

I-DIVERGENCE-BASED DEREVERBERATION METHOD WITH AUXILIARY FUNCTION APPROACH

Naoki Yasuraoka,[†] Hirokazu Kameoka,[‡] Takuya Yoshioka,[‡] Hiroshi G. Okuno[†]

[†] Graduate School of Informatics
Kyoto University
Yoshida-honmachi, Sakyo-ku
Kyoto 606-8501, Japan

[‡] NTT Communication Science Laboratories
NTT Corporation
3-1 Morinosato Wakamiya, Atsugi,
Kanagawa, 243-0198, Japan

ABSTRACT

This paper presents a dereverberation method based on I-divergence minimization, which is particularly suitable for music signals. Existing dereverberation methods, including one designed for music, sometimes distort instrument sounds and make staccato-like tones. The problems with the Itakura-Saito-divergence-based formulation of the existing methods are 1) their tendency to excessive suppression of direct sound and 2) the difficulty of incorporating and optimizing sophisticated source models suitable for music signals. The proposed I-divergence-based method can mitigate these problems. Employing the I-divergence measure enables us to avoid the direct sound suppression problem and to use powerful music spectrum models without complicating its optimization. We develop a convergence-guaranteed parameter estimation algorithm based on the auxiliary function approach. Experimental results reveal the effectiveness of the proposed dereverberation method.

Index Terms—dereverberation, musical audio signal, I-divergence, auxiliary function method

1. INTRODUCTION

Audio effects such as reverberation, delay, and phase shift are important factors for enriching the perceptual quality of music. On the other hand, they are very harmful for automatic analysis (e.g. melody extraction [1] and chord detection [2]) of musical audio signals. Therefore, technologies for removing and controlling such audio effects will be vital for development of emerging music processing applications including automatic music analyzers [3] and active music listening systems [4].

Dereverberation for music signals, namely music dereverberation, is a much more challenging task than speech dereverberation. This is because musical tones tend to last long unlike moras in natural voices, which makes it difficult to distinguish direct sound from reverberation. Indeed, existing speech dereverberation methods such as [5] are not necessarily effective for music dereverberation. Recently, [6] proposed a music dereverberation method that used a harmonicity-based source model by taking advantage of strong harmonic structures inherent in music signals. This method was shown to be effective in removing reverberation from music signals.

However, we have recently experienced a serious problem even with the method of [6]: dereverberated signals with this method, especially for struck string instruments, tend to be distorted and sound like staccato tones.

We find the main cause of this problem to be that this method (as well as other speech dereverberation methods) is based on the

minimization of the Itakura-Saito (IS) divergence. The main drawback of employing the IS-divergence is twofold. Firstly, with the IS-divergence, we sometimes regard sustained parts of musical tones as reverberation and suppress the sustain mistakenly, which results in the staccato-like sounds. Secondly, because of technical inconveniences in the analysis of the IS-divergence, the parameters of an assumed source model are difficult to estimate. To avoid the latter, [6] settled for using different criteria for dereverberation filter optimization and source model parameter estimation. But, because of this settlement, [6] had to use a non-convergent iterative algorithm.

This paper presents a new dereverberation method based on the I-divergence criterion allowing for the joint estimation of dereverberation filter coefficients and source model parameters. In audio processing, the I-divergence is often used as a goodness of fit measure between an observed spectrum and a source model. The I-divergence measure allows us to use powerful music spectrum models such as harmonic Gaussian mixture model [7, 8] and nonnegative matrix factorization (NMF) model [9, 10]. To realize the I-divergence-based dereverberation, a new convergence-guaranteed optimization algorithm is proposed based on the auxiliary function concept (e.g., [10]).

The remainder of this paper is organized as follows: Section 2 describes the dereverberation problem and qualitatively compares existing and proposed reverberation methods. Section 3 introduces the proposed dereverberation method with I-divergence. Section 4 presents the evaluation results, and Section 5 concludes the paper.

2. COMPARISON OF EXISTING AND PROPOSED DEREVERBERATION METHODS

This section first defines the dereverberation problem. Then a new I-divergence-based dereverberation approach is introduced and compared with the existing IS-divergence-based one.

