MODAL DECOMPOSITION OF MUSICAL INSTRUMENT SOUND VIA ALTERNATING DIRECTION METHOD OF MULTIPLIERS

Yoshiki Masuyama, Tsubasa Kusano, Kohei Yatabe and Yasuhiro Oikawa

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

ABSTRACT

For a musical instrument sound containing partials, or modes, the behavior of modes around the attack time is particularly important. However, accurately decomposing it around the attack time is not an easy task, especially when the onset is sharp. This is because spectra of the modes are peaky while the sharp onsets need a broad one. In this paper, an optimization-based method of modal decomposition is proposed to achieve accurate decomposition around the attack time. The proposed method is formulated as a constrained optimization problem to enforce the perfect reconstruction property which is important for accurate decomposition. For optimization, the alternating direction method of multipliers (ADMM) is utilized, where the update of variables is calculated in closed form. The proposed method realizes accurate modal decomposition in the simulation and real piano sounds.

Index Terms— Constrained filtering, Fourier transform, perfect reconstruction, causality, piano.

1. INTRODUCTION

Modal decomposition is one of the most fundamental tools for analyzing a musical instrument sound containing partials, or modes, because the decaying processes of the modes greatly affect the timbre of musical instruments [1, 2]. Each mode may decay with a complicated decay process which characterizes the sound. For example, in the piano, transfer of energy among coupled strings, bridge and soundboard causes special decay patterns of the modes called "double decay" and "beats" which make the piano sound distinctive [3–7]. Since modes contain such significant information of a musical instrument sound, modal decomposition plays an important role in the studies on a musical instrument sound [7–9].

Modal decomposition also has an important role in synthesizing musical instrument sound based on models described later. As modes contain significant information of the corresponding sound, parametric modeling of each mode is often considered in the context of sound synthesis. Many models have been proposed in this respect, including exponentially damped sinusoidal (EDS) model [10–13], and damped and delayed sinusoidal (DDS) model [14–17] as the extension of EDS model. In order to represent complex decaying process such as "double decay" and "beats", recent years, adaptive harmonic model (AHM) has been applied to modeling of musical instruments sound [18–25]. In AHM, each mode is represented by the product of time-varying amplitude and a frequency modulated sinusoid. For using these models, modal decomposition is needed for estimating model parameter [26], especially in AHM. Therefore, a modal decomposition method accurately separating each mode is required.

For percussive instruments such as bells and plucked/struck string instruments (e.g. guitars and pianos), the behavior of modes around the attack time is an important factor for characterizing their timbre. However, accurately decomposing these instrument sounds into modes around the attack time is not an easy task because these instruments sound has sharp onset at the attack time. The spectra of modes are peaky but wideband owing to the sharp onset, and this obstructs accurate modal decomposition around the attack time. Hence, modal analysis around the attack time is often ignored due to this difficulty, and it takes into account only the decay part.

In this paper, an optimization-based method of modal decomposition is proposed to overcome the accuracy deficient around the attack time. It is formulated as an optimization with constraints of perfect reconstruction and causality in order to eliminate the phase delay and pre-ringing. For optimization, the alternating direction method of multipliers (ADMM) is utilized, where the update of variables is calculated in closed form. The performance of the proposed method is shown by simulation and the real piano sound.

2. MODAL DECOMPOSITION BASED ON FILTERBANK

A mode of a musical instrument sound corresponds to a single spectral peak. In this paper, a sound which consists of attack followed by decaying partials without noticeable frequency modulation is considered, such as percussive, plucked string, or struck string instruments. That is, signals with strong frequency modulation like vibrato are outside the scope of this paper. Based on this assumption, modal decomposition using a filterbank is considered here.

Let a signal of given musical instrument sound be denoted by $\mathbf{s} \in \mathbb{R}^L$, and its Fourier spectrum be represented by $\hat{\mathbf{s}} \in \mathbb{C}^L$, where L is the length of the signal. Hereafter, Fourier transform of $\mathbf{z}$ is denoted by $\hat{\mathbf{z}}$ $(= \mathbf{F}\mathbf{z})$, where $\mathbf{F} \in \mathbb{C}^{L\times L}$ is the Fourier transform matrix. Assuming the above condition to the signal $\mathbf{s}$, ith mode can be extracted by linear filtering in the frequency domain as

$$\hat{\mathbf{x}}_i = \hat{\mathbf{H}}_i \hat{\mathbf{s}}, \tag{1}$$

where $\hat{\mathbf{H}}_i \in \mathbb{C}^{L\times L}$ is a diagonal matrix whose diagonal elements are the frequency response of a predefined filter $\hat{\mathbf{h}}_i \in \mathbb{C}^L$ designed specifically for the ith mode $\mathbf{x}_i$. By preparing N bandpass filters corresponding to the N modes, linear filtering given by Eq. (1) approximately obtains the modes. The accuracy of this decomposition depends on the design strategy of $\hat{\mathbf{h}}_i$.

