AN EFFICIENT MODAL-BASED APPROACH TOWARDS GUZHENG SOUND SYNTHESIS

Enda Zhang* Gopal Gupta* Charles Greif Andrew Paplinski*

* Monash University

ABSTRACT

Encouraged by recent interest in traditional Chinese instruments, this work proposes a computational sound synthesis model for a traditional Chinese instrument, the guzheng. Digital waveguide model and modal synthesis are the most popular physical modelling methods. However, the excitation signal is usually over-simplified and the design of some digital filters is inadequate for real-time synthesis. We propose an accurate and efficient approach to synthesise the string signal of the guzheng. Our proposed model enables accurate parametric control of the pluck because the parameters are based on physical model of the instrument. The sound of the guzheng is then computed as the convolution of the synthesised string signal and the impulse response of the resonant body. As proof of concept, the 21st string of the guzheng is synthesised using the proposed model. The sound computed using our model shows improved accuracy with an acceptable computational complexity.

Index Terms— Sound synthesis, modal synthesis, digital waveguide model, guzheng

1. INTRODUCTION

The guzheng is a traditional Chinese plucked string musical instrument. It is widely agreed that the guzheng firstly appeared during the Qin Dynasty of Imperial China (221-206 BCE), and it evolves from another traditional Chinese musical instrument se. More strings have been added and its structure has been evolving gradually to the modern guzheng.

The guzheng has 21 strings which are made of either nylon or steel. The strings are mounted on the top plate of the body through the front nut and back nut. The vibration of the string is transmitted to the body through a bridge, which is located in a line on the top plate as shown in Fig. 1. The length of the string and therefore the fundamental frequency depends on both the tension of the string and the position of the bridge. The body of the guzheng is a sound box designed to be curved to balance the tension of strings. There are three or four soundholes located on the back plate to couple the sound of the guzheng to the air.[1][2]

When playing the guzheng, the string segment between the bridge and the front nut (right-half of the string) is to be plucked, while the other segment (left-half of the string) is used for pitch variation by pressing the string. Players wear plastic nails on their right hand when playing, in order to create sharper and more dynamic sound.

To date, very little research of the sound and acoustics of the guzheng has been published. Xiaowei Deng, a researcher from Shanghai Jiao Tong University, completed his master thesis based on a project of the acoustics of the guzheng, which was set up with Shanghai National Musical Instrument Factory. Deng [3] both studied the mechanical structure and performed lots of measurements on the physical properties of the guzheng. Also, a set of partial differential equations was proposed to synthesise the vibration signal and the sound of the 21st string.[1][2]

Digital Waveguide (DWG) model and Modal-based Synthesis are two most popular methods to efficiently synthesise the sound of a string instrument. Digital Waveguide Model (DWG), proposed by Julius O. Smith, can be interpreted as an efficient implementation of d’Alembert’s formula using delay lines in the one-dimensional case.[4] Low-pass filters and all-pass filters could be added to the delay line to simulate the damping effect and the stiffness of the string.[4] [5] The input of DWG model can be a wavetable or an excitation signal generated by a parametric synthesis model. Parametric synthesis frameworks have been applied to the synthesis of the acoustic guitar [6], electric guitar [7] and the clavinet [8]. However, it is challenging to accurately synthesise an excitation signal with control parameters since it mimics the effect of a pluck. It would be more intuitive to consider a “physical” way to produce the excitation signal which comes directly from a physical model.

Another approach, modal synthesis, produces the sound as the summation of exponentially decaying sinusoids, which represent the eigenmodes of a linear system. Trautmann and Rabenstein proposed and applied functional transformation methods (FTM) to convert partial differential equations into modal representations. As a proof of the concept, they applied FTM to the extended 1-dimensional wave equation for the synthesis of the classic guitar [9]. FTM is usually computationally expensive when the number of modes is large. In 2010, Bank et al. [10] proposed a modal-based piano synthesiser where all the hammer, string and soundboard are modelled by the modal synthesis. After optimisation, the model is guaranteed to run in real-time.

