Parameter Estimation of a Plucked String Synthesis Model with Genetic Algorithm

Janne Rionheimo\textsuperscript{1}, Vesa Välimäki\textsuperscript{1,2}
\textsuperscript{1}Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, Finland
\textsuperscript{2}Pori School of Technology and Economics, Tampere University of Technology, Pori, Finland
email: janne.rionheimo@hut.fi

Abstract

We describe a technique for estimating control parameters for a plucked string synthesis model using genetic algorithms. The model has been intensively used for sound synthesis but the fine tuning of the parameters has been carried out with a semi-automatic method that requires some hand adjustment with human listening. An automated method for extracting the parameters from recorded tones is described in this paper.

2 Plucked String Synthesis Model

The model used for plucked string synthesis in this study is shown in Figure 1. It is based on digital waveguide synthesis theory (Smith 1992) that is extended in accordance with commuted waveguide synthesis approach (Smith 1993) (Karjalainen, Välimäki, and Jäñosy 1993) to include also the body modes of the instrument to string synthesis model.

1 Introduction

Model-based sound synthesis is a powerful tool for creating natural sounding synthesized tones by simulating the sound production mechanisms and physical behavior of real instruments. These mechanisms are often too complex to simulate in every detail, so simplified models are used for synthesis. The aim is to generate a perceptually indistinguishable model for real instruments.

One workable method for physical modelling synthesis is based on digital waveguide theory proposed by Smith (Smith 1992). In the case of the plucked string instruments that can be extended to model also the plucking style and instrument body. The synthesis model of this kind can be applied to synthesize various plucked string instruments by changing the control parameters and using different body and plucking models. Our interest in this paper is the parameter estimation of the model. We use recorded tones as a target sound with which the synthesized tones are compared. Genetic algorithm is used to minimize the perceptual error between the tones.

This paper is sectioned as follows. The plucked string synthesis model and the control parameters to be estimated are described in Section 2. Parameter estimation problem and methods for solving it, particularly genetic algorithm, are discussed in Section 3. Experiments and results are analyzed in Section 4 and conclusions are made in Section 5.

Different plucking styles and body responses are stored as wavetables in the memory and used to excite the two string models $S_b(z)$ and $S_c(z)$ that simulate the effect of the two polarizations of the transversal vibratory motion. A single string model $S(z)$ in Figure 2 consists of a lowpass filter $H(z)$ that controls the decay rate of the harmonics and a delay line $z^{-L}$ that controls the fundamental frequency in conjunction with a fractional delay filter $F(z)$ (Jaffe and Smith 1983). Two string models are typically slightly mistuned to produce a natural sounding beat effect.

A one-pole filter

\[ H(z) = \frac{1 + a}{1 + az^{-1}} \]  \hspace{1cm} (1)
is used as a loop filter in the model. Parameter \( g (0 < g < 1) \) in equation 1 determines the overall decay rate of the sound while parameter \( a (-1 < a < 0) \) controls the frequency-dependent decay. Excitation signal is divided by mixing coefficient \( m_p \) before sending it to two string models. Coefficient \( g_c \) enables coupling of the two polarizations. Mixing coefficient \( m_o \) defines the proportion of the two polarization in the output sound. All parameters \( m_p, g_c \) and \( m_o \) are chosen to have values between 0 and 1. The transfer function of the entire model is written as

\[
M(z) = m_p m_o S_h(z) + (1 - m_p) (1 - m_o) S_v(z) + m_p (1 - m_o) g_c S_h(z) S_v(z),
\]

(2)

where the string models \( S_h(z) \) and \( S_v(z) \) for two polarizations can be written as an individual string model

\[
S(z) = \frac{1}{1 - z^{-L} F(z) H(z)}.
\]

(3)

Synthesis model of this kind is proposed by Karjalainen et al. (Karjalainen, Välimäki, and Tolonen 1998) and has been intensively used for sound synthesis of different plucked string instruments. Different methods for estimating the parameters has been used, but in consequence of interaction between the parameters, systematic methods are at least troublesome but probably impossible. Eight parameters that are used to control the synthesis model are listed in Table 1.

