partials with larger deviation which are wrongly evaluated can be repaired. From the results, two observations can be made. Firstly, the inharmonicity for lower strings is larger than for higher strings. This result can be explained by the inharmonicity coefficient equation as:

$$B = \frac{\pi^3 Q d^4}{64 l^2 T_s} \quad (4.1)$$

where d is the diameter of the string, l its length, Q is Young's modulus, and T_s is the string tension [7]. A larger diameter will cause the higher inharmonic characteristic. Table 4.1 lists the referred diameter values of Shanghai strings. Secondly, the inharmonicity increases as the length of the string decreases. From Eq. (4.1), the fact can be found that the length decreases

while the diameter is fixed.

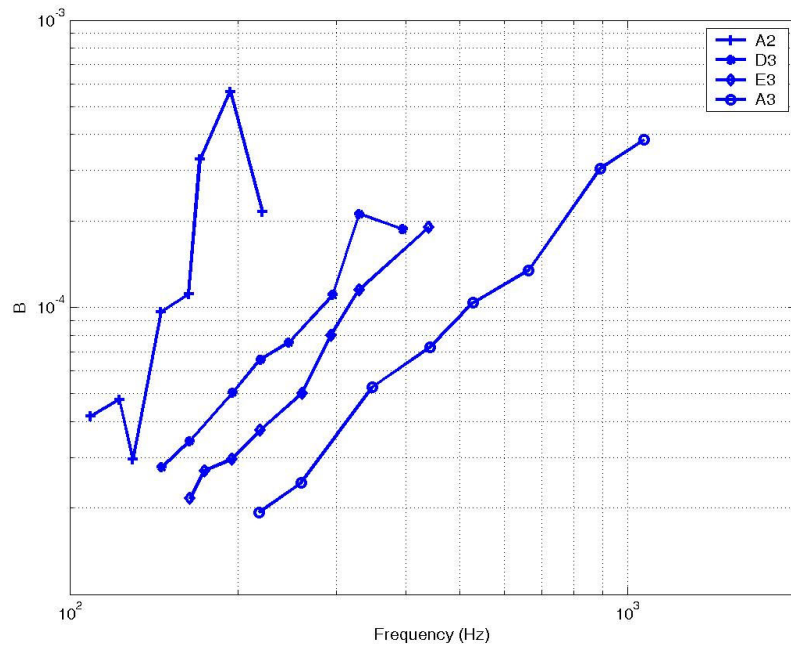


Figure 4.4: Estimates of the inharmonicity coefficients of different strings and frets.

Table 4.1
Diameters of Shanghai strings for reference [22]

String	A2	D3	E3	A3
Diameter	120.67 μ m	101.0 μ m	94.67 μ m	79.5 μ m

4.3 Behavior of Push and Pull Tones

One important technique of the pipa pushing the string inside or pulling the string outside is similar to the technique of string bending in playing the electric guitar. While doing the motions, 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. 4.5 shows the spectrogram

of two push and pull tones. The beginning of pitch slide is around 0.6 second in the pull tone later than the push tone, which is around 0.25 second. 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 second. The tone frequency can be also decreased by reversing the string movement. On the other hand, the spectrogram of Fig. 4.5 (b) is not as regular as Fig. 4.5 (a) because there is some unexpected small tremolo effect happening in this tone.

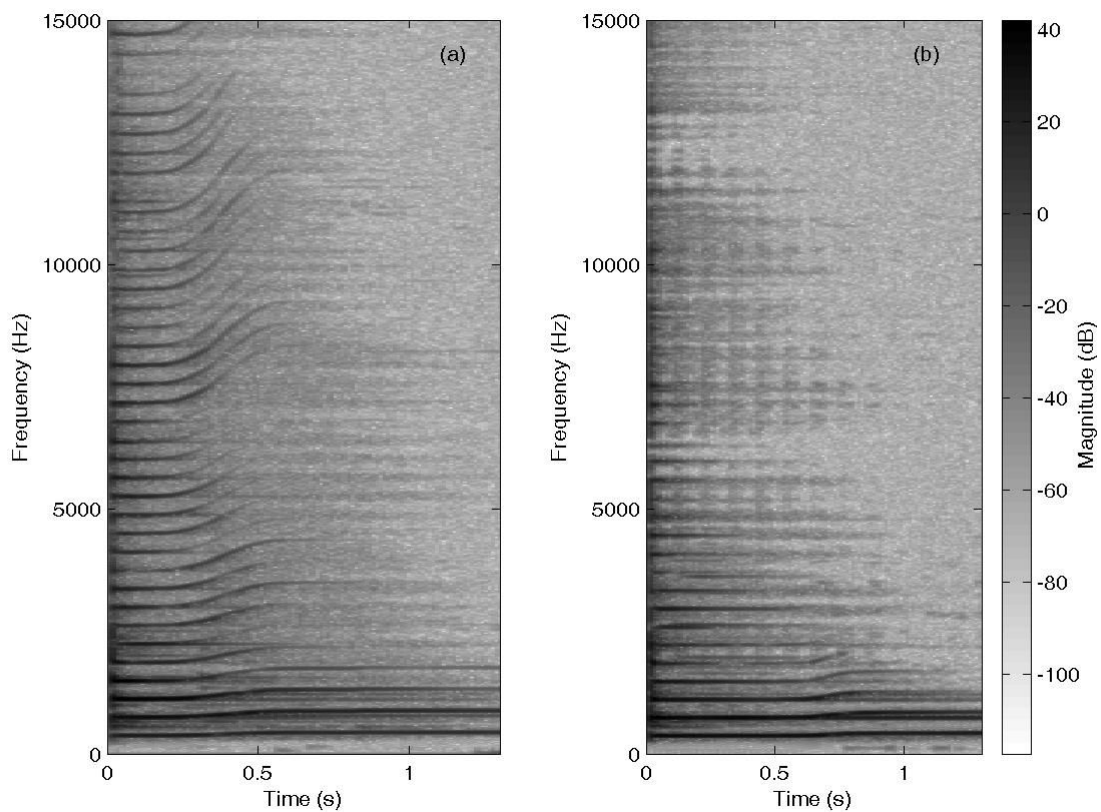


Figure 4.5: Spectrograms of the push tone with the fundamental frequency from 372 Hz to 444 Hz (a), and the pull tone with the fundamental frequency from 372 Hz to 419 Hz (b).