2.1. Problem Statement

Let $s_{n,l}$ denote an anechoic signal of speech or music in the short-time Fourier transform (STFT) domain, where, n and l correspond to the time frame and frequency bin indices, respectively. We refer to $s_{n,l}$ as a source signal. The source signal is assumed to be unobservable, and only its reverberated version, denoted by $x_{n,l}$, is available. Dereverberation is the process of estimating the source signal, $s_{n,l}$, by using the reverberant signal, $x_{n,l}$, without precise knowledge of the reverberation properties. We assume that the observed signal is monaural and dereverberation is performed in a batch processing manner.

We adopt the inverse filtering approach to achieve dereverberation. As in [5], we assume that the reverberant signal is generated

from the source signal $s_{n,l}$ by an auto regressive (AR) system as

$$x_{n,l} = \sum_{d=k}^K g_{d,l} x_{n-d,l} + s_{n,l}. \quad (1)$$

k and K are the initial delay time and the filter length, respectively. Then the source signal can be recovered as $x_{n,l} - \sum_{d=k}^K g_{d,l} x_{n-d,l}$. Thus, our goal is to estimate $g_{d,l}$. In this paper we refer to $g_{d,l}$ as a reverberation filter coefficient.

In order to estimate $g_{d,l}$, we introduce a source model characterized by a set of parameters. The source model is used to measure the degree to which a dereverberated signal is good. Since the source model parameters are unavailable in advance, they are also estimated in the course of the reverberation filter estimation.

To put this idea into shape, we have to set up an optimization problem, which is specified by the following two components:

1. source model reflecting the source characteristics such as harmonicity, which is described in Section 3.2, and
2. cost function between the dereverberated signal, $r_{n,l} = x_{n,l} - \sum_{d=k}^K g_{d,l} x_{n-d,l}$, and this source model.

2.2. IS-divergence-based existing method

Yoshioka et al. proposed a dereverberation method [5], which minimizes a cost function defined as follows:

$$Q^{(IS)} = \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} \left(\log p_{n,l} + \frac{|r_{n,l}|^2}{p_{n,l}} \right), \quad (2)$$

where $p_{n,l}$ is the source model that defines the power spectral envelope of the source. From the viewpoint of the estimation of $p_{n,l}$, this cost function consists of the Itakura-Saito (IS) divergence between $|r_{n,l}|^2$ and $p_{n,l}$. Although it is not the IS-divergence in terms of the estimation of $g_{d,l}$, we simply call it the IS-divergence. Note that the cost function is quadratic form with respect to the filter coefficient $g_{d,l}$, and thus $g_{d,l}$ can be estimated easily.

However, minimizing Eq.(2) with respect to $g_{d,l}$ sometimes leads to excessive suppression of direct sound. This can be accounted for as follows. We note that $r_{n,l}$ is the prediction error of the AR system representing reverberation (see Eq.(1)), and Eq.(2) is a weighted sum of the errors. Hence, this cost function tends to confuse the sustain part of direct sound as reverberation because the sustain part is almost predictable from its preceding part. Furthermore, sophisticated source models that were previously proposed for music spectrum modeling are difficult to optimize with the IS-divergence.

2.3. Proposed I-divergence-based method

In this paper, we consider the I-divergence between the magnitude of dereverberated signal $|r_{n,l}|$ and the source model $p_{n,l}$ as:

$$Q^{(I)} = \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} \left(|r_{n,l}| \log \frac{|r_{n,l}|}{p_{n,l}} - (|r_{n,l}| - p_{n,l}) \right). \quad (3)$$

Note that $p_{n,l}$ represents in turn the magnitude spectrum of the source. Advantages of the proposed approach are twofold. Firstly, with this cost function $g_{d,l}$ is estimated in a way that directly minimize the difference between the dereverberated signal and the source estimate, which avoids sound distortion. Secondly, various magnitude spectrum modeling techniques for music signals based on the I-divergence (e.g., Harmonic GMM [7, 8] designed for F0 estimation) can be incorporated into the present dereverberation system.

