PARAMETRIC APPROXIMATION OF PIANO SOUND BASED ON KAUTZ MODEL WITH SPARSE LINEAR PREDICTION

Kenji Kobayashi, Daiki Takeuchi, Mio Iwamoto, Kohei Yatabe and Yasuhiro Oikawa

Department of Intermedia Art and Science, Waseda University, Tokyo, Japan

ABSTRACT

The piano is one of the most popular and attractive musical instruments that leads to a lot of research on it. To synthesize the piano sound in a computer, many modeling methods have been proposed from full physical models to approximated models. The focus of this paper is on the latter, approximating piano sound by an IIR filter. For stably estimating parameters, the Kautz model is chosen as the filter structure. Then, the selection of poles and excitation signal rises as the questions which are typical to the Kautz model that must be solved. In this paper, sparsity based construction of the Kautz model is proposed for approximating piano sound.

Index Terms— IIR filter design, sparse optimization, ℓ_0 constrained least squares, autoregressive (AR) spectrum estimation, difference of convex (DC) algorithm.

1. INTRODUCTION

The piano is one of the most popular and attractive musical instruments, and therefore a lot of research has been conducted including piano sound synthesis [1–15]. For synthesizing the piano sound in a computer, full physical models are considered that aim to model every physical component of the piano by differential/integral equations which are solved numerically [1,2]. On the contrary, many approximation methods have also been proposed [3-15] because the full physical models are computationally expensive that restricts their practical applications. For synthesizing piano sound approximately, signal processing techniques are utilized [3,4]. A representative example would be digital waveguide [5-13] which takes advantage of the simple form of the d'Alembert's solution to the one-dimensional wave equation that can be approximated by delay lines. The modal based methods [13-15] using IIR (infinite impulse response) filters [16] are also the popular method to approximate the piano sound. These methods based on delay lines and filters are preferable for real-time synthesis owing to their computational efficiency.

In general, estimating parameters of IIR filters is not easy because of the non-linearity of their parameters. Estimated parameters may result in unstable IIR filters, and therefore special cares are necessary for utilizing them as the model. On the other hand, fixed-pole filters including the Kautz model [17–19] can be optimized quite easily because of linearity in parameters. Applying Kautz model, estimating parameters of IIR filter can be easier. Although such easiness is a very attractive property, it comes with prices that poles have to be fixed in advance. That is, the Kautz model is not flexible that may require more degrees of freedom than necessary. This property is not preferable for real-time synthesis because it ends up with either high computational cost or low approximation quality. It might make the Kautz model less attractive for piano sound approximation even though the stability and easiness of estimating parameters are quite beneficial.

In this paper, a method for reducing such unwanted properties of the Kautz model is proposed for approximating piano sound. The issues of the Kautz model are on the selection of poles and input signal as described in Section 2.1. That is, the poles have to be decided beforehand to estimate the parameters, where obtaining the poles crucial to the approximation is itself a difficult problem. In addition, for sound synthesis, the excitation signal inputted to the Kautz model can also be arbitrary. Then, better methods for selecting the poles and the exitation signal are necessary. In the proposed method, these two issues are resolved by sparsity based methods: the sparse linear prediction and ℓ_0 constrained least squares method. By applying the proposed method, the Kautz model becomes easier to approximate piano sound compared with previous modal based method.

2. SIGNAL APPROXIMATION BY KAUTZ MODEL

A Kautz model is a fixed-pole filter designed so that its parameters can be easily optimized. Let a set of poles $\{p_1, \dots, p_I\}$ and input signal e(n) be fixed first. Then, the transfer functions of a pair of orthonormal basis functions are given by

$$\Psi_{i}^{\pm}(z) = \frac{\sqrt{1 - |p_{i}|^{2}} |1 \pm p_{i}| (z^{-1} \mp 1)}{\sqrt{2} (1 - p_{i}z^{-1})(1 - p_{i}^{*}z^{-1})} \times \prod_{j=1}^{i-1} \frac{(z^{-1} - p_{j})(z^{-1} - p_{j}^{*})}{(1 - p_{j}z^{-1})(1 - p_{j}^{*}z^{-1})}, \quad (1)$$