2.1. Potential issues of linear filtering

It is well-known that linear filtering cannot achieve causality without phase delay. If a filter is causal, i.e., no component is generated before the onset, then there exists some phase delay which shifts the waveform. Such delay in phase greatly reduces the accuracy of modal decomposition. On the other hand, if a filter does not have phase delay (zero-phase), some components so-called pre-ringing

exist before the attack. This trade-off between phase delay and pre-ringing indicates that a linear filter cannot avoid deformation of an extracted mode. Steep attack of a mode is always corrupted by a linear filter, and thus accurate modal decomposition around the attack time cannot be accomplished by linear filtering. In other words, an accurate decomposition method must be a non-linear process.

2.2. Interpretation of linear filtering as least squares method

Before proceeding to the proposed method, an interpretation of linear filtering as the least squares method is introduced here. This interpretation will illustrate the relation between the proposed method and the ordinary linear filtering.

Let a filter $\hat{\mathbf{H}}_i$ admit the inverse $\hat{\mathbf{H}}_i^{-1}$. Then, linear filtering in Eq. (1) can be rewritten as

$$\hat{\mathbf{H}}_i^{-1}\hat{\mathbf{x}}_i = \hat{\mathbf{s}}, \tag{2}$$

which can be interpreted as the least squares method,

$$\min_{\hat{\mathbf{x}}_i} \ \frac{1}{2}\left\|\hat{\mathbf{H}}_i^{-1}\hat{\mathbf{x}}_i - \hat{\mathbf{s}}\right\|_2^2, \tag{3}$$

whose solution $\hat{\mathbf{x}}_i$ coincides with the original filtering. This interpretation indicates that a linear filtering can be recast to an optimization problem which is a more flexible form. Let every mode be denoted by $\mathbf{x}$. Then, in order to consider all modes simultaneously, modal decomposition by a linear filterbank is formulated as

$$\min_{\hat{\mathbf{x}}} \ \frac{1}{2}\left\|\hat{\mathbf{H}}^{-1}\hat{\mathbf{x}} - \hat{\mathbf{d}}\right\|_2^2, \tag{4}$$

where $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_1^T, \ldots, \hat{\mathbf{x}}_N^T]^T \in \mathbb{C}^{NL}$, $\hat{\mathbf{x}}^T$ is the transpose of $\hat{\mathbf{x}}$, $\hat{\mathbf{d}} \in \mathbb{C}^{NL}$ is the vector concatenated N copies of $\hat{\mathbf{s}}$, and $\hat{\mathbf{H}} \in \mathbb{C}^{NL\times NL}$ is the diagonal matrix whose diagonal elements are given by $[\hat{\mathbf{h}}_1^T, \ldots, \hat{\mathbf{h}}_N^T]^T$. This representation allows a compact notation of the proposed method in the next section.

3. PROPOSED METHOD

As in the previous section, linear filtering can be interpreted as the least squares problem. This point of view allows us to incorporate additional constraints into the filtering process. In this section, we propose a modal decomposition method by adding constraints into the least squares problem so that the undesirable trade-off discussed in Section 2.1 is avoided, which results in much higher accuracy comparing to the linear filterbank. The proposed method consists of two constraints, perfect reconstruction and causality, and each constraint is explained one-by-one in the preceding subsections.

3.1. Constraint of perfect reconstruction condition

For accurate modal decomposition, a perfect reconstruction property is considered first. We say that the decomposed modes satisfy the perfect reconstruction condition when $\mathbf{s} = \sum_{i=1}^N \mathbf{x}_i$ holds. That is, the original signal can be perfectly reconstructed by adding the decomposed modes. By imposing this property into Eq. (4), a filtering problem with perfect reconstruction constraint is defined as

$$\min_{\hat{\mathbf{x}}} \ \frac{1}{2}\left\|\hat{\mathbf{H}}^{-1}\hat{\mathbf{x}} - \hat{\mathbf{d}}\right\|_2^2 \ \text{ s.t. } \ \hat{\mathbf{s}} = \sum_{i=1}^N \hat{\mathbf{x}}_i. \tag{5}$$

After solving this problem, filtered signals satisfying the perfect reconstruction condition can be obtained.