It is important to note here that the cost function given by Eq.(3) is no longer a quadratic form with respect to the reverberation filter

Table 1. Algorithm for parameter estimation

- | |
|---|
| 1. Initialize source magnitude model parameters assuming $ r_{n,l} $ to be $ x_{n,l} $, and initialize $g_{d,l}$ by the IS-divergence based dereverberation given by Eq.(2) with $p_{n,l} = x_{n,l} ^2$. |
| 2. Update source magnitude model parameters using current $g_{d,l}$. |
| 3. Update $g_{d,l}$ using a current source magnitude model estimate: <ol style="list-style-type: none"> 3.a Update auxiliary variables $\{t_{n,l}, b_{n,l}, \omega_{n,l}\}$ by Eq.(13). 3.b Update $g_{d,l}$ by Eq.(14). 3.c Iterate from 3.a for several times. |
| 4. Iterate 2. and 3. until convergence. |

coefficients $g_{d,l}$. We are thus concerned with deriving an algorithm for estimating $g_{d,l}$, as will be discussed in the following section.

3. DEREVERBERATION WITH I-DIVERGENCE

This section describes the proposed parameter estimation algorithm using the auxiliary function method. The principle of the auxiliary function method is to iteratively minimize an upper bound of a cost function, which is easier to minimize than the original cost function. The upper bound is called an auxiliary function, and it is generally characterized by both the parameters of interest and additional parameters called auxiliary parameters. See [10] for more details.

3.1. Parameter estimation algorithm

Now our objective is to introduce an iterative algorithm for estimating $g_{d,l}$ efficiently on the basis of I-divergence. The key to the development of the proposed algorithm is its use of the auxiliary function method to deduce from Eq.(3) a quadratic form with respect to $g_{d,l}$, which can be optimized analytically. The auxiliary function method is used twice before obtaining the quadratic form. The overall parameter estimation algorithm is summarized in Table 1.

First we introduce one inequality to eliminate nonlinearity of a logarithmic function in the first term in Eq.(3) as follows:

$$\log |r_{n,l}| \leq \log t_{n,l} + \frac{1}{t_{n,l}} (|r_{n,l}| - t_{n,l}), \quad (4)$$

where $t_{n,l} > 0$ is an auxiliary parameter. The right-hand side of the inequality is a linear function that is tangent to $\log |r_{n,l}|$ at $|r_{n,l}| = t_{n,l}$, and thus can be considered as an auxiliary function of $\log |r_{n,l}|$. By using this, we obtain an auxiliary function Q^+ for $Q^{(I)}$ as:

$$Q^+ = \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} \left(\frac{|r_{n,l}|^2}{t_{n,l}} + \left(\log \frac{t_{n,l}}{p_{n,l}} - 2 \right) |r_{n,l}| + p_{n,l} \right). \quad (5)$$

However, $\frac{\partial}{\partial g_{d,l}} Q^+ = 0$ still cannot be solved analytically because of the second term in Eq.(5).

Next, we introduce the following two inequalities, which are chosen depending on $\log(t_{n,l}/p_{n,l}) - 2$, appearing in Eq.(5).

$$\text{in case } D_{n,l} \geq 0: |r_{n,l}| \leq \frac{|r_{n,l}|^2}{2b_{n,l}} + \frac{b_{n,l}}{2}, \quad b_{n,l} > 0 \quad (6)$$

$$\text{in case } D_{n,l} < 0: |r_{n,l}| \geq \text{Re} [\omega_{n,l}^* r_{n,l}], \quad |\omega_{n,l}| = 1 \quad (7)$$

where $D_{n,l} = \log(t_{n,l}/p_{n,l}) - 2$ and superscript $*$ stands for complex conjugate. $b_{n,l}$ and $\omega_{n,l}$ are auxiliary parameters for Eq.(6) and Eq.(7), respectively. The right-hand side of the first inequality is a convex quadratic function that is tangent to $|r_{n,l}|$ at $|r_{n,l}| = b_{n,l}$. The right-hand side of the second inequality means inner product of $r_{n,l}$ and an directional vector $\omega_{n,l}$.