4.4. Analysis of Wheel Tone

A wheel tone made by continuously plucking with all fingers is a unique timbre. Fig. 7 shows the five-plucks wheel tone of 372 Hz waveform and the earlier four partials spectrum extracted from the whole tone duration FFT result. The dynamics of each pluck cannot be the same in fact. The time interval between two plucks can be smaller than 0.06 second. 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.

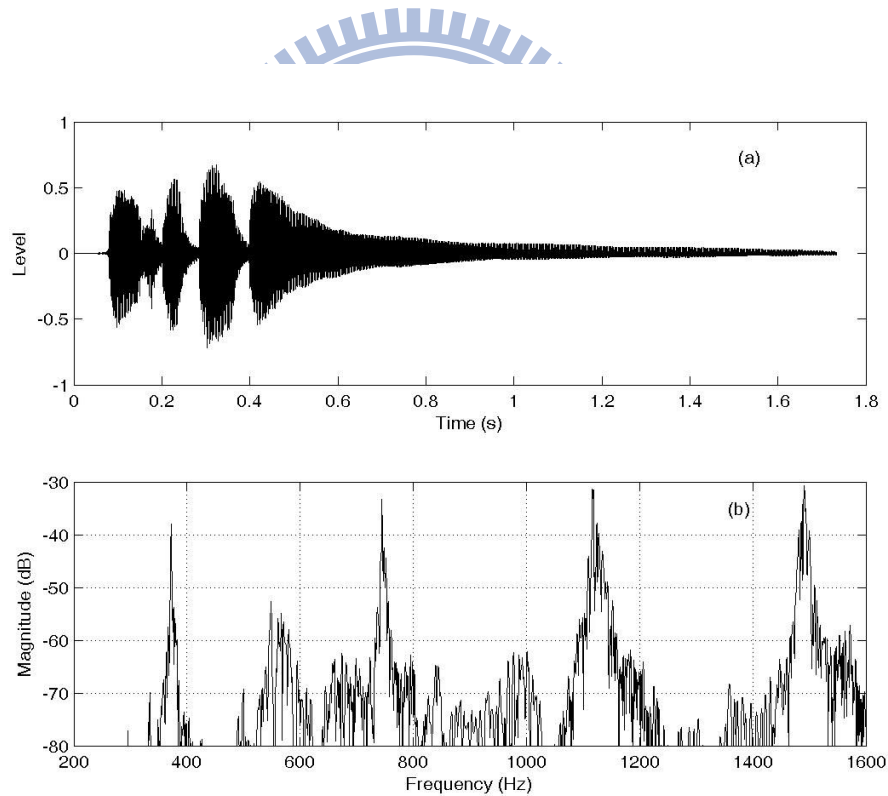


Figure 4.6: Waveform and the early four partials of a wheel tone.

Chapter 5

Sound Synthesis of the Pipa

5.1 Description of the Synthesis Model

The block diagram of the synthesis model is as in Fig. 5.1. The pipa model essentially constitutes of a delay loop (DL) digital waveguide string model $S(z)$, a body model filter $H_b(z)$ described in next section, and a comb filter placed at the input for simulating the effect of plucking position. 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.

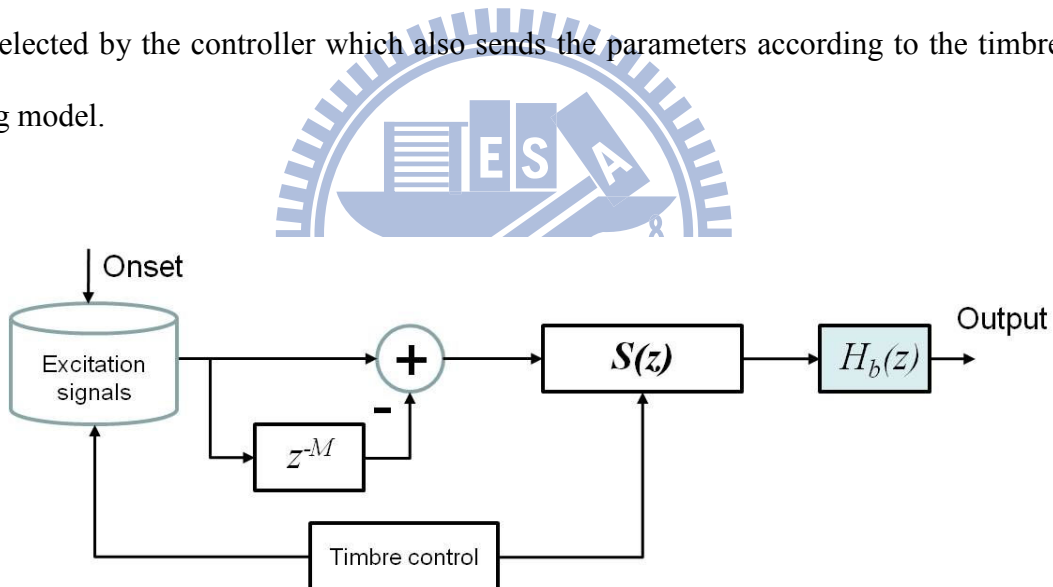


Figure 5.1: Block diagram of the pipa synthesizer.

5.2 String Model and Excitation Signals Generation

The DL string model, $S(z)$, synthesizes the transversal vibrations with the inharmonicity coefficients shown as Fig. 5.2. In this model, the z^{-L} block implements the integral delay of the DL. The remainder fractional part of the loop delay is realized by $H_f(z)$, a third-order Lagrange filter. A FIR filter can gain the advantage of stability and computation speed. Using

higher order filter with some help to higher frequency response is needed, but it is increasingly more difficult to compute. After making a trade-off, the third-order filter is used for the string model. The transfer function $H_s(z)$ is a cascaded of four first order allpass dispersion filter. According to the design method described in section 2.4, the filter generates attached delays D_1 for the desired frequency, therefore, the values of N used in the integral delay and the remainder fractional part must be modified because of the extra delay. The damping filter is a one-pole one-zero IIR structure because of A4 string timbre consideration. Equation (5.1) is the final version of the damping filter where g_l 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.

