석사 학위논문
Master’s Thesis

디지털 도파관 이론에 기반한 거문고의 소리
합성에 관한 연구

Sound Synthesis of the Geomungo Using Digital Waveguide
Modeling

김승훈 (金承勳  Kim, Seung-hun)
문화기술대학원
Graduate School of Culture Technology

KAIST

2011
디지털 도파관 이론에 기반한 거문고의 소리 합성에 관한 연구

Sound Synthesis of the Geomungo Using Digital Waveguide Modeling
Sound Synthesis of the Geomungo Using Digital Waveguide Modeling

Advisor : Professor Yeo, Woon Seung

by

Kim, Seung-hun

Graduate School of Culture Technology

KAIST

A thesis submitted to the faculty of KAIST in partial fulfillment of the requirements for the degree of in the Graduate School of Culture Technology. The study was conducted in accordance with Code of Research Ethics.¹

2011. 6. 1.

Approved by

Professor Yeo, Woon Seung

[Advisor]

¹Declaration of Ethical Conduct in Research: I, as a graduate student of KAIST, hereby declare that I have not committed any acts that may damage the credibility of my research. These include, but are not limited to: falsification, thesis written by someone else, distortion of research findings or plagiarism. I affirm that my thesis contains honest conclusions based on my own careful research under the guidance of my thesis advisor.
디지털 도파관 이론에 기반한 저문고의 소리 합성에 관한 연구

김승훈

위 논문은 한국과학기술원 석사학위논문으로 학위논문심사위원회에서 심사 통과하였음.

2011년 6월 1일

심사위원장  여운승 (인)
심사위원  노준용 (인)
심사위원  이교구 (인)
This paper presents a sound synthesis method for the geomungo, a Korean traditional plucked-string instrument, based on physical modeling. Commuted waveguide synthesis method is used as a basic synthesis algorithm in this work. Recorded geomungo tones are analyzed by estimating frequency responses and fundamental frequency curves. A synthesis model proposed here consists of a fractional delay allpass filter, a loss filter which combines a ripple filter and an one-pole filter, and a digital delay line. Calibration of parameters in the model is done by minimizing the difference between the magnitude response of the designed filter and estimated losses of harmonic partials. An time-varying model which has Lagrange interpolation FIR filter as a fractional delay filter, time-varying loss filter and delay line, and a gain factor to preserve energy is proposed and evaluated to generate geomungo tones which have fluctuating pitches. Gain control using a sinusoidal function in the time-varying loss filter is also discussed. In addition, a hybrid-modal synthesis model which synthesizes the signal with the resonators at low frequencies and with the digital waveguide model at high frequencies is proposed. Finally, real-time sound synthesis application is implemented based on The Synthesis toolkit(STK). The former shows that physical modeling based on digital waveguide theory is an effective method to synthesize geomungo sounds.
### Contents

Abstract ................................................. i

Contents .............................................. ii

Chapter 1. Introduction 1

1.1 the geomungo ........................................ 4

1.1.1 Structure ....................................... 5

1.1.2 Technique ....................................... 5

1.2 Background theory ................................. 6

1.2.1 Physical modeling ................................. 6

1.2.2 MSW algorithm ................................... 7

1.2.3 Karplus-Strong(KS) algorithm .................. 8

1.2.4 Digital Waveguide Theory ....................... 9

1.2.5 Commuted Waveguide Synthesis ................. 13

1.2.6 Fractional delay filter .......................... 14

1.3 Related work ...................................... 19

1.3.1 Physical modeling of plucked string instruments .... 19

1.3.2 Physical modeling of asian string instruments ... 20

Chapter 2. Analysis 21

2.1 Extraction of geomungo tones .................... 21

2.2 Frequency response ............................... 21

2.3 Estimation of fundamental frequency ............. 23
Chapter 1. Introduction

The aim of this paper is to propose a model-based sound synthesis method for the geomungo, a Korean traditional plucked-string instrument and a geomungo synthesizer based on the digital waveguide synthesis algorithm.

The first attempt to synthesize musical sounds artificially with a machine was considered as "The Musical Telegraph" by Elisha Gray in 1876. He developed an oscillator which could produce a sound of a musical note through an electromagnetic vibration, and a machine transmitting the sound through telephone line based on the oscillator[37].

The advance in computer technology encouraged the synthesis of digital sound, but the earliest experiments about sound synthesis were tried for the purpose of creating speech sounds. The initial experiments for synthesis of non-speech analog sound were creating waveforms which generated the basic signals such as sinusoid or square wave from an oscillator periodically. Thus, in order to generate the non-speech sounds in computer, a new software was necessary because it required different algorithms from the softwares for speech sound[8]. The first experiment to synthesize musical sounds using a computer was done through MUSIC I program developed by Max Mathew in 1957[44]. He used IBM 704 computer in IBM World Headquarters, NYC. MUSIC I could generate triangle wave signals and control pitch and duration.

Then synthesizer was invented through a continuous advance in computer speed and capability. It is an electrical musical instrument which generated sound signals using oscillator and filter. Typical synthesizers consist of a piano-shaped keyboard for users to generate musical notes, knob and switch to control several parameters. The first synthesizer was Moog synthesizer invented by Robert Moog in 1964, and it used voltage controlled oscillator (VCO) which controlled pitch of the generated sounds as the magnitude of supplied voltage. Keyboard was also used as a controller and subtractive synthesis was the way of sound synthesis in Moog synthesizer[35][36].

Historically, various sound synthesis techniques for synthesizers have been tried to produce sounds of musical instruments[43]. Sampling synthesis is a method used in early synthesizers. It has samples of the instruments by recording, and the four separated portions which correspond to attack, decay, sustain, and release are designated in the samples. Instrumental sound is generated
by controlling the four portions depending on the users’ input.

Additive synthesis forms a complex waveform by adding the waveform of basic signals such as sinusoid signal. Theoretically, it can generate any complex waveforms with many basic waveforms. Hammond organ is a popular analog synthesizer using additive synthesis.

Frequency modulation (FM) synthesis is one of widely-used sound synthesis methods, which is based on FM in communication systems. A basic principle of FM synthesis is that an oscillator called modulator oscillator is modulated by another oscillator called carrier oscillator. The greatest advantage of FM synthesis is that it can generate a complex waveform using only a small number of oscillators. John Chowing proposed the use of FM in music at first[11]. DX7 synthesizer made by Yamaha is a commercially successful synthesizer based on the FM synthesis[19].

These sound synthesis techniques have one thing in common: digital recordings are edited to re-synthesize the sound. They controls the input signal and function to get a desired spectrum. On the other hand, physical modeling synthesis is based on mathematical models of the physical acoustics. A small change in producing sounds can be controlled by adjusting parameters, so it does not demand to record all sound samples. Despite these advantage, physical modeling has been developed recently because sound synthesis methods based on the mathematical model require high computational costs. VL1 synthesizer made by Yamaha uses digital waveguide theory which is an improved model based on Karplus-Strong (KS) algorithm and is one of the efficient algorithms in physical modeling[55][43].

The synthesizers discussed above are based on the hardware which consists of limitless electronic components, so they are heavy and expensive. Thus, a software synthesizer called virtual instrument was proposed as an alternative. The functions in hardware synthesizers are realized as a computer program, so a virtual instrument is a relatively cheap and portable method to generate the sounds of instruments. Most software synthesizers exist as plug-in programs built in digital audio workstation (ADW) which is a computer program for users to make music by recording and editing the sounds. By using the virtual instruments, composers can make music without recording the sounds of real instruments. VSTi in Virtual Studio Technology (VST) developed by Steinberg company is a popular virtual instrument.

These software synthesizers can generate the sounds of various western musical instruments. However, there is lack of the programs for Korean traditional musical instruments. VST for Korean traditional music developed by Hanyang University is the only known virtual Korean traditional
instruments so far, but it cannot create the instrument sounds due to the unique acoustical characteristics such as a great pitch fluctuation since it is based on sampling synthesis.

Therefore, I want to propose an algorithm for a virtual geomungo synthesizer based on the acoustical analysis in this paper. As I mentioned above, it requires to record limitless samples to establish a sound synthesis algorithm for the geomungo based on sampling synthesis, so physical modeling is a proper synthesis method which can reflect complex acoustical characteristics of the geomungo. Especially, pitch control in the synthesis model is important because pitch of the geomungo sounds varies more than 20Hz. Therefore, sound synthesis models based on the acoustical analysis were proposed and the time-varying synthesis model which could create the vibrato tones of the geomungo was implemented as a real-time sound synthesis program in order to make a virtual instrument. This work is expected to be a foundation for sound synthesis models of Korean traditional musical instruments and be an extension of the possibilities of composing the geomungo musics.

This work is organized as follow. Firstly, the geomungo is introduced. Then, review of background theories such as digital waveguide synthesis, and previous works about physical modeling of stringed instruments and asian instruments are discussed in next chapter. Analysis of geomungo tones is discussed in chapter 2. Based on the analysis, a general sound synthesis model is proposed and parameters for the model are calibrated in chapter 3. In chapter 4, the time-varying synthesis model is proposed and the synthesized geomungo tones having fluctuant pitches are introduced. Hybrid synthesis model for a more accurate synthesis through the control of the harmonic partials is proposed in chapter 5. At last, in chapter 6, the real-time sound synthesis application which is implemented based on STK is proposed.
1.1 the geomungo

The geomungo is a string musical instruments of zither families (Fig. 1.1). It is known as a remolded Guqin instrument imported from China before 5th century. The name comes from "geomun" (meaning "black") and "go" (meaning "zither"). It is one of three major traditional string instruments in Silla, an ancient kingdom in Korea (the gayageum, the bipa, and the geomungo). The volume of the sound is not loud and timbre is not gorgeous, but the sound is low and sonorous. Thus, the geomungo was played by scholars in Chosun dynasty and was also widely used in Buddhist music and court music. It has about 3 octaves pitch range which is the largest pitch range of Korean traditional musical instruments. The geomungo is used to play bass in ensemble, but it can play all the music range in the solo virtuoso genre called Sanjo. In addition, a percussion sound is added because the plectrum hit the body when plucking the string strongly.

Since the 1980s, new songs for the geomungo have been composed, and technique and structure of the instrument have been improved. However, because of lack of systematic research in unique timbre and sound, the advantages of the geomungo are still not be taken in modern orchestral music.
1.1.1 Structure

The length is approximately 162cm and the width is 22cm. The curved front of the body is made of Paulownia tree and the back of the body is made of chestnut. Inside of the body is hollow to amplify the vibration from the strings. Six strings of twisted silks are fastened on the body. Names of each strings are *moonhyun, yuhyun, daehyun, gwaesangcheong, gwachacheong*, and *moohyun*. The second, third, and fourth strings are stretched over sixteen frets, and first, fifth, and sixth strings are stretched over three movable bridges called *Anjok*. The plectrum which is 20cm long and made of bamboo. The geomungo does not have standard frequencies for tuning, but typically six strings are tuned to Eb2, Ab2, Db2, Bb1, Bb1, Bb2.