To reduce the computation complexity of FTM, a combination method was proposed by Trautmann et al. which calculates the parameters in DWG using the result of FTM instead of estimating from measurement.[11] Rectangular functions were used as the excitation signal of the DWG. The model is more efficient than using FTM di-
rectly while keeps the precision at a desirable level. However, using a rectangular function as the excitation signal is inherently inaccurate since the vibration after propagating on the string has a more structured waveform. The rectangular excitation signal does not allow effective parametric control of a physical model. Furthermore, the original design of the dispersion filter is not a closed form, therefore is hard to implement in real-time.

To the best of our knowledge, Deng’s approach is the first computational model for the guzheng synthesis. As Deng modelled the string using a 1-dimensional ideal wave equation, the vibration signal was produced based on the physical parameters only. Therefore the model can be adjusted easily by the physical parameters and the synthesised sound is relatively accurate. However, as sound synthesis was not Deng’s main focus, his approach has some drawbacks and therefore needs to be improved. First, his model synthesises the sound based on numerical analysis of partial differential equations. It is hard to develop a sound synthesiser using his approach as his synthesis model cannot run in real-time. Second, the damping effect and the stiffness of the string were not included in his model. Third, the plucking was modelled by the initial displacement and initial velocity, which are caused by the plucking force and needs further measurement.

Some similar string instruments have been modelled in previous studies. Kantele is a traditional Finish musical instrument, which has a similar acoustic structure as the zither and the guzheng \[12\]. Guqin, as a similar instrument compared to the guzheng, has been thoroughly studied in its acoustics \[13\]. Both Guqin \[14\] and Kantele \[15\] have been synthesised using DWG model and pre-recorded excitation signals. Consequently, in this paper, we propose a computational synthesis model for the guzheng. The model is as efficient as typical DWG-based models while keeps the accuracy at a desirable level by adopting modal synthesis. The stiffness of the string and pitch varying are also considered in our proposed model. The novelty of this work includes:

- Using modal synthesis to produce the excitation signal to achieve high accuracy
- Allowing parametric control of the pluck using parameters and physical characteristics of the string
- Enabling real-time synthesis by changing the design of the dispersion filter
- Applying our proposed method to the guzheng, of which a real-time synthesiser is yet to be created.

The method for the estimation of parameters and the design of each part of the model are shown in Section 2.2. The results and discussion are in Section 3 and Section 4, respectively. Finally, the conclusions and discussion are in Section 5.

2. PROPOSED METHOD

The model consists of three main parts: modal synthesis, Digital Waveguide Model and the convolution. Using functional transformation method, modal synthesis part produces an excitation signal for DWG part, where the physical and damping parameters are the control parameters. The damping parameters, describing the damping effect of the particular string, can be estimated empirically or retrieved from a recorded vibration signal of the string. Next, the DWG model processes the excitation signal by a delay line and multiple digital filters. Such filters handles both the tone and the pitch of the synthesised sound. The final sound of the guzheng is produced by the convolution of the string signal and the impulse response of the body.

The detail of each part will be discussed in the following sections. The process and calculation of modal synthesis part are shown in section \[2.1\]. The method used for estimating the damping parameters in the PDE will be provided in section \[2.2\]. Details are given in the following subsections. The design of the low-pass filter and all-pass dispersion filter used in DWG part is listed in section \[2.3\] and \[2.4\] respectively. The consideration for fine-tuning is discussed in section \[2.5\].

2.1. Construction of Excitation Signal

As introduced previously, we will feed a short clip of wave signal into one-dimensional digital waveguide model. To achieve high accuracy, we synthesise the wave signal by FTM as a modal approach. Modal synthesis has been widely used in the synthesis of string instruments, and the detailed description can be found in many research works and textbooks, especially in \[9\] \[10\] \[16\].

