MULTI-RESOLUTION SIGNAL DECOMPOSITION WITH TIME-DOMAIN SPECTROGRAM FACTORIZATION

Hirokazu Kameoka

The University of Tokyo / Nippon Telegraph and Telephone Corporation

ABSTRACT

This paper proposes a novel framework that makes it possible to realize non-negative matrix factorization (NMF)-like signal decompositions in the time-domain. This new formulation also allows for an extension to multi-resolution signal decomposition, which was not possible with the conventional NMF framework.

Index Terms— Audio source separation, non-negative matrix factorization (NMF), majorization-minimization (MM), auxiliary function, multi-resolution representation

1. INTRODUCTION

Many sound recordings are mixtures of multiple sound sources. Audio source separation, i.e., the process by which individual sound sources are separated from a mixture signal, has long been a formidable challenge in the field of audio signal processing.

In recent years, non-negative matrix factorization (NMF) has attracted a lot of attention after being proposed as a powerful approach for music transcription [1] and audio source separation [2]. With this approach, the magnitude (or power) spectrogram of a mixture signal, interpreted as a non-negative matrix $\boldsymbol{Y}$, is factorized into the product of two non-negative matrices $\boldsymbol{H}$ and $\boldsymbol{U}$. This can be interpreted as approximating the observed spectra at each time frame as a linear sum of basis spectra scaled by time-varying amplitudes, and amounts to decomposing the observed spectrogram into the sum of rank-1 spectrograms. An important feature of NMF is that its non-negativity constraint usually induces sparse representations, i.e., $\boldsymbol{U}$ with a relatively large number of zero entries. This means that each observed spectrum is parsimoniously represented using only a few active basis spectra. In such situations, the sequence of observed spectra can be approximated reasonably well when each basis spectrum expresses the spectrum of an underlying audio event that occurs frequently in the entire observed range. Thus, with music signals, each basis spectrum usually becomes the spectrum of a frequently used pitch in the music piece.

Although the concept of the NMF-based audio source separation approach has been shown to be successful, one limitation is that it does not take account of phase information: the additivity of magnitude (or power) spectra is assumed, which holds only approximately. To overcome this limitation, we have previously proposed a framework called the "complex NMF" [3], where the observed complex spectrum at each time frame is modeled as the sum of components, each of which is described by the multiplication of a static basis spectrum, a time-varying amplitude and a time-varying phase spectrum. Unlike NMF, this model allows the components to cancel each other out, and so without any constraints, it does not naturally produce sparse representations. However, it has been shown that an additional sparsity constraint yields sparse representations similar to NMF. With a similar motivation, two groups (Parry et al. and Févotte et al.) have independently proposed a generative model of complex-valued coefficients of the short-time Fourier transform (STFT) of a mixture signal, where the power spectrogram of each underlying component is modeled as a rank-1 matrix (similarly to NMF) and the phase spectrogram is treated as uniformly distributed latent variables [4, 5]. They showed that the maximum likelihood estimation of the power spectral density parameters amounts to fitting the NMF model to an observed power spectrogram using the Itakura-Saito (IS) divergence as a goodness-of-fit criterion. This approach is called IS-NMF.

To the best of our knowledge, complex NMF and IS-NMF are among the first phase-aware NMF variants to be proposed. Although both approaches treat each element of the phase spectrogram as an independent parameter (or latent variable), the phases of time-frequency components are, in fact, constrained and dependent on each other. This is because the spectrograms obtained with typical time-frequency transforms (such as the STFT and the wavelet transform) are redundant representations (see Fig. 1 regarding STFT spectrograms). For example, the STFT spectrogram is computed by concatenating the Fourier transforms of overlapping short-time frames of the signal. Hence, all the elements of the STFT spectrogram must satisfy a certain condition to ensure that the waveforms within the overlapping segment of consecutive frames are consistent [6]. The shortcomings of the complex NMF and IS-NMF frameworks are that they fail to take account of this kind of redundancy. One possible way to further develop improved variants of these models would be to incoporate into the models the explicit condition that spectrograms must satisfy. However, this would make these models overcomplicated and difficult to optimize.