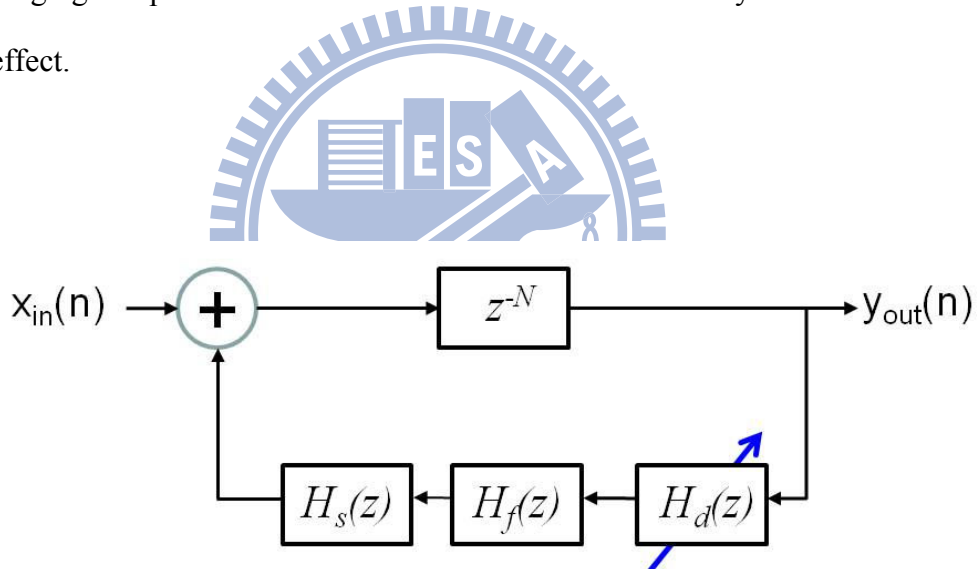


Figure 5.2: Block diagram of the string model.

$$H_{d1}(z) = \frac{g_1(0.9015 + 0.1275z^{-1})}{1 + 0.0331z^{-1}} \quad (5.1)$$

It is found that the string model blocks can be integrated to one transfer function $S_i(z)$ as (5.2), and the simulation instruction from the string input to output is reduced to one.

$$S_i(z) = \frac{z^{-N}}{1 - z^{-L}H_d(z)H_f(z)H_s(z)} \quad (5.2)$$

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 $A(z) = 1/S_i(z)$. Fig. 5.3 shows one case of $A(z)$ s, with the fundamental frequency 372 Hz. It works as a comb filter to depress the partial of the notes according to the gain of the damping filter. 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 or playing techniques, can affect the timbre. As in the discussion before, 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, using just one-pole filter as Eq. (5.3) instead of the damping filter as Eq. (5.2) can achieve the requirement of these two situations, as in the transfer function:

$$H_{d2}(z) = \frac{g_2 0.6502}{1 - 0.3412z^{-1}} \quad (5.3)$$

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. Finally, the output signals are sent through a dynamic control to make the tones change smoothly and naturally.

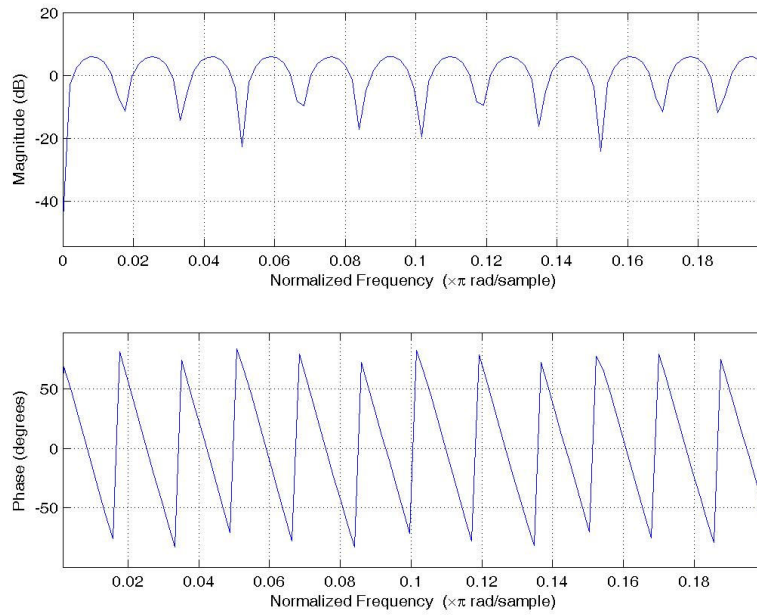


Figure 5.3: Frequency and phase response of the $A(z)$.

5.3 Body Model

As mentioned in section 2.6, it is better to create an independent body model for the sound synthesis. Fig. 5.4 shows the two spectra representing the pipa body which are 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 one is better one to be used because it contains some phenomena not included in the impulse response of knocking plate, one of which is the open strings sympathetic effect [3]. Four cascaded biquad shelving and peak-notch filters are appropriate to implement the response. The frequency response and the coefficients of the filters are shown in Fig. 5.5 and listed in Table 5.1. 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.