where p_i^* is complex conjugate of p_i . The conjugate pairs of the poles are considered for making the filter response real valued. By inputting the signal e(n) into the 2*I* filters $\{\Psi_i^+(z), \Psi_i^-(z)\}_{i=1}^I$, the corresponding 2I filtered signals $\{\kappa_i^+(n), \kappa_i^-(n)\}_{i=1}^I$ are obtained. Then, the output signal y(n) is defined by the linear combination of them,

$$y(n) = \sum_{i=1}^{I} \left[\theta_i^+ \kappa_i^+(n) + \theta_i^- \kappa_i^-(n) \right],$$
 (2)

where $\{\theta_i^+, \theta_i^-\}_{i=1}^I$ is the set of coefficients. This model is linear in these coefficients, and therefore they can be easily optimized in contrast to the ordinary rational filters which are usually non-linear in their parameters.

Let s(n) be a piano sound to be modeled. Then, the aim of this research is to approximate s(n) by a parametric model which has a few degrees of freedom. Usually, the parameters of the Kautz model are estimated by the least squares method:

$$\underset{\{\theta_{i}^{+},\theta_{i}^{-}\}_{i=1}^{I}}{\text{Minimize}} \sum_{n=1}^{N} \left| s(n) - \sum_{i=1}^{I} \left[\theta_{i}^{+} \kappa_{i}^{+}(n) + \theta_{i}^{-} \kappa_{i}^{-}(n) \right] \right|^{2}, (3)$$

where N is length of the piano signal s(n). After the parameters $\{\theta_i^+, \theta_i^-\}_{i=1}^I$ are estimated, the approximated signal can be obtained by (i) inputting the signal e(n) into the IIR filters $\{\Psi_i^+(z), \Psi_i^-(z)\}_{i=1}^I$ in Eq. (1), and (ii) taking the linear combination of them as in Eq. (2) with the estimated parameters.

2.1. Two questions arise from Kautz modeling

The easiness of the Kautz model on its parameter estimation comes with a price of restriction on two important factors: poles $\{p_i\}_{i=1}^{I}$ and input signal e(n). Since the poles must be fixed in advance, their suitable selection is crucial in the Kautz model. Moreover, the input signal also has to be fixed beforehand as it can be arbitrary for our purpose. In addition, input signal whose energy is concentrated at the beginning is preferable for less memory usage. That is, the following two questions must be resolved:

- How to decide the effective poles $\{p_i\}_{i=1}^{I}$ while the number of them should be small as possible.
- How to decide the input signal e(n) whose energy should be concentrated within the small time interval.

In the next section, an answer to these questions in terms of sparsity is proposed.

3. PROPOSED METHOD

As mentioned in the introduction, the order of the filter should be low for implementation. However, this requirement contradicts with the Kautz model because it cannot optimize the poles. That is, the Kautz model usually requires more poles than necessary to achieve a certain approximation accuracy since preparing poles critical for the approximation in advance is generally difficult. In addition, for saving the computational resources, the input signal for exciting the filter response should have small number of non-zero elements, which must be fixed in advance for estimating parameters of the Kautz model. In this section, methods of selecting the poles and the excitation signal are proposed to circumvent the above difficulties.

3.1. Sparse selection of poles after candidates generation

As stated in the previous section, the Kautz model requires fixed poles in advance, and they cannot be optimized within the framework of that model. Therefore, to obtain a better set of poles, we propose a method of generating several candidates of the poles and a method of sparsely selecting the prominent poles from them, which is an often-utilized strategy in acoustics [20–24].