In order to eliminate the potential issues of linear filtering into the proposed method, the formulation is modified in three parts. Firstly, let $\hat{\mathbf{H}}^{-1}$ be replaced by an arbitrary diagonal matrix $\mathbf{W}$. This modification allows a zero in the diagonal entries of $\mathbf{W}$, while $\hat{\mathbf{H}}^{-1}$ does not allow it owing to the inversion. Secondly, to handle the complicated component caused by the attack, let a residual $\mathbf{x}_{N+1} \in \mathbb{R}^L$, which is expected to be a pulse at the attack time, be also considered. Then, the perfect reconstruction condition is relaxed to $\mathbf{s} = \sum_{i=1}^{N+1} \mathbf{x}_i$, where a non-modal component is allowed in the $\mathbf{x}_{N+1}$. Thirdly, although the fidelity to the data is considered in both the first and second terms of Eq. (5), the data in the first term is omitted (data fidelity is considered only in the constraint). Based on these three modifications, a modal decomposition problem of the following form in the frequency domain is considered:

$$\min_{\hat{\mathbf{x}}} \ \frac{1}{2}\left\|\mathbf{W}\hat{\mathbf{x}}\right\|_2^2 \ \text{ s.t. } \ \hat{\mathbf{s}} = \sum_{i=1}^{N+1} \hat{\mathbf{x}}_i, \tag{6}$$

where $\hat{\mathbf{x}} = [\hat{\mathbf{x}}_1^T, \ldots, \hat{\mathbf{x}}_{N+1}^T]^T \in \mathbb{C}^{(N+1)L}$, the diagonal matrix $\mathbf{W} \in \mathbb{C}^{(N+1)L\times(N+1)L}$ whose diagonal elements are given by $[\mathbf{w}_1^T, \ldots, \mathbf{w}_{N+1}^T]^T$, and $\mathbf{w}_i \in \mathbb{C}^L$ is a given weight for ith mode in the frequency domain.

3.2. Closed form solution to Eq. (6)

Let $\mathbf{W}_i \in \mathbb{C}^{L\times L}$ be a diagonal matrix whose diagonal elements are given by $\mathbf{w}_i$. Then Eq. (6) can be rewritten into an unconstrained optimization problem,

$$\min_{\hat{\mathbf{x}}_1,\ldots,\hat{\mathbf{x}}_N} \ \frac{1}{2}\sum_{i=1}^N \left\|\mathbf{W}_i\hat{\mathbf{x}}_i\right\|_2^2 + \frac{1}{2}\left\|\mathbf{W}_{N+1}(\hat{\mathbf{s}} - \sum_{i=1}^N \hat{\mathbf{x}}_i)\right\|_2^2, \tag{7}$$

which can be solved for each frequency separately:

$$\min_{\hat{x}_{1\xi},\ldots,\hat{x}_{N\xi}} \ \frac{1}{2}\sum_{i=1}^N \left|w_{i\xi}\hat{x}_{i\xi}\right|^2 + \frac{1}{2}\left|w_{(N+1)\xi}(\hat{s}_\xi - \sum_{i=1}^N \hat{x}_{i\xi})\right|^2, \tag{8}$$

where ξ is the frequency index, $w_{i\xi}$ is the ξth element of $\mathbf{w}_i$, $\hat{x}_{i\xi}$ is the ξth element of $\hat{\mathbf{x}}_i$, and $\hat{s}_\xi$ is the ξth element of $\hat{\mathbf{s}}$. The solution to Eq. (8) is obtained by

$$\hat{x}_{i\xi} = \frac{\prod_{j\neq i}|w_{j\xi}|^2}{\sum_{k=1}^{N+1}\prod_{j\neq k}|w_{j\xi}|^2}\hat{s}_\xi, \tag{9}$$

if the denominator is not zero. By denoting the fraction with $g_{i\xi}$ which represents the gain, the modal decomposition defined by Eq. (6) is given as

$$\hat{\mathbf{x}}_i = \mathbf{G}_i\hat{\mathbf{s}}, \tag{10}$$

where $\mathbf{G}_i \in \mathbb{R}^{L\times L}$ is the diagonal matrix whose diagonal entries are $\hat{\mathbf{g}}_i$, and $\hat{\mathbf{g}}_i = [g_{i1}, \ldots, g_{iL}] \in \mathbb{R}^L$ is gain for extracting ith mode.