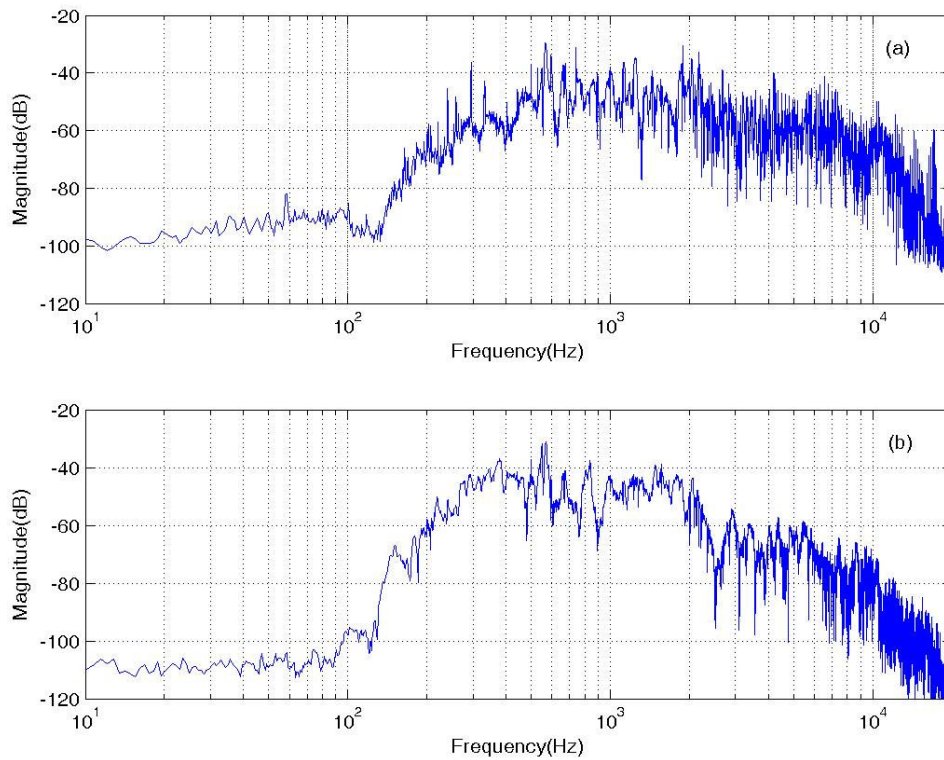


Figure 5.4: Spectra represent the pipa body including that (a) is the average of the excitation spectra of open strings, and (b) is the impulse response of a knocking sound.

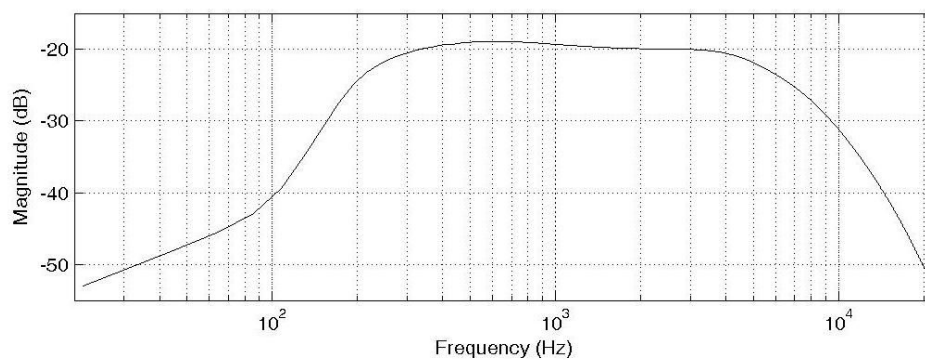


Figure 5.5: Frequency response of the body filter.

Table 5.1
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
$H_b = (b_0 + b_1z^{-1} + b_2z^{-2}) / (a_0 + a_1z^{-1} + a_2z^{-2})$				

5.4 Time-Variant String Model

The time-variant string model with the circular buffer delay line is shown in Fig. 5.5. The length of the buffer is set with a little longer than the result of the sampling rate divided by the lowest pipa pitch frequency so that the possible elongation can be accounted for. After using the structure, the string model is then separated to several blocks and run with a one by one sample.

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 Eq. (2.4), if the Lagrange filter with $N_f = 1$, $h(0) = 1-D$ and $h(1) = D$ is a linear interpolator. Gradually changing the $h(0)$ and $h(1)$ for the shift integer delay can reduce the glitch. Except the exact delay of initial and final frequency, the fractional parts (D) of the delay line between the two frequencies can be only set as 1 or 0 for frequency increased or decreased respectively. However, in order to avoid the noise generated because of an integral delay jump, it is much better to insert a 0.5 value between two integral delays.

The degradation of high frequency response is serious issue while using linear interpolation. To overcome the disadvantage, the damping filter gain (g_1 and g_2 in Eq. (4.2) and (4.3)) has to

be increased to compensate the additional attenuation. According to [33], the sound energy is also decreased by changing the delay line length. Therefore, an energy enhancement function should be implemented for this sound synthesis. According to 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 way is easy and the tone can be sustained to an enough long length.

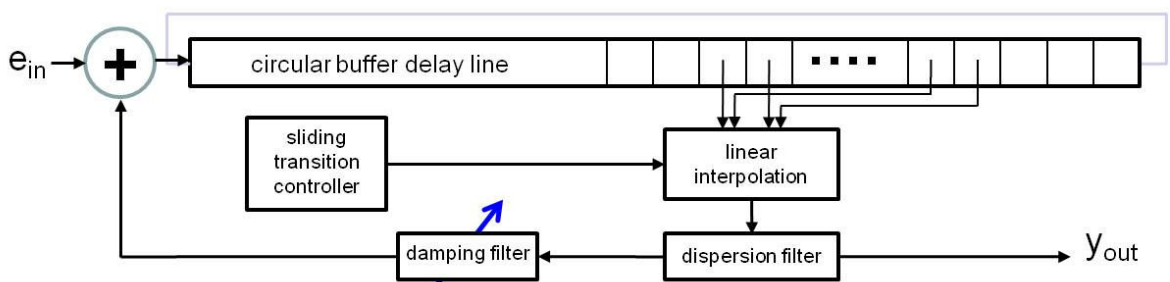


Figure 5.5: String model with circular buffer delay line and linear interpolation.

5.5 Wheel Tone and Other Pipa Techniques Models

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 triggering again and again. All plucks could be mixed up and the timbre of the attack moment was consequently weakened in the output result. Doing off-line 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.



Chapter 6

Synthetic Results

The waveforms, spectra, and spectrogram of the recorded and synthesized signals are compared as follows. All synthesized results are from the Matlab simulation.

6.1 Plucked Tones

Fig. 6.1 depicts the time responses of synthesized tone of 372 Hz with A3 string, for pressing beside the fret (a) and on the fret (b) terminations. The decay time responses correspond fairly well (compared with Fig. 4.2). Since the tones of latter one 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 untypical pluck tone is an average of two adjacent pitch tones.

One the other hand, due to different types of strings, two general excitation signals of typical plucked technique 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. Fig. 6.2 and Fig. 6.3 illustrate the characteristics of the three synthetic tones along with recorded ones. The attack-decay-sustain-release (ADSR) 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.