Firstly, candidates of the poles must be generated to construct the filters. This process is important because the final accuracy of the approximation is determined by the quality of candidates. Therefore, we propose to generate the candidate from the signal to be approximated. By using an autoregressive (AR) spectral estimation technique [25], a set of poles suitable for the approximation can be estimated from the signal. Multiple data (including the target signal) may be utilized for the same note to generate multiple sets of poles because a single data only gives a single set of poles which may not be optimal in terms of the Kautz modeling. By this process, similar but slightly different poles, which are all suitable for approximating a signal of that note, are obtained. Note that the best set of poles for approximating the signal s(n) depends on the input signal e(n) which excites the filters.

Then, the poles prominent for approximating the target signal must be selected from them. In this paper, formulation by an ℓ_0 constrained optimization problem is proposed to accomplish such selection. Let Eq. (3) be shortly written as

$$\underset{\boldsymbol{\theta}}{\text{Minimize}} \quad \sum_{n=1}^{N} |s(n) - \boldsymbol{\kappa}(n)^{T} \boldsymbol{\theta}|^{2}, \tag{4}$$

where $\boldsymbol{\kappa}^{T}$ denotes transpose of $\boldsymbol{\kappa}$, and

$$\boldsymbol{\theta} = [\theta_1^+, \theta_1^-, \theta_2^+, \theta_2^-, \dots, \theta_I^+, \theta_I^-]^T,$$
(5)

$$\boldsymbol{\kappa}(n) = [\kappa_1^+(n), \kappa_1^-(n), \dots, \kappa_I^+(n), \kappa_I^-(n)]^T.$$
(6)

Our proposal is to impose ℓ_0 constraint into this problem as

$$\underset{\boldsymbol{\theta}}{\text{Minimize }} \sum_{n=1}^{N} \left| s(n) - \boldsymbol{\kappa}(n)^{T} \boldsymbol{\theta} \right|^{2} \text{ subject to } \left\| \boldsymbol{\theta} \right\|_{0} \leq P, (7)$$

where $\|\cdot\|_0$ is the number of non-zero elements so-called ℓ_0 -norm which is regarded as an ideal measure of sparsity, and P is the desired number of selected poles. That is, the selection of the poles is recast into selection of positions of non-zero elements. A solution to this problem gives least squares approximation of the signal by P pole pairs while

the selected poles are automatically determined. Therefore, by solving this problem defined for a large number of candidates obtained as in the previous paragraph, the best P poles are expected to be selected afterward.

The difficulty associated with this proposal is basically owing to the non-convexity of the above optimization problem. Therefore, a recent algorithm based on the difference of convex (DC) programming called proximal DC algorithm [26] is employed in this paper for solving Eq. (7).

3.2. Generating excitation signal based on sparse LPC

In the above proposal, a large number of poles are obtained in advance, and the optimization problem is considered. However, the intermediate signals denoted by $\kappa(n)$ in Eq. (7) not only depend on the poles but also depend on the inputted excitation signal e(n). That is, the input signal e(n) have to be defined before running the algorithm for solving Eq. (7). Here, we propose to utilize a linear prediction technique to generate such excitation signal.

Linear prediction is the well-known parametric signal modeling technique which approximates the signal by a prediction filter. It allows to synthesize the original signal from the prediction residual through the corresponding synthesis filter. Therefore, residual of linear prediction can be regarded as the component of signal that is difficult to be approximated by a filter. In other words, predictable part of the signal should be easy to be approximated by the Kautz model. Then, it can be presumed that such residual may be suitable for the input excitation signal e(n) of the Kautz model.