Instead of using mixing models defined in the time-frequency domain, this paper proposes introducing a time-domain model that makes it possible to realize NMF-like signal decompositions. Moreover, we will show that the time-domain formulation also allows for an extension to multi-resolution signal decompositions, that was not possible with the conventional frameworks.

2. FORMULATION

Let us denote an observed signal at time t_n by $y[n]$, and the signal of the entire period by $\boldsymbol{y} = (y[1], \ldots, y[N])^{\mathsf{T}} \in \mathbb{R}^N$. While the NMF approach considers decomposing an observed magnitude spectrogram into the sum of rank-1 spectro-

This work was supported by JSPS KAKENHI Grant Number 26730100.

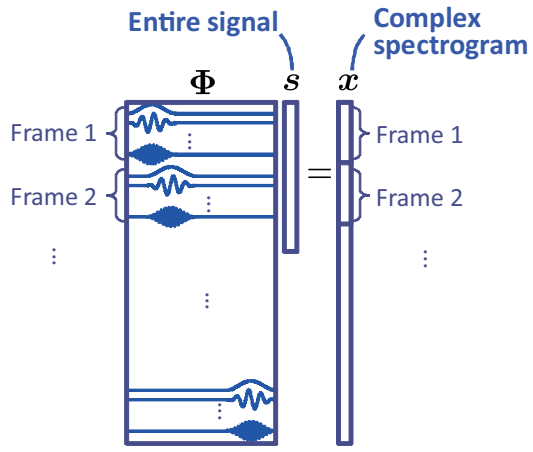

Fig. 1. Redundancy of STFT spectrograms.

grams, we consider decomposing the observed time-domain signal $\boldsymbol{y}$ into the sum of L signal components:

$$\boldsymbol{y} = \sum_{l=1}^{L} \boldsymbol{s}_l, \tag{1}$$

such that the magnitude spectrogram of each component is as close to a rank-1 structure as possible. Here, let us use $\boldsymbol{\psi}_{k,m} \in \mathbb{C}^N$ to denote arbitrary basis functions for time-frequency analysis, where k and m are the frequency and time indices, respectively. See Fig. 2 for a graphical illustration. With STFT, $\boldsymbol{\psi}_{k,m}$ is a windowed complex sinusoid. By using $\boldsymbol{\psi}_{k,m}$, the magnitude spectrogram of $\boldsymbol{s}_l$ can be written as $|\boldsymbol{\psi}_{k,m}^{\mathsf{H}} \boldsymbol{s}_l|$. Thus, the problem of interest can be cast as the optimization problem of minimizing

$$\mathcal{I}(\boldsymbol{\theta}) = \sum_{l} \sum_{k,m} (|\boldsymbol{\psi}_{k,m}^{\mathsf{H}} \boldsymbol{s}_l| - H_{k,l} U_{l,m})^2 + \mathcal{R}(\boldsymbol{U}),$$
$$\text{subject to} \quad \sum_{l} \boldsymbol{s}_l = \boldsymbol{y}, \tag{2}$$

with respect to $\boldsymbol{\theta} = \{\boldsymbol{H}, \boldsymbol{U}, \boldsymbol{S}\}$ where $\boldsymbol{H} = \{H_{k,l}\}$, $\boldsymbol{U} = \{U_{l,m}\}$ and $\boldsymbol{S} = \{\boldsymbol{s}_l\}$. Note that $H_{k,l} \geq 0$ and $U_{l,m} \geq 0$ are analogous to the basis spectrum and the time-varying amplitude in the NMF model. The first term of the above objective function becomes 0 when the magnitude spectrograms of all the members of $\boldsymbol{S}$ have exactly rank-1 structures. The second term $\mathcal{R}(\boldsymbol{U})$ is a regularization term for $\boldsymbol{U}$. It is important to note that as with complex NMF, this model allows the components to cancel each other out, and so some constraint is needed to induce the sparsity of $\boldsymbol{U}$. For this purpose, we define $\mathcal{R}(\boldsymbol{U})$ using the ℓ_p norm