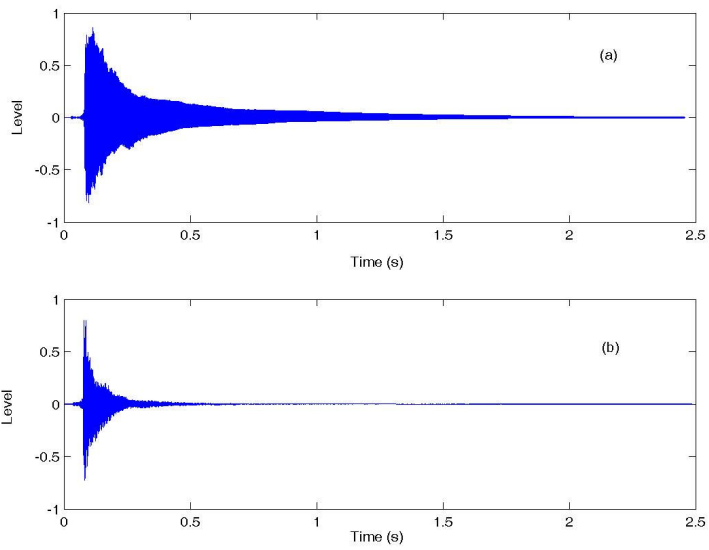


Figure 6.1: Pluck synthetic results with (a) touching beside the fret and (b) pressing on the fret terminations.

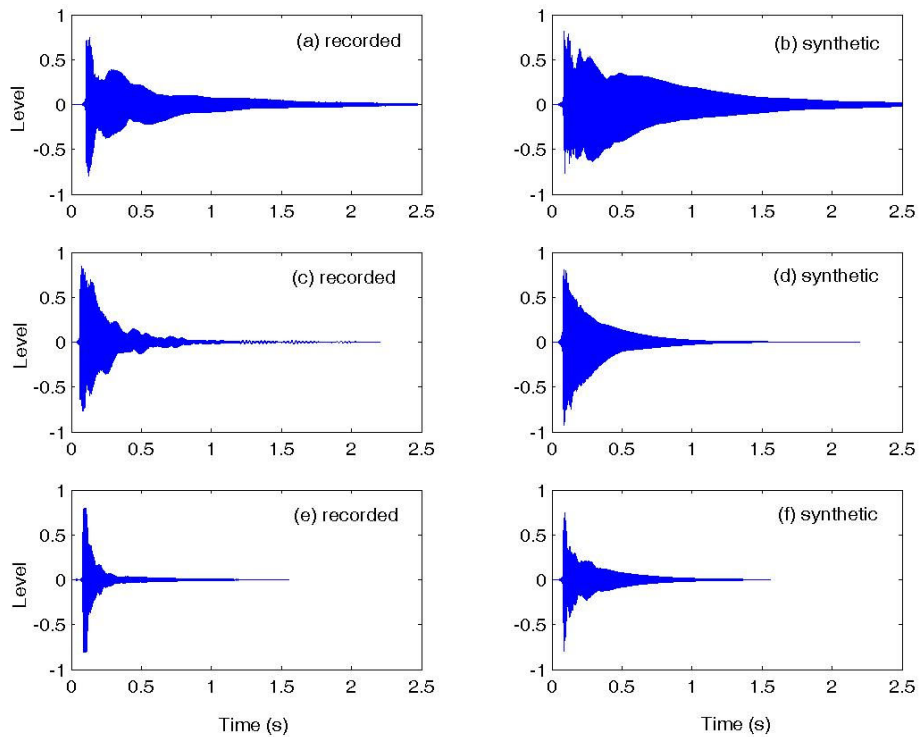


Figure 6.2: Waveforms of synthetic and recorded typical pluck tones of 166 Hz (a) (b), 660 Hz (c) (d), and 1109 Hz (e) (f).

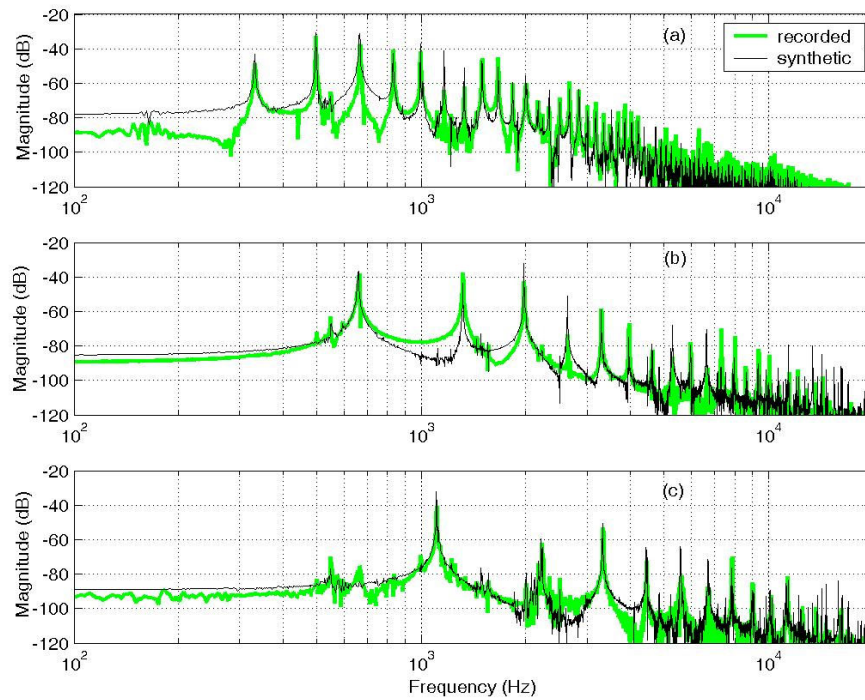


Figure 6.3: Spectra of three synthetic and recorded tones of 166 Hz (a), 660 Hz (b), and 1109 Hz (c).



6.2 Push Tones

The synthetic push tone is generated by the typical pluck excitation signal compared with the recorded one shown in Fig. 6.4. If without the energy compensation, 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 in the decay stage. The total rate of the gain increased is 1.10855 times according to the extra damping. Fig. 6.5 illustrates the spectrogram of this synthetic tone. Comparing with Fig. 4.5 (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 Fig. 6.6.

Fortunately, first-order interpolation and dispersion filters seem to work well on the partial frequencies match.