The second and third strings are used very frequently in playing, and they have the largest pitch range. Since the second string is thin, it produces clear sound, and third string produces low and rich sound. They change the sound by pushing and releasing the strings by left hand.

One of the biggest problems of the geomungo is that there is little deformed forms. Thus, improvement of the structure of the geomungo have been tried recently. To solve the problems because of small number of strings, some deformed geomungos which had more strings from 7 to 10 have been proposed and demonstrated. Also, an attempt was made to modify the structure to improve the timbre by having more frets.

1.1.2 Technique

The performer sits by placing left foot under right thigh and puts the geomungo on right thigh. The instrument is played propped up on edge and angled away from the performer. Right hand is used to pluck the strings both downward and upward with a plectrum held between the index finger and middle finger. Microtones are produced by pushing and releasing the string with left hand.

There are three ways to pluck the string with the plectrum in the geomungo: *daejeom, joongjeom, sojeom*. *Daejeom* is a method to hit the string very strongly, so it gives a strong accent on playing because of the sound radiated from banging the body. *Joongjeom* is the method to hang the plectrum on the string and then pluck the string. *Sojeom* is the way to pluck the string weakly. Because of these various methods, amplitude of the sound may not be simply expressed as a number.
Playing techniques are divided into left hand techniques and right hand techniques. Left hand techniques include daejeom, igim, tteul, and tteuldong and right hand techniques include nonghyun, choosung, toesung, jeonsung, and chachool. Table 1.1 explains each playing techniques.

<table>
<thead>
<tr>
<th>Technique</th>
<th>Hand</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>daejeom</td>
<td>left</td>
<td>hold plectrum above the strings and pluck it vertically</td>
</tr>
<tr>
<td>igim</td>
<td>left</td>
<td>pluck the strings downward with plectrum pressed</td>
</tr>
<tr>
<td>tteul</td>
<td>left</td>
<td>pluck the strings inward to outward</td>
</tr>
<tr>
<td>choosung</td>
<td>right</td>
<td>press on the strings to produce a little high sound</td>
</tr>
<tr>
<td>toesung</td>
<td>right</td>
<td>pull the strings down</td>
</tr>
<tr>
<td>jeonsung</td>
<td>right</td>
<td>vibrate plectrum around the position on a string</td>
</tr>
<tr>
<td>nonghyun</td>
<td>right</td>
<td>similar to vibrato and include jeonsung, toesung</td>
</tr>
</tbody>
</table>

Table 1.1: Playing techniques of the geomungo

These various techniques are related to pitch fluctuation. It tells the importance of pitch fluctuation in sound synthesis of the geomungo.

### 1.2 Background theory

#### 1.2.1 Physical modeling

As mentioned above, physical modeling is a sound synthesis method based on the mathematical model. The most fundamental concept is combining excitation and resonance. Excitation is an action to cause vibration, so it corresponds to plucking a string with a plectrum in the string instrument. Resonance means that the vibration from excitation is transmitted and vibrated through the string and body.

Most physical modeling methods are based on the wave equation, a second-order linear partial differential equation of waves expressed as

$$Ky'' = \epsilon \ddot{y}$$

where

$$K \triangleq \text{string tension}$$

$$\epsilon \triangleq \text{linear mass density}$$

\[ y'' \triangleq \frac{\partial}{\partial t} y(t, x) \]

\[ \ddot{y} \triangleq \frac{\partial}{\partial x} y(t, x). \]

Hiller and Ruiz proposed a method to synthesize instrumental sounds from the wave equation. This method derives difference equations from the wave equation for vibrating objects. Then the
difference equations are solved by an iterative approximation procedure, and discrete values from the equations represent a sound pressure wave. In this work, some conditions are necessary. First, in order to decide the environment in which the sound produces, constants for vibrating objects are need to specified. Second, the boundary conditions are required to limit values of the variables. Third, the initial state such as the position in the string needs to be specified. Then, the excitation acts as a force in the vibrating objects.

The early physical modeling synthesis methods based on the wave equation required at least one arithmetic operation for each point on a grid the interval of which is less than half a wavelength. They did not allow real-time sound synthesis because of a high computational cost, so an improved algorithm was required.

1.2.2 MSW algorithm

McIntyre, Schumacher, and Woodhouse proposed an alternative simplified method for physical modeling synthesis called MSW synthesis [34]. Fig. 1.2 shows the method. It consists of a nonlinear excitation which occurs between a string and a plectrum in plucked-string instrument and a linear resonator which represents string, tube, and body. Energy source is an input in this model, it passes through a nonlinear excitation model and a linear resonator. The output is used as an output waveform and a feedback into the nonlinear excitation. Since MSW synthesis method was a very simplified model, the sound from the model was not realistic. Thus, an improved model from the pure MSW model was required and Keefe proposed an extension MSW synthesis model[28].
1.2.3 Karplus-Strong(KS) algorithm

Karplus-Strong(KS) algorithm is a simple sound synthesis method for plucked string instruments [27] (Fig. 1.3) derived from MSW synthesis method. Actually, KS algorithm is a refined MSW algorithm. This algorithm is composed of three elements: a burst of white noise, a delay line, and a low-pass filter. The low-pass filter used in the MSW algorithm is the form of averaging two adjacent samples, so original KS algorithm is defined by

\[
y(n) = x(n) + \frac{y(n - N) + y(n - N - 1)}{2}
\]  

where \(N\) is the length of the delay line, \(x_n\) is the input signal made from a noise generator, and \(y_n\) is the output signal. This discrete time-domain signal can be converted into \(z\) transform equations as

\[
Y(z) = X(z) + \frac{Y(z)z^{-N} + Y(z)z^{-(N+1)}}{2}
\]

\[
Y(z)(1 - \frac{z^{-N} + z^{-(N+1)}}{2}) = X(z)
\]

Thus, the transfer function of the equation is

\[
H(z) \triangleq \frac{Y(z)}{X(z)} = \frac{1}{1 - \frac{1 + z^{-1}}{2}z^{-N}} = \frac{1}{1 - H_a(z)H_b(z)}
\]

where

\[
H_a(z) \triangleq \frac{1 + z^{-1}}{2}
\]

\[
H_b(z) \triangleq z^{-N}
\]

\(H_a(z)\) and \(H_b(z)\) mean the low-pass filter and the delay line used in this algorithm.

Jaffe and Smith proved that KS algorithm could be used for simulating stringed instruments [23]. Because the effective loop length of the filter used in KS algorithm is \(N + 1/2\) (phase delay of the
low-pass filter is 0.5 and of the delay line is $N$), so the period of the signal is $T_s(N + 1/2)$ and the fundamental frequency is $F_s/(N + 1/2)$. As the magnitude spectrum of the produced signal, harmonics which are integer multiples of the fundamental frequency exist and decay over time.

Additionally, Jaffe and Smith extended the algorithm by introducing a filter contributing a small delay and a factor controlling decay time.

### 1.2.4 Digital Waveguide Theory

**Introduction**

Based on KS algorithm, Smith proposed a new sound synthesis method for a physical model called digital waveguide modeling. In this method, a traveling wave is simulated by a digital delay line. Also, damping and dispersion are lumped at specific points on the assumption that the commutativity of linear time-invariant system is valid in this system. This reduces the computational cost enough to allow real-time sound synthesis.

**Sampling**

Digital waveguide modeling is based on the wave equation of the ideal vibrating string (Eq. 1.1). The latter can be expressed by two separate traveling waves:

$$
y(x, t) = y_r(x - ct) + y_l(x + ct) \quad (1.3)
$$

$$
y(t, x) = y_r(t - x/c) + y_l(t + x/c) \quad (1.4)
$$

where $y_r(x - ct)$ is a right-going traveling wave, $y_l(x - ct)$ is a left-going traveling wave, and $c$ is the speed of wave. Then, this equation can be sampled with $t = nT, x = mX$, and $X = cT$:

$$
y(nT, mX) = y_r(nT - mX/c) + y_l(nT + mX/c)
$$

$$
= y_r((n - m)T) + y_l((n + m)T)
$$

$$
= y^+(n - m) + y^-(n + m)
$$

where

$$y^+(n) \triangleq y_r(nT) \quad y^-(n) \triangleq y_l(nT).$$

In this equation, $y^+(n - m)$ can be regarded as the output of m-sample delay of $y^+(n)$, and $y^-(n + m)$ also can be regarded as the input of m-sample delay of $y^-(n)$. Fig. 1.4 shows a
The ideal vibrating string can be simulated as a bidirectional delay line by integrating one-sample delay components which do not have any output of the waveguide model. Fig. 1.5 shows a more simplified diagram.

Damping

Damping and dispersion factors are required to be added to the ideal string model in order to simulate a real vibrating string. Damping means that the vibration magnitude decreases over time due to several reasons such as friction with the string. The loss actually exists complicatedly in the frequency domain, but it can be approximated as a constant value independent of frequency in the simplest case. In the wave equation, the loss can be expressed by adding damping factor $u$:

$$K'y'' = \epsilon \ddot{y} + uy.$$  

The modified wave equation also makes the solution of two traveling waves have decay factors which decreases exponentially as

$$y(t, x) = e^{-(u/2c)x/c} y_l(t - x/c) + e^{(u/2c)x/c} y_l(t + x/c).$$
This causes traveling waves to decay with a factor $g$:

$$\begin{align*}
y(t, x) &= g^{-m}y^+(n-m) + g^m y^-(n+m) \\
g &= e^{-uT/2}\epsilon.
\end{align*}$$

Because linear time-invariant system allows commutativity, each $g$ can be lumped into $g^m$ at a specific point (Fig. 1.6). In real environment, losses increase with the frequency, so $g^m$ can be replaced by a frequency-dependent factor $G(w)$. Zero-phase FIR filter is mostly used for $G(w)$ to have zero phase delay [42].

![Figure 1.6: Simulation of the ideal waveguide having decay factors to describe damping](image)

Dispersion

Dispersion means that the speed of wave depends on frequency because of stiffness in a vibrating string which occurs a restoring force. In the wave equation, it can be expressed by adding the moment constant $\kappa$:

$$Ky'' - \kappa y''' = c\ddot{y}$$

This makes the factor of wave speed $c$ depend on the frequency as $c(w)$ as

$$c(w) \equiv c_0(1 + \frac{\kappa w^2}{2Kc_0^2})$$

where $c_0$ is the wave speed when the stiffness does not exist. The wave speed depending on the frequency is related to the delay in the frequency domain, so allpass filter is used to implement the dispersion because allpass filter has a constant gain and non-uniform phase delay for all frequencies.