$$\mathcal{R}(\boldsymbol{U}) = 2\lambda \sum_{l,m} |U_{l,m}|^p, \tag{3}$$

where $\lambda > 0$ weighs the importance of the sparsity cost relative to the fitting cost. When $0 < p < 2$, $\mathcal{R}(\boldsymbol{U})$ promotes sparsity if the norm of $\boldsymbol{U}$ is bounded. To bound $\boldsymbol{U}$, we assume

$$\sum_{k} H_{k,l}^2 = 1. \tag{4}$$

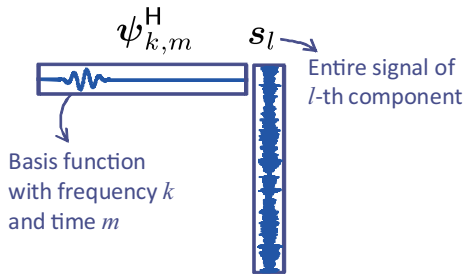

Fig. 2. Illustration of time-frequency basis functions.

3. OPTIMIZATION ALGORITHM

3.1. General principle of auxiliary function approach

Although it is difficult to solve the above optimization problem analytically, we can develop a computationally efficient algorithm for finding a locally optimal solution based on the auxiliary function concept. Here we introduce the general principle of the auxiliary function approach (also called the majorization-minimization approach).

Let us use $D(\theta)$ to denote an objective function that we want to minimize with respect to θ. $G(\theta, \alpha)$ is defined as an auxiliary function for $D(\theta)$ if it satisfies

$$D(\theta) = \min_{\alpha} G(\theta, \alpha). \tag{5}$$

We call α an auxiliary variable. By using $G(\theta, \alpha)$, $D(\theta)$ can be iteratively decreased according to the following theorem:

Theorem 1. *$D(\theta)$ is non-increasing under the updates, $\theta \leftarrow \operatorname{argmin}_\theta G(\theta, \alpha)$ and $\alpha \leftarrow \operatorname{argmin}_\alpha G(\theta, \alpha)$.*

3.2. Designing auxiliary function

When applying the auxiliary function approach to a certain minimization problem, the first step is to design an auxiliary function that bounds the objective function from above. The main difficulty with the present optimization problem lies in the discontinuity of the gradients of $|\boldsymbol{\psi}_{k,m}^{\mathsf{H}} \boldsymbol{s}_l|$ and $|U_{l,m}|^p$. We can design an auxiliary function for $\mathcal{I}(\boldsymbol{\theta})$ by invoking the following two inequalities.

Lemma 1. *For any complex number z and any complex number c satisfying $|c| = 1$, we have*

$$-|z| \leq -\mathrm{Re}(c^* z). \tag{6}$$

Equality holds when $c = z/|z|$.

Lemma 2. *When $0 < p < 2$, for any real (or complex) number x, we have*

$$2|x|^p \leq p|v|^{p-2}|x|^2 + 2 - p|v|^p. \tag{7}$$

Equality holds when $v = x$.