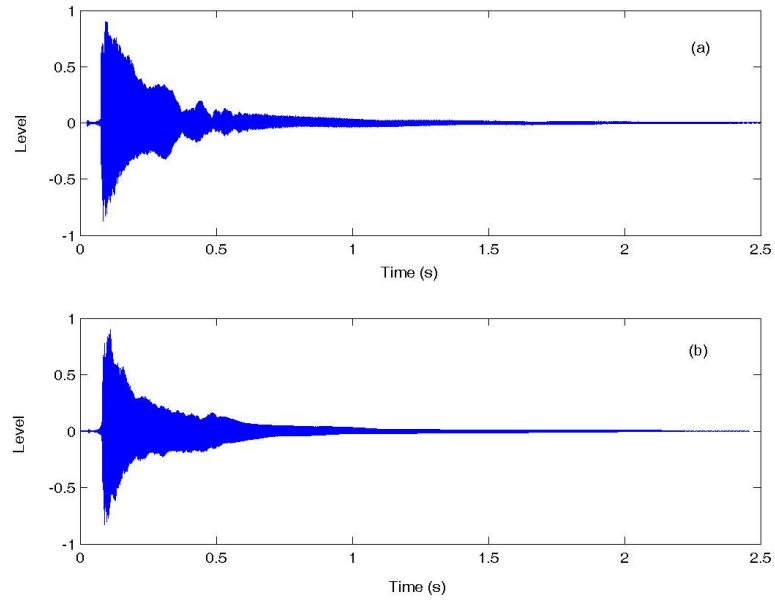


Figure 6.4: Waveform of a synthetic push tone (from 370 Hz to 444 Hz) (b) compared with the recorded one (a).

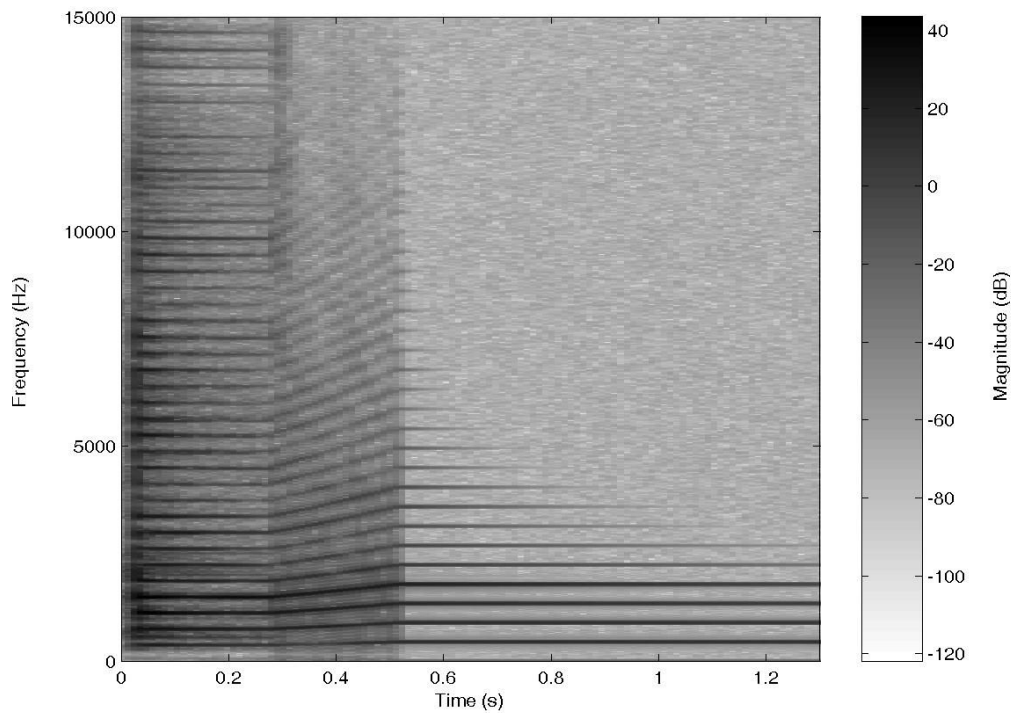


Figure 6.5: Spectrogram of the synthetic push tone.

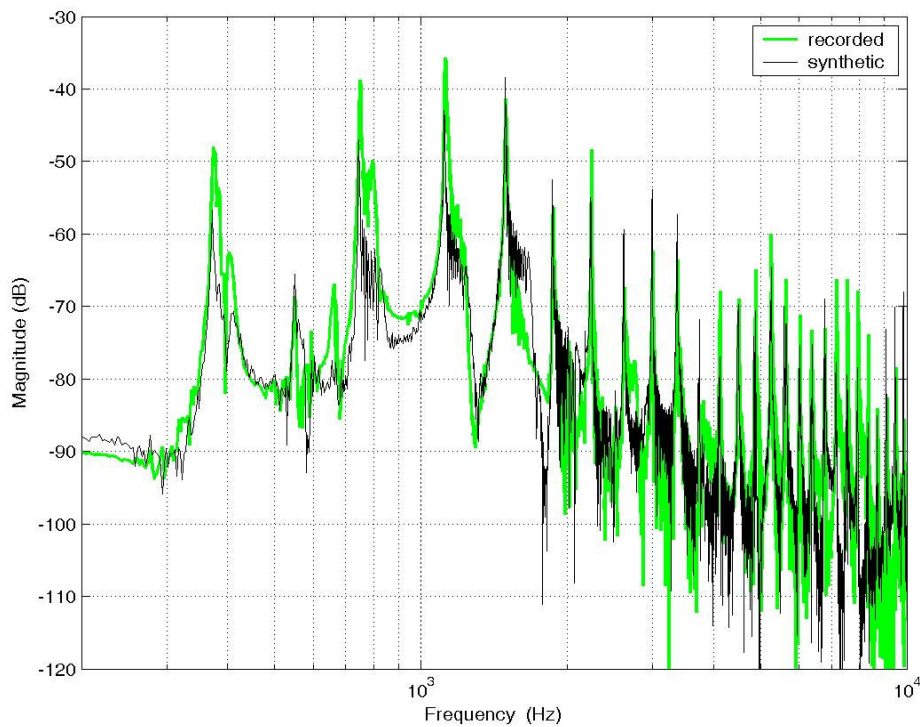


Figure 6.6: Spectra of synthetic and recorded push tones (from 370 Hz to 444 Hz).