Rigid termination

Also, rigid termination meaning that the string is tightly bound at the termination can be designed in this model. The string cannot vibrate at the termination, so it provides the boundary conditions
\( y(t, L \triangleq \frac{N}{2}) = 0 \) and \( y(t, 0) = 0 \) in the wave equation. From the conditions, the equation of traveling wave becomes

\[
y^+(n) = -y^-(n)
y^-(n + \frac{N}{2}) = -y^+(n - \frac{N}{2})
\]

Therefore, sign inversion occurs during reflection at each termination (Fig. 1.7).

**Figure 1.7:** Simulation of the ideal vibrating string with rigid terminations

**SDL model**

Karjalainen et al. showed that the bidirectional digital waveguide model could be reduced to a generalized form of KS algorithm called the single-delay loop(SDL) model\[26\]. Based on the partial transfer functions, the overall transfer function from the input, the excitation in the digital waveguide model having a bidirectional delay line, to the output in the bridge is solved. It results in several filters forming a cascade structure, so the SDL model is more efficient and easy to implement. Fig. 1.8 shows the basic structure of the SDL model. KS algorithm is a special case of the SDL model that random numbers fills the initial contents of the delay line and loop filter averages two adjacent samples as

\[
y(n) = \frac{x(n) + x(n - 1)}{2}.
\]

**Figure 1.8:** Structure of the single-delay loop(SDL) model
1.2.5 Commuted Waveguide Synthesis

In plucked string instruments, plucking causes the string to vibrate and the vibration is transferred to body through the bridge, then the sound is radiated through vibration of the air around the body of the instrument. Thus, the plucked string instrument model has basically three components which represent an excitation signal, a string model, and a body model (Fig. 1.9). Each of them is expressed as a filter in the SDL model and is denoted by $e(n)$, $s(n)$, and $b(n)$.

![Figure 1.9: Structure of plucked string model model](image)

In this model, the transfer function $h(n)$ is expressed as

$$h(n) = e(n) \ast s(n) \ast b(n)$$  \hspace{1cm} (1.6)

where the asterisk denotes discrete convolution. In order to produce sounds from the model, the filters corresponding to each component need to be approximated. However, the order of FIR filter model for modeling the body is very high, so it requires too high computational cost to implement a real-time sound synthesis system.

One solution for the problem is to implement the body filter with sampled impulse response of the body of the instrument and a small number of resonators[6], but quality is not enough to provide the features of the instrument[25].

An alternative solution called commuted waveguide synthesis was proposed by Smith. It provides a simple and efficient synthesis model[51]. Because the instrument model can be considered as linear time-invariant(LTI), Eq. 1.6 allows to commute the components $s(n)$ and $b(n)$ as

$$h(n) = e(n) \ast b(n) \ast s(n).$$

Then, convolution of the excitation signal $x(n)$ and the body model $b(n)$ gives

$$h(n) = x(n) \ast s(n)$$

$$x(n) \triangleq e(n) \ast b(n)$$

where $x(n)$ can be used as the input of the string model and the string model $s(n)$ is built upon digital waveguide theory. In this way, the filter corresponding to the body model need not
be approximated. $s(n)$ exists as a cascade form of several filters, and it can be estimated by analyzing the gain of partials from recorded samples of the instrument. $x(n)$ which combines the excitation signal and the body model is stored as a form of excitation table and used to produce sound through $s(n)$. From digital recordings of the instruments, it can be estimated by inverse filtering of the string model $s(n)$.

1.2.6 Fractional delay filter

Length of the digital delay line $L$ in the SDL model decides the fundamental frequency as

$$L = \frac{f_s}{f_0}$$

$$f_0 = \frac{f_s}{L}$$

where $f_s$ is sampling frequency and $f_0$ is fundamental frequency. Since $L$ is always an integer, $f_0$ must be tuned as a number which make $L$ an integer. It means that not all numbers can be $f_0$. Thus, in order to set $f_0$ as any number, it requires an additional filter called fractional delay filter which has a small phase delay and does not change the loop gain.

Ideal fractional delay filter has a uniform gain for all frequencies and a constant phase delay (linear phase response). To implement the filter in digital domain, first-order allpass filter and Lagrange interpolation FIR filter have been used[23][24].

First-order allpass filter

The transfer function of the first-order allpass filter is given as

$$F(z) \triangleq \frac{C + z^{-1}}{1 + Cz^{-1}}. \quad (1.7)$$

The magnitude response of the allpass filter is

$$G(f) = |F(e^{j\omega T_s})|$$

$$= \frac{|C + e^{-j\omega T_s}|}{|1 + Ce^{-j\omega T_s}|}$$

$$= \frac{|C + 1|}{|1 + C|}$$

$$= 1.$$  

Thus, it has unity gain for all frequencies. Then low-frequency phase delay is approximated as
\[ P(f) = -\frac{\angle F(e^{jwT_s})}{wT_s} = -\frac{1}{wT_s} \angle \frac{C + e^{-jwT_s}}{1 + Ce^{-jwT_s}} = -\frac{1}{wT_s} \{ \angle (C + e^{-jwT_s}) - \angle (1 + Ce^{-jwT_s}) \} = -\frac{1}{wT_s} \{ \tan^{-1}(-\frac{\sin(wT_s)}{C + \cos(wT_s)}) - \tan^{-1}(-\frac{C\sin(wT_s)}{1 + C\cos(wT_s)}) \} \].

By using Maclaurin series expansion \[2\], \( \tan^{-1}(x) \) can be expressed as
\[
\tan^{-1}(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{2n+1} x^{2n+1} \quad \text{for} \quad |x| < 1
\]
\[
= x - \frac{x^3}{3} + \frac{x^5}{5} + \cdots
\]
\[
\approx x \quad \text{for} \quad x \rightarrow 0.
\]
Thus, the low-frequency phase delay can be approximated as
\[
P(f) \approx \frac{1}{wT_s} \left( \frac{\sin(wT_s)}{C + \cos(wT_s)} - \frac{C\sin(wT_s)}{1 + C\cos(wT_s)} \right) \approx \left( \frac{1}{C+1} - \frac{C}{1+C} \right) = \frac{1-C}{1+C}.
\]

Fig. 1.10 shows the magnitude and phase delay responses of an allpass filter for 0.1 sample delay. It has unity gain on frequency domain, and phase delay is 0.1 at low frequency.

**Larange interpolation FIR filter**

The equation of Lagrange interpolation FIR filter is given as
\[
F(z) \triangleq \sum_{i=0}^{N} h_i z^{-i}
\]
where the filter coefficients \( h_n \) of desired delay \( D \) is given as
\[
h_n = \prod_{k=0, k \neq n}^{N} \frac{D - k}{n - k}, n = 0, 1, ..., N.
\]
The filter is derived by the design of maximally flat filter \[58\]. \( N \) derivatives of an error function is zero at the frequency \( w_0 \) as
\[
\frac{d^k E(e^{jw})}{dw^k} \big|_{w=w_0} = 0
\]
Figure 1.10: (a) Magnitude response and (b) phase delay of allpass filter for the delay of 0.1 sample

where \( E(e^{jw}) \) is the difference between the FIR filter and the ideal fractional delay filter of desired delay \( D \) as

\[
E = \sum_{i=0}^{N} h_i e^{-jwi} - e^{-jwD}.
\]

To set the \( E \) as 0 at zero frequency, \( w_0 \) is set as 0. For \( k = 0 \), Eq. 1.10 becomes

\[
\sum_{i=0}^{N} h_i - 1 = 0
\]

\[
\sum_{i=0}^{N} h_i = 1.
\]

Thus, sum of the coefficients of the FIR filter should be equal to 1. Based on this fact, it can be said that gain of the filter at zero frequency is unity.

For \( k = n \), Eq. 1.10 becomes

\[
\sum_{i=0}^{N} i^n h_i - D^n = 0
\]

\[
\sum_{i=0}^{N} i^n h_i = D^n.
\]
To solve the Eq. 1.10, \( N + 1 \) linear equations needs to be collected as

\[
\sum_{i=0}^{N} i^k h_i = D^k, \quad k = 0, 1, 2, \ldots, N.
\]

These can be expressed as a matrix form as

\[
Ah = D \quad (1.11)
\]

where

\[
A = \begin{bmatrix}
0^0 & 1^0 & 2^0 & \cdots & N^0 \\
0^1 & 1^1 & 2^1 & \cdots & N^1 \\
0^2 & 1^2 & 2^2 & \cdots & N^2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0^N & 1^N & 2^N & \cdots & N^N
\end{bmatrix},
\]

\[
h = \begin{bmatrix}
h_0 \\
h_1 \\
h_2 \\
\vdots \\
h_N
\end{bmatrix}^T,
\]

\[
D = \begin{bmatrix}
D^0 & D^1 & D^2 & \cdots & D^N
\end{bmatrix}^T.
\]

Eq. 1.11 is solved as

\[
h = A^{-1}D
\]

and by Cramer’s rule, \( h_i \) is

\[
h_i = \frac{\det D_i}{\det A}
\]

where \( D_i \) is the matrix which replaces \( i \)th column of \( A \) with the vector \( D \) as

\[
D_i = \begin{bmatrix}
0^0 & 1^0 & D^0 & \cdots & N^0 \\
0^1 & 1^1 & D^1 & \cdots & N^1 \\
0^2 & 1^2 & D^2 & \cdots & N^2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0^N & 1^N & D^N & \cdots & N^N
\end{bmatrix}.
\]
Therefore, the solution is given as a form of Eq. 1.9. For example, the filter coefficients of the Lagrange interpolation FIR filter for $N = 3$ are given as

\[
\begin{align*}
    h_0 &= -\frac{1}{6}(D-1)(D-2)(D-3), \\
    h_1 &= \frac{1}{2}D(D-2)(D-3), \\
    h_2 &= \frac{1}{2}D(D-1)(D-3), \\
    h_3 &= -\frac{1}{6}D(D-1)(D-2).
\end{align*}
\]

As Fig. 1.11(a), magnitude response of the Lagrange interpolation FIR filter is flat around a particular frequency (low frequency). Because odd-order interpolators have flat group delay and flat magnitude response, typically third or fifth order interpolators are used[40].