These gains depend on the ratio of the weights. Simply observing that $g_{i\xi} = 1$ and $g_{(j\neq i)\xi} = 0$ if $w_{i\xi} = 0$ for any frequency index ξ, Eq. (6) can be interpreted as a zero-phase filterbank with the perfect reconstruction property which makes the modes exclusive of each other in the frequency domain.

3.3. Proposed forulation with causality constraint

Although the perfect reconstruction property is indispensable for the accurate decomposition, it does not eliminate the pre-ringing which

deteriorates the accuracy around the attack time. Therefore, an additional constraint corresponding to causality is considered:

$$\min_{\hat{\mathbf{x}}} \ \frac{1}{2}\left\|\mathbf{W}\hat{\mathbf{x}}\right\|_2^2 \ \text{ s.t. } \ \hat{\mathbf{s}} = \sum_{i=1}^{N+1}\hat{\mathbf{x}}_i, \ \ \left[\mathbf{F}^{-1}\hat{\mathbf{x}}_i\right]_n = 0 \ \ (n < \tau_A), \tag{11}$$

where n is the time index and τ_A is time index corresponding to the attack time. Since this causality constraint explicitly eliminates the pre-ringing, modal decomposition without phase delay and/or pre-ringing is realized by solving Eq. (11).

3.4. ADMM algorithm for solving Eq. (11)

In this paper, ADMM [27, 28] is adopted for solving Eq. (11). ADMM is an algorithm which can solve the following convex optimization problem:

$$\min_{\mathbf{x}\in\mathbb{C}^L,\mathbf{z}\in\mathbb{C}^L} f(\mathbf{x}) + g(\mathbf{z}) \quad \text{s.t.} \quad \mathbf{x} = \mathbf{z}, \tag{12}$$

where f and g are proper and lower-semicontinuous convex functions. For any $\mathbf{z}_0$, $\mathbf{u}_0$ and $\rho > 0$, ADMM is given by

$$\mathbf{x}_{k+1} = \mathrm{prox}_{\rho f}(\mathbf{z}_k - \mathbf{u}_k), \tag{13}$$
$$\mathbf{z}_{k+1} = \mathrm{prox}_{\rho g}(\mathbf{x}_{k+1} + \mathbf{u}_k), \tag{14}$$
$$\mathbf{u}_{k+1} = \mathbf{u}_k + \mathbf{x}_{k+1} - \mathbf{z}_{k+1}, \tag{15}$$

where k is the iteration index, and $\mathrm{prox}_{\rho f}(\cdot)$ is the proximity operator of f defined by [29]

$$\mathrm{prox}_{\rho f}(\mathbf{y}) = \arg\min_{\mathbf{x}} \ f(\mathbf{x}) + \frac{1}{2\rho}\left\|\mathbf{y} - \mathbf{x}\right\|_2^2. \tag{16}$$

For applying the ADMM algorithm to the proposed method in Eq. (11), it is reformulated as the following equivalent problem:

$$\min_{\hat{\mathbf{x}},\hat{\mathbf{z}}} \ \frac{1}{2}\left\|\mathbf{W}\hat{\mathbf{x}}\right\|_2^2 + \chi_{C_1}(\hat{\mathbf{x}}) + \chi_{C_2}(\hat{\mathbf{z}}) \quad \text{s.t.} \quad \hat{\mathbf{x}} = \hat{\mathbf{z}}, \tag{17}$$

where χ_C is the indicator function of a closed nonempty convex set C (i.e., $\chi_C(\mathbf{x}) = 0$ if $\mathbf{x} \in C$, and $\chi_C(\mathbf{x}) = \infty$ otherwise), C_1 and C_2 are the sets corresponding to each constraint in Eq. (11),

$$C_1 = \left\{ \hat{\mathbf{x}} \in \mathbb{C}^{(N+1)L} \mid \hat{\mathbf{s}} = \textstyle\sum_{i=1}^{N+1}\hat{\mathbf{x}}_i \right\}, \tag{18}$$
$$C_2 = \left\{ \hat{\mathbf{z}} \in \mathbb{C}^{(N+1)L} \mid [\mathbf{F}^{-1}\hat{\mathbf{z}}_i]_n = 0 \ \ (n < \tau_A) \right\}. \tag{19}$$