Guzheng has thicker strings compared to popular string instrument such as guitar. Therefore, in our case, the stiffness and damping effect on the string of the guzheng should not be neglected \[1\]. We take the dispersion and damping factors into account and consider the following extended one-dimensional wave equation

\[
\begin{align*}
\epsilon \frac{\partial^2 \psi}{\partial t^2} + E I \frac{\partial^4 \psi}{\partial x^4} - T s \frac{\partial^2 \psi}{\partial x^2} + d_1 \frac{\partial \psi}{\partial t} + d_3 \frac{\partial^2 \psi}{\partial t \partial x^2} &= F_e(x,t), \\
y(x,0) &= 0, \quad \frac{\partial y}{\partial t}(x,0) = 0, \\
y(x_0,t) &= 0, \quad \frac{\partial y}{\partial t}(x_0,t) = 0, \\
\end{align*}
\]

where \(x \in [0, l]\) is the spatial variable, \(t \in [0, +\infty)\) is the temporal variable, \(y(x,t)\) is the displacement of the string, \(\epsilon\) is the linear density, \(T_s\) is the tension of the string, \(E\) is the Young’s modulus of the string and \(l\) is the moment of inertia. \(d_1\) and \(d_3\) are the frequency dependent and independent decay parameter, respectively. The method for estimating \(d_1\) and \(d_3\) will be discussed in section \[2.2\].

On the guzheng, the pluck point is fixed when plucking, hence we can separate the temporal and spatial variables in general cases. In practice, the excitation force can be simplified as \(F_e(x,t) = F_{e1}(x)F_{e2}(t)\). These two parts can be further expressed as \(F_{e1}(x) = f(x-x_c)\) and \(F_{e2}(t) = 1_{[0,t_p]}\) where \(f\) is the magnitude of the excitation function, \(x_c\) is the pluck point, \(t_p\) is the time interval of the pluck and \(\delta(x-x_c)\) is a pulse at \(x = x_c\). The notation \(1_{[0,t_p]}\) represents an rectangular function, which takes 1 when \(x \in [0,t_p]\) and 0 otherwise. It is a reasonable approximation because the pluck point is always fixed when the string is plucked, and only the excitation force varies with time. Also, we assume the displacement \(y(x,t)\) is sufficiently smooth.

By applying a sequence of functional transforms to Eq(1), the PDE is transformed into a multi-dimensional discrete transfer function \[9\]. The discrete transfer function consists multiple second-order IIR filters which represent the modes in the excitation signal. Let \(\mu\) denote the mode number and \(x_a\) denote the point on string we calculate, then the discrete transfer function can be represented as

\[
H(x, z) = \sum_{\mu} a(\mu, x_a) \frac{b_1(\mu)z^{-1}}{1 + c_1(\mu)z^{-1} + c_0(\mu)z^{-2}},
\]
The coefficients of the IIR filters are calculated as
\[ a(\mu, x_0) = \frac{1}{2} f I \sin \left( \frac{\mu \pi x_0}{l} \right) \sin \left( \frac{\mu \pi x}{l} \right), \]
\[ b_1(\mu) = \frac{1}{c_0 \mu} \sin (\omega_0 T), \]
\[ c_1(\mu) = -2 \exp (\sigma_0 T) \cos (\omega_0 T), \]
\[ c_0(\mu) = \exp (2 \sigma_0 T), \]
where
\[ \mu = \frac{\pi}{l}, \]
\[ \sigma_0 = \frac{1}{2\epsilon} (d_3 \gamma_0^2 - d_3), \]
\[ \omega_0^2 = \frac{4 \epsilon EI - d_3^2}{4 \epsilon ^2} \gamma_0^2 + \frac{2 \epsilon T_s + d_1 d_3}{2 \epsilon ^2} \gamma_0^2 - \frac{d_3^2}{2 \epsilon}. \]

In the formulae above, \( T = 1/f_s \) denotes the sampling interval. Computationally, \( \mu \) can be implemented as the weighted average of the output of a large numbers of second-order IIR filters in parallel \([9]\).

### 2.2. Estimating Damping Parameters \( d_1 \) and \( d_3 \)

The damping parameters, controlling the decay rate of the string vibration, can be either estimated from a recorded string vibration signal or guestimated based on experience. If a pre-recorded string signal is available, the damping parameter \( d_1 \) and \( d_3 \) can be derived by linear regression.