By aplying (6) and (7) to $\mathcal{I}(\boldsymbol{\theta})$, we obtain

$$\mathcal{I}(\boldsymbol{\theta}) \leq \sum_{l} \sum_{k,m} |\boldsymbol{\psi}_{k,m}^{\mathsf{H}} \boldsymbol{s}_l - H_{k,l} U_{l,m} c_{l,k,m}|^2$$

$$+\lambda\sum_{l,m}\big(p|V_{l,m}|^{p-2}U_{l,m}^2+(2-p)|V_{l,m}|^p\big). \quad (8)$$

The right-hand side of this inequality can be used as an auxiliary function for $\mathcal{I}(\boldsymbol{\theta})$. Here, $\boldsymbol{C}=\{c_{l,k,m}\}$ and $\boldsymbol{V}=\{V_{l,m}\}$ are the auxiliary variables. This auxiliary function is minimized with respect to $\boldsymbol{C}$ and $\boldsymbol{V}$ (equality of the above inequality holds) when

$$c_{l,k,m}=\boldsymbol{\psi}_{k,m}^{\mathsf{H}}\boldsymbol{s}_l/|\boldsymbol{\psi}_{k,m}^{\mathsf{H}}\boldsymbol{s}_l|, \quad (9)$$
$$V_{l,m}=U_{l,m}. \quad (10)$$

It should be noted that (9) indicates the phase spectrogram of the signal estimate $\boldsymbol{s}_l$, updated at the previous iteration.

As the update rule for the auxiliary variables is given by (9) and (10), all we need is to derive the update rule for $\boldsymbol{\theta}$.

3.3. Update rules

We can use the method of Lagrange multipliers to derive the update equations for $\boldsymbol{S}$ and $\boldsymbol{H}$. By setting the partial derivative of the Lagrangian

$$\sum_{l,k,m}|\boldsymbol{\psi}_{k,m}^{\mathsf{H}}\boldsymbol{s}_l-H_{k,l}U_{l,m}c_{l,k,m}|^2+\boldsymbol{\gamma}_s^{\mathsf{T}}\Big(\sum_l\boldsymbol{s}_l-\boldsymbol{y}\Big),$$

with respect to $\boldsymbol{s}_l$ at zero, we obtain $\boldsymbol{s}_l=\boldsymbol{\Psi}^{-1}(\boldsymbol{d}_l-\boldsymbol{\gamma}_s)$, where

$$\boldsymbol{\Psi}=\sum_k\sum_m\mathrm{Re}(\boldsymbol{\psi}_{k,m}\boldsymbol{\psi}_{k,m}^{\mathsf{H}}), \quad (11)$$
$$\boldsymbol{d}_l=\sum_k\sum_m\mathrm{Re}(H_{k,l}U_{l,m}c_{l,k,m}\boldsymbol{\psi}_{k,m}). \quad (12)$$

(12) amounts to performing time-frequency synthesis to obtain the l-th signal component, using the current estimates of the time-frequency coefficients, $H_{k,l}U_{l,m}c_{l,k,m}$. This process corresponds to the inverse STFT when $\boldsymbol{\psi}_{k,m}$ are defined as the STFT basis functions. By substituting this result into (1) and solving $\boldsymbol{\lambda}_s$, we can write the update equation for $\boldsymbol{s}_l$ as:

$$\boldsymbol{s}_l=\boldsymbol{\Psi}^{-1}\bigg\{\boldsymbol{d}_l+\frac{1}{L}\Big(\boldsymbol{\Psi}\boldsymbol{y}-\sum_l\boldsymbol{d}_l\Big)\bigg\}. \quad (13)$$

It may appear that $\boldsymbol{\Psi}\in\mathbb{R}^{N\times N}$ must be inverted to compute (13). However, this can be avoided under the following setting. First, the time-frequency basis function can be set so that $\sum_m\boldsymbol{\psi}_{k,m}\boldsymbol{\psi}_{k,m}^{\mathsf{H}}$ becomes a circulant matrix (see Fig. 3). Thus, $\boldsymbol{\Psi}$ can be diagonalized using the discrete Fourier transform matrix: $\boldsymbol{\Psi}=\boldsymbol{F}\boldsymbol{F}^{\mathsf{H}}\boldsymbol{\Psi}\boldsymbol{F}\boldsymbol{F}^{\mathsf{H}}=\boldsymbol{F}\boldsymbol{D}\boldsymbol{F}^{\mathsf{H}}$, where $\boldsymbol{F}$ denotes the discrete Fourier tranform matrix and $\boldsymbol{D}$ denotes a diagonal matrix whose diagonal elements are given as the sum of the power spectra of $\boldsymbol{\psi}_{1,m},\ldots,\boldsymbol{\psi}_{K,m}$. It can be shown (without proof, owing to space limitations), for example, that with STFT using the square-root Hanning window, $\boldsymbol{D}$ becomes an identity matrix, in which case $\boldsymbol{\Psi}$ also becomes an identity matrix. Thus, (13) can be simplified to