6.3 Wheel Tones

The synthetic wheel tone 660 Hz compared with the recorded one is shown in Fig. 6.7. Although the synthetic one still has a phenomenon of the energy accumulation, each plunk can be identified from the waveform and heard clearly from the tone and the tone color is also bright. However, there is a little difference of the pluck speed between synthetic tone and recorded one of 660 Hz. This is because the excitation signal used for this synthesis is not from the specific recorded one but a general one.

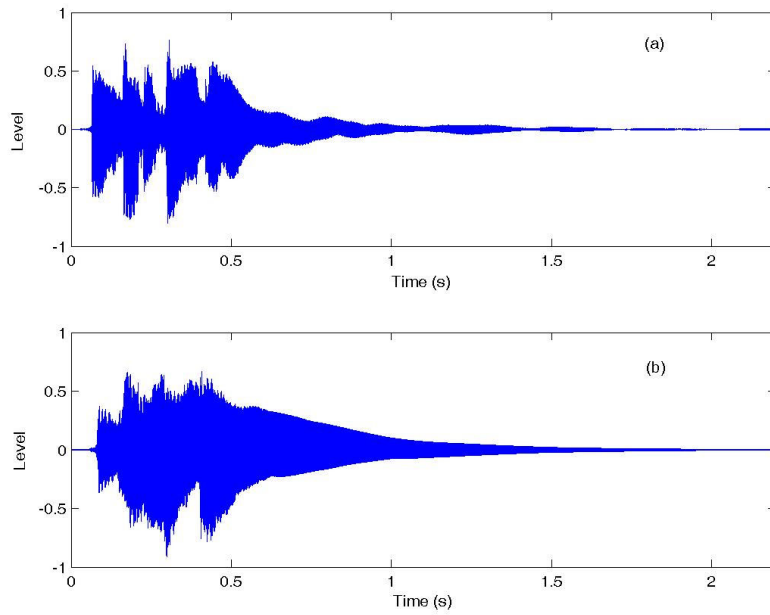
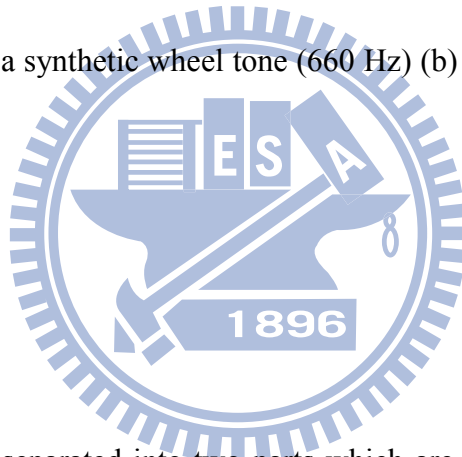


Figure 6.7: Waveform of a synthetic wheel tone (660 Hz) (b) compared with the recorded one (a).



6.4 Listening Tests

The listening tests are separated into two parts which are the similarity between synthetic and recorded tones and the synthesis acceptability. The SPSS software is used to get the statistical results. Among the total 13 subjects, there are 69.2% who have more than 5 years of music experience and 53.8% who have practiced the piano. Although there are 15.4% who listen to Chinese music frequently, the reliability coefficient (Cronbach's α) is 0.856, which shows the good quality of the subjects' timbre discrimination. Table 6.1 and 6.2 list the statistical results with a Likert 5-point scale. It can be found that the pipa tones synthesis has got an acceptable estimation, especially for the typical pluck tone. All tones for listening tests are available at http://web2.cc.nctu.edu.tw/~eamusic/music_tech_lab/pipa/pipa.html.

Table 6.1
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 6.2
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

6.5 Discussions

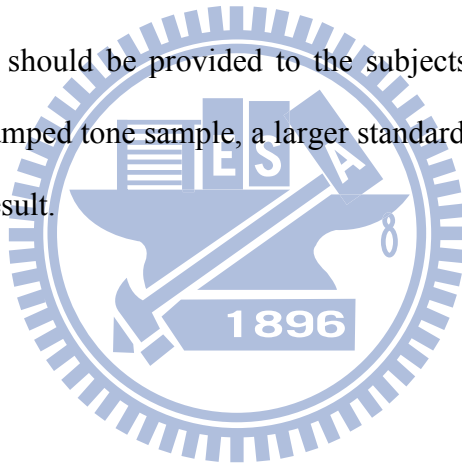
The recorded tones are played by human fingers while the synthetic ones are generated with the general excitation signals, there are more variations in the former. Therefore, there are some variances between the synthetic and recorded tones in the magnitude and the level of the spectra and the waveforms as shown in Fig. 6.2 and Fig. 6.3. But because the model has included most room acoustic characteristics of the pipa, the auditory timbres of the two are with a little difference. 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. Not only for the typical pluck tones, the method may be also helpful for other techniques.

Another one can be discussed is that there is a larger timbre variance between the synthetic and recorded push tones after the pitch sliding. Because of the remaining partials of 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 [34]. But to solve the issue, the string model still needs to be

divided to two paths for generating two types of partial somewhat similarly as the guqin model [11]. Or, the resonant characteristic of the pipa body should be investigated more so that it can preserve the sound reverberation. Besides, although the just-noticeable-difference (JND) for pitch change is typically around 1/10th of a semitone [35], 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 whether 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 statistic result.



Chapter 7

Conclusions and Future Works

The acoustic characteristics and a synthetic method of performing the traditional Chinese instrument pipa have been discussed in detail. The proposed synthesis model has demonstrated the important techniques of the instrument such as pluck 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 modified according to the discussion in last section, and improved further to cover more physical phenomena no matter the straightforward nonlinearity characteristic of the playing dynamics or more complex mixing between the resonator and the strings. Besides, more play techniques should be included. 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 [36], [37].

References

- [1] Julius O. Smith III, "Physical modeling using digital waveguides," *Computer Music Journal*, vol. 16, no. 4, pp.74-91,1992.
- [2] Julius O. Smith III, *Physical Audio Signal Processing for Virtual Musical Instruments and Audio Effects*, printed in the United States of America by W3K Publishing, August 2007 edition. Available: <https://ccrma.stanford.edu/~jos/pasp/pasp.html>
- [3] D. Jaffe and J. Smith, "Extensions of the Karplus-Strong plucked-string algorithm," *Computer Music Journal*, vol.7, no. 2, pp. 56-69, 1983, reprinted in [38].
- [4] M. Karjalainen, V. Välimäki, and Z. Janosy, "Towards high-quality sound synthesis of guitar and string instruments," in *Proceedings of International Computer Music Conference*, Tokyo, Japan, Sep. 1993, pp. 56-63. 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 *Proceedings of International Computer Music Conference*, 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," *Journal of the 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 Processing Letters*, vol. 13, no. 5, pp. 253-256, May. 2006.
- [8] CCRMA website: <https://ccrma.stanford.edu/~jos/>
- [9] HUT Acoustics Lab website: <http://www.acoustics.hut.fi/>
- [10] K. Karplus and A. Strong, "Digital synthesis of pluck-string and drum timbres," *Computer Music Journal*, vol. 7, no. 2, pp. 43-55, 1983, reprinted in [38].