### 2.3. Design of the Low-pass Filter

In our proposed model, the propagation of the wave on the string is modelled by a digital waveguide model. The damping effect is caused by the energy loss of the string in the air, and the transmission from string to bridge. As suggested by Rabenstein and Trautmann \([16]\), a one-pole filter can be designed based on the result from mathematical transforms to the original physical model. The filter coefficients are calculated by the physical parameters mentioned in \( \frac{[6]}{} \) only. To be specific, \( g = 1 - \pi d_1/\omega_1, a = -1 + c_0 - \sqrt{c_0^2 - 2c_0}, \) where \( c_0 = -c_0 \omega_1^2/\epsilon \).

### 2.4. Design of the Dispersion Filter

Similar to the damping effect, the dispersion effect of the string can be modelled by adding a dispersion filter into the loop of DWG model. Usually the design of the dispersion filter requires a lot a computation and can hardly run in real-time. To enable on-line processing, Thiran filter can be used to model the dispersion effect of piano and other string instruments \([17]\). The Thiran filter \([13]\) is a second-order all-pass filter
\[ A_{disp}(z) = \frac{a_2 + a_1 z^{-1} + z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \]
where \( a_k = (-1)^k (N/k) \prod_{n=0}^{k-1} \frac{D - 2 + n}{D - 2 + k + n}, \quad k = 1, 2. \) \( \) (3)

We use four cascade Thiran filters as an all-pass dispersion filter to model the stiffness of the string of the guzheng. Let \( D_{disp} \) denote the parameter used for the dispersion filter design, the coefficient \( D = D_{disp} \) is estimated based on the pitch of the sound and the physical parameters, including Young’s modulus, diameter, tension and length of the string \([17]\).

### 2.5. Consideration for Tuning

To produce the sound with fundamental frequency \( f_0 \), a total delay of \( f_s/f_0 \) is needed in the DWG part to create the periodical signal. Therefore, except for an integer-size delay unit for coarse tuning, it is necessary to have a fractional delay for fine-tuning. The Thiran filters can create group phase delay of \( \frac{D_{frac}}{D} \), and the dispersion filter \( A_{frac}(z) \) can create \( 4D_{frac} \) delay of delay. Thus, the integer-size delay should take \( L = \lfloor f_s/f_0 - 4D_{frac} \rfloor \) units, and the delay we need for fine tuning is calculated as \( D_{frac} = f_s/f_0 - L - 4D_{frac} \). The coefficients of \( A_{frac}(z) \) are calculated by \( \) with \( D = D_{frac} \) \([19]\) where \( D_{frac} \) denote the parameter used for the design of the fraction delay filter.

### 3. RESULTS

We firstly estimate the damping parameters. The recorded string signal was obtained from Deng \([2]\), who measured physical and acoustic parameters in his master thesis. By applying short-time Fourier transform to the recorded string signal and using the method introduced in \([2]\) we calculated the damping parameters as \( d_1 = 2.9390 \) and \( d_3 = -9.1 \times 10^{-4} \) with goodness-of-fit \( R^2 \) = 0.74 and \( p = \)
Table 3. The error comparison of each mode in the spectrum

| Method         | 1st  | 2nd  | 3rd  | 4th  | 5th  |
|----------------|------|------|------|------|------|
| Proposed method| 0.0411 | 0.1061 | 0.0005 | 0.1831 |
| FTM            | 0.0360 | 0.9287 | 0.6568 | 0.0746 |
| DWG model      | 0.1931 | 0.4757 | 0.3140 | 0.2653 |

| Method         | 6th  | 7th  | 8th  | 9th  | 10th |
|----------------|------|------|------|------|------|
| Proposed method| 0.11056 | 0.03297 | 0.0397 | 0.1408 | 0.0033 |
| FTM            | 0.2156 | 0.0392 | 0.1265 | 0.0684 | 0.1678 |
| DWG model      | 0.1279 | 0.1775 | 0.1148 | 0.0045 | 0.0889 |

2.84 × 10⁻⁷.