### Table 1: Parameters of the synthesis model.

<table>
<thead>
<tr>
<th>parameter</th>
<th>effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_{0h} )</td>
<td>fundamental frequency of the horizontal string model</td>
</tr>
<tr>
<td>( f_{0v} )</td>
<td>fundamental frequency of the vertical string model</td>
</tr>
<tr>
<td>( g_h )</td>
<td>overall gain of the horizontal string model</td>
</tr>
<tr>
<td>( a_h )</td>
<td>frequency dependent gain of the horizontal string model</td>
</tr>
<tr>
<td>( g_v )</td>
<td>overall gain of the vertical string model</td>
</tr>
<tr>
<td>( a_v )</td>
<td>frequency dependent gain of the vertical string model</td>
</tr>
<tr>
<td>( m_p )</td>
<td>input mixing coefficient</td>
</tr>
<tr>
<td>( m_o )</td>
<td>output mixing coefficient</td>
</tr>
<tr>
<td>( g_c )</td>
<td>coupling gain of the two polarizations</td>
</tr>
</tbody>
</table>

by Välimäki et al. (Välimäki, Huopaniemi, Karjalainen, and Jánosy 1996). The fundamental frequency of a recorded tone is first estimated, and then amplitude trajectories for partials are analyzed using the short-time Fourier transform (STFT). The decay rate of each partial is measured, and the one-pole loop filter is designed to approximate the decay behavior.

Another way to approach the parameter estimation problem is to consider it as a general search process where one tries to find an optimal set of parameters in the parameter space. Each point in the parameter space corresponds to a set of parameters and our aim is to find a set of parameters that produce desired output sound when used with the synthesis model. Thus, one has to define a quality value for each point in the parameter space, that denotes "how good" is the sound synthesized by the parameter set related to that particular point. This performance metric is usually called a fitness function or inversely error function, that denotes the error between the candidate solution and the desired sound. These functions give a numerical grade to each solution by means of which we are able to classify all possible parameter sets.

The search process can be carried out with several methodologies but evolutionary algorithms have shown a good performance in parameter estimation of synthesis models. Vuori and Välimäki (Vuori and Välimäki 1993) tried a simulated evolution algorithm for the flute model and Horner et al. proposed an automated system for parameter estimation of FM synthesizer using a genetic algorithm (Horner, Beuchamp, and Haken 1993). Garcia also used a genetic algorithm for designing sound synthesis techniques automatically (Garcia 1998).
3.1 Genetic Algorithm

Genetic algorithms mimic the evolution of the nature and take advantage of the principle of "survival of the fittest" (Mitchell 1998). Genetic algorithms operate on a population of potential solutions improving characteristics of the individuals from generation to generation. Each individual called a chromosome is made up of an array of genes that contain in our case the actual parameters to be estimated. Parameters can be represented as binary or floating point numbers. A simple genetic algorithm can be implemented as follows:

1. Initialization: Create a population of popSize individuals (chromosomes). The initial values of the genes (estimated parameters) are randomly assigned.

2. Fitness calculation: Calculate the fitness of each individual in the current population.

3. Parent selection: Select individuals from current population for mating based upon the fitness values of individuals such that the better individuals have an increased chance of being selected.

4. Crossover: Create a new population (generation) by mating the selected parents to produce offspring. The fittest parents can also survive to new generation.

5. Mutation: Mutate a few individuals in the new population.

6. Replace the current population with the new one.

7. Repeat steps 2-6 until termination.

The algorithm is normally terminated when a specified number of generations is produced or when the sum of the deviations among the individuals becomes smaller than some specified threshold. Several schemes are feasible for selection, crossover and mutation processes depending on which chromosome representation is used. The most critical and application dependent stage of operation is the calculation of the fitness value. Care has to be taken when considering how to rank the candidate solutions.

3.2 Fitness Calculation

There is no known objective method for measuring the similarity between two sounds that would perfectly match with human perception. Common method used by Horner (Horner, Beauchamp, and Haken 1993) and Garcia (Garcia 1998) is to measure the least squared error of the short-time spectrum analyses of the two sounds. The STFT of signal $y(n)$ as defined in is a sequence of discrete Fourier transforms (DFT):

$$Y_m(k) = \sum_{n=0}^{N-1} w(n) y(n + mH) e^{-j2\pi kn / N}, \quad m = 0, 1, 2, \ldots$$

with

$$w_k = \frac{2\pi k}{N}, \quad k = 0, 1, 2, \ldots, N - 1$$

where $N$ is the length of the DFT, $w(n)$ is a window function, and $H$ is the hop size or time advance (in samples) per frame. If $o(n)$ is the output sound of our model and $t(n)$ is the target sound that is desired then the inverse of the fitness (error) of the candidate solution is calculated as follows:

$$\frac{1}{F} = \frac{1}{L} \sum_{m=0}^{L-1} \sum_{k=0}^{N-1} (|O_m(k)| - |T_m(k)|)^2,$$

where $O_m(k)$ and $T_m(k)$ are the STFT sequences of $o(n)$ and $t(n)$ and $L$ is the length of the sequences $m = 0, 1, 2, \ldots, L - 1$.