Fractional delay filter of a first-order allpass filter has a simple form, and it gives sufficient approximation of delay in high sampling frequency. However, when parameters of a IIR filter vary while calculating the values, the transient effect occurs as output of the filter is similar to the output due to rapid modification in the input signal[60]. Because it causes an adverse effect on quality of the synthesized sound, Lagrange interpolation is used when variable digital delay line is required.
1.3 Related work

Since the digital waveguide theory was proposed, many studies have been proposed to synthesize the sounds of various instruments. They include various kinds of instruments such as wind instrument [28][45][46] and bowed string instrument [64][48], but I focused on plucked string instrument because the geomungo fall into the category.

1.3.1 Physical modeling of plucked string instruments

Välimäki et al. proposed a sound synthesis model based on the digital waveguide theory for several plucked string instruments such as guitar, the banjo, the mandolin, and the kantele [59]. They adopted commuted waveguide synthesis and Lagrange interpolation for variable delay line, and proposed a real-time synthesis method with a signal processor.

An improved model was proposed by Välimäki et al. to implement a guitar synthesizer [62]. Instead of the cascade model of commuted waveguide synthesis, they used the parallel model in which two resonators producing low resonances were separated from the excitation signal. This method could reduce the length of the excitation signal and parametrize the low body resonances.

Piano has a similar structure to other plucked string instruments, but a hammer instead of a finger is used to strike the string. Based on the structure of piano, Bank et al. proposed a synthesis model for piano sounds [4]. They discussed several models for interaction between the hammer and the string, loss filter design for minimization of the decay time error, high-order dispersion filter design for inharmonicity, the multirate resonator bank for coupled piano strings, and the multirate soundboard model. A commuted waveguide model for piano was also proposed [54]. It is a simplified synthesis model reducing the computational cost.

Välimäki et al. also tried to synthesize the sounds of Harpsichord based on modification of commuted waveguide synthesis [61]. In this model, a second-order resonator was coupled with the string model to produce the beating effect. Also, a ripple filter combined with an one-pole filter was used to design the loss filter, and a soundboard filter as well as release samples were added to the basic digital waveguide model.

A sound synthesis model for the Finnish Kantale was proposed based on digital waveguide modeling [14]. It considered tension modulation which was caused by string elongation because of transverse wave [57]. This nonlinear phenomenon was realized by approximating elongation and controlling
the parameter of fractional delay filter in real time.

Recent works combined the digital waveguide model and the plectrum model that the player physically interacted with the string\textsuperscript{[20]}[18]. The string model used transverse displacement caused by the plectrum as an input.

1.3.2 Physical modeling of asian string instruments

Most studies about physical modeling focused on synthesizing the sounds of western musical instruments, but some synthesis models of asian stringed instruments have been proposed. Sound production mechanisms of western and eastern plucked string instruments are generally similar, but difference in structures and playing styles creates different timbres.

Erkut et al. proposed a synthesis model of the Ud and the Renaissance Lute which were spread from the Middle East\textsuperscript{[15]}. In this work, they showed the implementation of the glissando effect, a glide from one pitch to another, in fretted/fretless instruments.

Model-based sound synthesis algorithm of the guqin, a Chinese plucked string instrument, was proposed by Penttinen et al.\textsuperscript{[41]}. Based on the commuted digital waveguide synthesis method, the synthesis model of the guqin included a body model filter, a ripple filter for flageolet tones, an additional SDL string model for inharmonic partials called phantom partials, and a friction model.

Recently the model-based sound synthesis of the Dan Tranh, a Vietnames plucked string instrument, has been reported\textsuperscript{[10]}. Estimated loop gain values and loop coefficient values of the Dan Tranh were compared with those of the Gayageum, a Korean traditional plucked string instrument, presented in \textsuperscript{[9]}. The result showed that loop coefficient values were different due to the differences in structure and playing style.
Chapter 2. Analysis

In order to synthesize the sounds of the geomungo using digital waveguide theory, analysis of the acoustical characteristics of geomungo tones was done first. After extracting geomungo tones from geomungo songs, frequency response and behavior of fundamental frequency over time were estimated.

2.1 Extraction of geomungo tones

Geomungo tones with or without fluctuations in pitch by different right hand techniques were recorded in the studio of Graduate School of Culture Technology, KAIST. Some geomungo tones were also extracted from several geomungo songs: dalmoori (ring around the moon), ilchool (sunrise), etc. Fig. 2.1 shows four examples of geomungo tones.

Tone of Fig. 2.1a makes a sound of a typical plucked string instrument like guitar. A big difference between Fig. 2.1a and Fig. 2.1b ∼ 2.1d is that fluctuations of pitch are very great in the latter (more than 20Hz). Without listening to the tones, the difference can be easily found in the figures because the amplitude of an ideal tone of a plucked string instrument decreases exponentially. Fig 2.1a exemplifies the characteristic well, whereas amplitudes decrease irregularly in Fig 2.1b ∼ 2.1d. Thus, I need to consider these acoustical characteristics of the geomungo in the sound synthesis model.

2.2 Frequency response

Analysis of the signals in frequency domain is required to design a loop filter. At first, I measured the magnitude spectrums of the extracted geomungo tones by the discrete Fourier transform (DFT) [53], which is defined by

\[ X_k = \sum_{n=0}^{N-1} x(t_n)e^{-j\omega_k t_n}, \quad k = 0, 1, 2, ..., N - 1. \]

Fig. 2.2 shows the results of the transform computed with the fast Fourier transform (FFT) algorithm of the tones of Fig. 2.1. From the figures, I could roughly find the distribution of
harmonic partials in frequency domain. In addition, the fluctuations of pitch are also shown in Fig. 2.2b ∼ 2.2d as the peaks in the figures are less clear than in Fig 2.2a.

As discussed in 1.2.4, each decay factor $g$ is lumped into $g^n$ at one point and is replaced by

Figure 2.1: Extracted geomungo tones
a filter $G(w)$. Thus, I need to design the filter by estimating the damping factors which mean the transition of the magnitude at harmonic frequencies of sound signals. The estimation can be achieved by the discrete-time short-time Fourier transform (STFT) \[3\], which is defined by

$$Y_m(k) = \sum_{n=0}^{N-1} w(n)y(n + mH)e^{-2\pi jkn/N},$$

$m = 0, 1, 2, ..., k = 0, 1, 2, ..., N - 1$

where $N$ is FFT length, $w(n)$ is a window function, and $H$ is the number of overlapping segments. Fig. 2.3 shows STFT results of the geomungo tones of Fig. 2.1. In this work, hamming window is used as a window function, $N$ is the next power of 2 greater than the length of each tone, and $H$ is 0.5 times the window size of 8192.

Similar to the FFT results, pitch fluctuations are easily found in Fig. 2.2b $\sim$ 2.2d as the peak width of each harmonic partial becomes wider. I could also find that amplitudes of the harmonic partials of the frequencies more than about 1000Hz decrease rapidly in the beginning. The phenomenon is common to all geomungo tones, so it needs to be considered in the design of the loss filter.

### 2.3 Estimation of fundamental frequency

As mentioned in 1.2.6, the fundamental frequency $f_0$ is used to calculate length of the delay line $L$ and coefficients of the fractional delay filter. In order to estimate the fundamental frequency, cumulative mean normalized difference function proposed by Cheveigné et al.\[13\] is used.

When the signal $x_t$ is periodic with $T$, it can be mathematically defined by

$$x_t - x_{t+T} = 0$$

for all $t$. Thus, sum of the squares of the difference over a window of length $W$ is also 0 as

$$\sum_{j=t}^{t+W-1} (x_j - x_{j+T})^2 = 0.$$

From the equation, the difference function $d_t(\tau)$ can be defined as

$$d_t(\tau) = \sum_{j=t}^{t+W-1} (x_j - x_{j+\tau})^2.$$

Using the difference function $d_t(\tau)$, I can find the value of $\tau$ for which $d_t(\tau)$ is the lowest (it cannot be zero in all cases because signals are not perfectly periodic). However, as Cheveigné et al., without a lower limit on the search range, it may choose zero-lag. Also, in imperfect periodic
cases, it can choose low lag instead of desired lag. Thus, they proposed an alternative called cumulative mean normalized difference function:

\[
d'_t(\tau) = \begin{cases} 
1, & \text{if } \tau = 0 \\
\frac{d_t(\tau)}{(1/\tau) \sum_{j=1}^{\tau} d_t(j)} & \text{otherwise.}
\end{cases}
\]

It avoids selection of zero-lag and low lags by defining the value of the function as 1 at zero-lag. Thus, it reduces error rate and removes the necessity of upper limit because \(d'_t(\tau)\) remains below 1. The fundamental frequency at \(\tau\) is estimated as

\[
f_0(\tau) = \frac{f_s}{d'_t(\tau)}.
\]

Using the cumulative mean normalized difference function, Fig. 2.4 shows fundamental frequency curves of the geomungo tones of Fig. 2.1. The x-axis presents time and the y-axis shows estimated fundamental frequency. Compared with Fig. 2.4a, Fig. 2.4b ～ 2.4d show great pitch fluctuations in nearly 30Hz.
Figure 2.2: Frequency response of geomungo tones
Figure 2.3: STFT (short-time Fourier transform) analysis of geomungo tones
Figure 2.4: Fundamental frequency estimation of geomungo tones
Chapter 3. General sound synthesis model

3.1 Synthesis model

Fig. 3.1: Overall structure of the sound synthesis model for geomungo

Fig. 3.1 shows the structure of a sound synthesis model for the geomungo. It is similar to the SDL model of Fig. 1.8 but excitation samples are used as input sound signals and the loop filter in the SDL model is composed of fractional delay filter and loss filter. Because this model does not consider variable pitch, re-synthesis of the constant pitch tone of Fig. 2.1a is the purpose in this chapter.

From Fig. 3.1 the transfer function of the synthesis model $T(z)$ can be expressed as

$$
Y(z) = Y(z)S(z) + X(z)
$$

$$
Y(z)(1 - S(z)) = X(z)
$$

$$
T(z) = \frac{Y(z)}{X(z)} = \frac{1}{1 - S(z)}
$$

where

$$
S(z) = z^{-1}H(z)F(z).
$$

(3.1)

3.1.1 Loss filter

Fig. 3.2: Structure of the loss filter $H(z)$
The loss filter in the synthesis model is implemented as a combination of an one-pole filter and a ripple filter suggested by Välimäki et al. [61]. One-pole filter is a simplified IIR low-pass filter which is designed to simulate the damping. The equation is expressed as

$$H_1(z) = g(1+a) \frac{1}{1 + az^{-1}}$$

where $g$ is the gain at 0Hz (overall loop gain) and $a$ is the feedback gain which decides cutoff frequency of the filter. Fig. 3.3 shows the magnitude response of the one-pole filter for $a = -0.01$ and $g = 1$.

$$H_2(z) = r + z^{-R}$$

where $r$ is the ripple depth, and $R$ is the length of the delay line in the ripple filter. Fig. 3.4 shows the magnitude response of the ripple filter for $r = 0.002$ and $R = 20$.