$$\boldsymbol{s}_l=\boldsymbol{d}_l+\frac{1}{L}\Big(\boldsymbol{y}-\sum_l\boldsymbol{d}_l\Big). \quad (14)$$

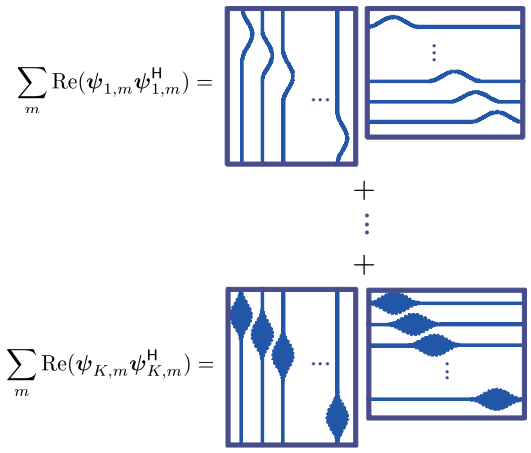

Fig. 3. Graphical illustration of $\boldsymbol{\Psi}$. $\sum_m\boldsymbol{\psi}_{k,m}\boldsymbol{\psi}_{k,m}^{\mathsf{H}}$ can be set at a circulant matrix in both cases of STFT and wavelet transform.

By setting the partial derivatives of the auxiliary function with respect to $H_{k,l}$ and $U_{l,m}$ at zeros, the update rules for $\boldsymbol{H}$ and $\boldsymbol{U}$ are given as

$$H_{k,l}=\frac{\sum_m U_{l,m}|\boldsymbol{\psi}_{k,m}^{\mathsf{H}}\boldsymbol{s}_l|}{\sqrt{\sum_k(\sum_m U_{l,m}|\boldsymbol{\psi}_{k,m}^{\mathsf{H}}\boldsymbol{s}_l|)^2}}, \quad (15)$$
$$U_{l,m}=\frac{\sum_k H_{k,l}|\boldsymbol{\psi}_{k,m}^{\mathsf{H}}\boldsymbol{s}_l|}{\sum_k H_{k,l}^2+\lambda p|V_{l,m}|^{p-2}}. \quad (16)$$

The iterative algorithm is summarized as follows:

1. Initialize $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{S}$.
2. Update $\boldsymbol{C}$ and $\boldsymbol{V}$ using (9) and (10).
3. Update $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{S}$ using (14)–(16) and return to 2.

For the initialization (step 1), we can use conventional NMF algorithms followed by Wiener filtering to obtain the estimates of $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{S}$.

3.4. Relation to complex NMF algorithm

It is interesting to compare the update rules of the present algorithm with those of the complex NMF algorithm [3]. The aim with Complex NMF is to approximate an observed complex spectrogram $Y_{k,m}$ with a model of the form

$$F_{k,m}=\sum_l H_{k,m}U_{l,m}e^{j\phi_{l,k,m}}, \quad (17)$$

where $\phi_{l,k,m}$ denotes the phase spectrogram of the l-th signal component. Under a certain condition, the complex NMF algorithm can be described as follows:

1. Initialize $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{\phi}$.
2. Update $\boldsymbol{X}=\{X_{l,k,m}\}$ and $\boldsymbol{V}=\{V_{l,m}\}$ using

$$X_{l,k,m}=H_{k,l}U_{l,m}e^{j\phi_{l,k,m}} + \tfrac{1}{L}\big(Y_{k,m}-\textstyle\sum_{l'}H_{k,l'}U_{l',m}e^{j\phi_{l',k,m}}\big), \quad (18)$$
$$V_{l,m}=U_{l,m}. \quad (19)$$

3. Update $\boldsymbol{H}$, $\boldsymbol{U}$ and $\boldsymbol{\phi}$ using

$$H_{k,l} = \frac{\sum_m U_{l,m}|X_{l,k,m}|}{\sqrt{\sum_k (\sum_m U_{l,m}|X_{l,k,m}|)^2}}, \quad (20)$$

$$U_{l,m} = \frac{\sum_k H_{k,l}|X_{l,k,m}|}{\sum_k H_{k,l}^2 + \lambda p |V_{l,m}|^{p-2}}, \quad (21)$$

$$e^{j\phi_{l,k,m}} = X_{l,k,m}/|X_{l,k,m}|. \quad (22)$$

and return to 2.

Interestingly, there is a similarity between the present and complex NMF algorithms, even though the models and the objective functions are different. Specifically, $X_{l,k,m}$ and $e^{j\phi_{l,k,m}}$ of the complex NMF algorithm are analogous to $\boldsymbol{s}_l$ and $c_{l,k,m}$ of the present algorithm, respectively.

$X_{l,k,m}$ can be viewed as an estimate of the complex spectrogram of the l-th signal component. In step 2, $X_{l,k,m}$ is updated by adding the portion of the error between the observed spectrogram and the model to the current estimate of $H_{k,l}U_{l,m}e^{j\phi_{l,k,m}}$. $\boldsymbol{\phi}$ is then updated at its argument $\arg(X_{l,k,m})$, and $\boldsymbol{H}$ and $\boldsymbol{U}$ are updated using its magnitude $|X_{l,k,m}|$. Similarly, according to (14), $\boldsymbol{s}_l$ is updated by adding the portion of the error between an observed signal and the sum of $\boldsymbol{d}_l$ to the current estimate of $\boldsymbol{d}_l$, where $\boldsymbol{d}_l$ is the signal converted from the set $\{H_{k,l}U_{l,m}c_{l,k,m}\}_{k,m}$. According to (9), (15) and (16), $c_{l,k,m}$ is then updated at the argument of $\boldsymbol{\psi}_{k,m}^{\mathsf{H}}\boldsymbol{s}_l$ (the complex spectrogram of $\boldsymbol{s}_l$), and $\boldsymbol{H}$ and $\boldsymbol{U}$ are updated using the magnitude of $\boldsymbol{\psi}_{k,m}^{\mathsf{H}}\boldsymbol{s}_l$.

One drawback with complex NMF is that $X_{l,k,m}$ is not guaranteed to satisfy the explicit condition that complex spectrograms must satisfy. This implies that the complex NMF algorithm is designed to search an unnessarily large solution space for the optimal parameters. By contrast, the present algorithm always ensures that the estimate of the complex spectrogram of the l-th latent component, $\boldsymbol{\psi}_{k,m}^{\mathsf{H}}\boldsymbol{s}_l$, is associated with a time-domain signal $\boldsymbol{s}_l$, keeping the search within a proper solution space.

4. MULTI-RESOLUTION DECOMPOSITION

Another important advantage of the present model is that it enables multi-resolution signal decompositions.

The optimal time-frequency resolution may depend on the kinds of sound sources. For example, temporal resolution is more important than frequency resolution for percussive sounds, with the opposite being true for low-pitched sounds. Thus, it is desirable to be able to arbitrarily set different time-frequency resolutions for individual sources. Of course, this was not possible with the NMF framework and its variants including complex NMF and IS-NMF, since in the conventional frameworks, the decompositions are carried out in the time-frequency domain with a particular resolution. By contrast, the proposed framework allows us to model the time-frequency structure of each component with a different resolution, thanks to the time-domain formulation.