However, residual generated by the ordinary linear prediction may not improve the approximation accuracy of piano sound as shown in the next section. Instead, we propose to use the sparse linear prediction technique [27, 28] for generating the excitation signal. The linear prediction model of L coefficients is defined as

$$x(n) = \sum_{l=1}^{L} a_l x(n-l) + e(n),$$
(8)

where e(n) is the prediction residual, and a_l is the coefficient of the linear prediction model. The ordinary linear prediction is based on the ℓ_2 -norm as it is formulated by the least squares method,

$$\underset{\{a_l\}_{l=1}^{L}}{\text{Minimize}} \sum_{n=1}^{N} \left| s(n) - \sum_{l=1}^{L} a_l s(n-l) \right|^2.$$
(9)

In contrast, the sparse linear prediction is based on the ℓ_1 -norm which induces sparsity,

$$\underset{\{a_l\}_{l=1}^L}{\text{Minimize}} \quad \sum_{n=1}^N \left| s(n) - \sum_{l=1}^L a_l s(n-l) \right|, \tag{10}$$

Therefore, its residual $e(n) = s(n) - \sum_{l=1}^{L} a_l s(n-l)$ is more sparse than the ordinary one, and its energy is concentrated



Fig. 1. Piano sound of A4 (441 Hz) and its residual of sparse linear prediction in Section 3.2. The enlarged view shows the first 100 samples which illustrate that most of the energy of the residual is concentrated in the first few samples.

at certain time instances. This residual of the sparse linear prediction is used as the excitation signal of the Kautz model.

This non-smooth variant of least squares method is called as least absolute deviations, and there exist many algorithms for solving it. As it is non-smooth convex optimization problem, the well-known alternating direction method of multipliers (ADMM) [29, 30] is adopted in this paper to solve it.

3.3. Summary of the proposed method

The proposed method for approximating piano sound by using the Kautz model is summarized as follows:

- 1. Get audio signal of the target piano sound s(n).
- 2. Apply sparse linear prediction to the target sound to obtain a prediction residual e(n) which is utilized as the excitation signal of the Kautz filters.
- 3. Generate a large number of candidates of poles for the Kautz modeling from piano sound by an AR spectral analysis technique.
- 4. Construct the Kautz filter $\{\Phi_i^{\pm}\}$ from obtained poles in the previous step.
- 5. Input the excitation signal e(n) obtained by sparse linear prediction in Step 2 to the filters for calculating $\kappa(n)$ in Eq. (7).
- 6. Solve Eq. (7) by the proximal DC algorithm to obtain P poles and the corresponding parameters $\{\theta_i^{\pm}\}$.

After executing these steps, the result ends up with the Kautz model $\{\Phi_i^{\pm}\}$ with the coefficients $\{\theta_i^{\pm}\}$, and the excitation signal e(n). Then, the approximated piano sound is generated by inputting e(n) to the filters and taking the linear combination using estimated coefficients as introduced in Section 2. Since the number of poles P can be set to a desired number, the accuracy and the computational cost can be adjusted.



Fig. 2. Waveforms of the target and approximated signals. P is the number of selected poles and RE denotes the relative error.

4. EXPERIMENTS

The proposed method was applied to a real piano sound (C. Bechstein) of A4 (441 Hz). Since the proposed method consists of several steps, each step is explained in each preceding paragraph.

Firstly, sparse linear prediction was applied to obtain the excitation signal e(n). The obtained residual is shown in Fig. 1, where the length of prediction filter L in Eq. (8) was 1000. As shown in the figure, the residual has quite small values at most of the duration, and the remaining energy was concentrated around the first 10 samples. Therefore, these samples are expected to contain the components difficult to be represented by a filter.

Next, by using first 10 samples of this residual as the excitation signal, the parameters of the Kautz model were estimated. For generating candidates of the effective poles, AR model of 1000 order was applied to 12 piano sounds of the same note. Then, only P important poles were selected from those poles by solving Eq. (7). The target signal and its approximated version are shown in Fig. 2 for $P \in \{10, 20, 50\}$, where RE denotes the relative error of the approximation:

RE =
$$\left(\sum_{n=1}^{N} |s(n) - \hat{y}(n)|^2\right)^{\frac{1}{2}} / \left(\sum_{n=1}^{N} |s(n)|^2\right)^{\frac{1}{2}}$$
, (11)

where s(n) is the target signal and $\hat{y}(n)$ is its approximation. From this result, it can be seen that the approximation error becomes smaller as the number of selected pole increases. The enlarged view on the figure (c) confirms that the approximated sound has the similar phase as the target signal. Note that the number of selected poles P trades the computational complexity and approximation accuracy. Therefore, the number of poles can be choosen depending on desired computational cost, and the approximation error can be checked to confirm whether that amount of error is acceptable.