Then the loss filter is determined by combining the one-pole filter and the ripple filter. The transfer function of the loss filter is

$$H(z) = H_1(z)H_2(z)$$

$$= g(1+a) \frac{r + z^{-R}}{1 + az^{-1}}.$$  \hspace{1cm} (3.2)

$$= g(1+a) \frac{r + z^{-R}}{1 + az^{-1}}.$$  \hspace{1cm} (3.3)

Fig. 3.2 shows the structure of loss filter. Fig. 3.5 shows the magnitude response of the loss filter. Actually, it is not stable because the gain at 0Hz exceeds 1.
Figure 3.4: Magnitude response of the ripple filter $H_2(z)$ for $r = 0.002$ and $R = 20$

Figure 3.5: Magnitude response of the loss filter $H(z)$ for $a = -0.01, g = 1, r = 0.002$ and $R = 20$

To ensure the stability, $g$ is slightly less than 1 and $a$ is slightly less than 0. $a$ is used to implement a low-pass filter because high-frequency partials decay faster than low-frequency partials in low-pass filters (it is a typical phenomenon in plucked string instruments). Also, $r$ is used for ripple filter which oscillates the magnitude response of the one-pole filter. $R$ is calculated as

$$R = \text{round}(r_{rate}L)$$

where $r_{rate}$ is the ripple rate parameter.

3.1.2 Fractional delay filter

Figure 3.6: Structure of the fractional delay filter $F(z)$

In 1.2.6 I introduced two types of fractional delay filters: first-order allpass filter and Lagrange
interpolation FIR filter. In this work, the allpass filter is used as a fractional delay filter because it has a simple structure to approximate the fractional delay and unity magnitude response for all frequencies. Fig. 3.6 shows the structure of the fractional delay filter.

3.1.3 Delay line

In the original digital waveguide theory, the length of delay line $L$ is decided as

$$L = \frac{f_s}{f_0}.$$  

In the proposed sound synthesis model, I estimate the fundamental frequency $\hat{f}_0$ and use it instead of $f_0$ as

$$L = \frac{f_s}{\hat{f}_0}.$$  

Other filters used in the model also have phase delays, so the delay line length $L_1$ can be defined as

$$L_1 = L - P_F - P_L = \frac{f_s}{\hat{f}_0} - P_F - P_L$$

where $P_F$ and $P_L$ are the phase delay of the fractional delay filter and the loss filter. In this model, $P_F$ is $\frac{1-C}{1+C}$ and $P_L$ is $R + 1$, so

$$L_1 = \frac{f_s}{\hat{f}_0} - R - 1 - \frac{1-C}{1+C}.$$  

3.2 Calibration

In this section, I estimate the parameters and extract an excitation sample for the synthesis model. Based on the delay line length, frequency-dependent gains at the harmonic frequencies are estimated, and they are used to approximate the coefficients of the loop filter. Then, input signal is gathered by the inverse loop filter.

In 2.3, I estimated the fundamental frequency of geomungo tones. I decided 63.1Hz as the fundamental frequency of Fig. 2.2a by average. Because $f_s$ is 44100Hz in this work, $L$ is 698.89.

3.2.1 Estimation of the frequency-dependent loss

In order to know the loop gains at the harmonic frequencies, the envelope curves meaning the sequences of magnitude values of the harmonic partials over time need to be approximated as
Figure 3.7: Envelope curves and approximated straight lines of 10 lowest harmonics of Fig 2.1a

straight lines on a decibel scale because the magnitudes of the harmonics decrease exponentially
(thus, they decrease linearly on a decibel scale). Thus, I used STFT results of the geomungo
tones in 2.2 and a polynomial curve fitting function to approximate the curve as a linear function.
Graph of the results is shown in Fig. 3.7 and slopes of the approximated line are shown in Table 3.1.

In addition, another method was proposed to estimate the frequency-dependent loss[33]. By
applying a bandpass filter, a single harmonic partial can be obtained. Local maximum values

<table>
<thead>
<tr>
<th>Harmonic index</th>
<th>$F_0 = 63.1$Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.1691</td>
</tr>
<tr>
<td>2</td>
<td>-0.1264</td>
</tr>
<tr>
<td>3</td>
<td>-0.0957</td>
</tr>
<tr>
<td>4</td>
<td>-0.0640</td>
</tr>
<tr>
<td>5</td>
<td>-0.0683</td>
</tr>
<tr>
<td>6</td>
<td>-0.0792</td>
</tr>
<tr>
<td>7</td>
<td>-0.1030</td>
</tr>
<tr>
<td>8</td>
<td>-0.0963</td>
</tr>
<tr>
<td>9</td>
<td>-0.0634</td>
</tr>
<tr>
<td>10</td>
<td>-0.1118</td>
</tr>
</tbody>
</table>

Table 3.1: Slopes $\beta_k$ of approximated straight lines of envelop curves
are obtained from the bandpassed signal and the amplitude envelope is acquired based on the values. Because this method was originally proposed for a hybrid waveguide model, so details are discussed later in the hybrid model chapter. Fig 3.8 compares the losses estimated by the two methods and it can be found that the results are almost same.

Figure 3.8: Comparison of the estimated losses of the (a) 1st partial and (b) 7th partial. Blue lines are the losses estimated by STFT, green lines are maximum values of the bandpassed signals, and red dot lines are approximated straight lines

The loop gains at the harmonic frequencies are calculated by using the slopes of the envelopes as

\[ G_k = 10^{\beta_k L/20H}, \ k = 1, 2, ..., N \]

where \( \beta_k \) are the slopes of the approximated lines, \( H \) is the number of overlapping segments used in STFT of the geomungo tones, and \( L \) is \( f_s/f_0 \). Values of the estimated loop gains are shown in Table 3.2 and they are used to design the loss filter in next section.

### 3.2.2 Design of loss filter

In order to design the loss filter of Eq. 3.3, the loop coefficients \( a \) and \( g \) for the one-pole filter \( H_1(z) = g(1 + a)/(1 + az^{-1}) \) are estimated first, then \( r \) and \( R \) for the ripple filter \( H_2(z) = r + z^{-R} \) are calculated.
Table 3.2: Estimated loop gains $G_k$

<table>
<thead>
<tr>
<th>Harmonic index</th>
<th>$F_0 = 63.1$Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9901</td>
</tr>
<tr>
<td>2</td>
<td>0.9852</td>
</tr>
<tr>
<td>3</td>
<td>0.9832</td>
</tr>
<tr>
<td>4</td>
<td>0.9856</td>
</tr>
<tr>
<td>5</td>
<td>0.9807</td>
</tr>
<tr>
<td>6</td>
<td>0.9731</td>
</tr>
<tr>
<td>7</td>
<td>0.9594</td>
</tr>
<tr>
<td>8</td>
<td>0.9565</td>
</tr>
<tr>
<td>9</td>
<td>0.9674</td>
</tr>
<tr>
<td>10</td>
<td>0.9373</td>
</tr>
</tbody>
</table>

**One-pole filter design**

To design the filter that matches with the desired magnitude at the harmonic frequencies, a high-order filter instead of an one-pole filter needs to be designed, but it cannot be used in real-time applications. Thus, I approximated the gains by using the weighted least-squares method in which the solution minimizes the sum of products of an error weighting function and squares of the errors (differences between the magnitude responses of the filter and the estimated loop gains). The equation is shown as

$$E = \sum_{k=1}^{N} W(G_k)[|H_1(w_k)| - G_k]^2$$  \hspace{1cm} (3.4)

where $N$ is the maximum order of harmonics used to design the filter, $W(G_k)$ is an error weighting function and $w_k$ means $k$th harmonic frequency. In this work, $N$ is 10 and $W(G_k)$ is used as $W(G_k) = \frac{1}{1-G_k}$ to provide a larger weight to the harmonics of larger gain.

The coefficient $g$ of the one-pole filter is generally selected as $G_1$ which is the loop gain value at the fundamental frequency. Then, $a$ is chosen from $-1$ to $0$ which minimizes the error $E$. Fig. 3.9 shows the value of $E$ on the coefficient $a$. From the graph, I can find that $E$ is minimized at $a = -0.7453$ on Fig. 3.9a.

**Ripple filter design**

I used the design method of the ripple filter proposed by Välimäki et al. From the second partial, the partial that has the largest $G_k$ value is selected and I denote the harmonic index by
In this work, $k_{max} = 4$ ($G_4 = 0.9856$).

Then the absolute value of $r$ is determined as

$$|r| = G_{k_{max}} - |H_1(z_{k_{max}})|$$

where $H_1(z_{k_{max}})$ is the value of the one-pole filter $H_1(z)$ at the harmonic of $k_{max}$ ($|r| = 0.0025$).

The sign of $r$ is determined by comparison of the first partial’s gain and the magnitude of the one-pole filter at the first harmonic frequency. If the first partial’s gain is bigger, then the sign of $r$ is positive. Otherwise, the sign of $r$ is negative. Thus, the sign of $r$ is positive in this work because $G_1 = 0.9901 > H_1(w_1) = 0.9896$.

When checking a stability, the filter is stable because $g + r = 0.9926 < 1$. Thus, the sign of $r$ is positive. If the filter was unstable, the sign of $r$ would be inverted. Fig. 3.10 shows the magnitude response of the filter and $G_k$ values.

Finally, in order to set the value of $R$, the ripple rate parameter $r_{rate}$ is determined as $1/k_{max}$ for positive $r$ and $1/(2k_{max})$ for negative $r$. Thus, in this work, $r_{rate}$ is $1/4$ because $r$ is 0.0025.

### 3.2.3 Design of allpass filter and delay line

Because $L$ is 698.89, the fractional delay is 0.89 and the length of the delay line is 698. The value $C$ in the allpass filter of in Eq. 1.7 is decided as
where $fd$ is the value of fractional delay. From the equation, $C$ is 0.058. Fig. 3.11 shows the phase delay of the allpass filter for $C = 0.058$. Phase delay is about 0.089 at the low frequency.