The multi-resolution extension is straightforward. By additionally introducing a "resolution index" r, we obtain an objective function for the multi-resolution signal decomposition problem:

$$\mathcal{I}(\boldsymbol{\theta}) = \sum_{r,l}\sum_{k,m}(|\boldsymbol{\psi}_{r,k,m}^{\mathsf{H}}\boldsymbol{s}_{r,l}| - H_{r,k,l}U_{r,l,m})^2 + \mathcal{R}(\boldsymbol{U}),$$

Table 1. SNR improvements (dB) obtained by NMF and TSF.

Track #	NMF	TSF
1	8.53	11.03
2	4.96	8.20
3	12.27	13.05
4	10.41	11.23
5	10.19	11.53

subject to $\sum_{r,l}\boldsymbol{s}_{r,l} = \boldsymbol{y}$. Here, each r indicates a different time-frequency resolution. The above formulation can be explained as follows. All the signal components are divided into R groups, each of which consists of L_r members, and $\boldsymbol{s}_{r,l}$ denotes the l-th signal component within the group r. Since $\boldsymbol{\psi}_{r,k,m}$ is indexed by r, the spectrograms of the signal components in a different group r'will have a different time-frequency resolution.

It should be noted that the parameter update rules are derived in the same way as in the previous section.

5. EXPERIMENTS

We quantitatively compared the source separation performance of NMF (followed by Wiener filtering) and the proposed method (hereafter, time-domain spectrogram factorization (TSF)) by conducting supervised source separation experiments. We used professionally produced music recordings from the SiSEC 2013 database, available at `https://sisec.wiki.irisa.fr/`, as the experimental data. Each recording is a mixture of multiple tracks, each of which is produced by a single instrument or singer. The separated tracks are also available. We divided each recording into two segments, namely a test data segment and a training data segment. With all these methods, the basis spectra were pretrained using the individual tracks of the training data, and then source separation was performed on the test data. All the audio samples were monaural and sampled at 22.05kHz. STFT was computed using a square-root Hanning window that was 32ms long with a 16ms overlap. With both methods, 6 basis spectra were assigned to each track. Thus, for 5-track recordings, a total of 30 basis spectra were used for the separation. Tab. 1 shows the signal-to-noise Ratio (SNR) improvements after the separations with the two methods. From these results, we confirmed that TSF performed better than NMF.

6. CONCLUSIONS

This paper proposed time-domain spectrogram factorization (TSF), which makes it possible to realize NMF-like signal decompositions in the time domain. This new formulation also allows for an extension to multi-resolution signal separation, which was not possible with the conventional NMF framework. We confirmed through supervised source separation experiments that the proposed method outperformed NMF.

7. REFERENCES

[1] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for music transcription," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2003, pp. 177–180.

[2] T. Virtanen, "Sound source separation using sparse coding with temporal continuity objective," in *Proceedings of the International Computer Music Conference (ICMC)*, 2003, pp. 231–234.

[3] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A new sparse representation for acoustic signals," in *Proceedings of 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2009)*, 2009, pp. 3437–3440.

[4] R. M. Parry and I. Essa, "Phase-aware non-negative spectrogram factorization," in *Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA)*, 2007, pp. 536–543.

[5] C. Févotte, N. Bertin, and J. -L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computation*, vol. 21, no. 3, 2009.

[6] J. Le Roux, H. Kameoka, N. Ono, and S. Sagayama, "Fast signal reconstruction from magnitude stft spectrogram based on spectrogram consistency," in *Proceedings of The 13th International Conference on Digital Audio Effects (DAFx-10)*, 2010, pp. 397–403.