Then we calculated the coefficients of the low-pass filter $H_{10}$ based on the estimate of damping parameters. The coefficients of the dispersion filter are calculated based on the physical parameters of the 21st string of the guzheng. The pitch of 21st string is D2 (73.42 Hz), therefore the total delay needed is 48000/73.42 = 653.77 samples. Both the integer delay and the filters contribute to the total delay, and the fraction delay needed is 1.8566. The coefficients of the fraction delay filter are calculated using Eqn. (14) by substituting $D = D_{frac} = 1.8566$. All filter coefficients are listed in the Table 1.

The excitation function $F_e(t)$ is chosen to produce a discrete rectangular pulse of length 144 samples as the plucking time is chosen to be 3 milliseconds. The pluck point is set to be at the 1/7 of the length of the string. The excitation force is chosen to produce a discrete rectangular pulse of length 144 samples as the plucking time is chosen to be 3 milliseconds. The pluck point is set to be at the 1/7 of the length of the string. The excitation force is chosen to produce a discrete rectangular pulse of length 144 samples as the plucking time is chosen to be 3 milliseconds. The pluck point is set to be at the 1/7 of the length of the string. The excitation force can also be applied by convolving the impulse response with the excitation force $F_e(t)$.

4. DISCUSSION

4.1. Accuracy

Generally speaking, the accuracy of a sound synthesis system is evaluated by the similarity between the synthesised sound and the sound from the real instrument. The evaluation is subjective and aesthetic because it depends on the preference of timbre, musical experiences and so on. Subjective evaluation of the sound quality is worth studying and related experiments are left for future work. In this study, we mainly focus on the comparison on spectrum as the technical aspect of the model accuracy. But again, we by no means consider the spectrum comparison is the only way to prove accuracy and the quality of a synthesis method.

We compared our proposed method to other two state-of-art approaches. One is the ordinary Functional Transformation Method (FTM) as a modal synthesis method from directly from physical model. Another one is a Digital Waveguide model with the considerations of the damping and dispersion effect proposed by Trautmann and Rabenstein [11].

We synthesise the string signal of the guzheng using Functional Transformation Method and Digital Waveguide. Comparing the spectrum of both synthesised string signals, our proposed method generally yields higher accuracy in the spectrum. The comparison of errors is shown in Table 3.

4.2. Performance

We also measured the processing time of our proposed method with FTM and DWG model as two state-of-art approaches. The two comparison approaches are in the same parameter settings as in the accuracy evaluation.

We synthesise the string signal for different lengths using all these three methods and record the processing time for each. When generating different length of synthesised signal, every method was evaluated 10 times to reduce the error. As shown in Fig. 3, the processing time is linear for every methods. The FTM model has a significant high computational complexity compared to our proposed method. Compared to DWG model, our proposed method takes almost the same time to produce the signal but requires a bit more computation, because of the synthesis of the excitation signal. However, as the length of the excitation signal is usually fixed, the computation will not increase much and the slope of the two lines in Fig 3 are almost the same.

5. CONCLUSION

In this work, we proposed an efficient modal-based approach for the sound synthesis of the guzheng which keeps the accuracy from the physical model and takes advantages from the use of DWG by simulating the propagation of the vibration on the string. We adopt a new design of the dispersion filter to be more applicable and practical. The proposed model is more flexible as it enables parametric control of the pluck. Our proposed model was implemented in MATLAB. By comparison with DWG model and FTM, our proposed model is proved to be acceptably accurate and efficient.

By now, only the 21st string of the guzheng is synthesised because of limited time, while other strings can be easily simulated by similar models given particular physical parameters. The simulation of other strings and the modelling of non-linearities of the string could also be done as future work.

6. ACKNOWLEDGEMENT

We would like to express special thanks to Professor Zhengyue Yu and his formal master student Xiaowei Deng, who kindly provided lots of experiments data and continuously supported this work.
7. REFERENCES

[1] X. Deng, Z. Yu, W. Yao, and M. Chen, “Dynamic analysis on Guzheng string vibration and bridge transmission,” Journal of Vibration and Shock, vol. 34, no. 18, pp. 166–170, 2015.

[2] X. Deng, “Analysis of structure vibration and acoustic characteristics of chinese musical instrument Guzheng,” Master’s thesis, Shanghai Jiao Tong University, 2015.