Figure 3.11: Phase delay of the allpass filter for $C = 0.058$
3.2.4 Inverse filtering

As discussed in 1.2.5, an excitation sample which is used as an input signal in the synthesis model can be extracted by applying inverse filtering to the extracted geomungo tone. Thus, this section starts with implementation of the inverse filter of the string model $S(z)$ in Eq.3.1 as

$$S^{-1}(z) = 1 - z^{-L_1} F(z) H(z)$$

$$= 1 - z^{-L_1} \frac{c + z^{-1}}{1 + cz^{-1}} * g(1 + a) \frac{r + z^{-R}}{1 + az^{-1}}$$

$$= \frac{A(z)}{B(z)}$$

where

$$A(z) = 1 + (c + a)z^{-1} + acz^{-2} - g(1 + a)cz^{-L_1} - g(1 + a)rz^{-(L_1+1)} - g(1 + a)cz^{-(L_1+R)} - g(1 + a)z^{-(L_1+1+R)}$$

$$B(z) = 1 + (c + a)z^{-1} + acz^{-2}.$$ 

Fig. 3.12 shows the residual signal extracted by inverse filtering of the geomungo tone. Magnitude spectrum of the residual is illustrated in Fig. 3.13. Compared with Fig. 2.2a, it can be found that the inverse filtering canceled the harmonic partials of the geomungo tone.

![Figure 3.12: Excitation signal extracted from the geomungo tone of Fig. 2.1a](image)

![Figure 3.13: Magnitude spectrum of Fig. 3.12](image)
3.3 Synthesis

The synthesized geomungo tones are shown in Fig. 3.14. It sounds similar to the original signal. In addition, through the magnitude spectrum and the STFT analysis result of the synthesized signal (Fig. 3.15 and 3.16), I can find re-created peaks at harmonic frequencies. Magnitudes of the created peaks are a little different than the original ones due to the difference between loop gains of the original tone at the harmonic frequencies and magnitude of the designed loss filter.

To evaluate the synthesized tone, envelope curves of the tone was estimated (Fig. 3.17). Based on the curves, loop gains of the synthesized tone were estimated and compared with the loop gains of the original tone (Table 3.3 and Fig. 3.18).

Figure 3.14: Synthesized geomungo tone from Fig. 3.12

Figure 3.15: Magnitude spectrum of the synthesized geomungo tone of Fig. 3.14
Figure 3.16: STFT analysis of the synthesized geomungo tone of Fig. 3.14

Figure 3.17: Envelope curves of the synthesized tone of Fig. 3.14
<table>
<thead>
<tr>
<th>Harmonic index</th>
<th>Original tone</th>
<th>Synthesized tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9901</td>
<td>0.9892</td>
</tr>
<tr>
<td>2</td>
<td>0.9852</td>
<td>0.9925</td>
</tr>
<tr>
<td>3</td>
<td>0.9832</td>
<td>0.9960</td>
</tr>
<tr>
<td>4</td>
<td>0.9856</td>
<td>0.9832</td>
</tr>
<tr>
<td>5</td>
<td>0.9807</td>
<td>0.9909</td>
</tr>
<tr>
<td>6</td>
<td>0.9731</td>
<td>0.9707</td>
</tr>
<tr>
<td>7</td>
<td>0.9594</td>
<td>0.9651</td>
</tr>
<tr>
<td>8</td>
<td>0.9565</td>
<td>0.9575</td>
</tr>
<tr>
<td>9</td>
<td>0.9674</td>
<td>0.9684</td>
</tr>
<tr>
<td>10</td>
<td>0.9373</td>
<td>0.9420</td>
</tr>
</tbody>
</table>

Table 3.3: Estimated loop gains $G_k$ between the original tone (Fig. 2.1a) and the synthesized tone (Fig. 3.14).

Figure 3.18: Comparison of estimated loop gains $G_k$ between the original tone and the synthesized tone.
Chapter 4. Time-varying synthesis model

4.1 Synthesis model

The synthesis model discussed in chapter 3.1 works well enough to re-synthesize the sounds of the recorded geomungo tones. However, it needs to be improved to synthesize the tones of fluctuating pitches. Thus, in this section, I propose a revised model to synthesize the geomungo tones more naturally.

The previous model cannot reflect pitch variation because the length of the delay line was fixed in the previous model. However, as discussed in 2.3, the geomungo tones have a great pitch fluctuations. Since a geomungo technique called *Nonghyun* (similar to vibrato and trill, but more dynamic) occurs the phenomenon, the pitch fluctuation is distinguished from pitch glide (pitch variation) meaning that the fundamental frequency of the instrument decreases after plucking due to the transverse vibration\(^\text{[32]}\). It is rather more similar to the glissando effect for the fretless instrument discussed in \([15]\).

![Figure 4.1: Overall structure of the revised sound model](image)

Fig. 4.1 shows the revised model to reflect the fluctuating pitch. All filters used in this model are time-varying filters to synthesize variable pitch sounds. The fractional delay filter is changed to the Lagrange interpolation FIR filter discussed in \([1.2.6]\) because it is more favorable for variable delay. Length of the delay line is changed to be time-varying. Also, a gain factor \(g_c\) is added to the synthesis model to preserve the energy when the pitch is altered \([40]\). The equation of the \(g_c\)
is expressed as
\[
g_c = \sqrt{1 - \Delta x} \\
\approx 1 - \frac{\Delta x}{2} \text{ for small } x \\
(\Delta x = x_n - x_{n-1})
\]
where \(x_n\) is the position in the string at time index \(n\) and \(\Delta x\) is the change of the delay line length.

### 4.1.1 Loss filter

![Diagram of Loss Filter](image)

Fig. 4.2 shows the structure of the time-varying loss filter \(H(z)\). A ripple filter used in the previous loss filter model is removed in this model because it decreases the sound quality of the synthesis model when changing the filter coefficients. Also, the coefficients in the filter need to be changed as time dependent values because amplitudes of the variable pitch sounds such as Fig. 2.1b and 2.1d do not decrease exponentially as discussed in 2.1. There are two coefficients in this filter: the loop gain \(g\) and the feedback gain \(a\). However, changing the feedback gain results in transients effects which create audible clicks [16]. Thus, the loop gain coefficient \(g\) is changed to depend on time as \(g(n)\) and a transfer function of the time-varying loss filter is expressed as
\[
H(z) = g_n (1 + a) \frac{1}{1 + az^{-1}}.
\]

### 4.1.2 Fractional delay filter

As discussed in 1.2.6, Lagrange interpolation FIR filter is better for variable delay line. Because the frequency changes over time, the desired fractional delay factor \(D\) is changed as a time-dependent value \(D(n)\) as
\[
D(n) = \frac{f_s}{f_0(n)} - \text{floor} \left[ \frac{f_s}{f_0(n)} \right].
\]
Therefore, the fractional delay filter coefficients are altered as \(h_0(n), h_1(n), h_2(n),\) and \(h_3(n)\) (Fig. 4.3).
4.1.3 Delay line

In this time-varying sound model, the length of the delay line also depends on time as

\[ L_1(n) = \frac{f_s}{f_0(n)} - D(n) = \text{floor}\left(\frac{f_s}{f_0(n)}\right) \]

where \(D(n)\) is the phase delay of the fractional delay filter.

4.2 Calibration

The purpose of this chapter is to synthesize the geomungo tone of \(2.1\). Thus, filter coefficients need to be estimated. As Fig. \(2.3b\), magnitudes of the harmonics with frequencies more than 1000Hz are very small and decrease rapidly, so I measured the envelope curves of 5 lowest harmonics (fundamental frequency is 170Hz) to design the loss filter (Fig. \(4.4\)). From the estimated loop gains, \(g\) and \(a\) in the loss filter were calculated as 0.9923 and -0.6635.

Two different ways can be used to extract the excitation signals. The first method is to extract the residual signal (Fig. \(4.6\)) from a geomungo tone which has similar timbre and constant pitch (Fig. \(4.6\)).
The second method is to use the target signal itself. Since the frequency of the inverse filter does not depend on time but the frequency of the geomungo tone varies depending on time, the vibrating sound after the plucking sound cannot be inverse filtered. Thus, length of the residual signal should be shortened by deamplifying and removing the vibrating sound after the plucking sound. Fig. 4.7 shows the residual signal generated by the second method.

Calibration method used in this section is same as 3.2. Because magnitudes of the harmonics with frequencies more than 1000Hz are also very small, loop gains of 5 lowest harmonics were estimated for inverse filtering.

![Figure 4.5: A geomungo tone having constant pitch for generating an excitation signal](image)

![Figure 4.6: Excitation signal extracted from the geomungo tone of Fig. 4.5](image)

![Figure 4.7: Excitation signal extracted from the geomungo tone of Fig. 2.1b](image)
4.3 Synthesis

Fig. 4.8 shows the synthesized tone from the excitation signal of Fig. 4.6 without the gain factor \( g_c \) and variable loop gain in the loss filter. As the figure, magnitude decreases exponentially, so gain control is required because magnitude of the target geomungo tone of Fig. 2.1b does not decrease exponentially. Gain control is important since it decides the degree of vibrato when hearing the synthesized geomungo tones. With the gain factor, the tone is generated as Fig. 4.9. Because it does not alter the amplitude sufficiently, variable loop gain in the loss filter is required in this case.

![Figure 4.8: Synthesized geomungo tone without gain control](image)

![Figure 4.9: Synthesized geomungo tone with control of the gain factor](image)

The variable loop gain coefficient can be estimated as a sinusoid function. Fig. 2.1b is considered as sum of exponentially decreasing signal and sinusoidally oscillating signal. It can be easily found by Fig. 4.11 which subtracts the synthesized signal with the constant gain from the target geomungo tone. The waveform oscillates frequently in the sustain part which starts at 0.5 second after the attack. Thus, by estimating the distance between peaks (Fig. 4.12), period of the sinusoid function could be calculated and amplitude was decided from Fig. 4.11. In this case, period is about 0.25s and phase is 0 at 1.05s, so the sinusoid function could be decided as

\[
C \cos(2\pi ft + \theta) = C \cos(8\pi t - 26.39)
\]
Because the loop gain $g_n$ means the decay rate of the synthesized signal, so differentiation of the sinusoid function was added to the constant loop gain value to define $g_n$ as

$$g_n = g - C' \sin(8\pi t - 26.39).$$

where $g$ means the constant loop gain value estimated by inverse filtering and $C$ is used as 0.05 in this case. Fig. 4.13 shows the calculated loop gain coefficient.

The variable loop gain coefficient can also be compared with the envelop curve. Fig. 4.10 shows the envelope curves of the synthesized tone without gain control, and the curves do not fluctuate enough to hear the vibrato. However, in Fig. 4.14 it can be found that the red solid curve, the envelop curve of the second lowest harmonic, oscillates about 10dB while decaying. Positions of the curve are almost consistent with the positions of the peaks chosen in Fig. 4.12 (1.05 and 1.3 second).