4 Experimentation and Results

To study the efficiency of the proposed method we first tried to estimate the parameters from the sound produced by the model itself. First the same excitation signal extracted from the real recorded tone by the method described in (Väliläni and Tolonen 1998) is used for target and output sounds. More realistic case is simulated when the excitation for resynthesis is extracted from the target sound. Original and estimated parameters are shown in Table 2.

<table>
<thead>
<tr>
<th>parameter</th>
<th>original parameter</th>
<th>estimated parameter (original excitation)</th>
<th>estimated parameter (extracted excitation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f_{0h}$</td>
<td>330</td>
<td>330.406</td>
<td>330.5598</td>
</tr>
<tr>
<td>$f_{0v}$</td>
<td>331</td>
<td>331.412</td>
<td>331.5355</td>
</tr>
<tr>
<td>$g_h$</td>
<td>0.989</td>
<td>0.9888</td>
<td>0.9893</td>
</tr>
<tr>
<td>$a_h$</td>
<td>-0.18</td>
<td>-0.1699</td>
<td>-0.1971</td>
</tr>
<tr>
<td>$g_v$</td>
<td>0.991</td>
<td>0.9901</td>
<td>0.9914</td>
</tr>
<tr>
<td>$a_v$</td>
<td>-0.22</td>
<td>-0.2270</td>
<td>-0.1173</td>
</tr>
<tr>
<td>$m_p$</td>
<td>0.5</td>
<td>0.4414</td>
<td>1.0000</td>
</tr>
<tr>
<td>$m_o$</td>
<td>0.5</td>
<td>0.6061</td>
<td>0.9310</td>
</tr>
<tr>
<td>$g_c$</td>
<td>0.1</td>
<td>0.1474</td>
<td>0.4738</td>
</tr>
</tbody>
</table>

Table 2: Original and estimated parameters.
When using the original excitation signal for resynthesis almost the exact parameters are estimated. Only small differences in mixing parameters $m_o$, $m_p$, and coupling coefficient $g_c$ can be noticed. When trying to optimize only the mixing parameters with different values of $g_c$ we find that maximal fitness values are very equal. Synthesized tones produced by corresponding parameter values are also indistinguishable. That is to say the parameters $m_p$, $m_o$, and $g_c$ are not orthogonal.

Similar behavior is found when using an extracted excitation. Although the parameter values differ from the original ones the corresponding synthesized tone is indistinguishable from the target sound. Purpose of the two slightly mistuned string models is to simulate the beating effect. Here the beating effect is obtained by closing totally the straight path to the vertical polarization string $(1 - m_p)$ and increasing the coupling gain value $g_c$. Because not only the excitation but the whole tone produced by horizontal polarization is now exciting the vertical string model the mixing coefficient $m_o$ is adjusted to attenuate the output of the vertical string model. The function of the gain parameters of the vertical string model differs now from the basic situation so they are not comparable with the original ones in the terms of efficiency of the optimization method. The cause of this functional rearrangement of the blocks of the synthesis model is the small dissimilarities between the excitation signals. Optimization method tries to adapt the model for best performance with the particular excitation signal.

Convergence of the genetic algorithm when estimating the parameters for real sound can be seen in Figure 3. The perceptual quality of the resulting synthesized sounds is comparable or better than with our previous methods and it does not require any hand tuning after the estimation procedure.

![Figure 3: Convergence of the algorithm.](image)

## 5 Conclusions and Future Work

Our method showed good performance in estimating parameters for the plucked-string synthesis model and it takes into account the difference between the fundamental frequencies of the two string models that was previously adjusted by hand. Through this method we are able to create parameter sets that corresponds to recorded real instrument tones with certain features. Final judgement of quality of a synthesized tones created with these parameter sets, is always done by human so the most critical field is the perceptual error calculation. In the future we will include more perceptual behavior to the fitness function. Also the automated arrangement of the functional blocks of the synthesis model is in our interest.

## References