[3] X. Deng, Z. Yu, W. Yao, and M. Chen, “Simulation analysis of vibro-acoustic characteristics of traditional Guzheng,” Journal of Shanghai Jiao Tong University, vol. 50, no. 2, pp. 300–305, 2016.

[4] J. O. Smith, “Efficient synthesis of stringed musical instruments,” 1993.

[5] ——, Physical Audio Signal Processing: For Virtual Musical Instruments and Audio Effects. http://ccrma.stanford.edu/˜jos/pasp/, Accessed 1st Mar 2018, online book, 2010 edition.

[6] V. Välimäki and T. Tolonen, “Development and calibration of a guitar synthesizer,” Journal of the Audio Engineering Society, vol. 46, no. 9, pp. 766–778, 1998.

[7] N. Lindroos, H. Penttinen, and V. Välimäki, “Parametric electric guitar synthesis,” Computer Music Journal, vol. 35, no. 3, pp. 18–27, 2011.

[8] L. Gabrielli, V. Välimäki, H. Penttinen, S. Squartini, and S. Bilbao, “A digital waveguide-based approach for clavinet modeling and synthesis,” EURASIP Journal on Advances in Signal Processing, vol. 2013, no. 1, p. 103, 2013.

[9] R. Rabenstein and L. Trautmann, “Digital sound synthesis of string instruments with the functional transformation method,” Signal Processing, vol. 83, no. 8, pp. 1673–1688, 2003.

[10] B. Bank, S. Zambon, and F. Fontana, “A modal-based real-time piano synthesizer,” IEEE transactions on audio, speech, and language processing, vol. 18, no. 4, pp. 809–821, 2010.

[11] L. Trautmann, B. Bank, V. Valimaki, and R. Rabenstein, “Combining digital waveguide and functional transformation methods for physical modeling of musical instruments,” in Audio Engineering Society Conference: 22nd International Conference: Virtual, Synthetic, and Entertainment Audio. Audio Engineering Society, 2002.

[12] C. Erkut, M. Karjalainen, P. Huang, and V. Välimäki, “Acoustical analysis and model-based sound synthesis of the kantele,” The Journal of the Acoustical Society of America, vol. 112, no. 4, pp. 1681–1691, 2002.

[13] C. Waltham, Y. Lan, and E. Koster, “An acoustical study of the qin,” The Journal of the Acoustical Society of America, vol. 139, no. 4, pp. 1592–1600, 2016.

[14] H. Penttinen, J. Pakarinen, V. Välimäki, M. Laurson, H. Li, and M. Leman, “Model-based sound synthesis of the Guqin,” The Journal of the Acoustical Society of America, vol. 120, no. 6, pp. 4052–4063, 2006.

[15] M. Karjalainen, J. Backman, and J. Polkki, “Analysis, modeling, and real-time sound synthesis of the kantele, a traditional finnish string instrument,” in Acoustics, Speech, and Signal Processing, 1993. ICASSP-93., 1993 IEEE International Conference on, vol. 1. IEEE, 1993, pp. 229–232.

[16] L. Trautmann and R. Rabenstein, Digital sound synthesis by physical modeling using the functional transformation method. Springer Science & Business Media, 2003.

[17] J. Rauhala and V. Valimaki, “Tunable dispersion filter design for piano synthesis,” IEEE Signal Processing Letters, vol. 13, no. 5, pp. 253–256, 2006.

[18] J.-P. Thiran, “Recursive digital filters with maximally flat group delay,” IEEE Transactions on Circuit Theory, vol. 18, no. 6, pp. 659–664, 1971.

[19] T. Laakso, V. Valimaki, M. Karjalainen, and U. Laine, “Splitting the unit delay [FIR/all pass filters design],” IEEE Signal Processing Magazine, vol. 13, no. 1, pp. 30–60, 1996.

[20] L. Trautmann and R. Rabenstein, “Sound synthesis with tension modulated nonlinearities based on functional transformations,” Acoustics and Music: Theory and Applications (AMTA), NE Mastorakis, Ed., Jamaica, pp. 444–449, 2000.