![Figure 4.10: Envelope curves of the synthesized geomungo tone without gain control](image1)

![Figure 4.11: Magnitude difference between the target signal and Fig 4.9](image2)

The sinusoid function for the variable loop gain coefficient is also consistent with the loop gain curve at the second lowest harmonic frequency after 0.5 second (Fig. 4.14), which is calculated based on a series of the slopes of the lines approximated in a narrow range of the envelope curve.
It may use the loop gain curve itself as a variable loop gain coefficient, but it makes the filter unstable and amplitude of the generated signal exceeds unity and the envelop curves of other harmonic frequencies do not fluctuate significantly (the loop gain value decides the decay rates of all harmonic partials). It means that frequency of the approximated sinusoidal function should be similar to the loop gain of the second lowest harmonic, but amplitude of the function should be smaller.

The fundamental frequency curve measured in 2.3 also needs to be modified for controlling the length of delay line and the fractional delay filter because there was a slight difference in the fundamental frequency of the synthesized tone. The modification is applied a little differently for
each sound, but the frequency curve was pushed back by 1000 samples and increased by 5Hz in this case. Finally, the waveform of the re-synthesized geomungo tone of Fig. 4.1b from Fig. 4.5 is shown in Fig. 4.15. Also, Fig. 4.16 shows the synthesized geomungo tone from Fig. 4.7.

![Figure 4.15: Synthesized geomungo tone from the residual signal of Fig. 4.6](image1)

![Figure 4.16: Synthesized geomungo tone from the residual signal of Fig. 4.7](image2)

Envelope curves of Fig. 4.15 is shown in Fig. 4.17. Compared with the envelope curves of the original geomungo tone (Fig. 4.4), the blue curve, the envelope curve of the first lowest harmonic, oscillates significantly while decaying instead of the red curve, the second lowest harmonic. The difference shows the limitation of the digital waveguide modeling that each harmonic cannot be controlled separately because harmonics are created using a digital delay line.

![Figure 4.17: Envelope curves of the synthesized geomungo tone of Fig. 4.15](image3)
Fig. 4.21 shows another re-synthesized geomungo tone of Fig. 4.18 by using Fig. 4.19 (Excitation signal is shown in Fig. 4.20). In this case, gain factor $g_c$ was required but variable loop gain coefficient in the loss filter was not necessary.

Figure 4.18: Target geomungo tone having fluctuating pitch

Figure 4.19: Geomungo tone having constant pitch for re-synthesis of Fig. 4.18

Figure 4.20: Excitation signal extracted from the geomungo tone of Fig. 4.19
Figure 4.21: Synthesized geomungo tone based on Fig. 4.18
Chapter 5. Hybrid modal-waveguide synthesis model

Lee et al. proposed hybrid approach which uses the additive synthesis for generating the harmonic partials at low frequencies and digital waveguide synthesis for generating the harmonic partials at high frequencies [33]. This model can create more accurate partials because the additive synthesis generates each partial separately with sinusoidal oscillators and it makes it possible to control gain and frequency of each partial. On the other hand, the harmonic partials created using a delay line in the digital waveguide synthesis have the integer multiple frequencies and same decay rate, but the actual geomungo tone shows inharmonicity that the fundamental frequencies of harmonic partials are a little different. Therefore, the hybrid synthesis method is considered as a more effective method to synthesize the geomungo tones which have irregular decays and fluctuant frequencies.

5.1 Synthesis model

Fig. 5.1 shows the hybrid model to synthesize the geomungo tone of Fig. 2.1c which has fluctuant frequencies. The harmonic partials at high frequencies are synthesized based on the digital waveguide model in 3.1 from the high-pass filtering of the excitation sample, and the formers at low frequencies are synthesized with six digital resonators. The additive synthesis part is similar to the model proposed by Lee et al. [33], but I added a gain factor $g_c$ for each resonator for irregular decay of the geomungo sound and the resonant frequency of each resonator was controlled based on the fundamental frequency estimation of each harmonic partial.

5.1.1 Resonator

Fig. 5.2 shows a second-order resonator used to synthesize each partial. The transfer function of the resonator is defined as

$$H_{RES}(z) = \frac{a}{1 - p(n)z^{-1}} + \frac{\bar{a}}{1 - \bar{p}(n)z^{-1}}$$

(5.1)
In this equation, $\sigma, \theta, r$ mean the exponential decay rate, initial phase, and initial magnitude of the partial. The frequency of the resonator depends on discrete time as $f(n)$. With the factors, Eq. 5.2 is solved as

$$H_{RES}(z) = \frac{re^{j\theta} \left( \frac{1}{2} r ge^{j(\theta-w(n))} + \frac{1}{2} r ge^{j(w(n)-\theta)} \right) z^{-1}}{1 - 2g\cos(w(n))z^{-1} + g^2 z^{-2}}.$$
Fig. 5.3 shows a signal generated by the resonator. A clear peak at 200Hz can be found in the magnitude response (Fig. 5.4).

Figure 5.3: Waveform of a signal generated by the resonator when the initial magnitude $r = 1$, initial phase $\theta = 0$, exponential decay rate $g = 0.9999$, and frequency is 200Hz

Figure 5.4: Magnitude response of the signal Fig. 5.3

5.2 Calibration

5.2.1 Extraction of harmonic partials

In order to generate harmonic partials, parameters of the resonators are required to be estimated from the harmonic partials of the geomungo tone of Fig. 2.1c. Each partial can be extracted by a band-pass filter, which was designed using Kaiser window FIR filter design approach, a technique to design a multiband FIR filter with cutoff frequencies and ripple parameters [39]. Fig. 5.5 shows the magnitude response of a bandpass filter to extract the third harmonic partial. Passband exists around 480Hz, and other stopband regions are attenuated below -100dB. Using the designed bandpass filter, 6 lowest harmonic partials were extracted as shown in Fig. 5.6. The peaks are shown in the magnitude spectrums of the partials (Fig. 5.7) and revealed better when compared
the magnitude response of the sum of the extracted partials with the magnitude response of the
geomungo tone (Fig. 5.8).

Figure 5.5: Magnitude response of the bandpass filter designed using Kaiser window FIR filter
design approach for the third harmonic partial of the geomungo tone.

5.2.2 Estimation of decay rate and initial magnitude

The second method to estimate the decay rates of the harmonic partials discussed in 3.2.1 is
also used to estimate the exponential decay rate and initial magnitude of each partial. In the
waveforms, the positive local maximum points are obtained by taking the positive points of zero
derivative (Fig. 5.9). A straight line can be fitted from the positive local maximum points on a
decibel scale (Fig. 5.10). Then the slope of the fitted line becomes the exponential decay rate, and
the value of the line at $t = 0$ becomes the initial magnitude. To use the exponential decay rate
for the filter, it needs to be calculated as $\sigma = \log_{10}(f_s \cdot e^{\sigma_0/20})$ when $\sigma_0$ is the estimated decay rate
because $\sigma_0$ is defined on time unit and decibel scale, but $\sigma$ is defined on sample unit and linear
scale. Table 5.1 shows the estimated initial magnitudes and decay rates.

<table>
<thead>
<tr>
<th>Harmonic index</th>
<th>Initial magnitude $r$</th>
<th>Decay rate $\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0067</td>
<td>-1.4204</td>
</tr>
<tr>
<td>2</td>
<td>0.0055</td>
<td>-2.3893</td>
</tr>
<tr>
<td>3</td>
<td>0.0058</td>
<td>-3.8327</td>
</tr>
<tr>
<td>4</td>
<td>0.0030</td>
<td>-3.4683</td>
</tr>
<tr>
<td>5</td>
<td>0.0027</td>
<td>-3.4736</td>
</tr>
<tr>
<td>6</td>
<td>0.0026</td>
<td>-12.5970</td>
</tr>
</tbody>
</table>

Table 5.1: Estimated initial magnitudes and decay rates
5.2.3 Estimation of fundamental frequency

To control the resonant frequencies of the resonators, fundamental frequency curve of each harmonic partial needs to be estimated. The algorithm of fundamental frequency estimation is same as used previously in 2.3. Fig. 5.11 show the estimated fundamental frequency curves from the signals of Fig. 5.6.
Figure 5.7: Magnitude spectrum of the bandpassed signals

5.2.4 Estimation of gain factor

Because the amplitude of each harmonic partial does not decrease exponentially as discussed above, the gain factor $g_c$ should be estimated. The estimation method is similar to the method used in the time-varying waveguide model to estimate the variable loop gain of the loss filter. Sinusoidal functions which are used as the gain factors in the hybrid model are fitted with the local maximum points of the bandpassed signals. Because amplitude of the third harmonic partial is the biggest, so the periods of the sinusoidal functions of other harmonic partials are based on the period of the sinusoidal function of the third harmonic partials. The period approximated based on the third harmonic partial is about 0.22 second, and the formers for other harmonic partials are near 0.11
Figure 5.8: Comparison of the magnitude spectrums of the sum of the bandpassed signals (green curve) and the original geomungo tone (blue curve, Fig. 2.2c)

Figure 5.9: Comparison of the waveform of the bandpassed signal for the third partial (blue curve) and the positive local maximum points (green curve)

Figure 5.10: Approximated straight line (green curve) is fitted over the positive local maximum points (blue curve)

second. Fig. 5.13 shows the synthesized signal for the resynthesis of the third partial (Fig. 5.12) without the gain factor. With the gain factor of a sinusoidal function, waveform of the synthesized signal (Fig. 5.14) is more similar to the target signal.
Figure 5.11: Magnitude spectrum of the bandpassed signals

Figure 5.12: Waveform of the bandpassed signal for the third harmonic partial
5.3 Synthesis

Fig. 5.15 shows the excitation signal generated from the geomungo tone of Fig. 2.1c. Because the target geomungo tone has fluctuant frequencies, the length of the inverse filtered signal should be shortened by deamplifying and removing the vibrating sound after the pluck sound. Using the excitation signal, 6 signals were generated from the resonators (Fig. 5.16).

Also, high-pass filtered signal of the excitation signal (Fig. 5.17) was used as the input signal for digital waveguide synthesis part at high frequencies. Fig. 5.18 shows the synthesized signal from the digital waveguide model.

Finally, sum of the signals from the resonators and waveguide model generated an output signal of the hybrid model. Fig. 5.19 and Fig. 5.20 shows the waveform and the STFT result. By comparison of the magnitude spectrums (Fig. 5.21), the synthesized geomungo tone is found to be very similar to the target tone. In addition, Fig. 5.22 shows the envelope curves of the original tone (Fig. 5.22a) and the synthesized geomungo tone (Fig. 5.22b). Each harmonic fluctuates differently through separate gain control of harmonic partials, so the synthesized tone sounds...
more similar to the original tone.