By introducing the above two inequalities, we obtain another new auxiliary function Q^{++} :

$$Q^{++} = \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} Q_{n,l}^{++}, \quad (8)$$

$$Q_{n,l}^{++} = \begin{cases} \left(\frac{D_{n,l}}{2b_{n,l}} + \frac{1}{t_{n,l}} \right) |r_{n,l}|^2 + c & (D_{n,l} \geq 0) \\ \frac{1}{t_{n,l}} \left| r_{n,l} + \frac{t_{n,l}}{2} D_{n,l} \omega_{n,l} \right|^2 + c & (D_{n,l} < 0) \end{cases}, \quad (9)$$

where c represents terms that do not depend on $g_{d,l}$. For simplicity, let us set $\kappa_{n,l}$ and $\rho_{n,l}$ as

$$\kappa_{n,l} = \begin{cases} x_{n,l} & (D_{n,l} \geq 0) \\ x_{n,l} + \frac{t_{n,l}}{2} D_{n,l} \omega_{n,l} & (D_{n,l} < 0) \end{cases}, \text{ and} \quad (10)$$

$$\frac{1}{\rho_{n,l}} = \begin{cases} \frac{1}{t_{n,l}} + \frac{D_{n,l}}{2b_{n,l}} & (D_{n,l} \geq 0) \\ \frac{1}{t_{n,l}} & (D_{n,l} < 0) \end{cases}. \quad (11)$$

Then Eq.(8) can be rewritten as

$$Q^{++} = \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} \frac{\left| \kappa_{n,l} - \sum_{d=k}^K g_{d,l} x_{n-d,l} \right|^2}{\rho_{n,l}} + c. \quad (12)$$

This form is obviously quadratic with respect to $g_{d,l}$. Thus $g_{d,l}$ can be estimated in the same way as IS-divergence-based estimation.

$g_{d,l}$ is estimated by alternatively minimizing Q^{++} with respect to the auxiliary parameters, $\{t_{n,l}, b_{n,l}, \omega_{n,l}\}$, and reverberation filter coefficients $g_{d,l}$. The update rules for the auxiliary parameters are

$$t_{n,l} = |r_{n,l}|, \quad b_{n,l} = |r_{n,l}|, \quad \text{and} \quad \omega_{n,l} = r_{n,l} / |r_{n,l}|. \quad (13)$$

The first and the second rules are given by minimizing the right-hand side of Eqs. (4) and (6), respectively. The third rule is given by maximizing the right-hand side of Eq.(7). The update rule for $g_{d,l}$ is derived based on Eq.(12) and we obtain

$$[g_{k,l}, \dots, g_{K,l}]^T = R_l^{-1} r_l \quad (14)$$

where, R_l and r_l are called a modified correlation matrix and a modified correlation vector, respectively, and are structured as follows:

$$R_l = \begin{bmatrix} \sum_{n=0}^{N-1} \frac{x_{n-k,l} x_{n-k,l}^*}{\rho_{n,l}} & \dots & \sum_{n=0}^{N-1} \frac{x_{n-K,l} x_{n-k,l}^*}{\rho_{n,l}} \\ \vdots & \ddots & \vdots \\ \sum_{n=0}^{N-1} \frac{x_{n-k,l} x_{n-K,l}^*}{\rho_{n,l}} & \dots & \sum_{n=0}^{N-1} \frac{x_{n-K,l} x_{n-K,l}^*}{\rho_{n,l}} \end{bmatrix}, \quad (15)$$

$$r_l = \begin{bmatrix} \sum_{n=0}^{N-1} \frac{\kappa_{n,l} x_{n-k,l}^*}{\rho_{n,l}} & \dots & \sum_{n=0}^{N-1} \frac{\kappa_{n,l} x_{n-K,l}^*}{\rho_{n,l}} \end{bmatrix}^T. \quad (16)$$

Note that R_l and r_l can be calculated efficiently using the Fast Fourier Transform in the same way as [11].

3.2. Source model

This section introduces two kinds of source models that have estimation algorithms with both I and IS divergence. Convergence is guaranteed in case these models are estimated with the same divergence as one used for estimation of reverberation filter coefficients.