Then, by regarding the functions in Eq. (17) as

$$f(\hat{\mathbf{x}}) = \frac{1}{2}\left\|\mathbf{W}\hat{\mathbf{x}}\right\|_2^2 + \chi_{C_1}(\hat{\mathbf{x}}), \qquad g(\hat{\mathbf{z}}) = \chi_{C_2}(\hat{\mathbf{z}}), \tag{20}$$

the ADMM algorithm for Eq. (11) is obtained as follows:

$$\hat{\mathbf{x}}_{k+1} = \mathrm{prox}_{\rho f}(\hat{\mathbf{z}}_k - \hat{\mathbf{u}}_k), \tag{21}$$
$$\hat{\mathbf{z}}_{k+1} = P_{C_2}(\hat{\mathbf{x}}_{k+1} + \hat{\mathbf{u}}_k), \tag{22}$$
$$\hat{\mathbf{u}}_{k+1} = \hat{\mathbf{u}}_k + \hat{\mathbf{x}}_{k+1} - \hat{\mathbf{z}}_{k+1}, \tag{23}$$

where P_{C_2} is metric projection onto C_2 which can be calculated by

$$P_{C_2}(\hat{\mathbf{z}}) = \mathbf{F}P_{C_2'}(\mathbf{F}^{-1}\hat{\mathbf{z}}), \tag{24}$$

because the Fourier transform matrix $\mathbf{F}$ is unitary [30],

$$P_{C_2'}(z_n) = \begin{cases} 0 & n < \tau_A, \\ z_n & \text{otherwise}, \end{cases} \tag{25}$$

and $C_2' = \{ \mathbf{z} \in \mathbb{R}^{(N+1)L} \mid z_n = 0 \ \ (n < \tau_A) \}$. The proximity operator in the $\hat{\mathbf{x}}$-update is given by the solution to

$$\min_{\hat{\mathbf{x}}} \ \frac{1}{2}\left\|\mathbf{W}\hat{\mathbf{x}}\right\|_2^2 + \frac{1}{2\rho}\left\|\hat{\mathbf{x}} - \hat{\mathbf{y}}_k\right\|_2^2 \quad \text{s.t.} \quad \hat{\mathbf{s}} = \sum_{i=1}^{N+1}\hat{\mathbf{x}}_i, \tag{26}$$

which is a constrained optimization problem similar to Eq. (6). Therefore, Eq. (26) can be solved analytically in the same way as Eq. (9), which is written as

$$\hat{x}_{i\xi} = [\,\nu_{i\xi} + \rho(\mu_{i\xi} - \eta_{i\xi})\,]/\zeta_{i\xi}, \tag{27}$$

where each term is defined as follows,

$$\nu_{i\xi} = \textstyle\prod_{j\neq i}(|w_{j\xi}|^2 + \rho)\hat{s}_\xi, \tag{28}$$
$$\mu_{i\xi} = \Big(\textstyle\sum_{l\neq i}\prod_{j\neq l,i}(|w_{j\xi}|^2 + \rho)\Big)\hat{y}_{i\xi}, \tag{29}$$
$$\eta_{i\xi} = \textstyle\sum_{l\neq i}\Big(\prod_{j\neq l,i}(|w_{j\xi}|^2 + \rho)\hat{y}_{l\xi}\Big), \tag{30}$$
$$\zeta_\xi = \textstyle\sum_{l=1}^{N+1}\prod_{j\neq k}(|w_{j\xi}|^2 + \rho). \tag{31}$$

By substituting $\hat{y}_{i\xi} = \hat{z}_{i\xi} - \hat{u}_{i\xi}$ in the above formulas, $\hat{\mathbf{x}}$-update in Eq. (21) can be calculated easily.

A decomposed result of the proposed method is obtained by iterating Eqs. (21)–(23) from arbitrary initial values, where Eqs. (21) and (22) are calculated by Eqs. (27) and (24), respectively. While an arbitrary choice is allowable, one preferable choice for the initial value $\mathbf{z}_0$ is a solution to Eq. (6) which is given in Eq. (9).

3.5. Weighting rule for the proposed method

Choice of the weight in Eq. (11) is important since it determines the decomposed result. As the weight penalizes the energy of each frequency component, the ith mode is dominated by some frequencies at which the weight $\mathbf{w}_i$ contains small values. On the other hand, frequencies with large weights do not remain in the result much. That is, the weight should be set small around the center frequency of the target mode and large around that of the non-target modes. To do so, the center frequency of each mode is required.

For determining the center frequencies of modes, the autoregressive (AR) model is utilized. By calculating AR spectrum $\hat{\mathbf{s}}$ of a signal $\mathbf{s}$, the center frequency f_i of the ith mode is obtained by the complex argument of the selected poles p_i. The amplitude of the AR spectrum $a_i = |\hat{\mathbf{s}}(p_i/|p_i|)|$ is also utilized in the weight design.