Finally, to see the effect of sparse linear prediction, the relative error was compared with the ordinary linear prediction. After obtaining the residual, first M samples were used as the input excitation of the Kautz model. The relative error



Fig. 3. Relative error of the approximation in terms of the length of the excitation signal.

is shown in Fig. 3, where the horizontal axis represents the length of the utilized residual M. In this figure, the left most point (M = 1) corresponds to the unit impulse which may be the easiest choice of the excitation signal. From Fig. 3, the error can be reduced comparing to the unit impulse only when the sparse linear prediction is used, and the ordinary linear prediction does not decrease the error. This result confirms that sparsifying the residual is important for generating a better excitation signal based on the framework of linear prediction.

5. CONCLUSION

In this paper, a new modal based method for approximating piano sound by the Kautz model was proposed. The proposed method aims to resolve the two issues of the Kautz model by two sparsity-aware optimization. By applying the proposed method to a real piano sound, it was confirmed that the two kinds of sparsity are important for approximating it. For the future work, degree of sparsity should be enhanced to obtain a better excitation signal, and listening tests are necessary in the evaluation of the proposed method. Moreover, the proposed method should be compared with the previous methods.

6. REFERENCES

- A. Chaigne and A. Askenfelt, "Numerical simulations of piano strings. I. a physical model for a struck string using finite difference methods," *J. Acoust. Soc. Am.*, vol. 95, no. 2, pp. 1112–1118, 1994.
- [2] J. Chabassier, M. Duruflé, and P. Joly, "Time domain simulation of a piano. part 2: numerical aspects," *ESAIM: Math. Model. Numer. Anal.*, vol. 50, no. 1, pp. 93–133, 2016.
- [3] Y. Masuyama, T. Kusano, K. Yatabe, and Y. Oikawa, "Modal decomposition of musical instrument sound via alternating direction method of mutipliers," in *IEEE Int. Conf. Acoust.*, *Speech, Signal Process. (ICASSP)*, Apr. 2018.
- [4] T. Kusano, K. Yatabe, and Y. Oikawa, "Envelope estimation by tangentially constrained spline," in *IEEE Int. Conf. Acoust.*, *Speech, Signal Process. (ICASSP)*, Apr. 2018.
- [5] M. Aramaki, J. Bensa, L. Daudet, P. Guillemain, and R. Kronland-Martinet, "Resynthesis of coupled piano string vibrations based on physical modeling," *J. New Music Res.*, vol. 30, no. 3, pp. 213–226, 2001.
- [6] H.-M. Lehtonen, J. Rauhala, and V. Valimaki, "Sparse multistage loss filter design for waveguide piano synthesis," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2005, pp. 331–334.
- [7] J. Rauhala, H.-M. Lehtonen, and V. Välimäki, "Multi-ripple loss filter for waveguide piano synthesis," in *Proc. Int. Comput. Music Conf.*, 2005.
- [8] B. Bank, "Nonlinear interaction in the digital waveguide with the application to piano sound synthesis.," in *Proc. Int. Comput. Music Conf.*, 2000.
- [9] G. E. Garnett, "Modeling piano sound using waveguide digital filtering techniques," in *Proc. Int. Comput. Music Conf.*, Sep. 1987, pp. 88–95.
- [10] M. Karjalainen, "1-D digital waveguide modeling for improved sound synthesis," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2002, vol. 2, pp. II–1869–II– 1872.
- [11] J. O. Smith and S. A. Van Duyne, "Commuted piano synthesis," in *Proc. Int. Comput. Music Conf.*, Sep. 1995, pp. 335– 342.
- [12] J. Bensa, S. Bilbao, R. Kronland-Martinet, and J. O Smith III, "The simulation of piano string vibration: From physical models to finite difference schemes and digital waveguides," *J. Acoust. Soc. Am.*, vol. 114, no. 2, pp. 1095–1107, 2003.
- [13] B. Bank, F. Avanzini, G. Borin, G. De Poli, F. Fontana, and D. Rocchesso, "Physically informed signal processing methods for piano sound synthesis: A research overview," *EURASIP J. Adv. Signal Process.*, vol. 2003, no. 10, pp. 941– 952, Sep. 2003.
- [14] B. Bank, S. Zambon, and F. Fontana, "A modal-based realtime piano synthesizer," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 4, pp. 809–821, 2010.
- [15] S. Zambon, H. M. Lehtonen, and B. Bank, "Simulation of piano sustain-pedal effect by parallel second-order filters," in *Proc. 11th Int. Conf. Digit. Audio Eff. (DAFx-08)*, Sep. 2008, pp. 199–204.