3.2.1. Harmonic GMM with residual (HGMM)

HGMM is a source model that can represent harmonic structure of instrument sounds. The HGMM describes a magnitude or power

spectrum of musical tones by using a GMM where the means of each Gaussian component appear at the harmonic frequencies of the tone. Assuming I musical tones, each of which has J harmonics, are present at each time frame, the magnitude or power of source is modeled as

$$p_{n,l} = \sum_{i=1}^I \left(\sum_{j=1}^J u_{i,j,n} H_{i,j,n,l} \right) + \sum_{m=1}^M v_{m,n} I_{m,n,l}, \quad (17)$$

$$H_{i,j,n,l} = e^{-\frac{(\Omega_l - j\mu_{i,n})^2}{2\sigma^2}}, \quad I_{m,n,l} = e^{-\frac{(\Omega_l - v_m)^2}{2\gamma^2}}, \quad (18)$$

where $u_{i,j,n}$ is the intensity of j -th peak of the i -th tone, $\mu_{i,n}$ is the fundamental frequency, and σ^2 is the spectral spread of each harmonic component. Ω_l means a scaling function that maps the index of frequency bins to Hertz. Inharmonic components such as drum sound and noise floor, are modeled by an inharmonic GMM in the second term of Eq.(17), which has M Gaussians with fixed means v_m and fixed large variance γ^2 .

Parameters to be estimated are $\{u_{i,j,n}, v_m, \mu_{i,n}\}$. Below are update rules of these with I and IS divergence, which are also based on the auxiliary function method.

- Let us put auxiliary parameters, $\Psi_{i,j,n,l}^{(I)}$ and $\Psi_{i,j,n,l}^{(IS)}$ as

$$\Psi_{i,j,n,l}^{(I)} = \frac{u_{i,j,n} H_{i,j,n,l} |r_{n,l}|}{p_{n,l}}, \quad \text{and} \quad (19)$$

$$\Psi_{i,j,n,l}^{(IS)} = u_{i,j,n} H_{i,j,n,l} \left[1 + \frac{u_{i,j,n} H_{i,j,n,l} (|r_{n,l}|^2 - p_{n,l})}{p_{n,l}^2} \right]. \quad (20)$$

- Update rules of $u_{i,j,n}$ are:

$$u_{i,j,n} = \begin{cases} \frac{\sum_{l=0}^{L-1} \Psi_{i,j,n,l}^{(I)}}{\sum_{l=0}^{L-1} H_{i,j,n,l}} & \text{(I-divergence)} \\ \frac{1}{L} \sum_{l=0}^{L-1} \frac{\Psi_{i,j,n,l}^{(IS)}}{H_{i,j,n,l}} & \text{(IS-divergence)} \end{cases}. \quad (21)$$

Update rules of v_m, n are similar to those of $u_{i,j,n}$.

- An update rule of $\mu_{i,n}$ with I-divergence is:

$$\mu_{i,n} = \frac{\sum_{j=1}^J \sum_{l=0}^{L-1} \Omega_l j \Psi_{i,j,n,l}^{(I)}}{\sum_{j=1}^J \sum_{l=0}^{L-1} j^2 \Psi_{i,j,n,l}^{(I)}}. \quad (22)$$

$\mu_{i,n}$ is hard to update analytically with IS-divergence and obtained in numerical calculations such as the Newton method.

3.2.2. Nonnegative-Matrix-Factorization-based model (NMF)

NMF is a dimension reduction technique that has been successfully applied to musical signal processing such as source separation, transcription [9] and so on. Update rules of NMF with I and IS divergence are described in [12].

4. EXPERIMENTS

The accuracy of dereverberation and that of source modeling are evaluated by using simulated data to compare proposed and existing methods.

4.1. Experimental Condition

The proposed I-divergence-based dereverberation method is compared with the existing IS-divergence-based method using two different source models, the HGMM and the NMF model. A combina-

Table 2. Experimental results: The left-hand side in each cell shows the LSD between the true source and the dereverberated signal. The right-hand is the LSD between the true source and the signal given by the source model. Smaller values indicate higher quality. The best performing values are shown in boldface.