Figure 5.15: Excitation signal extracted and modified from the geomungo tone of Fig. 2.1c

Figure 5.16: Waveforms of the synthesized signals from the resonators
Figure 5.17: High-pass filtered excitation signal from Fig. 5.15

Figure 5.18: Waveform of the signal synthesized by the digital waveguide model from the signal of Fig. 5.17

Figure 5.19: Synthesized signal from the hybrid model
Figure 5.20: STFT result of the signal of Fig. 5.19

Figure 5.21: Comparison of magnitude spectrums of the original geomungo tone (green curve) and the synthesized geomungo tone (blue curve)
Figure 5.22: Comparison of envelope curves of (a) the original tone and (b) the synthesized tone.
Chapter 6. Real-time sound synthesis
application of the geomungo

Based on the algorithm discussed above, the geomungo was realized as a real-time sound synthesis application. Real-time synthesis is important to use the application as a virtual geomungo instrument that can be used in various ways such as sound editing softwares or mobile phones.

In this work, The Synthesis Toolkit (STK) [12], C++ programming environment to support sound synthesis and audio processing, was used to implement the synthesis program. Since STK provides various components used in audio processing such as IIR filter and RTAudio API that supports real-time audio streaming output [47], it is suitable for prototyping the synthesis algorithm as a real-time application.

6.1 Time-constant model

Fig. 6.1 shows the structure of the implemented program for the synthesis of geomungo sounds of constant pitch. It is similar to the structure of sound synthesis model of Fig. 3.1. FileWvIn is a class to store excitation signal files. Iir class supports IIR filter, so the loss filter of Eq. 3.3 is implemented using the class. DelayL class supports delay line and Lagrange interpolation fractional delay filter. Output signals are played as real-time audio streaming by RTAudio api. It can be used in conjunction with FileWvOut class to write the output signal as a wave file.

In the program, there are two parameters: fundamental frequency and excitation signal. Any floating number can be set as a fundamental frequency to synthesize a geomungo tone and the excitation signal can be selected from the sample array. Fig. 6.2 shows the command-line interface

![Figure 6.1: Structure of the sound synthesis program of the geomungo](image)

---

- 64 -
based program producing geomungo sounds. Fig. 6.3 shows the synthesized geomungo tones of 63.1Hz, 160Hz, 210Hz, and 320Hz.

Playing ... press [enter] to quit.
asurank:Geomungo seunghuns ./geo 440 0
Playing ... press [enter] to quit.
asurank:Geomungo seunghuns ./geo 63.1 0
Playing ... press [enter] to quit.
asurank:Geomungo seunghuns ./geo 148.5 1
Playing ... press [enter] to quit.
asurank:Geomungo seunghuns ./geo 148.5 0
Playing ... press [enter] to quit.
asurank:Geomungo seunghuns ./geo 222.5 1
Playing ... press [enter] to quit.
asurank:Geomungo seunghuns

Figure 6.2: Real-time sound synthesis program of the geomungo

6.2 Time-varying model

To synthesize the signals having fluctuant pitches, the program was implemented based on the algorithm discussed in 4. Fig. 6.4 shows the structure. The gain parameter used in Iir class is controlled by the approximated sinusoidal function, and the delay parameter in DelayL class is controlled by the estimated fundamental frequency curve.

Using the time-varying real-time sound synthesis program, I tried to re-synthesize the geomungo tone of Fig. 2.1d. Frequency curve of Fig. 2.4d is used to control the DelayL, and gain coefficient curve of Fig. 6.5 is used to control the Iir. Using the excitation signal of Fig. 6.6 a geomungo tone was generated from the real-time synthesis application (Fig. 6.7).
Figure 6.3: Synthesized (a) 63.1Hz (b) 160Hz (c) 210Hz (d) 320Hz geomungo tones using the real-time application
Figure 6.4: Structure of sound synthesis program of the geomungo based on the time-varying synthesis model.

Figure 6.5: Variable loop gain coefficient $g_n$ used for the Iir in the real-time synthesis application.

Figure 6.6: Excitation signal extracted from the geomungo tone of Fig. 2.1d.
Figure 6.7: Re-synthesized signal of Fig. 2.1d using the residual signal of Fig. 6.6 by the real-time synthesis application
Chapter 7. Conclusion

7.1 Evaluation

As a qualitative evaluation of the synthesized geomungo tones based on the digital waveguide modeling, I interviewed an expert in Korean traditional music. After playing the original tones and synthesized tones, I asked about the possibility of the proposed model to use in real music. She answered that timbres of the synthesized tones were very similar to the original tones, but it was better to add small noises because sounds played by the geomungo have some noises like friction. In addition, she said that it was also better to synthesize more geomungo tones to generate various kinds of vibrato effects.

In addition, a quantitative evaluation was performed based on an objective evaluation method. In previous studies, group listening tests were used for evaluation of audio quality, but they are subjective and time-consuming. Thus, research about object testing methods has been conducted. For example, Perceptual Evaluation of Audio Quality (PEAQ) is the standardized assessment method referred by the International Telecommunications Union (ITU). Perceptual Evaluation of Speech Quality (PESQ) and Noise-To-Mask Ratio (NMR) are also the methods for objective audio quality assessment[7]. However, these models were developed for digital speech coding for wireless telephony or music coding for compression, so focuses of the models are different because synthesized instrument tones sound different from a real instrument[17]. Nevertheless, there has been a research for evaluating the waveguide synthesis method based on objective algorithms such as Perceptual Audio Quality Measure (PAQM) and Noise-To-Mask Ratio (NMR)[63]. PAQM algorithm showed that 71% of the scores fell within the range of scores of the human subjective listening test. Thus, it is still doubtful whether the PAQM test assures the objective quality of the synthesized tones, but I used the PAQM test to show that it is better to use the control of the gain factor in the loop filter of the waveguide synthesis model based on a sinusoidal function. Because the PAQM method scores based on difference between an original signal and a target signal based on a human perception model, so getting high marks in the test means that the target signal is more similar to the original signal.

Evaluation algorithm based on the PAQM test was implemented. The PAQM test includes com-
putation of spectral power density, transfer function of human ear model, transformation of the samples from frequency to bark scale, time-domain spreading, and frequency-domain spreading. Instructions of the PAQM test were in [5]. The final result from the PAQM test is a score from 1 to 5, and a lower score means that difference is smaller. Using the algorithm, synthesized tones were scored based on the comparison of the original tones. As the test, the synthesized tone without gain control (Fig. 4.8) scores 1.92, and the former with gain control (Fig. 4.15) scores 1.67. Thus, synthesis of fluctuating geomungo tone with gain control is better. In addition, the synthesized geomungo tone with another excitation signal (Fig. 4.16) scores 3.33, synthesized tone with the hybrid model (Fig. 5.19) scores 2.3, and synthesized tone from the real-time application (Fig. 6.7) scores 2.97.

7.2 Summary and conclusion

This paper discussed the sound synthesis of the geomungo, a Korean traditional plucked string instrument. Physical modeling was used as a synthesis method for the instrument because it allowed to generate various sounds based on mathematical models. Description of the geomungo was discussed at first, and background theories about physical modeling were introduced.

From the extracted geomungo samples, frequency responses of the signals were estimated by FFT and SFTF, and the harmonic partials were found from the magnitude spectrum. Then the behaviors of fundamental frequency in time were estimated and some of them showed great pitch fluctuations.

The model proposed in this work was based on the digital waveguide theory which was a generalized Karplus-Strong algorithm. A ripple filter was used with a one-pole filter to design a loss filter, and allpass filter was used to tune the fundamental frequency. Filter parameters were calibrated by calculating the loss of the harmonic partials. Then input signals in the synthesis model were calculated by inverse filtering from the recorded samples.

Synthesized geomungo tones sounds similar to original signals, and the peaks were found in the magnitude spectrum. Then, in order to produce vibrato sounds having fluctuating pitches, an time-varying synthesis model was proposed. The gain factor and loop gain coefficient in the loss filter were controlled based on the magnitude analysis of the target tone, and the length of delay line and fractional delay filter were controlled based on the frequency analysis of the target tone. The variable gain coefficient was approximated as a sinusoid function. Excitation signal could be
extracted from a geomungo tone which had constant pitch and similar timbre to the target tone or from the target tone itself, but the latter needed to be modified by deamplifying and removing the vibrating part in the sound.

For a more accurate synthesis, hybrid modal-waveguide synthesis which generated harmonic partials at low frequencies with second-order resonators and at high frequencies with the digital waveguide model. For fluctuant pitches, frequency and gain factor of each resonator was also controlled by the frequency curve and the approximated gain coefficient curve of a sinusoidal function. The synthesized geomungo tone was very similar to the target tone in frequency domain comparison.

Finally, a real-time sound synthesis application was realized by using STK. Using the application, it could be possible to synthesize the geomungo tones of any fundamental frequency. The time-varying model was also implemented for the vibrato sounds.

The model proposed in this work showed that physical modeling synthesis based on digital waveguide theory was an effective method to generate the geomungo sounds which had various pitch fluctuations. I expect that proposing sound synthesis models based on physical modeling for other Korean traditional plucked string instruments is also highly effect. Ultimately, implementation of a virtual synthesizer program which can be used in sound recording software is required for composers to make use of the sounds of Korean traditional instruments without recording.

7.3 Future works

Future work includes the analysis of physical properties of each string and the design of a more elaborate model considering nonlinear acoustic characteristics of geomungo tones. In real instrument, the relation between string and body is nonlinear. The proposed model in this work assumed that it was linear and time invariant (LTI) because this made it possible to reduce the computational cost. However, it may be better to design the body filter for realistic sounds. Also, design of a model which generates an input signal is also necessary to synthesize model only based on mathematical equations without any recorded samples. Finally, evaluation model for sound synthesis methods of instruments needs to be established. Separation between music and noise is important in audio compression, but generation and dissolution of harmonic partials are more important in instrumental sounds. Thus, previous evaluation models need to be changed to result in more accurate test scores.
References


[37] 120 Years of Electronic Music. Elisha gray and “the musical telegraph”(1876). http://120years.net/machines/telegraph/.


Acknowledgement
이 록 서

이 름 : 김승훈
생 년 월 일 : 1988년 4월 23일
주 소 : 부산광역시 해운대구 반여4동 우방신계타운 105-1806
E-mail 주소 : seunghun.kim@kaist.ac.kr

학 력

2004. 3. – 2006. 2. 부산과학고등학교
2006. 2. – 2009. 2. 한국정보통신대학교 (ICU) 전자통신공학과 (B.S.)
2009. 2. – 2011. 8. 한국과학기술원 (KAIST) 문화기술대학원 (M.S.)

경 력

2007. 8. – 2008. 2. 핀란드 탐페레공과대학교 (Tampere University of Technology) 교환학생

연구 업적