- [11] H. Penttinen, J. Pakarinen, V. Välimäki, and et al, "Model-based sound synthesis of the guqin," *The Journal of the Acoustical Society of America*, 120(6), pp. 4052-4063, Dec, 2006.
- [12] S.-J. Cho, U.-P. Chong, and S.-B. Cho, "Synthesis of the Dan Trahn Based on a Parameter Extraction System," *Journal of the Audio Eng. Soc.*, vol. 58, no. 6, pp. 498-507, June 2010.
- [13] V. Välimäki, J. Pakarinen, C. Erkut, and M. Karjalainen, "Discrete-time modelling of musical instruments," *Reports on Progress in Physics*, vol. 69, no. 1, pp. 1-78, Jan. 2006.
- [14] M. Laurson, V. Norilo, and M. Kuuskankare, "PWGLSynth: a visual synthesis language for virtual instrument design and control," *Computer Music Journal*, vol. 29, pp. 29–41. 2005.
- [15] V. Välimäki, *Discrete-Time Modeling of Acoustic Tubes Using Fractional Delay Filters*, Dr. Tech. thesis, Helsinki University of Technology, Lab. of Acoustics and Audio Signal Processing, Report 37, 1995.
Available: http://www.acoustics.hut.fi/~vpv/publications/vesa_phd.html
- [16] J. S. Abel, V. Välimäki, and J. O. Smith, "Robust, efficient design of allpass filters for dispersive string sound synthesis," *IEEE Signal Processing Letters*, vol. 17, no. 4, pp. 406-409, April 2010.
- [17] 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 *Proceedings of International Computer Music Conference*, Thessaloniki, Greece, Sep. 1997, pp. 252–255.
- [18] Erkut, C., Valimaki, V., Karjalainen, M. and Laurson, M., "Extraction of physical and expressive parameters for model-based sound synthesis of the classical guitar", *AES108th Convention*, preprint no. 5114, Paris, February 2000.
- [19] A. Horner, "Nested modulator and feedback FM matching of instrument tones," *IEEE Transactions on Speech and Audio Processing*, vol.6, no. 4, pp. 398-409, July 1998.

- [20] Alvin W. Y. Su and Sheng-Fu Liang, "A class of physical modeling recurrent networks for analysis/synthesis of plucked string instruments," *IEEE Transactions on Neural Networks*, vol. 13, no. 5, pp.1137-1148, Sep. 2002.
- [21] Sheng-Fu Liang and Alvin W. Y. Su, "Modeling and Analysis of Acoustic Musical Strings Using Kelly-Lochbaum Lattice Networks," *Journal of Information Science and Engineering*, vol. 20, pp. 1161-1182, 2004.
- [22] Shin Hui Lin Chin and J. Berger, "Analysis of pitch perception of inharmonicity in pipa string using response surface methodology," *Journal of New Music Research*, vol. 39, no.1, pp. 63-73, 2010.
- [23] Anne-Marie Burns, "Karplus-Strong plucked-string synthesis algorithm or how to create string instruments out of noise," available on website:
<http://www.music.mcgill.ca/~amburns/physique/introduction.html>
- [24] V. Välimäki and T. I. Laakso, "Principles of fractional delay filters", *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, pp. 3870 – 3873, vol.6, 2000.
- [25] Julius O. Smith III, *Introduction to Digital Filters*, printed in the United States of America by W3K Publishing, May 2008 edition.
- [26] J. Rauhala and V. Välimäki, "Dispersion modeling in waveguide piano synthesis using tunable allpass filters," in *Proceeding of International conference on digital audio effects*, pp. 71-75, 2006.
- [27] S. Sanders and R. Weiss, "Synthesizing a guitar using physical modeling techniques," available: <http://www.ee.columbia.edu/~ronw/dsp/>
- [28] S. A. Van Duyne and J. O. Smith III: "Physical Modeling with the 2-D Digital Waveguide Mesh," *Proceedings of the International Computer Music Conference*, pp. 40-47, 1993.
- [29] S. Feng, "Some acoustical measurements on the chinese musical instrument p'i-p'a,"

Journal of the Acoustical Society of America, vol. 75, no. 2, pp. 599–602, 1984.

[30] 鄭德淵, 中國樂器學, 生韻出版, 台北, 1984.

[31] Liu Fang website: <http://www.liufangmusic.net/>

[32] J. Rauhala, H.M. Lehtonen, and V. Välimäki, “Fast automatic inharmonicity estimation algorithm,” *Journal of the Acoustical Society of America*, vol. 121, no. 5, pp. EL184-EL189, 2007.

[33] J. Pakarinen, M. Karjalainen, V. Valimaki, and S. Bilbao, “Energy behavior in time-varying fractional delay filters for physical modeling synthesis of musical instruments,” *IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. iii/1 - iii/4 Vol. 3, 18-23 March 2005.

[34] H. A. Conklin, “Generation of partials due to nonlinear mixing in a stringed instrument,” *Journal of the Acoustical Society of America*, vol. 105, no.1, pp. 536-545, 1999.

[35] Thomas D. Rossing, *The Science of Sound 2nd Ed*, Addison-Wesley, 1990.

[36] V. Välimäki, H. Penttinen, and et al, “Sound synthesis of the harpsichord using a computation ally efficient physical model,” *Journal on applied signal processing*, pp. 934-948, 2004.

[37] M. Karjalainen, T. Tolonen, V. Välimäki, and et al, “An overview of new techniques and effects in model-based sound synthesis,” *Journal of New Music Research*, vol. 30, no. 3, pp. 203-212, 2001.

[38] C. Roads, ed., *The Music Machine*, Cambridge, MA: MIT Press, 1989.