		$K = 80$		$K = 160$		$K = 240$	
Derev.-I / HGMM-I (Proposed)	$k = 3$	8.72	8.68	8.68	8.71	8.75	8.78
	$k = 10$	8.65	8.70	8.63	8.72	8.68	8.69
	$k = 20$	8.66	8.68	8.63	8.66	8.62	8.65
Derev.-I / NMF-I (Proposed)	$k = 3$	8.74	9.06	8.78	8.96	8.88	9.03
	$k = 10$	8.64	8.98	8.62	9.08	8.64	8.82
	$k = 20$	8.64	8.78	8.62	8.78	8.63	8.78
Derev.-IS / HGMM-I [6]	$k = 3$	9.82	9.66	9.98	9.82	10.05	9.89
	$k = 10$	9.07	8.98	9.14	9.19	9.16	9.09
	$k = 20$	8.79	8.76	8.82	8.78	8.85	8.82
Derev.-IS / HGMM-IS	$k = 3$	10.06	13.53	10.27	13.61	10.34	13.64
	$k = 10$	9.24	12.69	9.34	12.80	9.32	13.37
	$k = 20$	8.91	13.37	8.98	13.38	9.01	13.41
Derev.-IS / NMF-IS	$k = 3$	10.14	10.06	10.34	10.21	10.40	10.26
	$k = 10$	9.24	9.80	9.36	10.15	9.22	9.32
	$k = 20$	8.90	9.16	8.94	9.22	8.96	9.10

tion of IS-divergence-based dereverberation and I-divergence-based HGMM, which is our conventional method [6], is also evaluated.

Twelve musical pieces in RWC Music Database, Jazz and Classic [13] synthesized with a MIDI tone generator, were used as the true source signals. The signals were truncated up to 20 seconds, and reverberated by being convolved with two impulse responses. The reverberation times of these impulse responses were about 1 and 1.5 seconds, respectively. We measured the dereverberation performance using the log spectral distance (LSD), which is defined as:

$$\text{LSD}(\mathbb{Y}, \mathbb{Z}) \equiv \sum_{n=0}^{N-1} \sum_{l=0}^{L-1} \left| 20 \log_{10} \frac{|y_{n,l}|}{|z_{n,l}|} \right| / NL, \quad (23)$$

We calculated LSD between 1) the true source, $\mathbb{S} = \{s_{n,l}\}$, and the dereverberated signal, $\mathbb{R} = \{\hat{r}_{n,l}\}$, and 2) the true source and the signal given by the source model, $\mathbb{P} = \{\hat{p}_{n,l}\}$, where $\hat{\cdot}$ means the estimated result. The former, $\text{LSD}(\mathbb{S}, \mathbb{R})$, stands for dereverberation accuracy, and the latter, $\text{LSD}(\mathbb{S}, \mathbb{P})$, stands for source modeling accuracy. Sampling rate was 44.1kHz. STFT is computed using a 1024 point Gaussian window that has a 256-point overlap.

4.2. Results and Discussion

Table 2 shows the evaluation results with several conditions about initial delay time k and filter length K . Each left-hand value in each cell shows $\text{LSD}(\mathbb{S}, \mathbb{R})$ and each right-hand value shows $\text{LSD}(\mathbb{S}, \mathbb{P})$. First, we can see that the accuracy of the I-divergence-based source model estimation was better than that of the IS-divergence-based source model estimation. Furthermore, if we compare the results of I-divergence-based dereverberation and the method based on [6], we find that this improvement in source model estimation led to the dereverberation performance improvement only when the reverberation filter is also estimated based on the I-divergence. This clearly indicates the advantage of the use of the I-divergence both for source model estimation and reverberation filter estimation.

Figure 1 shows the piano spectrograms of the dereverberated signal obtained with the proposed (c) and existing (d) methods. It can be found that (c) could successfully cancel the reverberation and avoid distorting the sustain of piano tones.