Based on the information obtained by AR modeling, a weighting rule for the modal decomposition is proposed. Firstly, a resonance filter $\hat{\mathfrak{h}}_i$ with a conjugate pair of poles, p_i and $\overline{p}_i$, is constructed. Then, it is normalized to have the unit amplitude $\tilde{\mathfrak{h}}_i = \hat{\mathfrak{h}}_i / \max_\theta |\hat{\mathfrak{h}}_i(e^{i\theta})|$. Utilizing these elements, the weighting matrix corresponding to the ith mode is designed by

$$\mathbf{W}_i = \mathbf{W}_i^{\text{dip}}\mathbf{W}_i^{\text{peaks}}, \tag{32}$$

where each matrix $\mathbf{W}_i^{(\cdot)} \in \mathbb{C}^{L\times L}$ is a diagonal matrix of $\mathbf{w}_i^{(\cdot)} \in \mathbb{C}^L$ whose elements are calculated as

$$w_{i\xi}^{\text{dip}} = \frac{1}{|\tilde{\mathfrak{h}}_{i\xi}|} - 1, \qquad w_{i\xi}^{\text{peaks}} = \sum_{j\neq i}^{N} a_j|\tilde{\mathfrak{h}}_{j\xi}|. \tag{33}$$

From the construction, $\mathbf{w}_i^{\text{dip}}$ consists of a single dip at the center frequency of ith mode, and $\mathbf{w}_i^{\text{peaks}}$ consists of $N-1$ peaks at the

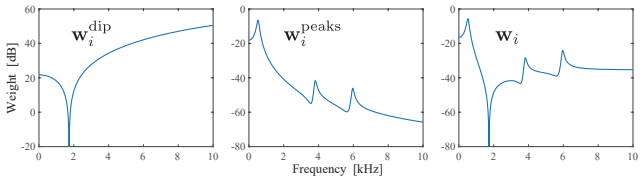

Fig. 1. An example of the proposed weights $\mathbf{w}_i^{\text{dip}}$, $\mathbf{w}_i^{\text{peaks}}$, $\mathbf{w}_i$

Table 1. SDR of the decomposed modes of the simulated signal.

	SDR [dB]			
Modes	1st	2nd	3rd	4th
Causal filter	50.6	2.8	2.0	2.7
Zero-phase filter	33.9	2.7	2.0	2.7
STFT(a)	35.4	26.4	24.3	24.0
STFT(b)	43.4	28.7	27.2	26.3
Proposed method	**111.1**	**97.4**	**96.2**	**93.5**

other frequencies. An example of the proposed weights is illustrated in Fig. 1. For the residual, a special weight $\mathbf{w}_{N+1}^{\text{allPeaks}}$ is proposed,

$$w_{(N+1)\xi}^{\text{allPeaks}} = \lambda \sum_{j=1}^{N} a_j |\tilde{\mathfrak{h}}_{j\xi}|, \tag{34}$$

which eliminates all modes from the residual, where $\lambda > 0$ is a parameter adjusting the energy of the residual.

4. EXPERIMENTS

4.1. Simulation

The proposed method was applied to a simulated musical instrument sound which was synthesized by adding four impulse responses of resonance filters whose center frequencies were $f_1 = 527, f_2 = 1731, f_3 = 3798$, and $f_4 = 5952$ Hz, and the absolute value of the poles which corresponded to the resonance filters were $|p_1| = 0.99996, |p_2| = 0.99978, |p_3| = 0.99965$, and $|p_4| = 0.99959$. They were obtained by approximation of a bell sound with 500 order AR model using Burg's method [31]. Here, the residual was not considered which corresponds to the limit as the parameter of the residual λ to ∞. The proposed method was compared with four other methods: causal filters, zero-phase filters, and the short-time Fourier transform (STFT) with two kinds of parameters. The sampling rate was 44100 Hz, the absolute values of the poles of the two types of filters were 0.99, window and overlap length for STFT were respectively 1024 samples and 512 samples in STFT(a), and 512 samples and 256 samples in STFT(b). The performance of decomposition was evaluated by Signal-to-Distortion Ratio (SDR) [32].

SDR of the proposed method was higher than those of other methods, especially for higher order modes as shown in Table 1. This is because modal decomposition by filters cause mode-mixing in high order modes. On the other hand, the proposed method was able to eliminate such mode-mixing phenomena. In contrast to STFT which does not maintain the energy of modes, the proposed method maintains it by the perfect reconstruction constraint that leads to the higher performance of the proposed method. The causal filters resulted in the phase delay and the zero-phase filters caused the pre-ringing, while STFT also resulted in the pre-ringing depending on the window length. On the other hand, modes decomposed by the proposed method had no phase delay and pre-ringing.