- [16] B. Bank, "Perceptually motivated audio equalization using fixed-pole parallel second-order filters," *IEEE Signal Process. Lett.*, vol. 15, pp. 477–480, 2008.
- [17] B. Ninness and F. Gustafsson, "A unifying construction of orthonormal bases for system identification," *IEEE Trans. Autom. Control*, vol. 42, no. 4, pp. 515–521, 1997.
- [18] G. Vairetti, T. van Waterschoot, M. Moonen, M. Catrysse, and S. H. Jensen, "Sparse linear parametric modeling of room acoustics with orthonormal basis functions," in 22nd Eur. Signal Process. Conf. (EUSIPCO), Sep. 2014, pp. 1–5.
- [19] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 7, pp. 1547–1561, 2017.
- [20] K. Yatabe and Y. Oikawa, "Optically visualized sound field reconstruction based on sparse selection of point sound sources," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2015, pp. 504–508.
- [21] K. Yatabe and Y. Oikawa, "Optically visualized sound field reconstruction using Kirchhoff–Helmholtz equation," Acoust. Sci. & Tech., vol. 36, no. 4, pp. 351–354, 2015.
- [22] Y. Tamura, K. Yatabe, and Y. Oikawa, "Least-squares estimation of sound source directivity using convex selector of a better solution," *Acoust. Sci. & Tech.*, vol. 38, no. 3, pp. 128–136, 2017.
- [23] T. Tachikawa, K. Yatabe, Y. Ikeda, and Y. Oikawa, "Sound source localization based on sparse estimation and convex clustering," *Proc. Meet. Acoust.*, vol. 29, no. 1, pp. 055004, 2016.
- [24] T. Tachikawa, K. Yatabe, and Y. Oikawa, "Coherence-adjusted monopole dictionary and convex clustering for 3D localization of mixed near-field and far-field sources," in *IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2017, pp. 3191–3195.
- [25] S. Kay, Modern spectral estimation: Theory and application, Prentice Hall, New Jersey, 1988.
- [26] J. Gotoh, A. Takeda, and K. Tono, "DC formulations and algorithms for sparse optimization problems," *Math. Progr.*, Jul. 2017.
- [27] D. Giacobello, M. G. Christensen, M. N. Murthi, S. H. Jensen, and M. Moonen, "Sparse linear prediction and its applications to speech processing," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 20, no. 5, pp. 1644–1657, Jul. 2012.
- [28] D. Giacobello, M. G. Christensen, T. Jensen, M. Murthi, and S. Jensen, "Stable 1-norm error minimization based linear predictors for speech modeling," *IEEE/ACM Trans. Audio, Speech Lang. Process.*, vol. 22, pp. 912–922, May 2014.
- [29] N. Parikh and S. Boyd, "Proximal algorithms," Found. Trends Optim., vol. 1, no. 3, pp. 123–231, 2013.
- [30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.