Another small experiment for real reverberation on CDs showed similar property as the above results. The experiment also showed that the dereverberation performance did not vary much according to the signal length. The audio files of both simulated data and real recordings are available at <http://winnie.kuis.kyoto-u.ac.jp/members/yasuraok/ICASSP2011/>.

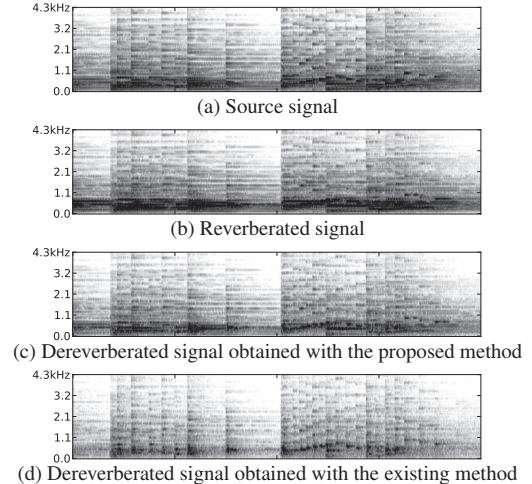


Fig. 1. Example of a dereverberated signal

5. CONCLUSION

This paper presented a new dereverberation method using I-divergence, which is often used for musical signal processing. The proposed method uses the auxiliary function method for convergence-guaranteed parameter estimation. Our experimental results revealed that the proposed method could perform music dereverberation more accurately and robustly than the existing IS-divergence-based one.

Future work includes integrating the proposed method into various types of musical signal processing such as sound source separation and sound analysis and synthesis. We will also extend our method for multi-channel processing, in particular, stereo input.

6. REFERENCES

- [1] A. P. Klapuri, "Multiple fundamental frequency estimation based on harmonicity and spectral smoothness," *IEEE Trans. Speech and Audio Process.*, vol. 11, no. 6, pp. 804–816, 2003.
- [2] K. Lee and M. Slaney, "Acoustic chord transcription and key extraction from audio using key-dependent HMMs trained on synthesized audio," *IEEE Trans. Audio, Speech and Lang. Process.*, vol. 16, no. 2, pp. 291–301, 2008.
- [3] J. A. Moorer, *On the segmentation and analysis of continuous musical sound by digital computer*, Ph.D. thesis, Stanford University, 1975.
- [4] M. Goto, "Active music listening interfaces based on signal processing," in *Proc. ICASSP, 2007*, vol. 4, pp. 1441–1444.
- [5] T. Yoshioka, T. Nakatani, and M. Miyoshi, "An integrated method for blind separation and dereverberation of convolutive audio mixtures," in *Proc. EUSIPCO, 2008*.
- [6] N. Yasuraoka, T. Yoshioka, T. Nakatani, A. Nakamura, and H. G. Okuno, "Music dereverberation using harmonic structure source model and Wiener filter," in *Proc. ICASSP, 2010*, pp. 53–56.
- [7] H. Kameoka, T. Nishimoto, and S. Sagayama, "Extraction of multiple fundamental frequencies from polyphonic music using harmonic clustering," in *Proc. ICA, 2004*, pp. 1–59–62.
- [8] H. Kameoka, T. Nishimoto, and S. Sagayama, "A multipitch analyzer based on harmonic temporal structured clustering," *IEEE Trans. Audio, Speech and Lang. Process.*, vol. 15, no. 3, pp. 982–994, 2007.
- [9] P. Smaragdis and J.C Brown, "Non-negative matrix factorization for polyphonic music transcription," in *Proc. WASPAA, 2003*, pp. 170–180.
- [10] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS, 2001*, pp. 556–562.
- [11] T. Yoshioka, T. Nakatani, and M. Miyoshi, "Fast algorithm for conditional separation and dereverberation," in *Proc. EUSIPCO, 2009*.
- [12] M. Nakano, H. Kameoka, J. Le Roux, Y. Kitano, N. Ono, and S. Sagayama, "Convergence-guaranteed multiplicative algorithms for nonnegative matrix factorization with beta-divergence," in *Proc. MLSP, 2010*.
- [13] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Popular, classical, and jazz music databases," in *Proc. ISMIR, 2002*, pp. 287–288.