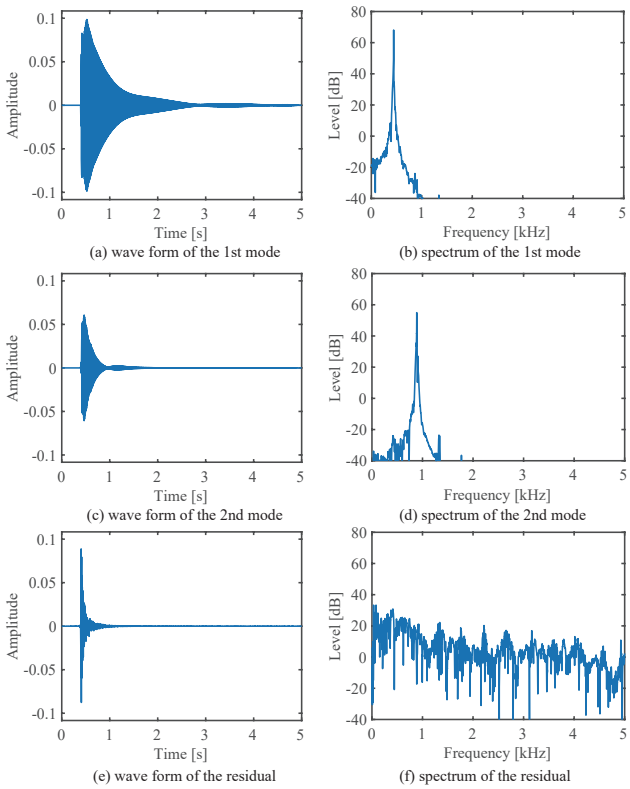

Fig. 2. Results of the proposed method applied to a real piano sound.

4.2. Application to piano sound decomposition

A piano sound of A4 was decomposed into 16 modes and the residual to see the applicability of the proposed method to the real data. The piano sound was also approximated with 1000 order AR model by Burg's method [31] where the sampling rate was 96000 Hz. The weight matrix was constructed by the proposed weighting rule in Section 3.5, and the parameter of the residual was set to $\lambda = 0.01$.

Waveforms and spectra of two decomposed modes and the residual obtained by the proposed method are shown in Fig. 2. According to Fig. 2 (a) and (c), decomposed modes are of long duration, and those spectra are peaky as shown in Fig. 2 (b) and (d). In addition, "double decay" and "beats" which are typical to the piano sound [3–7] can be seen in the decay processes of these modes. On the other hand, according to Fig. 2 (e) and (f), the residual was of short duration and wide band. Hence, the residual should be represented by the non-modal percussive component around the attack time which were produced by the hammer strike.

5. CONCLUSION

In this paper, the modal decomposition method of musical instrument sound is proposed. By interpreting a filtering process as the least squares method, the proposed method is formulated as a constrained optimization problem which enables to incorporate two constraints, so that undesired trade-off of linear filtering is circumvented. The proposed optimization problem is solved by the ADMM algorithm, where the closed form solution of the constrained quadratic problem allows an easy and fast update of the variables.

6. REFERENCES

[1] P. Iverson and L. C. Krumhansl, "Isolating the dynamic attributes of musical timbre," *J. Acoust. Soc. Am.*, vol. 94, no. 5, pp. 2595–2603, Nov. 1993.

[2] S. Handel, *Hearing*, chapter Timbre perception and auditory object identification, pp. 425–461, Moore, B., New York, NY, 1995.

[3] V. Välimäki, J. Huopaniemi, M. Karjalainen, and Z. Janosy, "Physical modeling of plucked string instruments with application to real-time sound synthesis," *J. Audio Eng. Soc.*, vol. 44, no. 5, pp. 331–353, 1966.

[4] G. Weinreich, "Coupled piano strings," *J. Acoust. Soc. Am.*, vol. 62, no. 6, pp. 1474–1484, 1977.

[5] C. T. Hundley, H. Benioff, and W. D. Martin, "Factors contributing to the multiple rate of piano tone decay," *J. Acoust. Soc. Am.*, vol. 64, no. 5, pp. 1303–1309, 1978.

[6] 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.

[7] T. Cheng, S. Dixon, and M. Mauch, "Modelling the decay of piano sounds," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 594–598.

[8] M. Karjalainen, V. Välimäki, and A. A. P. Esquef, "Efficient modeling and synthesis of bell-like sounds," in *Proc. 5th Int. Conf. Digit. Audio Eff. (DAFx-02)*, 2002, pp. 181–186.

[9] A. A. P. Esquef, M. Karjalainen, and V. Välimäki, "Frequency-zooming ARMA modeling for analysis of noisy string instrument tones," *EURASIP J. Adv. Signal Process.*, vol. 10, pp. 935–967, Dec. 2003.

[10] J. Nieuwenhuijse, R. Heusens, and F. E. Deprettere, "Robust exponential modeling of audio signals," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 1998, vol. 6, pp. 3581–3584.

[11] R. Badeau, R. Boyer, and B. David, "EDS parametric modeling and tracking of audio signals," in *Proc. 5th Int. Conf. Digit. Audio Eff. (DAFx)*, 2002, pp. 139–144.

[12] O. Derrien, R. Badeau, and G. Richard, "Parametric audio coding with exponentially damped sinusoids," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 7, pp. 1489–1501, Jul. 2013.

[13] H. Kris, V. Werner, L. Philippe, W. Patrick, and V. H. Sabine, "Perceptual audio modeling with exponentially damped sinusoids," *Signal Process.*, vol. 85, no. 1, pp. 163–176, Jan. 2005.

[14] R. Boyer and K. Abed-Meraim, "Audio transients modeling by damped & delayed sinusoids (DDS)," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2002, vol. 2, pp. 1729–1732.

[15] R. Boyer and K. Abed-Meraim, "Audio modeling based on delayed sinusoids," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 2, pp. 110–120, Mar. 2004.

[16] R. Boyer and K. Abed-Meraim, "Damped and delayed sinusoidal model for transient signals," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1720–1730, May 2005.

[17] K. Kobayashi, D. Takeuchi, M. Iwamoto, K. Yatabe, and Y. Oikawa, "Parametric approximation of piano sound based on Kautz model with sparse linear prediction," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018.

[18] G. P. Kafentzis, G. Degottex, O. Rosec, and Y. Stylianou, "Time-scale modifications based on a full-band adaptive harmonic model," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 8193–8197.

[19] G. Degottex and Y. Stylianou, "Analysis and synthesis of speech using an adaptive full-band harmonic model," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 10, pp. 2085–2095, Oct. 2013.

[20] P. G. Kafentzis and Y. Rosec, O.and Stylianou, "Robust full-band adaptive sinusoidal analysis and synthesis of speech," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*. May 2014, pp. 6260–6264, IEEE.

[21] V. Morfi, G. Degottex, and A. Mouchtaris, "Speech analysis and synthesis with a computationally efficient adaptive harmonic model," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 11, pp. 1950–1962, Nov. 2015.

[22] M. Caetano, P. G. Kafentzis, A. Mouchtaris, and Y. Stylianou, "Adaptive sinusoidal modeling of percussive musical instrument sounds," in *Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2013, pp. 1–5.

[23] C. Marcelo and K. George, "Adaptive modeling of synthetic nonstationary sinusoids," in *Proc. 18th Conf. Digit. Audio Eff. (DAFx-15)*, Nov. 2015, pp. 1–7.

[24] M. Caetano, G. Kafentzis, G. Degottex, A. Mouchtaris, and Y. Stylianou, "Evaluating how well filtered white noise models the residual from sinusoidal modeling of musical instrument sounds," in *IEEE Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2013, pp. 1–4.

[25] M. Caetano, P. G. Kafentzis, A. Mouchtaris, and Y. Stylianou, "Full-band quasi-harmonic analysis and synthesis of musical instrument sounds with adaptive sinusoids," *Appl. Sci.*, vol. 6, May 2016.

[26] T. Kusano, K. Yatabe, and Y. Oikawa, "Envelope estimation by tangentially constraint," in *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018.

[27] M. Fortin and R. Glowinski, *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems*, Elsevier, New York, NY, 1983.

[28] 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, 2010.

[29] N. Parikh and S. Boyd, "Proximal algorithms," *Found. Trends Opt.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.

[30] P. L. Combettes and J. Pesquet, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, chapter Proximal Splitting Methods in Signal Processing, pp. 185–212, Springer, New York, NY, 2011.

[31] Kay S., *Modern spectral estimation*, Prentice Hall, Upper Saddle River, NJ, 1988.

[32] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.